Unilateral Control in Repeated Games

Kai Li (kai.li@sjtu.edu.cn)

Shanghai Jiao Tong University

Nov 16, 2018

Outline

Iterated Prisoner's Dilemma (2-player)

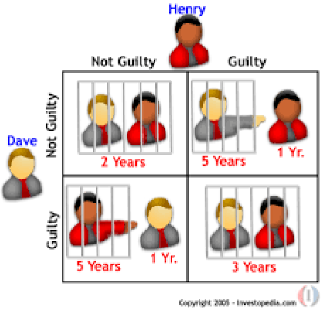
Prisoner's Dilemma: Intuition

PD is a symmetric two-player game.

- Payoff

  c: cooperation; d: defection

$$\begin{array}{cc} & \begin{array}{cc} c & d \end{array} \\ \begin{array}{c} c \\ d \end{array} & \begin{pmatrix} 3,3 & 0,5 \\ 5,0 & 1,1 \end{pmatrix} \end{array}$$



- Nash Equilibrium?

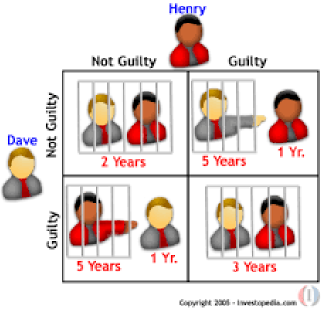
Prisoner's Dilemma: Intuition

PD is a symmetric two-player game.

- Payoff

  c: cooperation; d: defection

$$\begin{array}{cc} & \begin{array}{cc} c & d \end{array} \\ \begin{array}{c} c \\ d \end{array} & \begin{pmatrix} 3,3 & 0,5 \\ 5,0 & 1,1 \end{pmatrix} \end{array}$$



- Nash Equilibrium? Find the best response.

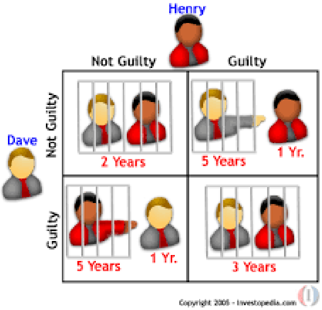
Prisoner's Dilemma: Intuition

PD is a symmetric two-player game.

- Payoff

  c: cooperation; d: defection

$$\begin{array}{cc} & \begin{array}{cc} \color{orange}{c} & \color{orange}{d} \end{array} \\ \begin{array}{c} \color{blue}{c} \\ \color{blue}{d} \end{array} & \begin{pmatrix} 3,3 & 0,\color{red}{5} \\ \color{red}{5},0 & \color{red}{1},\color{red}{1} \end{pmatrix} \end{array}$$



- Nash Equilibrium? Find the best response.

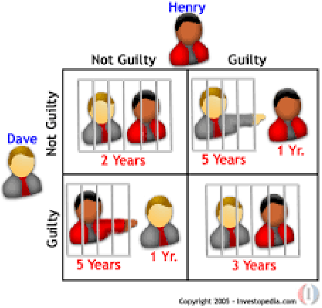
Prisoner's Dilemma: Intuition

PD is a symmetric two-player game.

- Payoff

  c: cooperation; d: defection

$$\begin{array}{cc} & \begin{array}{cc} c & d \end{array} \\ \begin{array}{c} c \\ d \end{array} & \begin{pmatrix} 3,3 & 0,5 \\ 5,0 & 1,1 \end{pmatrix} \end{array}$$



- Nash Equilibrium? (d, d)
- Social Optimum?

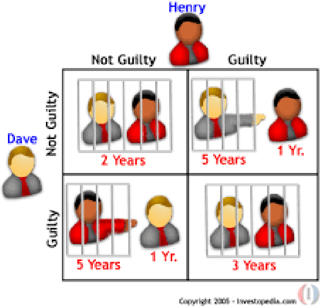
Prisoner's Dilemma: Intuition

PD is a symmetric two-player game.

- Payoff

  c: cooperation; d: defection

$$
\begin{array}{c c}
 & \begin{array}{cc} \quad c \quad & \quad d \quad \end{array} \\
\begin{array}{c} c \\ d \end{array} & \begin{pmatrix} 3,3 & 0,5 \\ 5,0 & 1,1 \end{pmatrix}
\end{array}
$$



- Nash Equilibrium? (d, d)
- Social Optimum? (c, c)

Iterated Prisoner's Dilemma: Finitely Repeated

Why is cooperation so common in society?

Iterated Prisoner's Dilemma: Finitely Repeated

Why is cooperation so common in society?

Long-run relationships – Repeated Games

Iterated Prisoner's Dilemma: Finitely Repeated

Why is cooperation so common in society?

Long-run relationships – Repeated Games

Consider that PD is finitely repeated (N stages)

- In N-th stage: Both players will defect, (d, d)

Iterated Prisoner's Dilemma: Finitely Repeated

Why is cooperation so common in society?

Long-run relationships – Repeated Games

Consider that PD is finitely repeated (N stages)

- In N-th stage: Both players will defect, (d, d)
- In $(N-1)$-th stage: (d, d) $\cdots$

Iterated Prisoner's Dilemma: Finitely Repeated

Why is cooperation so common in society?

Long-run relationships – Repeated Games

Consider that PD is finitely repeated (N stages)

- In N-th stage: Both players will defect, (d, d)
- In $(N-1)$-th stage: $(d, d) \cdots$
- Cannot get out of the dilemma

Iterated Prisoner's Dilemma: Finitely Repeated

Why is cooperation so common in society?

Long-run relationships – Repeated Games

Consider that PD is finitely repeated (N stages)

- In N-th stage: Both players will defect, (d, d)
- In $(N-1)$-th stage: (d, d) $\cdots$
- Cannot get out of the dilemma

What about the infinitely repeated games?

Iterated Prisoner's Dilemma: Infinitely Repeated

Strategies

Strategy: *History* $\rightarrow$ *Action*

- Simplification – Memory-one Strategy: Decisions based only on the previous stage outcome
- Conditional cooperation probability
  X: $\boldsymbol{p} = (p_{cc}, p_{cd}, p_{dc}, p_{dd})$,
  Y: $\boldsymbol{q} = (q_{cc}, q_{cd}, q_{dc}, q_{dd})$

Markov Chain $\boldsymbol{M}$

$$\begin{array}{c} \\ cc \\ cd \\ dc \\ dd \end{array} \begin{array}{c} \begin{array}{cccc} \quad cc \quad & \quad cd \quad & \quad dc \quad & \quad dd \quad \end{array} \\ \begin{pmatrix} p_1 q_1 & p_1(1-q_1) & (1-p_1)q_1 & (1-p_1)q_1 \\ p_2 q_3 & p_2(1-q_3) & (1-p_2)q_3 & (1-p_2)q_3 \\ p_3 q_2 & p_3(1-q_2) & (1-p_3)q_2 & (1-p_3)q_2 \\ p_4 q_4 & p_4(1-q_4) & (1-p_4)q_4 & (1-p_4)q_4 \end{pmatrix} \end{array}$$

Iterated Prisoner's Dilemma: Infinitely Repeated

Payoffs

Markov Chain

- Unique (in most cases) stationary distribution $\boldsymbol{v} = (v_{cc}, v_{cd}, v_{dc}, v_{dd})$

$$\boldsymbol{v}^T \cdot \boldsymbol{M} = \boldsymbol{v}$$

- Average distribution $\frac{1}{n}\sum_{k=1}^{n} \boldsymbol{v}^{(k)}$
- Krylov-Bogoliubov Argument

$$\boldsymbol{v} = \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \boldsymbol{v}^{(k)}$$

Payoff Vector $\boldsymbol{S}_X = (R, S, T, P)$, $\boldsymbol{S}_Y = (R, T, S, P)$
Average Payoff

$$s_X = \boldsymbol{S}_X \cdot \boldsymbol{v} \text{ and } s_Y = \boldsymbol{S}_Y \cdot \boldsymbol{v}$$

http://galton.uchicago.edu/~lalley/Courses/312/MarkovChains.pdf

Zero-determinant (ZD) Strategies: Insights (1)

Let $\boldsymbol{M}' = \boldsymbol{M} - \boldsymbol{I}$, as $\boldsymbol{v}^T\boldsymbol{M} = \boldsymbol{v}$, we have

$$\boldsymbol{v}^T\boldsymbol{M}' = 0.$$

$\boldsymbol{M}$ has an eigenvalue 1, $\det(\boldsymbol{M} - \mathbf{1}) = 0$. Thus,

$$\mathrm{Adj}(\boldsymbol{M}')\boldsymbol{M}' = \det(\boldsymbol{M}') = 0$$

When $\boldsymbol{v}$ is unique, the matrix $\boldsymbol{M}'$ has rank 3.
Therefore $\boldsymbol{v}$ is proportional to every row of $\mathrm{Adj}(\boldsymbol{M}')$.

Iterated Prisoner's Dilemma contains strategies that dominate any evolutionary opponent. PNAS 2012

Zero-determinant (ZD) Strategies: Insights (2)

Consider $\boldsymbol{M}'$ (the forth column is substituted by $\boldsymbol{f}$)

$$\begin{aligned}\det(D) &= \det\begin{bmatrix} m'_{11} & m'_{12} & m'_{13} & f_1 \\ m'_{21} & m'_{22} & m'_{23} & f_2 \\ m'_{31} & m'_{32} & m'_{33} & f_3 \\ m'_{41} & m'_{42} & m'_{43} & f_4 \end{bmatrix} \\ &= f_1 \mathrm{Adj}(\boldsymbol{M'})_{4,1} + f_2 \mathrm{Adj}(\boldsymbol{M'})_{4,2} + f_3 \mathrm{Adj}(\boldsymbol{M'})_{4,3} + f_4 \mathrm{Adj}(\boldsymbol{M'})_{4,4} \\ &= \boldsymbol{f} \cdot k\boldsymbol{v}\end{aligned}$$

Zero-determinant (ZD) Strategies: Insights (3)

$$\boldsymbol{f}\cdot k\boldsymbol{v} = \det\begin{bmatrix} p_1q_1-1 & p_1(1-q_1) & (1-p_1)q_1 & f_1 \\ p_2q_3 & p_2(1-q_3)-1 & (1-p_2)q_3 & f_2 \\ p_3q_2 & p_3(1-q_2) & (1-p_3)q_2-1 & f_3 \\ p_4q_4 & p_4(1-q_4) & (1-p_4)q_4 & f_4 \end{bmatrix}$$

$$= \det\begin{bmatrix} p_1q_1-1 & -1+p_1 & -1+q_1 & f_1 \\ p_2q_3 & -1+p_2 & q_3 & f_2 \\ p_3q_2 & p_3 & -1+q_2 & f_3 \\ p_4q_4 & p_4 & q_4 & f_4 \end{bmatrix}$$

Zero-determinant (ZD) Strategies: Definition

$$\boldsymbol{f} \cdot k\boldsymbol{v} = \det \begin{bmatrix} p_1q_1 - 1 & -1 + p_1 & -1 + q_1 & f_1 \\ p_2q_3 & -1 + p_2 & q_3 & f_2 \\ p_3q_2 & p_3 & -1 + q_2 & f_3 \\ p_4q_4 & p_4 & q_4 & f_4 \end{bmatrix}$$

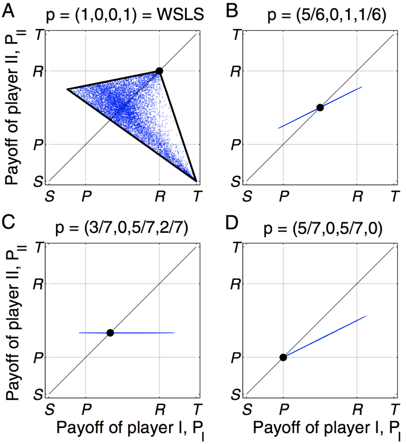
Let $\tilde{\boldsymbol{p}} = (-1 + p_1, -1 + p_2, p_3, p_4)^T$.
If we set $\boldsymbol{f} = \alpha \boldsymbol{S}_X + \beta \boldsymbol{S}_Y + \gamma \boldsymbol{1}$ and let $\tilde{\boldsymbol{p}}$ be proportional to $\boldsymbol{f}$, then we have

$$\alpha s_X + \beta s_Y + \gamma = 0.$$

We call such strategies Zero-determinant.

- Unilateral strategy
- Linear relation

Illustration



Evolution of extortion in Iterated Prisoner's Dilemma games. PNAS 2013

Akin's Lemma: Intuition

$$\boldsymbol{f} \cdot k\boldsymbol{v} = \det \begin{bmatrix} p_1 q_1 - 1 & -1 + p_1 & -1 + q_1 & f_1 \\ p_2 q_3 & -1 + p_2 & q_3 & f_2 \\ p_3 q_2 & p_3 & -1 + q_2 & f_3 \\ p_4 q_4 & p_4 & q_4 & f_4 \end{bmatrix}$$

If $\boldsymbol{f} \propto \tilde{\boldsymbol{p}}$, then $\boldsymbol{f} \cdot \boldsymbol{v} = 0$.

Lemma ([Akin, 2012])

$$\tilde{\boldsymbol{p}} \cdot \boldsymbol{v} = (\boldsymbol{p} - (1, 1, 0, 0)^T) \cdot \boldsymbol{v} = 0$$

Akin's Lemma: Proof

Lemma ([Akin, 2012])

$$\tilde{\boldsymbol{p}} \cdot \boldsymbol{v} = (\boldsymbol{p} - (1,1,0,0)^T) \cdot \boldsymbol{v} = 0$$

Proof.

- k-th stage:cooperating probability: $h_c(k) = (1,1,0,0)^T \cdot \boldsymbol{v}^{(k)}$
- $(k+1)$-th stage: cooperating probability: $h_c(k+1) = \boldsymbol{p} \cdot \boldsymbol{v}^{(k)}$
- $$\begin{aligned}(\boldsymbol{p} - (1,1,0,0)^T) \cdot \boldsymbol{v} &= \left((\boldsymbol{p} - (1,1,0,0)^T)\right) \cdot \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \boldsymbol{v}^(k) \\ &= \lim_{n\to\infty} \frac{1}{n}(h_c(n+1) - h_c(1)) = 0\end{aligned}$$

Akin's Lemma: Application

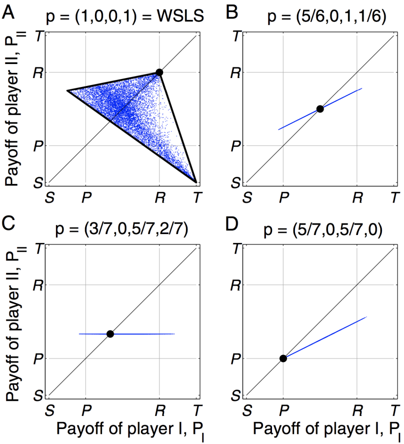
Lemma ([Akin, 2012])

$$\tilde{\boldsymbol{p}} \cdot \boldsymbol{v} = (\boldsymbol{p} - (1, 1, 0, 0)^T) \cdot \boldsymbol{v} = 0$$

If we let $\tilde{\boldsymbol{p}} = \alpha \boldsymbol{S}_X + \beta \boldsymbol{S}_Y + \gamma \mathbf{1}$, then we have

$$\alpha s_X + \beta s_Y + \gamma = 0.$$

Payoff Control



Evolution of extortion in Iterated Prisoner's Dilemma games. PNAS 2013

Payoff Control: Simple Case

If X wants to restrict Y's expected payoff:

$$s_Y \leq W$$

Calculation

$$s_Y - W = (\boldsymbol{S}_Y - W\mathbf{1}) \cdot \boldsymbol{v} = (R - W, S - W, T - W, P - W) \cdot \boldsymbol{v}$$
$$(1 - p_2)(s_Y - W) \leq 0$$

Akin's Lemma

$$(1 - p_2)v_2 = (-1 + p_1)v_1 + p_3 v_3 + p_4 v_4$$

Payoff Control in the Iterated Prisoner's Dilemma. IJCAI18

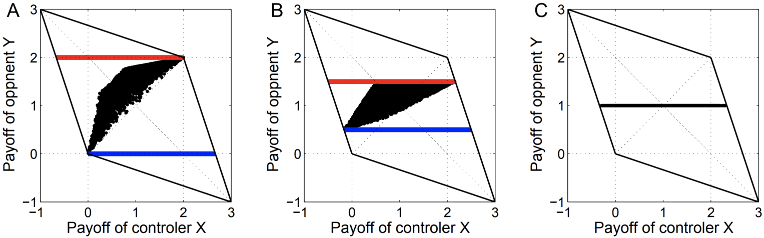
Payoff Control: Simple Case Cont'd

$$\alpha_1 v_1 + \alpha_3 v_3 + \alpha_4 v_4 \leq 0$$

v is a distribution.
We confine all $\alpha_i \leq 0$.

$$\begin{cases} 0 \leq p_2 < 1, \\ 0 \leq p_1 \leq \min\left(1 - \frac{R-W}{T-W}(1-p_2), 1\right), \\ 0 \leq p_3 \leq \min\left(\frac{W-S}{T-W}(1-p_2), 1\right), \\ 0 \leq p_4 \leq \min\left(\frac{W-P}{T-W}(1-p_2), 1\right). \end{cases}$$

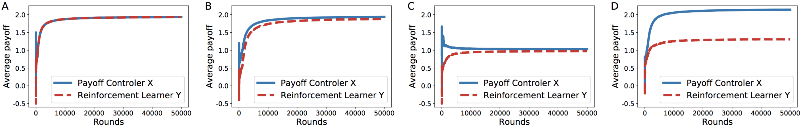
Payoff Control: Illustration



Reinforcement Learning Opponent

Regard the other player as Environment.
Average Reward Reinforcement Learning

$$q_\pi(s,a) = \sum_{k=1}^{\infty} \mathbb{E}_\pi \left[R_{t+k} - \bar{r}(\pi) \mid S_t = s \right]$$



Repeated Public Goods Games (multi-player)

Cooperation in Multi-agent Systems

Enforcing cooperation on agents is significant.

However, due to the social dilemmas in many systems, cooperation may be hard to achieve.

The Public Goods Game

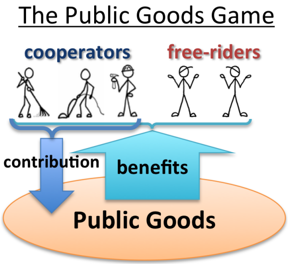
The public goods game is a classic model for social dilemmas.

- Cooperator (c) $\rightarrow$ contributes the endowment
- Defector (d) $\rightarrow$ contributes noting
- Endowments $\times$ r (*public goods*), then distribute $/n$
- Confronted with k cooperating opponents, a focal player obtains

$$R_{c,k} = \frac{r(k+1)}{n} - 1 \text{ or } R_{d,k} = \frac{rk}{n}.$$

Payoffs: free-riders $>$ contributors.

Nash Equilibrium?

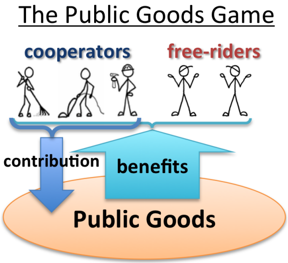


The Public Goods Game

The public goods game is a classic model for social dilemmas.

- Cooperator (c) $\rightarrow$ contributes the endowment
- Defector (d) $\rightarrow$ contributes noting
- Endowments $\times$ r (*public goods*), then distribute $/n$
- Confronted with k cooperating opponents, a focal player obtains

$$R_{c,k} = \frac{r(k+1)}{n} - 1 \text{ or } R_{d,k} = \frac{rk}{n}.$$



Payoffs: free-riders $>$ contributors.

Nash Equilibrium? All players choose to defect. (Tragedy of the commons)

Repeated Public Goods Games

Strategies

Repeated Games – Long-run Relationships

- Strategy: *History* $\rightarrow$ *Action*
- Simplification:
  1. Symmetric setting: ~~Who is playing~~ How many
  2. Memory-one strategy: Decisions based only on the previous stage outcome.

Repeated Public Goods Games

Strategies

Repeated Games – Long-run Relationships

- Strategy: *History* $\rightarrow$ *Action*
- Simplification:
  1. Symmetric setting: ~~Who is playing~~ How many
  2. Memory-one strategy: Decisions based only on the previous stage outcome.
- Formally, a memory-one strategy $\boldsymbol{p}$

$$\boldsymbol{p} = (p_{c,0}, \cdots, p_{c,k}, \cdots, p_{c,n-1}, \\ p_{d,0}, \cdots, p_{d,k}, \cdots, p_{d,n-1}),$$

where each p is a conditional probability
(Previous outcome $\rightarrow$ Cooperation probability in current stage)

Repeated Public Goods Games

Payoffs

All players adopt memory-one strategies:

Repeated Game $\rightarrow$ Markov Chain.

- Transition: *Previous outcome* $\rightarrow$ *Current outcome*
- Unique stationary distribution $\boldsymbol{v}$ over the outcomes (most cases)

Repeated Public Goods Games

Payoffs

All players adopt memory-one strategies:

Repeated Game → Markov Chain.

- Transition: *Previous outcome* → *Current outcome*
- Unique stationary distribution $\boldsymbol{v}$ over the outcomes

Payoffs: average payoff over all stages

- Payoff vector $\boldsymbol{\pi}$:

$$\boldsymbol{\pi} = (R_{c,0}, \cdots, R_{c,k}, \cdots, R_{c,n-1}, R_{d,0}, \cdots, R_{d,k}, \cdots, R_{d,n-1}).$$

- Average payoff π calculation

$$\pi = \boldsymbol{\pi} \cdot \boldsymbol{v}.$$

Cooperation Enforcing Strategy

Intuition

Enforcing cooperation

- Transitional methods: coordination algorithms, direct/indirect reciprocity, central institutions
- Disadvantages: hard to set up among substantial agents

Cooperation Enforcing Strategy

Intuition

Enforcing cooperation

- Transitional methods: coordination algorithms, direct/indirect reciprocity, central institutions
- Disadvantages: hard to set up among substantial agents

Via individual influence?

Cooperation Enforcing Strategy

Intuition

Via individual influence?

- Goal: design a unilateral strategy
- Property: The best response of all the opponents is to cooperate.

Cooperation Enforcing Strategy

Intuition

Via individual influence?

- Goal: design a unilateral strategy
- Property: The best response of all the opponents is to cooperate.

$$\Downarrow$$

Punishment: Any deviation from cooperation $\rightarrow$ Payoff decreases

Cooperation Enforcing Strategy

Intuition

Via individual influence?

- Goal: design a unilateral strategy
- Property: The best response of all the opponents is to cooperate.

$$\Downarrow$$

Punishment: Any deviation from cooperation $\rightarrow$ Payoff decreases

$$\Downarrow$$

Either for all players $l \in \{1, 2, \ldots n\}$, $\pi_l = R_{c,n-1}$,
or for any opponent $j \in \{1, 2, \ldots, n\} \setminus \{i\}, \pi_j < R_{c,n-1}$.

Cooperation Enforcing Strategy

Definition

Definition

$\boldsymbol{p}$ for player i is called cooperation enforcing:
(1) Player i cooperates in the first stage.
(2) $p_{c,n-1} = 1$.
(3) Either for all players $l \in \{1, 2, \ldots n\}$, $\pi_l = R_{c,n-1}$,
or for any opponent $j \in \{1, 2, \ldots, n\} \setminus \{i\}, \pi_j < R_{c,n-1}$.

Cooperation Enforcing Strategy

Definition

Definition

$\boldsymbol{p}$ for player i is called cooperation enforcing:
(1) Player i cooperates in the first stage.
(2) $p_{c,n-1} = 1$.
$\rightarrow$ for stable cooperation
(3) Either for all players $l \in \{1, 2, \dots n\}$, $\pi_l = R_{c,n-1}$,
or for any opponent $j \in \{1, 2, \dots, n\} \setminus \{i\}, \pi_j < R_{c,n-1}$.

Cooperation Enforcing Strategy

Definition

Definition

$\boldsymbol{p}$ for player i is called cooperation enforcing:
(1) Player i cooperates in the first stage.
(2) $p_{c,n-1} = 1$.
(3) Either for all players $l \in \{1, 2, \ldots n\}$, $\pi_l = R_{c,n-1}$,
or for any opponent $j \in \{1, 2, \ldots, n\} \setminus \{i\}, \pi_j < R_{c,n-1}$.
$\rightarrow$ restrict the best response

Cooperation Enforcing Strategy

Definition

Definition

$\boldsymbol{p}$ for player i is called <u>cooperation enforcing</u>:
(1) Player i cooperates in the first stage.
(2) $p_{c,n-1} = 1$.
(3) Either for all players $l \in \{1, 2, \ldots n\}$, $\pi_l = R_{c,n-1}$,
or for any opponent $j \in \{1, 2, \ldots, n\} \setminus \{i\}, \pi_j < R_{c,n-1}$.

Lemma (Property)

If every player i adopts a cooperation enforcing strategy $\boldsymbol{p}_i$, then $(\boldsymbol{p}_1, \boldsymbol{p}_2, \cdots, \boldsymbol{p}_n)$ is a Markov Perfect Equilibrium (MPE).

Cooperation Enforcing Strategy

Definition

Definition

$\boldsymbol{p}$ for player i is called cooperation enforcing:
(1) Player i cooperates in the first stage.
(2) $p_{c,n-1} = 1$.
(3) Either for all players $l \in \{1, 2, \ldots n\}$, $\pi_l = R_{c,n-1}$,
or for any opponent $j \in \{1, 2, \ldots, n\} \setminus \{i\}, \pi_j < R_{c,n-1}$.

Dose this kind of strategies exist?

Main Results

Theorem

In the repeated public goods game with $r > \frac{n}{2}$, if a memory-one strategy $\boldsymbol{p}$ cooperates in the first stage and satisfies the following constraints:

$$\begin{cases} p_{c,n-1} = 1 \\ p_{c,n-2} < 1 \\ p_{d,n-1} < \dfrac{(1-p_{c,n-2})(R_{c,n-1}-R_{c,n-2})}{R_{d,n-1}-R_{c,n-1}} \\ p_{d,n-2} < \dfrac{(1-p_{c,n-2})(R_{c,n-1}-R_{d,n-2})}{R_{d,n-1}-R_{c,n-1}} \\ \dots \\ p_{d,k} < \dfrac{(1-p_{c,n-2})(R_{c,n-1}-R_{d,k})}{R_{d,n-1}-R_{c,n-1}} \\ \dots \\ p_{d,0} < \dfrac{(1-p_{c,n-2})(R_{c,n-1}-R_{d,0})}{R_{d,n-1}-R_{c,n-1}} \end{cases},$$

then $\boldsymbol{p}$ is a cooperation enforcing strategy.

Proof Sketch

Recall condition (3)

> *Either for all players $l \in \{1, 2, \ldots n\}$, $\pi_l = R_{c,n-1}$,*
> *or for any opponent $j \in \{1, 2, \ldots, n\} \setminus \{i\}, \pi_j < R_{c,n-1}$.*

Rewrite it as

$$\begin{aligned} &(\forall j \neq i, \pi_j < R_{c,n-1}) \vee (\forall l,\ \pi_l = R_{c,n-1}) \\ \Leftrightarrow &\neg(\forall j \neq i, \pi_j < R_{c,n-1}) \rightarrow (\forall l,\ \pi_l = R_{c,n-1}) \\ \Leftrightarrow &\exists j \neq i, \pi_j \geq R_{c,n-1} \rightarrow v_{c^n} = 1 \end{aligned}$$

where $v_{c^n} = 1$ means stable cooperation.

Proof Sketch Cont'd

$$\exists j \neq i, \pi_j \geq R_{c,n-1} \rightarrow v_{c^n} = 1.$$

Control pipeline:

$$\boldsymbol{p} \quad \rightarrow \quad \boldsymbol{v} \quad \rightarrow \quad \pi_j$$

Relation between $\boldsymbol{p}$ and $\boldsymbol{v}$?

Proof Sketch Cont'd

$$\exists j \neq i, \pi_j \geq R_{c,n-1} \rightarrow v_{c^n} = 1.$$

Control pipeline:

$$\boldsymbol{p} \quad \rightarrow \quad \boldsymbol{v} \quad \rightarrow \quad \pi_j$$

Relation between $\boldsymbol{p}$ and $\boldsymbol{v}$?

Lemma ([Akin, 2012, Hilbe et al., 2014])

Let $\boldsymbol{p}^R$ denote the Repeat strategy, then

$$(\boldsymbol{p} - \boldsymbol{p}^R) \cdot \boldsymbol{v} = 0.$$

Proof Sketch Cont'd

$$\exists j \neq i, \pi_j \geq R_{c,n-1} \rightarrow v_{c^n} = 1.$$

Control pipeline:

$$\boldsymbol{p} \quad \rightarrow \quad \boldsymbol{v} \quad \rightarrow \quad \pi_j$$

Relation between $\boldsymbol{p}$ and $\boldsymbol{v}$?

Lemma ([Akin, 2012, Hilbe et al., 2014])

Let $\boldsymbol{p}^R$ denote the Repeat strategy, then

$$(\boldsymbol{p} - \boldsymbol{p}^R) \cdot \boldsymbol{v} = 0.$$

Insights

$$(1 - p_{c,n-2})(\pi_j - R_{c,n-1}) \geq 0 \rightarrow v_{c^n} = 1$$
$$\Leftrightarrow (1 - p_{c,n-2})(\boldsymbol{\pi}_j - R_{c,n-1}\mathbf{1}) \cdot \boldsymbol{v} \geq 0 \rightarrow v_{c^n} = 1$$

Case Study

x: fixed cooperation enforcing strategy;
y, z: all memory-one strategies

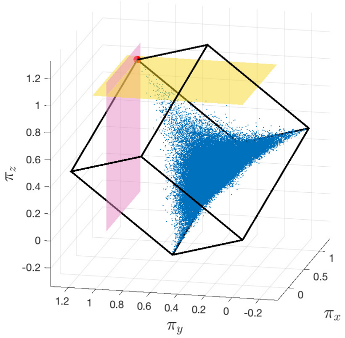


Figure 1: Excepted payoffs. $R_{c,n-1} = 1$.

Collusion Resistance

Multiple opponents wants to deviate?

Or even make collusion?

We prove that as long as our strategy exists,

$$\text{Average Payoff of Alliance} \leq R_{c,n-1}.$$

Theorem

If a cooperation enforcing strategy $\boldsymbol{p}$ exists, then it is collusion resistant.

Against Learning and Collusive Players

Theory

What if a player has no idea of Markov strategies? Learning!

- From her point of view: Repeated Games $\rightarrow$ Markov Decision Process (MDP) (environment: other players' strategies)
- Bellman optimality equation [Mahadevan, 1996]

$$Q^*(\boldsymbol{o}, a) = \max_{a' \in A} \mathbb{E}\left[R_{\boldsymbol{o}'} - R^* + Q^*(\boldsymbol{o}', a')\right].$$

- Learning mechanics: average-reward reinforcement learning algorithm

Against Learning and Collusive Players

Algorithm

Algorithm 1: A Learning Player's Strategy

Initialize a matrix: $Q(\boldsymbol{o}, a) \leftarrow 0$ for all $\boldsymbol{o} \in A^n, a \in A$;
Initialize an estimate of the average payoff $\bar{R} \leftarrow 0$;
Set outcome of the initial stage game $\boldsymbol{o}(0) \leftarrow c^n$;
Set the learning rate parameters α, β;
for $t = 1, 2, \cdots$ **do**
  Take action a with ϵ-greedy policy based on $Q(\boldsymbol{o}(t-1), a)$;
  Receive stage game outcome $\boldsymbol{o}(t)$ and payoff R;
  $\delta \leftarrow R - \bar{R} + \max_{a'} Q(\boldsymbol{o}(t), a') - Q(\boldsymbol{o}(t-1), a)$;
  $Q(\boldsymbol{o}(t-1), a) \leftarrow Q(\boldsymbol{o}(t-1), a) + \alpha\delta$;
  if $Q(\boldsymbol{o}(t-1), a) = \max_{a'} Q(\boldsymbol{o}(t-1), a)$ **then**
    $\bar{R} \leftarrow (1-\beta)\bar{R} + \beta[(t-1)\bar{R} + R]/t$;
  end
end

Against Learning and Collusive Players

Simulation

- Cooperation enforcing vs. Cooperation enforcing vs. Learning
- Cooperation enforcing vs. Learning vs. Learning
- Cooperation enforcing vs. Learning alliance (Stackelberg setting)

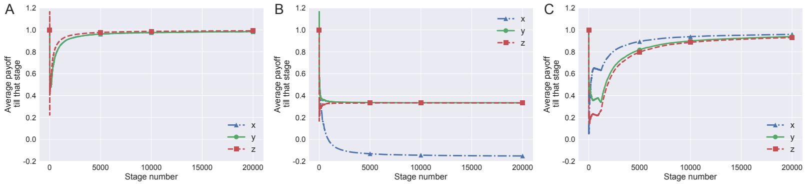


Figure 2: Illustration of average payoffs during learning. $R_{c,n-1} = 1$.

Conclusions and Future Work

Conclusions

- Define cooperation enforcing strategies
- Prove several properties (MPE, collusion resistant, et al.)
- Identify them in repeated public goods games
- Simulate with learning players

Future Work

- The effect of r/n (MPCR) on cooperation
- Generalization in more games (asymmetric, or with imperfect information)
- Larger action space

More Advanced Topics

Continuous Action Space

Theorem. (Autocratic Strategies). *Suppose that $\sigma_X[x,y]$ is a memory-one strategy for player X and let σ_X^0 be player X's initial action. If, for some bounded function, ψ, the equation*

$$\alpha u_X(x,y)+\beta u_Y(x,y)+\gamma=\psi(x)-\lambda\int_{s\in S_X}\psi(s)\,d\sigma_X[x,y](s) - (1-\lambda)\int_{s\in S_X}\psi(s)\,d\sigma_X^0(s) \qquad [4]$$

holds for each $x\in S_X$ and $y\in S_Y$, then σ_X^0 and $\sigma_X[x,y]$ together enforce the linear payoff relationship

$$\alpha\pi_X+\beta\pi_Y+\gamma=0 \qquad [5]$$

for any strategy of player Y. In other words, the pair $(\sigma_X^0,\sigma_X[x,y])$ is an autocratic strategy for player X.

Autocratic strategies for iterated games with arbitrary action spaces. PNAS 2016

References

Akin, E. (2012).
Stable cooperative solutions for the iterated prisoner's dilemma.
arXiv preprint, 1211.0969.

Hilbe, C., Wu, B., Traulsen, A., and Nowak, M. A. (2014).
Cooperation and control in multiplayer social dilemmas.
Proceedings of the National Academy of Sciences, 111(46):16425–16430.

Mahadevan, S. (1996).
Average reward reinforcement learning: Foundations, algorithms, and empirical results.
Machine Learning, 22(1-3):159–195.

Resources

Paper Collection: https://drive.google.com/open?id=1VePfjyh_1FNS5bBce1AnYh8QS_zz33of

Markdown Software Recommendation: Typora