

## 不完全観測繰り返しゲームのチーム生産モデルにおける均衡解析

## Analysis of Equilibria in Repeated Team-production Game with Imperfect Monitoring

重中 風奎 \*<sup>1</sup>      山本 駿 \*<sup>1</sup>      岩崎 敦 \*<sup>2</sup>      関口 格 \*<sup>3</sup>      横尾 真 \*<sup>1</sup>  
Fuuki Shigenaka      Shun Yamamoto      Atsushi Iwasaki      Tadashi Sekiguchi      Makoto Yokoo

\*<sup>1</sup>九州大学 システム情報科学府

Graduate School of Information Science and Electrical Engineering at Kyushu University

\*<sup>2</sup>電気通信大学 大学院情報システム学研究所

Graduate School of Information Systems, University of Electro-Communications

\*<sup>3</sup>京都大学 経済研究所経済戦略研究部門

Strategic Economic Studies Division, Institute of Economic Research, Kyoto University

Repeated games with imperfect monitoring represent long-term relationships, where players cannot observe others' actions exactly. In a team-production game, players obtain diminishing returns for their efforts. When a team-production game is repeated infinitely, players can maximize team's payoff by working in rotation. However, except for special cases, this strategy can constitute an equilibrium only when the working cost is low. In this paper, we identify several alternative strategies that can constitute an equilibrium even when the working cost becomes higher.

## 1. 序論

無限回繰り返しゲーム（以降繰り返しゲーム）はプレイヤー間の長期的な協力関係を考察する枠組みとして知られる。繰り返しゲームにおいては、プレイヤーが他のプレイヤーの行動をどの程度正確に観測できるかによって問題の性質が大きく変化する。他のプレイヤーの行動を直接観測できない代わりに、選択した行動に依存して発生するシグナルを観測する場合は不完全観測（imperfect monitoring）と呼ばれる。特に、全プレイヤーで共通のシグナルを観測する場合を公的観測（public monitoring）と呼ぶ。不完全観測の条件下においては、他のプレイヤーの行動を完全に知ることができないため、プレイヤー間の協調が困難となる。また、各プレイヤーによる他のプレイヤーの行動についての推測も考慮しなければならないため、完全観測に比べ解析することも困難なものとなる。

チーム生産ゲームは、収穫逓減を仮定したゲームである。各プレイヤーは労働/休憩の2通りの行動の選択肢が与えられる。労働はプレイヤー全体のために自らが労働コストを負担し、全プレイヤーで均一に与えられる報酬を受け取る行動を表す。休憩は自らコストを負担することなく、均一に与えられる報酬を受け取るのみの行動を表す。ただし、収穫逓減を仮定しているため、すべてのプレイヤー（ $n$ 人）の中の、ある人数（ $k^*$ 人）だけが労働を行うことですべてのプレイヤーの合計利得は最大化される。しかしながら、労働のコストが大きい場合に利得を最大化する戦略が均衡を構成しないことがあった。

本稿では、公的観測付繰り返しチーム生産ゲームにおいて効率的な利得を与えながら、より大きい労働コストにおいても均衡を構成する具体的な戦略を示す。第2節では、本稿で述べるチーム生産ゲームのモデルと公的観測付繰り返しゲームの説明を行う。第3節では、最大利得を与える既知の戦略を引用し、説明する。第4, 5節は本稿の主張であり、より大きい

労働コストにおいても均衡を構成する具体的な戦略を提示し、計算機実験によるデータの例示を行う。

## 2. モデル

本節では公的観測付繰り返しゲームの解説を行い、考察に用いる均衡概念を述べる。また、チーム生産ゲームを数学的に定義する。

### 2.1 公的観測付繰り返しゲーム

繰り返しゲームは、ステージゲームと呼ばれる同一のゲームを無限回繰り返すこと自体を1つのゲームとみなしたものである。 $t$ 回目のステージゲームをピリオド  $t$  ( $t = 0, 1, 2, \dots$ ) と呼ぶ。プレイヤーの集合を  $N$  とする。各ピリオドで、各プレイヤーは全プレイヤーで共通の行動集合  $A$  から自身が行う行動を選択する。各プレイヤーの選択した行動の組合せにより、ピリオド  $t$ の結果として各プレイヤー  $i \in N$  は全プレイヤーで共通のシグナル  $w^t \in \Omega$  と各プレイヤーごとの利得  $u_i^t$  が与えられる。ただし、 $\Omega$  はシグナルの集合である。

繰り返しゲームの利得について、割引因子  $\delta$  ( $0 < \delta < 1$ ) を導入する。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程度を表すパラメタである。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。

プレイヤー  $i$  のピリオド  $t^*$  における割引平均利得を次の式で定義する。

$$(1 - \delta) \sum_{t=t^*}^{\infty} \delta^{t-t^*} u_i^t \quad (1)$$

これはピリオド  $t^*$  以降で与える利得の平均値を表したものである\*<sup>1</sup>。各プレイヤーは各ピリオドにおいて、自身の割引平均利得を最大化するための行動を行う。以降、単に利得と述べた場合、割引平均利得を指す。

連絡先: 重中風奎, 九州大学大学院システム情報科学府, 812-0395 福岡県福岡市西区元岡 744 番地, (092)802-3576, shigenaka@agent.inf.kyushu-u.ac.jp

\*<sup>1</sup> プレイヤー  $i$  がピリオド  $t^*$  以降に与えられる利得の平均値を  $g_i$  とすると、ピリオド  $t^*$  以降に与えられる利得の合計について  $\sum_{t=t^*}^{\infty} \delta^{t-t^*} g_i = \sum_{t=t^*}^{\infty} \delta^{t-t^*} u_i^t$  が成立することから (1) が導かれる。

公的ヒストリ  $h^t \in \Omega^t$  をピリオド  $t$  までの各ピリオドの観測の結果を記述したものとす。  $H$  を起こりうるすべての公的ヒストリの集合とする。 プレイヤ  $i \in N$  の私的ヒストリ  $h_i^t \in \Omega^t \times A^t$  をピリオド  $t$  までの各ピリオドにおける観測とプレイヤ  $i$  が選択した行動の結果を記述したものとす。  $H_i$  をプレイヤ  $i$  の起こりうるすべての私的ヒストリの集合とする。

各プレイヤが各ピリオドでどの行動を行うかを記述したものを戦略と呼ぶ。 公的戦略は  $\sigma: H \rightarrow A^n$  であり、公的ヒストリ  $h^t$  からピリオド  $t$  における各プレイヤ  $i \in N$  の行動  $a_i^t \in A$  を指定する関数である。 プレイヤ  $i$  の私的戦略は  $\sigma_i: H_i \rightarrow A$  であり、私的ヒストリ  $h_i^t$  からピリオド  $t$  における自身の行動  $a_i^t \in A$  を指定する関数である。

均衡概念として完全公的均衡 (perfect public equilibrium) を考える。 これは、任意のピリオド  $t$  におけるすべての公的ヒストリ  $h^t \in \Omega^t$  について公的戦略  $\sigma$  がピリオド  $t$  以降ナッシュ均衡を構成することを表す。 すなわち、全プレイヤが完全公的均衡を構成する公的戦略に従っているとき、あるプレイヤ 1 人がその戦略の代わりにどのような私的戦略を行っても、そのプレイヤの利得は増大しない。

プレイヤ  $i \in N$  の公的戦略  $\sigma$  に対する一回逸脱戦略  $\sigma_i^{dev}$  を、プレイヤ  $i$  が特定の私的ヒストリ  $h_i^t$  においてのみ  $\sigma$  と異なる行動を選択し、その他のピリオドは  $\sigma$  に従う戦略とする。 プレイヤ  $i$  の  $\sigma$  に対する一回逸脱戦略の集合を  $Q_i(\sigma)$  とす。 完全公的均衡については一回逸脱の原理 (one-shot deviation principle) が成立することが知られている [Mailath 06]。 これは、公的戦略  $\sigma$  が完全公的均衡を構成するのは、すべてのプレイヤ  $i \in N$  について、  $\sigma$  に対するすべての一回逸脱戦略  $\sigma_i^{dev} \in Q_i(\sigma)$  はプレイヤ  $i$  の利得を増加しない場合、かつその場合に限るというものである。

## 2.2 公的観測付繰り返しチーム生産ゲーム

本稿では、公的観測付繰り返しチーム生産ゲームについて述べる。 このゲームは文献 [Kobayashi 14] に準ずるものである。 ステージゲームを次のように与える。 行動の集合を  $A = \{W, S\}$  とす。  $W/S$  は労働/休憩を表し、  $W$  を行うプレイヤは労働コスト  $c$  (定数) を負担する。 各ステージゲームの利得は生産の結果による。 生産の結果はシグナルとして観測される。  $\Omega = \{g, b\}$  とす。  $g$  (good) は生産の成功を表し、このとき各プレイヤは報酬  $x$  (定数) を獲得する。  $b$  (bad) は生産の失敗を表し、このとき報酬は発生しない。  $k$  人が行動  $W$  を行った場合に生産が成功する確率を  $\pi_k$  とす ( $k = 0, 1, \dots, n$ )。  $\pi_k$  は  $k$  について単調増加し、  $W$  を行う人数が多いほど生産成功確率が増加する。 各ピリオドにおける行動の組合せ  $a$  によるプレイヤ  $i \in N$  の期待利得を  $v_i(a)$  とす。 次の 2 つの仮定をおく。

仮定 1 ある整数  $k^* (0 < k^* < n)$  について次の式が成立する。

$$\pi_{k^*} n x - k^* c = \max_k \{\pi_k n x - k c\}$$

仮定 2 すべての  $k$  について次の式が成立する。

$$(\pi_{k+1} - \pi_k) x < c$$

仮定 1 は全プレイヤの合計利得が最大化されるとき  $W$  を行う人数を規定するものである。 このときの利得  $\pi_{k^*} n x - k^* c$  を社会的最適 (social optimum) な利得と呼ぶ。 仮定 2 は各プレイヤは他のプレイヤの行動に関わらず  $S$  を選択することで  $v_i$  を増大することができるという仮定である。 すなわち、1 度きりのステージゲームの利得を考えた場合、すべてのプレイヤは  $S$  を行う。

表 1: 公的観測付チーム生産の利得行列 ( $n = 2$ )

	2	W	S
1	W	$(\pi_2 x - c, \pi_2 x - c)$	$(\pi_1 x - c, \pi_1 x)$
	S	$(\pi_1 x, \pi_1 x - c)$	$(\pi_0 x, \pi_0 x)$

$n = 2$  のとき、利得表は表 1 で与えられる。  $k^* = 1$  のとき、社会的最適な利得を与えるのは  $a = (W, S), (S, W)$  である。

## 2.3 戦略 FSA

本稿では、公的観測付繰り返しチーム生産ゲームの戦略を有限オートマトン (FSA, finite state automaton) により表現することを考える。 戦略 FSA を  $\langle \Theta, \hat{\theta}, f, T \rangle$  により定義する。  $\Theta$  は状態の集合を表し、  $\Theta = \{\theta_1, \theta_2, \dots, \theta_s\}$  である。 ここで、  $s$  は FSA の状態数を表す。  $\hat{\theta} \in \Theta$  は初期状態を表す。  $f: \Theta \rightarrow A^n$  は各状態において各プレイヤが選択する行動を表す。  $T: \Theta \times \Omega \rightarrow \Theta$  は各状態におけるシグナルによる決定的な遷移を表す。 すなわち、戦略 FSA は各ピリオドにおける各プレイヤの行動を表す。 ここで、  $f(\theta)$  は状態  $\theta \in \Theta$  における各プレイヤの行動の組合せを表し、  $T(\theta, w) \in \Theta$  は状態  $\theta \in \Theta$  において観測が  $w \in \Omega$  であった場合の次のピリオドの状態を表すものとする。

戦略を FSA で記述することにより、各ピリオドにおける利得はそのピリオドの状態  $\theta \in \Theta$  のみによって決定する。 状態  $\theta^* \in \Theta$  についてのプレイヤ  $i \in N$  の利得  $V_{\theta^*}^i$  は次の式で表される。

$$V_{\theta^*}^i = (1 - \delta) v_i(f(\theta^*)) + \delta \sum_{\theta \in \Theta} p(\theta|\theta^*) V_{\theta}^i$$

ただし、  $p(\theta|\theta^*)$  は状態  $\theta^*$  から  $\theta$  に遷移する確率を表す。 また、プレイヤ  $i$  の繰り返しゲーム全体での利得は  $V_{\theta^*}^i$  である。

公的戦略を戦略 FSA で定めた場合の一回逸脱戦略について考える。 一回逸脱戦略の利得も、逸脱行動 (公的戦略によって決められた行動と異なる行動) を行うピリオドにおける公的戦略 FSA の状態のみによって決定する。 したがって、各状態に対する一回逸脱戦略によって利得の増加ができないとき、すべての一回逸脱戦略は利得を増加しない。 公的戦略 FSA の状態  $\theta^*$  に対するプレイヤ 1 の一回逸脱戦略の利得は、次の条件を満たす  $\theta_{dev}^*$  を元の公的戦略 FSA に追加したときの状態  $\theta_{dev}^*$  の利得として表せる。

- $f_{-i}(\theta_{dev}^*) = f_{-i}(\theta^*)$  かつ  $f_i(\theta_{dev}^*) \neq f_i(\theta^*)$
- $T(\theta_{dev}^*, w) = T(\theta^*, w)$  ただし、  $w \in \Omega$

ただし、  $f(\theta)$  について  $f_i$  はプレイヤ  $i \in N$  の行動を表し、  $f_{-i}$  は  $i$  を除くすべてのプレイヤの行動の組合せを示す。

## 3. キャタピラ戦略

本節では、文献 [Kobayashi 14] において示されているキャタピラ戦略について取り上げる。 キャタピラ戦略は以下の FSA で記述される戦略である。

- $\Theta = \{R_k (k = 1, 2, \dots, n)\}$
- $\hat{\theta} = R_1$
- $f_i(R_k) = \begin{cases} W & (k \leq i + \alpha < k + k^*, \alpha \in \{0, n\}) \\ S & (\text{otherwise}) \end{cases}$

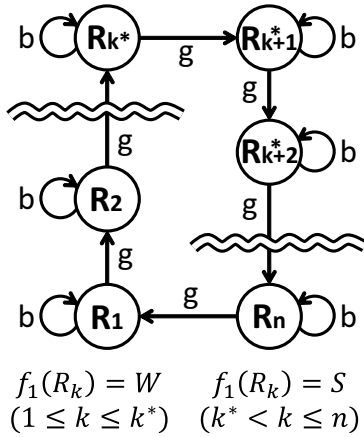


図 1: キャタピラ戦略 (行動はプレイヤー 1 のみ表記)

$$\bullet T(R_k, w) = \begin{cases} R_{k+1} & (k \neq n, w = g) \\ R_1 & (k = n, w = g) \\ R_k & (w = b) \end{cases}$$

キャタピラ戦略は常に  $k^*$  人のプレイヤーが  $W$  を行うため、社会的最適な利得を与える。また、同文献には次の 2 つの命題が示されている。

命題 1 (i) キャタピラ戦略は、以下の式が満たされるとき、かつそのときに限り均衡を構成する割引因子が存在する。

$$c - (\pi_{k^*} - \pi_{k^*-1})x < \frac{\pi_{k^*} - \pi_{k^*-1}}{\pi_{k^*}} \cdot \frac{n - k^*}{n} c \quad (2)$$

(ii)(2) が満たされないとき、社会的最適な利得を与える均衡を構成する戦略は存在しない。

命題 2  $\delta$  を除くすべてのパラメタを固定する。任意の  $\varepsilon > 0$  に対し、 $\delta > \underline{\delta}$  に対してチーム全体の利得が  $\pi_{k^*} n x - k^* c - \varepsilon$  となる均衡戦略を与える  $\underline{\delta}$  が存在する。

(2) が満たされない範囲において、大きい利得を得る具体的な戦略は未知であった。そこで、本稿ではより広い範囲のパラメタで均衡を構成し、利得の大きい具体的な戦略の提示を行う。

#### 4. 1MP-キャタピラ戦略

本節では、新しく発見された戦略である 1MP (1-mutual punishment) -キャタピラ戦略について述べる。1MP-キャタピラ戦略は以下の FSA で記述される戦略である。

- $\Theta = \{R_k, P_k (k = 1, 2, \dots, n)\}$
- $\hat{\theta} = R_1$
- $f: \begin{cases} f_i(R_k) = \begin{cases} W & (k \leq i + \alpha < k + k^*, \alpha \in \{0, n\}) \\ S & (\text{otherwise}) \end{cases} \\ f_i(P_k) = S \end{cases}$
- $T: \begin{cases} T(R_k, w) = \begin{cases} R_{k+1} & (w = g, k \neq n) \\ R_1 & (w = g, k = n) \\ P_k & (w = b) \end{cases} \\ T(P_k, w) = \begin{cases} P_k & (w = g) \\ R_k & (w = b) \end{cases} \end{cases}$

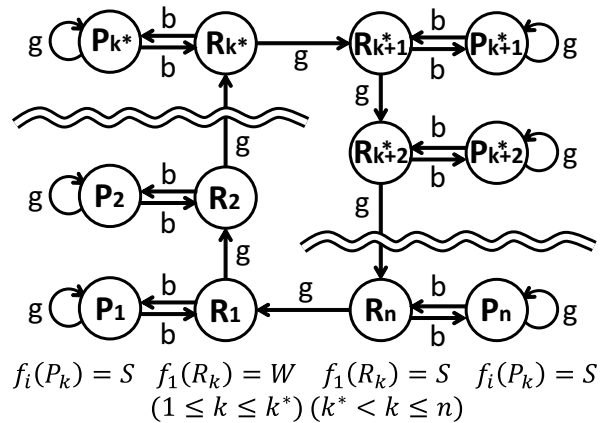


図 2: 1MP-キャタピラ戦略 (行動はプレイヤー 1 のみ表記)

生産が成功している間、各プレイヤーは社会的効率な割り当てられた行動を行う協力状態であるが、生産が失敗したときに両プレイヤーとも  $S$  を行う相互罰則状態になる。

定理 1 (3) が満たされるとき、1MP-キャタピラ戦略は均衡を構成する割引因子が存在する。

$$c - (\pi_{k^*} - \pi_{k^*-1})x < \frac{\pi_{k^*} - \pi_{k^*-1}}{\pi_{k^*}} \left\{ \frac{\pi_{k^*}(\pi_{k^*} - \pi_0)}{2 - (\pi_0 + \pi_{k^*})} x + \left\{ \frac{n - k^*}{n} - \frac{\pi_{k^*}}{2 - (\pi_0 + \pi_{k^*})} \frac{k^*}{n} \right\} c \right\} \quad (3)$$

証明 各状態について一回逸脱戦略との利得の比較を行う。仮定 1, 2 より、次の式が成立するとき一回逸脱戦略は利得を増大しないことが計算により導出される。

$$c - (\pi_{k^*} - \pi_{k^*-1})x < \frac{\pi_{k^*} - \pi_{k^*-1}}{\pi_{k^*}} \left\{ \frac{\delta \pi_{k^*}(\pi_{k^*} - \pi_0)}{1 + \delta(1 - \pi_0 - \pi_{k^*})} x + \left\{ \frac{1 - \beta^{n-k^*}}{1 - \beta^n} - \frac{\delta \pi_{k^*}}{1 + \delta(1 - \pi_0 - \pi_{k^*})} \frac{\beta^{n-k^*} - \beta^n}{1 - \beta^n} \right\} c \right\}$$

ただし、 $\beta = \frac{1}{\delta \pi_{k^*}} \left\{ 1 - \delta^2 \frac{(1 - \pi_0)(1 - \pi_{k^*})}{1 - \delta \pi_0} \right\}$  である。 $\delta \rightarrow 1$  とすると (3) を得る。 □

ここで、(2) と (3) の比較により次の系を得る。

系 1 均衡を構成する割引因子が存在するパラメタの値の範囲  $(x, c, \pi_k)$  の範囲) について、1MP-キャタピラ戦略はキャタピラ戦略を内包する。

次に、キャタピラ戦略、1MP-キャタピラ戦略の均衡を構成するパラメタの範囲と利得について計算機実験の結果を示し、各戦略の特徴を見る。

例 1 図 3 は  $n = 2, \delta = 0.99, \pi_0 = 0.3, \pi_1 = 0.7, \pi_2 = 0.9, x = 100$  としたときの割引平均利得についてのグラフである。横軸は労働コスト  $c$  であり、仮定 1, 2 の範囲全域を表す。縦軸は利得の値である。グラフは均衡を構成する場合に、その

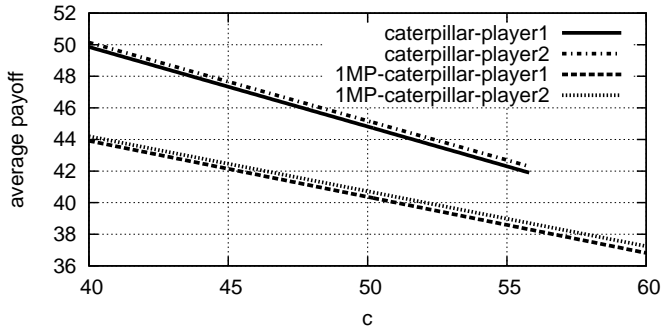


図 3: 割引平均利得 ( $n = 2, \delta = 0.99, \pi_0 = 0.3, \pi_1 = 0.7, \pi_2 = 0.9, x = 100$ )

戦略における各プレイヤーの割引平均利得を表す。カタピラ戦略は社会的最適な利得を与え、 $c \approx 55.84$  程度までの範囲で均衡を構成する。1MP-カタピラ戦略は仮定を満たす  $c$  の範囲全域で均衡を構成するが、利得はカタピラ戦略に劣る。

### 5. 1MP-カタピラ戦略の拡張性

1MP-カタピラ戦略は均衡を構成するパラメタの範囲を広げるが、両プレイヤーの利得は減少する。しかしながら、1MP-カタピラ戦略を拡張することで、利得を改善することができる。利得の改善は、 $k^*$  人だけが  $W$  を行うピリオド数を 1MP-カタピラ戦略よりも増加することで行う。そのため、改善戦略はカタピラ戦略と 1MP-カタピラ戦略の間の性質を持つ。

改善戦略の例として、下山戦略 (図 4) と登山戦略 (図 5) を挙げる。下山戦略は  $b$  を観測した場合に最初の  $m$  回だけ相互罰則状態への移行を防いだものであり、次の状態集合  $\Theta'$  を 1MP-カタピラ戦略に追加した FSA で記述される。

- $\Theta' = \{R_k^l (k = 1, 2, \dots, n, l = 1, 2, \dots, m)\}$

- $f_i(R_k^l) = f_i(R_k)$

- $T : \begin{cases} T(R_k^l, g) = \begin{cases} R_{k+1}^l & (k \neq n) \\ R_1^l & (k = n) \end{cases} \\ T(R_k^l, b) = \begin{cases} R_k^{l-1} & (l \neq 1) \\ R_k & (l = 1) \end{cases} \end{cases}$

ただし、初期状態  $\hat{\theta} = R_1^m$  とする。

登山戦略は下山戦略を元にして上の段へと戻る遷移を追加したものであり、下山戦略の  $\Theta'$  の遷移を表す関数  $T(R_k^l, g)$  を次のように置き換えた FSA により記述される。

- $T(R_k^l, g) = \begin{cases} R_{k+1}^n & (l = m, k \neq n) \\ R_1^n & (l = m, k = n) \\ R_{k+1}^{l+1} & (l \neq m, k \neq n) \\ R_1^{l+1} & (l \neq m, k = n) \end{cases}$

例 2 図 6 は例 1 (図 3) と同じパラメタにおいて、各プレイヤーの利得の平均値を各戦略ごとに表したものである。下山戦略は  $m = 5$ 、登山戦略は  $m = 1$  の場合について表す。このとき、下山戦略は 1MP-カタピラ戦略と同じく仮定を満たす  $c$  の範囲全域で均衡を構成するが、より大きい利得を与える。また、登山戦略は 1MP-カタピラ戦略よりも大きい利得を与え、カタピラ戦略よりも広い範囲で均衡を構成する。

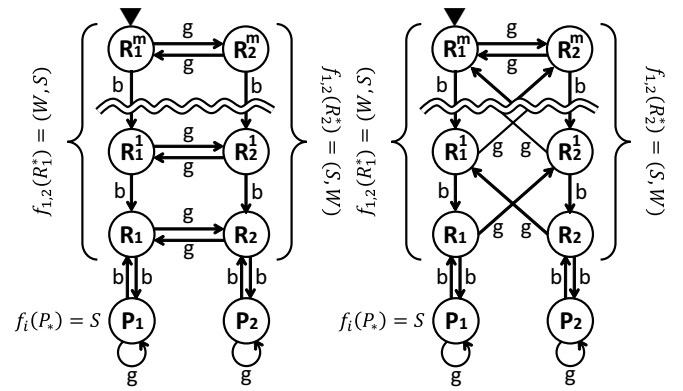


図 4: 下山戦略 ( $n = 2$ )

図 5: 登山戦略 ( $n = 2$ )

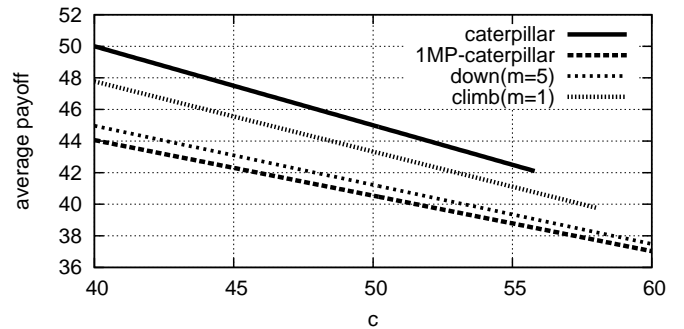


図 6: 各プレイヤーの割引平均利得の平均値 ( $n = 2, \delta = 0.99, \pi_0 = 0.3, \pi_1 = 0.7, \pi_2 = 0.9, x = 100$ )

### 6. 結論

本稿では無限回繰り返しチーム生産ゲームについて、両プレイヤーともが休憩  $S$  を行う状態を追加することにより均衡を構成するパラメタを広げる戦略として、カタピラ戦略の拡張である 1MP-カタピラ戦略を示した。さらに、1MP-カタピラ戦略を拡張することで利得の改善を行うことができることも示した。今後の課題として、命題 2 で示されるような近似的に社会的最適な利得を獲得する戦略の提示が挙げられる。

### 謝辞

本研究は JSPS 基盤研究 (S) (課題番号 24220003) の助成を受けました。ここに深く感謝いたします。

### 参考文献

[山本 14] 山本駿, 岩崎敦, 趙登吉, 横尾真: 不完全私的観測付き繰り返しゲームにおける均衡発見プログラム. 人工知能学会全国大会論文集 28, 1-4 (2014)

[Kobayashi 14] Hajime Kobayashi, Katsunori Ohta, and Tadashi Sekiguchi.: Repeated partnerships with decreasing returns. Public Economics Seminar, Keio University (2014).

[Mailath 06] George J Mailath and Larry Samuelson.: Repeated games and reputations: longrun relationships. OUP Catalogue (2006).