Vol.2017-AL-161 No.2
2017/1/17

マルコフ決定問題に対する多面体論的な最適解の存在証明

清藤 駿成[1]　コルマン マティアス[1]　小池 敦[1]　塩浦 昭義[2]　徳山 豪[1]

概要：本論文では, マルコフ決定問題を定式化した一般化線形相補性問題に対する二重描写法ベースアルゴリズムの挙動を解析する. また, 問題が構成する多面体の特徴から, 二重描写法ベースアルゴリズムのエッセンスを抽出する.

1. はじめに

意思決定では決定をした直後の利益を考えるだけでは不十分であり, その後の状況の変化を予想して将来を見据えた全体的に良い決定が必要である. また, 自身とは異なる目的をもつような意思決定者が関わる場合には, 彼らがどのような決定を行なうかを考慮する必要も出てくる. 以上のような, 動的で, 他者の意思決定が関与してくるような状況をモデル化したものが確率ゲーム (stochastic game)[1] である. 確率ゲームでは, 意思決定を行なう状況, その状況における行動の選択肢, その行動を選択した直後に得られる報酬とどのような状況に遷移するかを示す遷移確率が与えられていることを仮定し, それぞれのプレイヤーは将来得られる報酬も考慮して, 自身の利得を最大にするように戦略を決める.

筆者らは最近, N 人確率ゲームの特殊ケースである N 人完全情報確率ゲームを一般化線形相補性問題として定式化し, 二重描写法ベースアルゴリズムによってナッシュ均衡解が計算できることを明らかにした [2]. しかしながら, ナイーブなアルゴリズムでは計算時間の面で非常に効率が悪く, 二重描写法ベースアルゴリズムのエッセンスを抽出したアルゴリズムの設計が課題となっている.

本研究では, N 人完全情報確率ゲームのプレイヤーが 1 人の特殊ケースであるマルコフ決定問題を定式化した一般化線形相補性問題に対する二重描写法ベースアルゴリズムの挙動を解析し, N 人ゲームに対する解析の基盤を構築することを目的とする. また, アルゴリズムの出力として唯一の解が得られることから, 唯一の最適解が存在することの別証明を与えることができた. 本手法は 2 人ゼロ和完全情報確率ゲームにも容易に拡張することができ, 唯一のナッシュ均衡の存在を示すこともできる. そして, それらの問題が構成する多面体の構造から, 二重描写法ベースアルゴリズムのエッセンスが各頂点における独立戦略を列挙することだということを明らかにした.

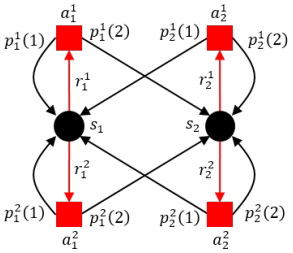

図 1 マルコフ決定問題の例. 状況を表す頂点 (黒丸) から選択できる行動 (赤四角) へは赤矢印がつながっており, 矢印上の値はその行動を選択した場合に得られる報酬である. 報酬を与えられた後には黒矢印でつながった頂点に遷移するが, 矢印上の値 (遷移確率) に従って頂点が選ばれる.

2. マルコフ決定問題

マルコフ決定問題は確率的な状況の変化をモデル化したグラフ上 (図 1) で割引総利得を最大化する問題である. グラフの頂点の集合を $\mathcal{S} = \{1, \cdots, n\}$ とする ($|\mathcal{S}| = n$). 各頂点 s にはプレイヤーが選択できる行動の集合 $\mathcal{A}_s = \{1, \cdots, m_s\}$ が与えられており, プレイヤーはその中から行動を決定する. 行動の決め方の様式として, 過去の頂点の履歴に依存せず, 現在の頂点のみを参考に決定する定常戦略, また複数の行動を確率的に選択するのではなく, 唯一つの行動のみを選択する純粋戦略を仮定する. 行動 a を選択した直後, プレイヤーに報酬 $r_s^a \geq 0$ が与えられ, 遷移確率 $p_s^a(s')$ に従って次の頂点 s' に遷移する. 各頂点でプレイヤーが実際に選択する行動の組を戦略 $\pi = \{\pi_1, \cdots, \pi_n\}$ と呼び, 特に頂点 s での行動を π_s とする. すべての戦略の集合は $\Pi = \prod_{s \in \mathcal{S}} \mathcal{A}_s$ で与えられる. 戦略 π が与えられたとき, 各頂点で与えられる報酬と遷移確率が定まり, 行動 π_s に対して $r_s^\pi, p_s^\pi(s')$ とする. このステップを無限回繰り返し, プレ

[1] 東北大学大学院情報科学研究科
[2] 東京工業大学工学院経営工学系

イヤーは各時刻で与えられた報酬に割引率 $\beta \in [0,1)$ を乗じたものの合計である割引総利得の最大化を目的とする．頂点 s から開始し，戦略 π を用いた場合の割引総利得を v_s^π とすると，以下のように与えられる．ただし，x_s^t は時刻 t に頂点 s にある確率分布である．

$$\begin{aligned} v_s^\pi &= \sum_{s' \in \mathcal{S}} x_{s'}^0 r_{s'}^\pi + \beta \sum_{s' \in \mathcal{S}} x_{s'}^1 r_{s'}^\pi + \cdots + \beta^t \sum_{s' \in \mathcal{S}} x_{s'}^t r_{s'}^\pi + \cdots \\ &= r_s^\pi + \beta \left(\sum_{s' \in \mathcal{S}} x_{s'}^1 r_{s'}^\pi + \cdots + \beta^{t-1} \sum_{s' \in \mathcal{S}} x_{s'}^t r_{s'}^\pi + \cdots \right) \end{aligned} \tag{1}$$

式 (1) の第 2 項以降の和は，時刻 0 で遷移した各頂点 s' から開始した場合の利得 $v_{s'}$ の重み付き平均であるから，利得 v_s^π は以下のように閉じた形で表すことができる．

$$v_s^\pi = r_s^\pi + \beta \sum_{s' \in \mathcal{S}} p_s^\pi(s') v_{s'} \tag{2}$$

各頂点での利得を最大化するような戦略 π^* とそのときの利得 $v_s^{\pi^*}$ がマルコフ決定問題の最適解であり，以下のように定義される．

定義 2.1. 戦略 π^* が以下の条件を満たすとき，それをマルコフ決定問題の最適解とする．

$$\begin{aligned} &\forall \pi \in \Pi, \forall s \in \mathcal{S} \\ &v_s^{\pi^*} \geq v_s^\pi \end{aligned} \tag{3}$$

また，式 (2) より以下の定理が得られる．

定理 2.2. 戦略 π が与えられたとき，それがマルコフ決定問題の最適解であるための必要十分条件は以下を満たすことである．

$$\forall s \in \mathcal{S}, \forall a \in \mathcal{A}_s$$

$$v_s^\pi = r_s^\pi + \beta \sum_{s' \in \mathcal{S}} p_s^\pi(s') v_{s'}^\pi \tag{4}$$

$$\geq r_s^a + \beta \sum_{s' \in \mathcal{S}} p_s^a(s') v_{s'}^\pi \tag{5}$$

マルコフ決定問題には唯一の最適解が存在することが知られている [1].

3. 一般化線形相補性問題

線形相補性問題は与えられた線形の等式系，相補性条件，非負条件を満たすようなベクトルを探す問題である．本研究ではその中でもより一般化された問題である一般化線形相補性問題を扱う [3]．一般化線形相補性問題は，行列 $M \in \mathbb{R}^{m \times n}$，ベクトル $\boldsymbol{q} \in \mathbb{R}^n$，$K$ 個のインデックス部分集合 $\mathcal{B}_k$ に対して，以下の条件を満たすベクトル $\boldsymbol{x} \in \mathbb{R}^n$ を探す問題である．

$$M\boldsymbol{x} = \boldsymbol{q} \tag{6}$$

$$\prod_{i \in \mathcal{B}_k} x_i = 0 \tag{7}$$

$$\boldsymbol{x} \geq \boldsymbol{0} \tag{8}$$

式 (7) が相補性条件であり，インデックス部分集合 $\mathcal{B}_k$ に含まれる変数 x_i の中に少なくとも 1 つ，0 が含まれていなければいけないことを意味している．また，人工変数 $\rho \in \mathbb{R}$ を用いて，

$$\begin{pmatrix} M & -\boldsymbol{q} \end{pmatrix} \begin{pmatrix} \boldsymbol{x} \\ \rho \end{pmatrix} = \boldsymbol{0} \tag{9}$$

$$\prod_{i \in \mathcal{B}_k} x_i = 0 \tag{10}$$

$$\boldsymbol{x} \geq \boldsymbol{0}, \rho \geq 0 \tag{11}$$

のように，均一化した問題を考えることもある．この場合，得られた解に対して $\rho = 1$ とした解が元々の相補性問題の解である．このように均一化した問題を考えることにより，等式系と非負条件によって定義される多面体が錐となる．

4. 二重描写法ベースアルゴリズム

一般化線形相補性問題は二重描写法ベースアルゴリズム (double description method based algorithm) によって，そのすべての解を列挙，または解が存在しないことを示すことができる [3]．二重描写法は元々不等式表現された多面体の端点を列挙するアルゴリズムである．二重描写法ベースアルゴリズムでは，相補性問題の非負条件からなる多面体的錐を初期多面体とし，相補性問題のそれぞれの等式を満たす超平面を 1 つずつ追加し，その交差からなる多面体的錐の辺を逐次的に更新する．ナイーブなアルゴリズムでは，辺の更新の際に超平面の正領域にある辺と負領域にある辺のすべてのペアに対して隣接かどうかを判定し，隣接の場合にはそれらの内分点かつ超平面上に新しい辺を形成する．ただし，2 つの辺の結合によって相補性条件を満たさない辺が作られる場合は削除する．2 つの端線の結合によってつくられる辺が相補性条件を満たすとき，2 つの辺は交差相補性を満たすと呼ぶ．すべての超平面を追加し終えた時点での多面体的錐を構成するベクトルの集合が一般化相補性問題の解であり，解が存在しない場合には空集合となる．二重描写法ベースアルゴリズムの挙動の概図を図 2 に，擬似コードを Algorithm 1 に示す．

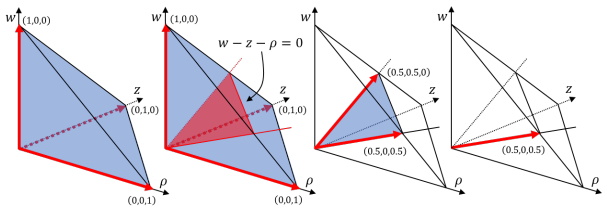

図 2 二重描写法ベースアルゴリズムの挙動の概図．満たすべき条件は $w - z - \rho = 0, wz = 0, w, z, \rho \geq 0$.

Algorithm 1 DDM Based Algorithm

```
R ← すべての次元の基本ベクトルの集合
for all 超平面 do
  R+, R−, R0 ← φ
  for all er ∈ R do
    if er が超平面の正領域 then
      R+ ← R ∪ {er}
    else if er が超平面の負領域 then
      R− ← R ∪ {er}
    else if er が超平面上 then
      R0 ← R ∪ {er}
    end if
  end for
  R′ ← R0
  for all (er1, er2) ∈ R+ × R− do
    if er1, er1 が隣接かつ交差相補性 then
      er1, er1 の内分点かつ超平面上に新しい辺 ern を形成
      R′ ← R′ ∪ {ern}
    end if
  end for
  R ← R′
end for
return R
```

5. マルコフ決定問題に対する一般化線形相補性問題

定理 2.2 より, 戦略 π と利得 v_s^π が最適解であるための必要十分条件は以下のように与えられていた.

$$v_s^\pi = r_s^\pi + \beta \sum_{s' \in \mathcal{S}} p_s^\pi(s') v_{s'}^\pi \tag{12}$$

$$\geq r_s^a + \beta \sum_{s' \in \mathcal{S}} p_s^a(s') v_{s'}^\pi \tag{13}$$

ここで, 不等式 (13) を等式に変換するために, 右辺に非負のスラック変数 w_s^a を加える.

$$v_s^\pi = r_s^a + \beta \sum_{s' \in \mathcal{S}} p_s^a(s') v_{s'}^\pi + w_s^a \tag{14}$$

$$w_s^a \geq 0 \tag{15}$$

スラック変数 w_s^a は戦略 π における利得 v_s^π と行動 a を選択した場合の利得 $r_s^a + \beta \sum_{s' \in \mathcal{S}} p_s^a(s') v_{s'}^\pi$ の差を意味している. したがって, 行動 π_s に対するスラック変数 w_s^π の値は 0 となり, 頂点 s のスラック変数 w_s^a は相補性条件を満たしている.

$$\prod_{a \in \mathcal{A}_s} w_s^a = 0 \tag{16}$$

以上より, マルコフ決定問題の最適解は一般化線形相補性問題として以下のように定式化できる. なお, 各報酬の値は非負であることを仮定しているため, 各頂点の利得 v_s も非負である.

定理 5.1. マルコフ決定問題の最適解は以下の一般化線形相補性問題の解と同値であり, $w_s^a = 0$ となる行動 a が頂点 s の最適行動である.

$$\forall s \in \mathcal{S}, \forall a \in \mathcal{A}_s$$

$$v_s = r_s^a \rho + \beta \sum_{s' \in \mathcal{S}} p_s^a(s') v_{s'} + w_s^a \tag{17}$$

$$\prod_{a \in \mathcal{A}_s} w_s^a = 0 \tag{18}$$

$$w_s^a, v_s, \rho \geq 0 \tag{19}$$

6. 二重描写法ベースアルゴリズムの挙動解析

マルコフ決定問題を定式化した一般化線形相補性問題を二重描写法ベースアルゴリズムで解く際の挙動を解析する. アルゴリズムの出力として唯一のベクトルが出力されることを示し, マルコフ決定問題に唯一の最適解が存在することの別証明を与える.

6.1 初期辺集合

アルゴリズムで扱う辺 $\boldsymbol{er}$ は各 w_s^a, v_s, ρ 成分からなる $nm+n+1$ 次元のベクトルである. そのベクトルのインデックス集合を $\mathcal{I}$ とし, $\boldsymbol{w}, \boldsymbol{v}$ 成分に対応しているインデックス集合をそれぞれ $\mathcal{I}_w, \mathcal{I}_v$ とする.

初期多面体は相補性問題の非負条件 (式 (19)) からなる多面体的錐である. これを構成する辺の集合は各次元の成分 i に対する基本ベクトル $\boldsymbol{e}_i$ の集合である. ただし, $\boldsymbol{e}_i$ は i 成分が 1 でそれ以外の成分が 0 のベクトルである. したがって, 初期多面体は $nm+n+1$ 個の辺によって構成されている.

すべての辺の集合を $\mathcal{R}$ とし, 以下のように辺部分集合を定義する.

定義 6.1. 辺集合 $\mathcal{R}$ に対して, 以下の辺部分集合を定義する.

- $\mathcal{R}_w$: $\boldsymbol{w}$ 成分のみが正である辺の集合
- $\mathcal{R}_v$: 少なくとも 1 つの $\boldsymbol{v}$ 成分が正で, かつ ρ 成分が 0 である辺の集合
- $\mathcal{R}_{v\rho}$: 辺集合 $\mathcal{R}$ から辺集合 $\mathcal{R}_w$ を除いた辺集合 $(\mathcal{R} \setminus \mathcal{R}_w)$

初期辺集合において, 辺集合 $\mathcal{R}_{v\rho}$ は $n+1$ 個の要素をもつ.

6.2 $v\rho$ 空間と多面体

本解析では $\boldsymbol{w}, \boldsymbol{v}, \rho$ 成分からなる $nm+n+1$ 次元の空間のうち, $\boldsymbol{v}, \rho$ 成分のみからなる $n+1$ 次元の空間 ($v\rho$ 空間) を考える. また, 辺集合 $\mathcal{R}_{v\rho}$ の辺のみを可視化し, 辺集合 $\mathcal{R}_w$ の辺は可視化しない. なぜなら, 本問題では各ステップで辺集合 $\mathcal{R}_w$ の辺 $\boldsymbol{e}_{w_s^a}$ の符号が 1 つずつ正となって辺集合 $\mathcal{R}_{v\rho}$ の符号が負の辺と結合し, その w_s^a 成分を加えていくように解釈することができるからである (後述).

さらに, 辺集合 $\mathcal{R}_{v\rho}$ から構成される多面体的錐のうち, $\boldsymbol{v}, \rho$ 成分の和が 1(一定) となる超平面で切断したファセッ

トの多面体 $\mathcal{G}$ を扱い, 今後は辺を端点と呼ぶ. 初期多面体では $n+1$ 個のすべての端点同士が隣接である単体であり, 次元は n である (図 3).

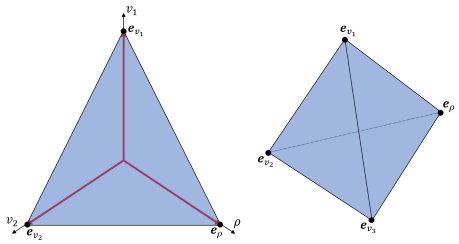

図 3 初期端点集合から構成される多面体. $n=2,3$ の場合を示す. $n=2$ では $v\rho$ 空間の軸も示すが, $n=3$ では多面体のみを可視化する.

6.3 超平面と端点の符号の性質

多面体に追加していく超平面は相補性問題のそれぞれの等式 (17) を満たすベクトルの集合である. 各等式は頂点 s の行動 a に対応するスラック変数 w_s^a を定義する式である. 頂点 s の行動 a に対応する超平面を $\mathcal{H}_s^a = \{\boldsymbol{x} \in \mathbb{R}^{nm+n+1} | A_s^a \boldsymbol{x} = 0\}$ とする. ただし, $A_s^a \boldsymbol{x} = 0$ は以下の等式を意味する.

$$w_s^a - v_s + \beta \sum_{s' \in \mathcal{S}} p_s^a(s') v_{s'} + r_s^a \rho = 0 \qquad (20)$$

アルゴリズムの各ステップでは各端点が超平面の正領域・負領域, または超平面上のどこに位置しているのか (符号) を判定する必要がある. 本相補性問題の超平面はその位置に特徴があり, 端点の符号の判定が容易である場合がある.

性質 6.2. 超平面 $\mathcal{H}_s^a$ に対して, 各端点の符号は以下のように定まる.

- w_s^a 成分の基本ベクトル $\boldsymbol{e}_{w_s^a}$ の符号は正である.
- $w \neq w_s^a$ 成分の基本ベクトル $\boldsymbol{e}_w$ の符号は 0 である.
- v_s 成分の基本ベクトル $\boldsymbol{e}_{v_s}$ の符号は負である.
- $v_{s'} \neq v_s$ 成分の基本ベクトル $\boldsymbol{e}_{v_{s'}}$ の符号は非負である.
- 各 $\boldsymbol{v}$ 成分のすべての基本ベクトルの集合を $\mathcal{E}$ とする. v_s 成分の基本ベクトル $\boldsymbol{e}_{v_s}$ を含む任意の部分集合 $\mathcal{E}'$ のすべての基本ベクトルから形成される重心ベクトル $\boldsymbol{er}_c$ の符号は負である.
- ρ 成分の基本ベクトル $\boldsymbol{e}_\rho$ の符号は正である.

証明.

$$\begin{aligned}
A_s^a \boldsymbol{e}_{w_s^a} &= w_s^a = 1 > 0 \\
A_s^a \boldsymbol{e}_w &= 0 \\
A_s^a \boldsymbol{e}_{v_s} &= -v_s + \beta p_s^a(s) v_s = -1 + \beta p_s^a(s) \\
&\leq -1 + \beta < 0 \\
A_s^a \boldsymbol{e}_{v_{s'}} &= \beta p_s^a(s) v_{s'} = \beta p_s^a(s) \geq 0
\end{aligned}$$

$$\begin{aligned}
A_s^a \boldsymbol{er}_c &= -v_s + \beta \sum_{s' \in \mathcal{S}} p_s^a(s') v_{s'} \\
&\leq -\frac{1}{|\mathcal{E}'|} + \beta \sum_{s' \in \mathcal{S}} p_s^a(s) \frac{1}{|\mathcal{E}'|} \\
&\leq -\frac{1}{|\mathcal{E}'|} + \beta \frac{1}{|\mathcal{E}'|} < 0 \\
A_s^a \boldsymbol{e}_\rho &= r_s^a \rho = r_s^a \geq 0 \quad \blacksquare
\end{aligned}$$

性質 6.2 より, 端点集合 $\mathcal{R}_w$ の端点 $\boldsymbol{e}_{w_{s'}^{a'}}$ はそれぞれが対応している頂点 s' と行動 a' が超平面 $\mathcal{H}_s^a$ のそれと一致している場合は正, そうでないならば 0 である. したがって, 各ステップで端点集合 $\mathcal{R}_w$ の端点 $\boldsymbol{e}_{w_s^a}$ が 1 つずつ正となり, 端点集合 $\mathcal{R}_{v\rho}$ の符号が負の端点 $\boldsymbol{er}$ と結合して, その w_s^a 成分を $\boldsymbol{er}$ に追加する.

端点集合 $\mathcal{R}_v$ の端点の符号に関しては以下の性質が成り立つ.

性質 6.3. 端点集合 $\mathcal{R}_v$ の端点 $\boldsymbol{er}$ は, その $\boldsymbol{v}$ 成分の最大値が v_s である場合, 超平面 $\mathcal{H}_s^a$ に対する符号は負である.

証明.

$$\begin{aligned}
A_s^a \boldsymbol{er} &= -v_s + \beta \sum_{s' \in \mathcal{S}} p_s^a(s') v_{s'} \\
&\leq -v_s + \beta v_s < 0 \quad \blacksquare
\end{aligned}$$

また, 超平面 $\mathcal{H}_s^a$ 上のベクトルは以下の性質を満たす.

性質 6.4. 超平面 $\mathcal{H}_s^a$ 上の ρ 成分が 0 である任意のベクトルは, $\boldsymbol{v}$ 成分の最大値が v_s ではない.

証明. 性質 6.3 より, v 空間上のベクトルのうち $\boldsymbol{v}$ 成分の最大値が v_s であるものは, 超平面 $\mathcal{H}_s^a$ に対して符号は負である. したがって, それは超平面 $\mathcal{H}_s^a$ 上にはない. $\blacksquare$

図 4 は性質 6.2, 6.3, 6.4 を $n=2,3$ の場合に対してまとめた概図である.

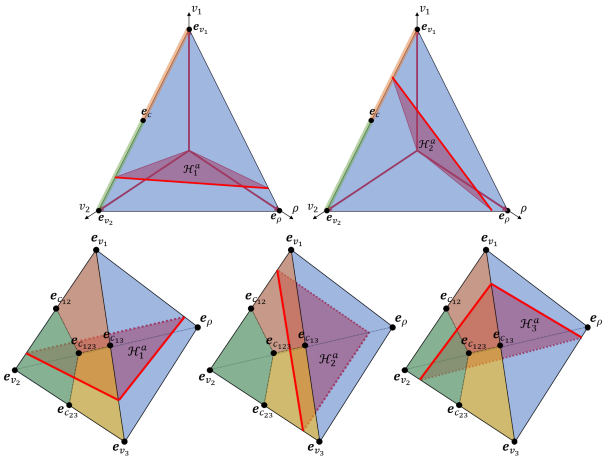

図 4 超平面の特徴と端点の符号. 同色の領域上のベクトルは最も近い v_s 成分の基本ベクトル $\boldsymbol{e}_{v_s}$ が共通であり, それぞれ s が対応する超平面 $\mathcal{H}_s^a$ に対して符号が負になる.

6.4 アルゴリズムの流れに沿って

初期多面体に超平面を順に追加していき, すべての超平面を追加し終えた時点で唯一の端点が残ることを示す. 図 5 は $n = 3$ の場合の多面体の挙動をまとめたものである.

二重描写法ベースアルゴリズムは超平面の追加の順番に任意性がある. 本証明ではアルゴリズムを n 個のフェイズにわけ, フェイズ s では頂点 s の各行動 a に対応する超平面 $\mathcal{H}_s^a$ を順に追加し, すべての超平面を追加し終えたならば次の頂点に対応する超平面を追加する.

アルゴリズムの各ステップでは, 各超平面 $\mathcal{H}_s^a$ に対して端点集合 $\mathcal{R}_{v\rho}$ の異符号な 2 つの端点のペアが隣接かつ交差相補性を満たすならば, それらの内分点かつ超平面上に新しい端点をつくる. また, 符号が負の端点は端点集合 $\mathcal{R}_{w_s^a}$ の符号が正の端点 $\boldsymbol{e}_{w_s^a}$ と結合して新しい端点をつくる. この場合, $\boldsymbol{v}, \rho$ 成分に変化はないため, $v\rho$ 空間における座標は変わらない.

各ステップの端点集合 $\mathcal{R}_v$ の端点にはあらかじめ符号が確定しているものが存在する.

補題 6.5. 任意のフェイズ s で追加されるすべての超平面 $\mathcal{H}_s^a$ に対して, 符号が常に負となる端点集合 $\mathcal{R}_v$ の端点が存在する. また, 各フェイズの 1 つ目の超平面 $\mathcal{H}_s^0$ に対しては非負となる端点集合 $\mathcal{R}_{v\rho}$ の端点が存在する.

補題の証明は次の節で行なう. 補題 6.5 より, 各フェイズで必ず少なくとも 1 度は多面体と超平面の交差が存在することが分かる (図 5(1 段目左, 4 段目左, 5 段目左)). 繰り返し超平面を追加していく際に, フェイズ内で符号が常に負の端点が存在しているため, 図 5(3 段目左) のようにそれら端点を囲むような構造をつくっていく.

$v\rho$ 空間における超平面 $\mathcal{H}_s^a$ 上の端線は $\mathcal{R}_v$ の端点同士の結合によってつくられた端点であるため, $\boldsymbol{e}_{w_s^a}$ とは結合しておらず, したがって w_s^a 成分の値は 0 である. 一方, 超平面上 $\mathcal{H}_s^a$ に属していない端線は $\boldsymbol{e}_{w_s^a}$ と結合してつくられた端点であり, w_s^a 成分の値は正である. フェイズ s で追加したどの超平面にも属していない端線が存在する場合, すべての行動 a に対する w_s^a 成分の値が 0 であり, これは相補性条件を満たしていない. これら端線はフェイズ終了時に削除される.

相補性条件を満たしていない端線を削除した後に残る多面体は, 図 5(3 段目右) のように, そのフェイズで追加した超平面が定義するファセットの集合体になっている. このような構造を複体と呼ぶことにし, 複体の次元は複体を構成しているファセットの次元とする. 初期多面体は n 次元の多面体 1 つからなる複体である. 各フェイズ終了時の複体は元の複体を構成していた多面体のファセットの集合体であるから, 次元が 1 低い複体である.

フェイズが変わると前のフェイズとは異なる領域を中心に, それを囲むように新しい複体を形成していく (図 5(4, 5 段目)). 図 5 で視覚的に分かるように, フェイズが進行するに従って複体の次元は 1 ずつ減少していく. n 個すべてのフェイズが終了した時点で元々 n 次元であった複体は 0 次元の点となっており, その点のベクトルが一般化線形相補性問題の唯一解である.

定理 6.6. マルコフ決定問題を定式化した一般化線形相補性問題は唯一解をもつ.

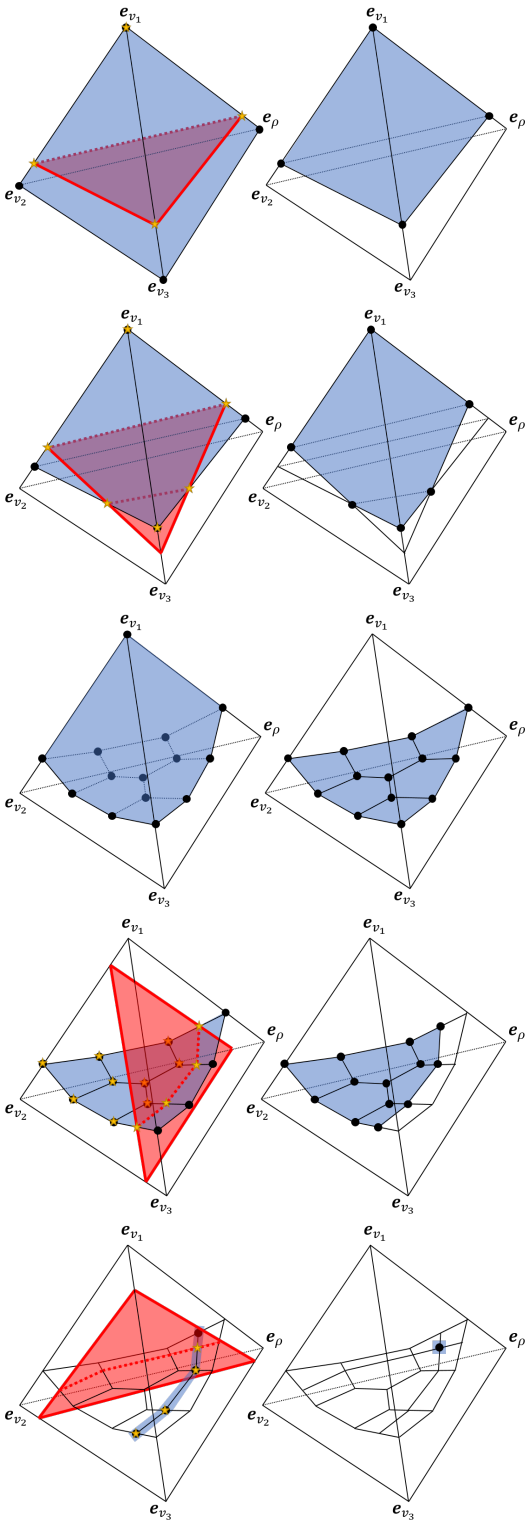

図 5 アルゴリズム中の複体の変化.

6.5 補題の証明

新しくインデックス集合 $\mathcal{I}'_v$, 端点集合 $\mathcal{R}_{\mathcal{I}'_v}$, 部分複体 $\mathcal{G}_{\mathcal{I}'_v}$ を定義し, 諸性質を導く.

定義 6.7. インデックス集合 $\mathcal{I}_v \cup \{\rho\}$ の任意の部分集合を $\mathcal{I}'_{v\rho}$ とする. 端点集合 $\mathcal{R}_{v\rho}$ の端点のうち, $\boldsymbol{v}, \rho$ 成分で $\mathcal{I}'_{v\rho}$ 成分のみが正である端点集合を $\mathcal{R}_{\mathcal{I}'_{v\rho}}$ とする. また, 端点集合 $\mathcal{R}_{\mathcal{I}'_{v\rho}}$ から構成される複体を部分複体 $\mathcal{G}_{\mathcal{I}'_{v\rho}}$ と呼ぶ.

性質 6.8. フェイズ $1, \cdots, s-1$ が終了した時点で, $\mathcal{I}'_{v\rho} = \{v_1, \cdots, v_{s-1}, v_s\}$ としたときの $\mathcal{R}_{\mathcal{I}'_{v\rho}}$ の端点は, 超平面 $\mathcal{H}^a_s$ に対して符号が負である.

証明. 前述の議論より, 各フェイズ $1, \cdots, s-1$ が終了した時点で残っている端点は, 超平面 $\mathcal{H}^a_1, \cdots, \mathcal{H}^a_{s-1}$ 上に必ずある. 性質 6.4 を繰り返し用いることで, $\boldsymbol{v}$ 成分の最大値は $v_1, \cdots, v_{s-1}$ ではないことが分かる. したがって, $\boldsymbol{v}$ 成分の最大値は v_s であり, 端点の符号の性質 6.3 より超平面 $\mathcal{H}^a_s$ に対する符号は負である. ■

性質 6.9. $\mathcal{I}'_v = \{v_1, \cdots, v_{s-1}, v_{s'}\}(s' > s)$ としたときの $\mathcal{R}_{\mathcal{I}'_{v\rho}}$ の端点 $\boldsymbol{er}$ の超平面 $\mathcal{H}^a_s$ に対する符号は非負である.

証明.

$$A^a_s \boldsymbol{er} = \beta \sum_{t \in \mathcal{S}} p^a_s(t) v_t \geq 0 \quad ■$$

ただし, 性質 6.8, 6.9 は $\mathcal{R}_{\mathcal{I}'_{v\rho}}$ が空集合でないことを保証してはいない. それを保証する性質として以下が成り立つ.

性質 6.10. 各端点集合 $\mathcal{I}'_{v\rho}$, 端点集合 $\mathcal{R}_{\mathcal{I}'_{v\rho}}$, 部分複体 $\mathcal{G}_{\mathcal{I}'_{v\rho}}$ に対して, 以下の性質が成り立つ.

- フェイズ $1, \cdots, s-1$ が終了した時点で,
  - $\{v_1, \cdots, v_{s-1}\}$ を含む任意の集合 $\mathcal{I}'_{v\rho} : dim\left(\mathcal{G}_{\mathcal{I}'_{v\rho}}\right) = |\mathcal{I}'_{v\rho}| - s$
- フェイズ s が終了した時点で,
  - $\mathcal{I}'_{v\rho} = \{v_1, \cdots, v_s\} : |\mathcal{R}_{\mathcal{I}'_{v\rho}}| = 0$
  - $\{v_1, \cdots, v_s\}$ を含む任意の集合 $\mathcal{I}'_{v\rho} : dim\left(\mathcal{G}_{\mathcal{I}'_{v\rho}}\right) = |\mathcal{I}'_{v\rho}| - s - 1$

証明. $s = 1$ の場合を確かめるのは容易である. $s = t-1$ で性質が成り立っていると仮定する. $s = t-1$ で 2 つ目の主張が成り立っていることから, $s = t$ において 1 つ目の主張も成り立つ. インデックス集合 $\mathcal{I}'_{v\rho} = \{v_1, \cdots, v_t\}$ に対して, 部分複体 $\mathcal{G}_{\mathcal{I}'_{v\rho}}$ の次元は $dim\left(\mathcal{G}_{\mathcal{I}'_{v\rho}}\right) = |\mathcal{I}'_{v\rho}| - t = 0$ より, 端点集合 $\mathcal{R}_{\mathcal{I}'_{v\rho}}$ の要素は 1 つであり, 性質 6.8 よりこの端点の超平面 $\mathcal{H}^a_t$ に対する符号は負である. この端点はすべての超平面を追加し終えた時点ですべての w^a_t 成分が正であり, 相補性条件を満たさないため削除される. 同様に, $\mathcal{I}'_{v\rho} = \{v_1, \cdots, v_{t-1}, v_{t'}\}(t' > t)$ に対する端点集合 $\mathcal{R}_{\mathcal{I}'_{v\rho}}$ の要素も 1 つであり, 性質 6.9 よりこの端点の超平面 $\mathcal{H}^0_t$ に対する符号は非負である. よって, $\mathcal{I}'_{v\rho} = \{v_1, \cdots, v_t, v_{t'}\}(t' > t)$ を含む任意の集合 $\mathcal{I}'_{v\rho}$ が構成する複体はフェイズ中に少なくとも 1 度は超平面との交差をもち, フェイズ終了時に超平面が定義するファセットの集合体 (複体) のみが残るため, 次元は 1 減少する. したがって, 2 つ目の主張も成り立つ. 図 6 は各部分複体の次元の変化をまとめたものである. ■

$\mathcal{I}'_{v\rho}$	初期複体	フェイズ1	フェイズ2	フェイズ3
$\{v_1\}$	0	---	---	---
$\{v_2\}$	0	?	?	?
$\{v_3\}$	0	?	?	?
$\{\rho\}$	0	?	?	?
$\{v_1, v_2\}$	1	0	---	---
$\{v_1, v_3\}$	1	0	?	?
$\{v_1, \rho\}$	1	0	?	?
$\{v_2, v_3\}$	1	?	?	?
$\{v_2, \rho\}$	1	?	?	?
$\{v_3, \rho\}$	1	?	?	?
$\{v_1, v_2, v_3\}$	2	1	0	---
$\{v_1, v_2, \rho\}$	2	1	0	?
$\{v_2, v_3, \rho\}$	2	?	?	?
$\{v_1, v_2, v_3, \rho\}$	3	2	1	0

図 6 各部分複体の次元の変化. 数字は各インデックス集合 $\mathcal{I}'_{v\rho}$ から構成される部複体 $\mathcal{G}_{\mathcal{I}'_{v\rho}}$ の次元. —は要素が存在しないことを意味し, ?は要素が存在するか定めることができないことを意味する.

これまでの議論は $\mathcal{R}_{v\rho}$ の隣接な 2 つの端点のペアは常に結合可能である, つまり交差相補性を満たすことを前提としてきたが, これはどのステップにおいても成り立っている.

性質 6.11. $\mathcal{R}_{v\rho}$ の隣接な 2 つの端点のペアは交差相補性を常に満たす.

証明. 隣接な 2 つの端点は, 任意の頂点 s の超平面に対して, 少なくとも 1 つは共通の超平面 $\mathcal{H}^a_s$ 上に属しており, それらの w^a_s 成分は共通して 0 である. したがって, 少なくとも w^a_s 成分が 0 であり, 結合後につくられる端点も頂点 s における相補性条件を満たしている. ■

以上の性質から, 補題 6.5 が導くことができ, したがって定理 6.6 が成り立つ.

7. 2 人ゼロ和完全情報確率ゲーム

マルコフ決定問題に対する最適解の存在証明は 2 人ゼロ和完全情報確率ゲームのナッシュ均衡の存在証明に容易に拡張できる. 2 人ゼロ和完全情報確率ゲームでは頂点集合 $\mathcal{S}$ を 2 つの部分集合 $\mathcal{S}_1, \mathcal{S}_2$ に分割し ($\mathcal{S}_1 \cup \mathcal{S}_2 = \mathcal{S}, \mathcal{S}_1 \cap \mathcal{S}_2 = \phi$), それぞれプレイヤー 1 とプレイヤー 2 が行動を決定する頂点の集合である. プレイヤーには共通の報酬が与えられ, プレイヤー 1 は割引総利得の最大化, プレイヤー 2 はそれの最小化を目的にコントロールできる頂点における行動を決

定する. 2 人ゼロ和完全情報確率ゲームのナッシュ均衡解 $\pi^*, v_s^{\pi^*}$ は以下のように定義され, 同様に必要十分条件も得られる.

定義 7.1. 戦略 π^* が以下の条件を満たすとき, それを 2 人ゼロ和完全情報確率ゲームのナッシュ均衡解とする.

$$\forall \pi \in \Pi, \forall s \in \mathcal{S}_1 \quad v_s^{\pi^*} \geq v_s^{\pi} \tag{21}$$

$$\forall \pi \in \Pi, \forall s \in \mathcal{S}_2 \quad v_s^{\pi^*} \leq v_s^{\pi} \tag{22}$$

定理 7.2. 戦略 π が与えられたとき, それが 2 人ゼロ和完全情報確率ゲームのナッシュ均衡解であるための必要十分条件は以下を満たすことである.

$$\forall s \in \mathcal{S}_1, \forall a \in \mathcal{A}_s$$

$$v_s^{\pi} = r_s^{\pi} + \beta \sum_{s' \in \mathcal{S}} p_s^{\pi}(s') v_{s'}^{\pi} \tag{23}$$

$$\geq r_s^{a} + \beta \sum_{s' \in \mathcal{S}} p_s^{a}(s') v_{s'}^{\pi} \tag{24}$$

$$\forall s \in \mathcal{S}_2, \forall a \in \mathcal{A}_s$$

$$v_s^{\pi} = r_s^{\pi} + \beta \sum_{s' \in \mathcal{S}} p_s^{\pi}(s') v_{s'}^{\pi} \tag{25}$$

$$\leq r_s^{a} + \beta \sum_{s' \in \mathcal{S}} p_s^{a}(s') v_{s'}^{\pi} \tag{26}$$

2 人ゼロ和完全情報確率ゲームも唯一のナッシュ均衡解をもつことが知られている [1]. また, 同様に一般化線形相補性問題として定式化することができる.

定理 7.3. 2 人ゼロ和完全情報確率ゲームのナッシュ均衡は以下の一般化線形相補性問題の解と同値であり, $w_s^a = 0$ となる行動 a が頂点 s のナッシュ均衡解における行動である.

$$\forall s \in \mathcal{S}_1, \forall a \in \mathcal{A}_s$$

$$v_s = r_s^a \rho + \beta \sum_{s' \in \mathcal{S}} p_s^a(s') v_{s'} + w_s^a \tag{27}$$

$$\prod_{a \in \mathcal{A}_s} w_s^a = 0 \tag{28}$$

$$w_s^a, v_s, \rho \geq 0 \tag{29}$$

$$\forall s \in \mathcal{S}_2, \forall a \in \mathcal{A}_s$$

$$v_s = r_s^a \rho + \beta \sum_{s' \in \mathcal{S}} p_s^a(s') v_{s'} - w_s^a \tag{30}$$

$$\prod_{a \in \mathcal{A}_s} w_s^a = 0 \tag{31}$$

$$w_s^a, v_s, \rho \geq 0 \tag{32}$$

ここで, マルコフ決定問題と 2 人ゼロ和完全情報確率ゲームの定式化で異なるのは, 式 (30) のスラック変数の符号のみである. したがって, マルコフ決定問題において端点集合 $\mathcal{R}_{v\rho}$ の “負”の端点と端点集合 $\mathcal{R}_w$ の “正”の端点が結合していた点を, 2 人ゼロ和完全情報確率ゲームのプレイヤー 2 がコントロールする頂点集合 $\mathcal{S}_2$ においては端点集合 $\mathcal{R}_{v\rho}$ の “正”の端点と端点集合 $\mathcal{R}_w$ の “負”の端点が結合するように置き換えることで, 同様の手法によってそのナッシュ均衡解の存在を証明することができる.

定理 7.4. 2 人ゼロ和完全情報確率ゲームを定式化した一般化線形相補性問題は唯一解をもつ.

8. 多面体の構造と二重描写法ベースアルゴリズムのエッセンス

すべての超平面の非負領域からなる多面体を図 7 に示す. 同じ頂点が対応する超平面から定義されるファセットは同じ色にしてある. マルコフ決定問題の多面体は $v_1, \cdots, v_n$ からなる面で開いていて, ρ 方向に進むにつれて徐々に閉じていく特徴的な構造をもっていることが分かる. そして, 各頂点のファセットの集合は 1 点で交わっている. また, 2 人ゼロ和完全情報確率ゲームはプレイヤー 2 がコントロールする頂点が対応する超平面ではマルコフ決定問題とは反対側の領域を用いた多面体となっていて, 同様に各頂点のファセットの集合は 1 点で交わっている.

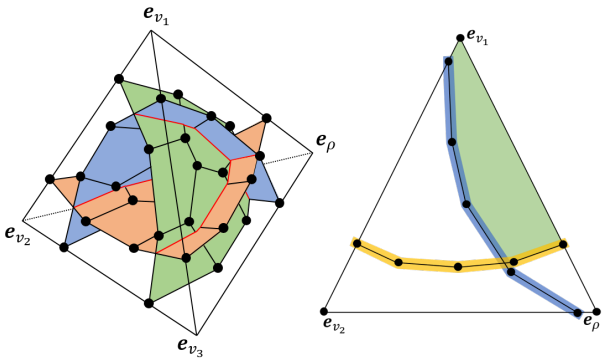

図 7 マルコフ決定問題と 2 人ゼロ和完全情報確率ゲームが構成する多面体の構造. 左がマルコフ決定問題 ($n = 3$). 右が 2 人ゼロ和完全情報確率ゲーム ($n = 2$) で, 頂点 1 ではプレイヤー 1, 頂点 2 ではプレイヤー 2 がコントローラー.

多面体のファセット, そして各頂点が対応するファセットの集合 (複体) 同士の交差は何を意味しているのだろうか. ここで, 明らかに他により良い戦略が存在している, つまり支配されている戦略を被支配戦略と呼び, 支配されていない戦略のことを独立戦略と呼ぶことにする. ある超平面 $\mathcal{H}_s^a$ が定義するファセット上のベクトルは w_s^a 成分が 0 であり, これは頂点 s において行動 a を選択しているという意味である. そして, 頂点 s に対応する複体は頂点 s で実際に用いる可能性のある行動の集合, つまり独立戦略の集合と解釈することができる. さらに次の頂点が対応する超平面 $\mathcal{H}_{s'}^a$ を加えていくことで頂点 s' の独立戦略の集合を構成していくが, これら複体同士の交差から成る新しい複体を構成する多面体は 2 つの超平面 $\mathcal{H}_s^a, \mathcal{H}_{s'}^{a'}$ 上にあり,

頂点 s では行動 a, 頂点 s' では行動 a' を選択していることを意味する. つまり, 新しい複体は頂点 s, s' を考慮した場合の独立戦略の集合である. したがって, 二重描写法ベースアルゴリズムは被支配戦略を削除して独立戦略を徐々に列挙していくアルゴリズムであることが分かる. フェイズ $1, \cdots, s$ まで終了した時点で頂点 $1, \cdots, s$ までを考慮した独立戦略を列挙していて, すべての頂点を考慮したとき, 独立戦略は唯一つに定まる.

9. まとめ

本研究ではマルコフ決定問題を定式化した一般化線形相補性問題に対する二重描写法ベースアルゴリズムの挙動の解析を行なった. そして, それらの問題が構成する多面体の構造から, 二重描写法ベースアルゴリズムが各フェイズにおける独立戦略を徐々に列挙するアルゴリズムだということを明らかにした. 今後は, 本研究で得られた多面体的特徴や二重描写法ベースアルゴリズムのエッセンスを生かした N 人完全情報確率ゲームに対するアルゴリズムの設計が課題である. また, マルコフ決定問題や 2 人ゼロ和完全情報確率ゲーム自体についても, その強多項式時間アルゴリズムが未解決問題となっており ([4], [5]), 以上の特徴を用いたアルゴリズムの設計も課題である.

謝辞 本研究の一部は, 文部科学省・未来社会実現のための ICT 基盤技術の研究開発「実社会ビッグデータ利活用のためのデータ統合・解析技術の研究開発」による.

参考文献

[1] Shapley, L. S.: Stochastic Games, *Proceedings of the National Academy of Sciences*, Vol. 39, No. 10, pp. 1095–1100 (1953).
[2] 清藤駿成，徳山　豪：一般化線形相補性問題を用いた N 人完全情報確率ゲームの定式化とアルゴリズム，情報処理学会東北支部研究会 (2016).
[3] Moor, B. D., Vandenberghe, L. and Vandewalle, J.: The Generalized Linear Complementarity Problem and an Algorithm to Find All its Solutions, *Math. Program.*, Vol. 57, pp. 415–426 (1992).
[4] Ye, Y.: The Simplex and Policy-Iteration Methods Are Strongly Polynomial for the Markov Decision Problem with a Fixed Discount Rate, *Math. Oper. Res.*, Vol. 36, No. 4, pp. 593–603 (2011).
[5] Hansen, T. D., Miltersen, P. B. and Zwick, U.: Strategy Iteration Is Strongly Polynomial for 2-Player Turn-Based Stochastic Games with a Constant Discount Factor, *J. ACM*, Vol. 60, No. 1, pp. 1:1–1:16 (2013).