数理解析研究所講究録
第611巻 1987年 13-25

On the closest multistrategy to the shadow minimum of a Markov game

ハルピン師範大　劉　兆華（Liu Zhaohua）
新潟大・理　田中　謙輔（Kensuke Tanaka）

1．序論

多目的決定問題では、ある凸錐によって導入される支配構造のもとでの最適解について多くの研究がなされている。我々は、協力マルコフ・ゲームの解の概念として、これに類似の解を論文［９］で導入した。このようなゲームの解に対応するPareto解の概念は、協力ゲームの最も弱い解としてAubinの著書［１］の中で既に論じられている。しかし、このような協力ゲームの解の集合は広くなりすぎる可能性がある。これが解の概念としての欠点になっている。よって、このような解の集合を狭くする必要があり、ここでは、shadow minimumから最も近い距離にあるPareto解を考えることにする。しかし、このようなノルム最小化問題の最適解を直接、一般的に求めることは、困難のように思われるので、これを双対型問題に変換して、最適解、即ち、弱最適解が、ある条

件のもとで、常に存在することを示す。さらに、主問題に対する最適解が存在するとき、この解と弱最適解の間の関係も述べる。

2. 協力m人マルコフ・ゲームの形式化

ここでは、割引因子をもつ協力m人マルコフ・ゲームを次のような2m+3個の組

$$(S,\ A^1,\ \cdots,\ A^m,\ q,\ r^1,\ \cdots,\ r^m,\ \beta) \qquad (1)$$

で与える。ただし、

(i) $S=\{1,\ 2,\ \cdots,\ s,\ \cdots\}$ は、ゲームの状態空間

(ii) A^i, $i=1,2,\cdots,m$, は、iプレイヤーの行動空間と呼ばれ、コンパクトな距離空間

(iii) qは、$(s,\ \bar{a})=(s,a^1,a^2,\cdots,a^m)\in S\times\prod_{i=1}^{m}A^i=S\times A$ に対応するS上の推移確率 $q(\cdot\mid s,\bar{a})$

(iv) r^i, $i=1,2,\cdots,m$, は、iプレイヤーの損失関数で、$S\times A$ 上で定義されている実数値関数

(v) β は、割引因子, $0<\beta<1$

この話を通して用いられる多重戦略は、$\vec{\pi}=(\pi^1,\pi^2,\cdots,\pi^m)$で表し、各$\pi^i$は、iプレイヤーの戦略で、各時点tの状態$S_t$に対応する$(A^i,\beta(A^i))$上の確率測度$\pi_t^i(\cdot|S_t)$の列によって与えられる、即ち、$\pi^i=(\pi_1^i,\pi_2^i,\cdots,\pi_t^i,\cdots)$、ただし、$\beta(A^i)$は、行動空間$A^i$上のBorel fieldである。特に、$\pi_t^i$が各時点tに無関係であれば、$\pi_t^i=\mu^i\in[P(A^i)]^S$となり、$\pi^i$を$\mu^i$と同一視して、定常戦略と呼ぶ。ただし、$P(A^i)$は、$(A^i,\beta(A^i))$上の確率測度の全体を表している。このような、各プレイヤーの戦略の全体を、Π^iと表し、多重戦略の全体を、$\Pi=\prod_{i=1}^{m}\Pi^i$と書くことにする。

このとき、初期確率分布$p_1\in P(S)$と多重戦略$\vec{\pi}=(\pi^1,\pi^2,\cdots,\pi^m)\in\Pi$に対して、ゲーム過程における状態の列$\{S_t\}$, $t=1,2,\cdots$, は、次のような確率分布$p_t(\cdot|\vec{\pi})$, $t=1,2,\cdots$,をもつマルコフ鎖を作っている:

$$p_t(s_t|\vec{\pi})=\sum_S q_{t-1}(s_t|s,\vec{\pi})p_{t-1}(s|\vec{\pi}) \qquad (2)$$

$$p_1(s_1|\vec{\pi})=p_1(s_1),\ \forall s_1\in S$$

ここで、

$$q_t(s'|s,\vec{\pi})=\int_A q(s'|s,\vec{a})\,d\vec{\pi}_t(\vec{a}|s)$$

$$d\vec{\pi}_t(\vec{a}|s)=\prod_{i=1}^{m}d\pi_t^i(a^i|s)$$

さらに、各iプレイヤーの総期待損失は、

$$I^i(\pi) = \sum_{t=1}^{\infty} \beta^{t-1} E_{\pi}[r^i(s_t, t, \pi)] \qquad (3)$$

で与えられ、各$I^i(\pi)$のベクトル表示は、

$$I(\pi) = (I^1(\pi), I^2(\pi), \cdots, I^m(\pi)) \qquad (4)$$

$$= \sum_{t=1}^{\infty} \beta^{t-1} E_{\pi}[r(s_t, t, \pi)]$$

で与える。ただし、

$$r^i(s_t, t, \pi) = \int_A r^i(s_t, \bar{a})\, d\pi_t(\bar{a} \mid s_t)$$

$$r(s_t, t, \pi) = (\cdots, r^i(s_t, t, \pi), \cdots)_{i=1}^{m}$$

今、$\alpha^i = \inf_{\pi} I^i(\pi)$ とおくことによって、このマルコフ・ゲームのshadow minimum

$$\bar{\alpha} = (\alpha^1, \alpha^2, \cdots, \alpha^m)$$

を定義する。このとき、

$$I(\pi) \in \bar{\alpha} + \mathbb{R}^m_+ , \qquad \forall \pi \in \Pi$$

が成立する。ただし、

$$\mathbb{R}^m_+ = \{x=(x_1, \cdots, x_m) \in \mathbb{R}^m \mid x_i \geqq 0, \forall i=1,2,\cdots,m\}$$

このとき、$\bar{\alpha} = I(\pi^*)$ となる $\pi^* \in \Pi$ が存在すれば、この多重戦略π^*は、各プレイヤーにとって最上の最適戦略と考えられるが、実際にこのような場合は殆ど起こらない。よって、各プレイヤーは協力することによって、

$$\| I(\pi) - \bar{\alpha} \|^2 = \sum_{i=1}^{m} (I^i(\pi) - \alpha^i)^2 \qquad (5)$$

を最小にする多重戦略を求めることを考える。

3. 補助定理と定義

必要な記号として、

$$K = \{ I(\bar{\pi}) ; \forall \bar{\pi} \in \prod_{i=1}^{m} \Pi^i = \Pi \} \subset R^m$$

を導入し、この集合K上に次の条件を課する。

（A1） KはR^mの中の凸集合とする;

$$\forall \bar{\pi}_1, \bar{\pi}_2 \in \Pi, \quad 0 < \forall \alpha < 1,$$

$$\alpha I(\bar{\pi}_1) + (1-\alpha) I(\bar{\pi}_2) \in K$$

この集合Kとshadow minimum $\bar{\alpha}$ に対して、記号

$$K - \bar{\alpha} = \{ x - \bar{\alpha} \in R^m ; \forall x \in K \}$$

を導入して、$K - \bar{\alpha}$の上の支持関数をR^m上に、

$$\delta(d | K - \bar{\alpha}) = \inf_{\bar{\pi}} \langle d, I(\bar{\pi}) - \bar{\alpha} \rangle$$

と定義する。

補助定理. 条件（A1）のもとで、

$$\rho = \inf_{\bar{\pi}} \| I(\bar{\pi}) - \bar{\alpha} \| > 0$$

と仮定する。

このとき、次の条件を満たす$d_* \in R^m_+$, $\| d_* \| = 1$が存在する:

$$\rho = \delta(d_* | K - \bar{\alpha})$$

証明. $\overline{K}$ をKの閉包とすると、ρ の定義より、$\rho = \| \overline{x} - \bar{\alpha} \|$ となる $\overline{x} \in \overline{K}$ が存在する。このとき、条件（A1）とK－$\bar{\alpha}$が凸集合であることから、次の不等式を満たす $d_* \in R^m_+$, $\| d_* \| = 1$ が作られる：

$$\langle d_*, x - \bar{\alpha} \rangle \geqq \langle d_*, \overline{x} - \bar{\alpha} \rangle = \rho, \qquad \forall x \in K$$

また、一方、$\overline{x} \in \overline{K}$ であるから、条件（A1）より、$\| x_n - \overline{x} \| \to 0, n \to \infty$ となる点列 $\{x_n\} \subset K$ が存在するので、

$$\delta(d_* | K - \bar{\alpha}) = \langle d_*, \overline{x} - \bar{\alpha} \rangle = \rho$$

を得る。

さらに、$K - \bar{\alpha} \subset R^m_+$ が成立しているので、$d_* \in R^m_+$ となり、補助定理が証明される。

次に、最小化ノルム問題と双対問題に対する最適多重戦略（最適解）の定義を次のように与える。

定義1. $\pi_* \in \Pi$ に関して、

$$\| I(\pi) - \bar{\alpha} \| \geqq \| I(\pi_*) - \bar{\alpha} \|, \qquad \forall \pi \in \Pi$$

が成立するとき、π_* を最適多重戦略（最適解）と呼ぶ。

定義2. $\pi_* \in \Pi$ に関して、

$$\rho = \delta(d_* | K - \bar{\alpha})$$

$= \langle d_{\cdot}, I(\bar{\pi}_{\cdot}) - \bar{\alpha} \rangle$

を満たす $d_{\cdot} \in R^{m}_{+}$、$\| d_{\cdot} \| = 1$ が存在するとき、$\bar{\pi}_{\cdot}$ を重み因子 $d_{\cdot}$ に関する弱最適多重戦略（弱最適解）と呼ぶ。

4. 弱最適解の存在

ここでは、推移確率 q と損失関数 r^{i}, $i=1,2,\cdots,m$, に関して、次のような条件が必要である。

（A２）　各 $(s',s) \in S \times S$ に対して、
$q(s'|s,\bar{a})$ は $\bar{a} \in A$ に関して連続である。

（A３）　各 i プレイヤーの損失関数 $r^{i}(s,\bar{a})$ は、
$S \times A$ 上で有界で、各状態 $s \in S$ に対して、
$\bar{a} \in A$ に関して連続である。

また、双対問題における損失関数は次のように変形される；

$$
\begin{aligned}
&\langle d_{\cdot}, I(\bar{\pi}) - \bar{\alpha} \rangle \\
&= \langle d_{\cdot}, \sum_{t=1}^{\infty} \beta^{t-1} E_{\bar{\pi}} [\, r(s_t, t, \bar{\pi})\,] - \bar{\alpha} \rangle \qquad (6) \\
&= \sum_{t=1}^{\infty} \beta^{t-1} \langle d_{\cdot}, E_{\bar{\pi}} [\, r(s_t, t, \bar{\pi})\,] - (1-\beta)\bar{\alpha} \rangle \\
&= \sum_{t=1}^{\infty} \beta^{t-1} E_{\bar{\pi}} [\, \langle d_{\cdot}, r(s_t, t, \bar{\pi}) - (1-\beta)\bar{\alpha} \rangle\,]
\end{aligned}
$$

よって、次のような双対型の協力ゲームを考えることになる；

$$(S,\ A,\ q,\ \langle d\cdot,\ r-(1-\beta)\bar{\alpha}\rangle,\ \beta) \qquad (7)$$

ただし、

$$A = \prod_{i=1}^{m} A^i$$

このとき、双対型協力ゲームについて、次の定理が成立する。

定理1. 条件（A1），（A2），（A3）のもとで、

$$\rho = \inf_{\bar{\pi}} \| I(\bar{\pi}) - \bar{\alpha} \| > 0$$

と仮定する。

このとき、

$$\rho = \langle d\cdot, I(\bar{x}\cdot) - \bar{\alpha} \rangle$$

を満たす $d\cdot \in R^m_+$, $\| d\cdot \| = 1$ に関する弱最適定常解 $\bar{x}\cdot \in \Pi$ が存在する。

証明. 今、C（S）を状態空間Sの上の全ての有界な実数値関数の集合とし、補助定理における重み因子 $d\cdot \in R^m_+$, $\| d\cdot \| = 1$ を用いて、次のようなC（S）上の

写像を定義する;

$$Tu(s)=\min_{\bar{\mu}\in P(A)}[<d\cdot, r-(1-\beta)\bar{\alpha}>+\beta\sum_{s'}u(s')q(s'|s,\bar{\mu})] \qquad (8)$$

明らかに、 $u\in C(S)$ ならば、 $Tu(s)\in C(S)$ が成立している。 話を簡単にするために、

$$L(\bar{\mu})=<d\cdot, r-(1-\beta)\bar{\alpha}>+\beta\sum_{s'}u(s')q(s'|s,\bar{\mu})$$

とおくと、 (8) 式は

$$Tu(s)=\min_{\bar{\mu}\in P(A)} L(\bar{\mu})u(s) \qquad (9)$$

と書ける。

このとき、 割引因子 β, $0<\beta<1$ であるから、 写像 T は、 C(S) の上の縮小写像になっており、 supnorm で C(S) はBanach空間になっているので、 T は不動点 $u^*\in C(S)$ を持っている。 即ち、

$$u^*(s)=Tu^*(s)$$

$$=\min_{\bar{\mu}\in P(A)} L(\bar{\mu})u^*(s) \qquad (10)$$

さらに、$L(\bar{\mu})u(s)$ は条件 (A2), (A3) より、 コンパクト集合 P(A) の上で連続となっているので、 次の等式を満たす定常多重戦略 $\bar{\mu}\cdot$ が存在する;

$$u^*(s)=L(\bar{\mu}\cdot)u^*(s) \qquad (11)$$

$$\leqq L(\bar{\mu})u^*(s), \qquad \forall\bar{\mu}\in P(A)$$

この結果、 (11) 式の最初の等式を u^* に対して繰り返し用

いることによって、初期状態 $s \in S$ に対して、

$$u^*(s) = \langle d_\cdot, I(\bar{\pi}_\cdot)(s) - \bar{\alpha} \rangle \qquad (12)$$

が得られる。

一方、同様の議論を（11）式の第2式に用いることによって、初期状態 $s \in S$ に対して、

$$u^*(s) \leqq \langle d_\cdot, I(\pi)(s) - \bar{\alpha} \rangle , \qquad \forall \pi \in \Pi \qquad (13)$$

を得る。よって、任意の $p \in P(S)$ によって、（12）式と（13）式の両辺を積分することによって、

$$\langle d_\cdot, I(\bar{\pi}_\cdot) - \bar{\alpha} \rangle \leqq \langle d_\cdot, I(\pi) - \bar{\alpha} \rangle , \qquad \forall \pi \in \Pi$$

を得ることが出来て、定理の証明は完成する。

定理2. 定理1と同じ条件のもとで、最小化ノルム問題の最適解 π_0 が存在すれば、弱最適定常解 $\bar{\pi}_\cdot$ と π_0 に関して次の関係が成立する；

$$\begin{aligned} \rho &= \| I(\pi_0) - \bar{\alpha} \| \\ &= \langle d_\cdot, I(\bar{\pi}_\cdot) - \bar{\alpha} \rangle \\ &= \langle d_\cdot, I(\pi_0) - \bar{\alpha} \rangle \end{aligned}$$

証明. $\pi_0 \in \Pi$ が最小化ノルム問題の最適解であることより、

$$\rho = \| I(\bar{\pi}_0) - \bar{\alpha} \| = \inf_{\bar{\pi}} \| I(\bar{\pi}) - \bar{\alpha} \|$$

ここで、$I(\bar{\pi}_0) \in K$ であるから、定理1より

$$\begin{aligned} <d_\cdot, I(\bar{\pi}_0) - \bar{\alpha}> &\geqq \inf_{\bar{\pi}} <d_\cdot, I(\bar{\pi}) - \bar{\alpha}> \qquad (14) \\ &= <d_\cdot, I(\bar{\mu}_\cdot) - \bar{\alpha}> \\ &= \inf_{\bar{\pi}} \| I(\bar{\pi}) - \bar{\alpha} \| \\ &= \rho \end{aligned}$$

が成立する。

このとき、$\| d_\cdot \| \leqq 1$ であるから、

$$\begin{aligned} <d_\cdot, I(\bar{\pi}_0) - \bar{\alpha}> &\leqq \| d_\cdot \| \, \| I(\bar{\pi}_0) - \bar{\alpha} \| \qquad (15) \\ &\leqq \inf_{\bar{\pi}} \| I(\bar{\pi}) - \bar{\alpha} \| \\ &= <d_\cdot, I(\bar{\mu}_\cdot) - \bar{\alpha}> \\ &= \rho \end{aligned}$$

よって、（14）式と（15）式より定理の結果が得られる。

References

[1] J.P.Aubin, "Mathematical Methods of Game and Economic Theory", North-Holland, Amsterdum, 1979.

[2] J.P.Aubin, "Applied Functional Analysis", Wiley-Interscience, New York, 1979.

[3] M.S.Bazaraa and C.M.Shetty, "Nonlinear Programming", John Wiley and Sons, New York, 1979.

[4] P.Billingsley, "Convergence of Probability Measures", Wiley-nterscience, New York, 1968.

[5] K.Fan, Fixed point and minimax theorem in locally convex topological linear spaces, Proc. Nat. Acad. Sci. U.S.A. 38 (1958), 121 - 126.

[6] H.C.Lai and K.Tanaka, Noncooperative n-person game with a stopped set, J. Math. Anal. Appl. 88 (1982), 153 - 171.

[7] H.C.Lai and K.Tanaka, A noncooperative n-person semi-Markov game with a separable metric space, Appl. Math. Optim. 11 (1984), 23 - 42.

[8] H.C.Lai and K.Tanaka, On an N-person noncooperative Markov game with a metric state space, J. Math. Anal. Appl. 101, (1984), 78 - 96.

[9] H.C.Lai and K.Tanaka, On a D-solution of a cooperative m-person discounted Markov game, J. Math. Anal. Appl. 115 (1986), 578 - 591.

[10] H.C.Lai and K.Tanaka, An N-person noncooperative discounted vector-valued dynamic game with a metric sapce, to appear in Appl. Math. Optim.

[11] H.C.Lai and K.Tanaka, An N-person noncooperative discounted vector-valued dynamic game with a stopped set, to appear in J. Computer and Math. with Applications (Special Issue, Pursuit-Evasion Games).

[12] H.C.Lai and K.Tanaka, A vector-minimization problem in a stochastic continuous-time n-person game, to appear in Proceeding of the Conference on Functional Analysis in honor of Professor Ky Fan, Lecture Notes in Pure and Applied Mathematics, Marcel Dekker.

[13] H.C.Lai and K.Tanaka, On vector reward n-person cooperative game, preprint.

[14] Z.Liu and K.Tanaka, On an optimal multistrategy and a weak optimal multistrategy of a Markov game, to appear in Sci. Rep. Niigata Univ., Ser.A, No. (1987),

[15] D.D.Luenberger, "Optimization by Vector Space Methods", Wiley-Interscience, New York, 1969.

[16] K.Tanaka, On the learning algorithm of 2-person zero-sum Markov game with expected average reward criterion, Bull. Inform. Cyber. 21 (1985), 1 - 17.

[17] K.Tanaka, On some vector valued Markov game, Japan, J. Appl. Math. 2 (1985), 293 - 308.

[18] K.Tanaka and H.C.Lai, A two-person zero-sum Markov game with a stopped set, J. Math. Anal. Appl. 86 (1982), 54 - 68.

[19] K.Tanaka and Y.Maruyama, The multiobjective optimization problem of set function, J. Information and Optimization Sciences, 5 (1984), 293 - 306.