数理解析研究所講究録
939巻 1996年 120-126

On an optimal strategy of ergodic control

新潟大学·理学部数学科　田中謙輔（Kensuke Tanaka）

In this paper, stochastic control processes have been investigated as dynamic programming models with an infinite horizon. Then, we want to seek for an optimal strategy under the expected average loss criterion. It has been shown to exist an optimal strategy under the various conditions. However, optimal strategy may not exist under weak conditions. Under some weak condition, we introduce a modified form of the dynamic model, which we want to call as dual dynamic one. By Fenchel's duality theorem, we show that optimal value of the original model is equal to one of the dual model. Moreover, we show that there exists an optimal strategy for the dual model.

Key words: dynamic programming, optimal policy, Fenchel inequality, and Fenchel's duality theorem

1　制御 D.P モデル の記号と構成

次のような簡単な D.P モデル (S, A, B, H, p, r) の平均期待損失基準のもとでの最適性について考察する. ただし,

(1) $S = \{1, 2, 3, \cdots, i, \cdots\}$, システムの状態空間
(2) A, B, システムの制御空間で共に Banach space
(3) $A(i)$, 各状態 $i \in S$ に対応する許容制御集合 $A(i) \subset A$
(4) H_i, 各状態 $i \in S$ に対応する連続な線形写像 $H_i \in L(A, B)$
(5) p, 各 $(i, H_i a) \in S \times B$ に対応する S 上の確率
(6) $r(i, a)$, 損失関数 $r : S \times A \to R \cup \{\infty\}$, $\not\equiv \infty$

この時, マルコフ制御戦略 $\pi = (f_1, f_2, \cdots, f_t, \cdots)$, $f_t : S \to A$, のみについて考察し, 各戦略 $f_t, t = 1, 2, 3, \cdots$, が現時点での状態のみに依存するのでマルコフ制御戦略, 又は単にマルコフ戦略と呼ばれており, このような制御戦略の全体を記号 Π で表す. 更に, 各初期状態 $i \in S$ と 制御戦略 $\pi \in \Pi$ に対応する平均期待損失を

$$I(\pi)(i) = \limsup_{n \to \infty} \frac{1}{n} \sum_{t=1}^{n} E_\pi[r(s_t, f_t(s_t))|s_1 = i]$$

で与える. この時, 確率過程 $\{s_t\}_{t=1,2,\cdots,}$ は戦略 π によって生成される S 上のマルコフ過程を構成している. ここで, 最適化問題を次のように表現する :

(MP)　　minimize　$I(\pi)(i)$　subject to　$\pi \in \Pi$.

Definition 1 この (MP) 問題で

$$\overline{I}(i) = \inf_{\pi \in \Pi} I(\pi)(i)$$

を満たす値 $\overline{I}(i)$ は初期状態 $i \in S$ に対応する最適損失値, 又は単に最適値と呼ぶ

Definition 2 この (MP) 問題で, もし $\overline{I}(i) = I(\overline{\pi})(i)$ を満たす制御戦略 $\overline{\pi} \in \Pi$ が存在すれば, この $\overline{\pi} \in \Pi$ を初期状態 $i \in S$ に対応する最適制御戦略と呼ぶ. 更にこの $\overline{\pi} \in \Pi$ が全ての初期状態 $i \in S$ に対応する最適制御戦略であるとき, 単に最適制御戦略 *(optimal control strategy)*, 又は最適戦略と呼ぶ.

2 補助定理と定理の証明

本論の主な補助定理と定理の証明を与える為に次の記号を用いる. この章を通して, $g : S \to R \cup \{\infty\}$, $\not\equiv \infty$, な関数の全体を $B(S)$ とおき, $u \in B(S)$ に対して次のような記号を導入する.

1. $T_a u(i) = r(i,a) + \sum_{j \in S} u(j) p(j|i, H_i a) = r(i,a) + G(H_i a, u)(i)$
2. $Tu(i) = \inf_{a \in A(i)} T_a u(i)$
3. $r^*(i,p) = \sup_{a \in A(i)} [\langle a, p \rangle - r(i,a)]$, $p \in A^*$, A^* は A の双対空間
4. $G^*(q,u)(i) = \sup_{b \in B} [\langle q, b \rangle - G(b,u)(i)]$, $q \in B^*$, B^* は B の双対空間
5. $T_q^* u(i) = -r^*(i, -H_i^* q) - G^*(q,u)(i)$, $H_i^* \in L(B^*, A^*)$
6. $T^* u(i) = \sup_{q \in B^*} T_q^* u(i)$

この時, 全ての $a \in A, q \in B^*$ と全ての $i \in S$ に対して, $T_a u(i), T_q^* u(i)$ に Fenchel の不等式を適用して, 次の補助定理が示される.

Lemma 1 全ての $i \in S$ に対して,

$$Tu(i) \geq T^* u(i)$$

Proof 各 $i \in S, u \in B(S)$ について, 全ての $a \in A, q \in B^*$ に対して, Fenchel の不等式を適用して

$$\begin{aligned} T_a u(i) - T_q^* u(i) &= r(i,a) + G(H_i a, u)(i) + r^*(i, -H_i^* q) + G^*(q,u)(i) \\ &\geq \langle -H_i^* q, a \rangle + \langle q, H_i a \rangle \\ &= -\langle q, H_i a \rangle + \langle q, H_i a \rangle \\ &= 0 \end{aligned}$$

よって, 全ての $a \in A, q \in B^*$ に対し,

$$T_a u(i) \geq T_q^* u(i)$$

上の不等式で, $a \in A$ に対して下限を $q \in B^*$ に対して上限を取る事より結論の不等式が得られる. □

次に, 各 $i \in S$ に対して, $\mathrm{dom}\, r^*(i,\cdot) = A_i^* \subset A^*, \mathrm{dom} G^*(\cdot,u)(i) = B_i^*(u) \subset B^*(u)$ に対して, 以下のような写像を導入する :

$$\Psi_i : A_i^* \times B_i^*(u) \to R \times A^*$$

$$\Psi_i(p,q) = (-r^*(i,p) - G^*(q,u)(i),\ p + H_i^* q)$$

上の写像を用い, Fenchel の双対定理を示す証明の流れの一部を適用し, 次の補助定理が得られる.

Lemma 2 全ての $i \in S$ と $u \in B(S)$ に対して, $\mathrm{dom} r(i,\cdot) = A_i \subset A(i), \mathrm{dom} G(\cdot,u)(i) = B_i(u) \subset B$ とおき,

$$\theta \in \mathrm{int}(H_i A_i - B_i(u)) \subset B,\ \theta \in H_i^* B_i^*(u) + A_i^* \subset A^*$$

を満たすとする. ただし記号 int は集合の内点の全体. この時全ての $i \in S$ に対して, 次の式を満たす写像 $f^* : S \to B^*$ が存在する

$$T^* u(i) = T_{f^*}^* u(i)$$

Proof 各状態 $i \in S$ と $u \in B(S)$ に対して, 結果の式を満たす B^* の対応する点 $f^*(i)$ が存在する事を示せば証明は終わる. そこで, $T^* u(i)$ の定義から $T^* u(i)$ に収束し, 次の条件を満たす実数列 $v_n^*(i)$ と $p_n \in A_i^*, q_n \in B_i^*(u)$ が存在する :

$$v_n^*(i) = -r^*(i,p_n) - G^*(q_n,u)(i),\ p_n = -H_i^* q_n$$

即ち,

$$(v_n^*(i), p_n + H_i^* q_n) \in \Psi_i(A_i^* \times B_i^*(u))$$

ただし, $r_n = p_n + H_i^* q_n$ とおくとき, 全ての $n = 1,2,\cdots,$ に対して, $r_n = \theta$ である. また, この補助定理の仮定より, 次の条件を満たす半径 $\varepsilon > 0$, 中心 θ の球 $B_\varepsilon \subset B$ が存在する :

$$B_\varepsilon \subset \mathrm{int}(H_i A_i - B_i(u))$$

ここで, 任意の $z \in B$ に対して, $\frac{\varepsilon}{||z||} z = H_i a - b$ と書ける $a \in A_i, b \in B_i(u)$ が存在する. この事より,

$$\begin{aligned} \frac{\varepsilon}{||z||}\langle z, q_n\rangle &= \langle H_i a - b, q_n\rangle \\ &= \langle H_i a, q_n\rangle - \langle b, q_n\rangle \end{aligned}$$

$$\begin{aligned} &= -\langle a, -H_i^* q_n\rangle - \langle b, q_n\rangle \\ &= -\langle a, p_n\rangle - \langle b, q_n\rangle \\ &\geq -\{r(i,a) + r^*(i,p_n)\} - \{G(b,u)(i) + G^*(q_n,u)(i)\} \\ &= v_n^*(i) - \{r(i,a) + G(b,u)(i)\} \end{aligned}$$

が成立する．この時に $v_n^*(i)$ は $T^*u(i)$ に収束する事から，全ての $z \in B$ に対して

$$\inf_{n\geq 1}\langle z, q_n\rangle > -\infty$$

だ成立する．よって，q_n の列は弱*コンパクトとなり，B^* のある点 q^* に弱*収束する部分列 $q_{n'}$ があり，同時に部分列 $p_{n'}$ は $p^* = -H_i^* q^*$ に弱*収束する．更に共役関数の構成法より $-r^*, -G^*$ は弱*上半連続であるから

$$\begin{aligned} -r^*(i,p^*) - G^*(q^*,u)(i) &\geq \limsup_{n\to\infty}\{-r^*(i,p_n)\} + \limsup_{n\to\infty}\{-G^*(q_n,u)(i)\} \\ &\geq \limsup_{n\to\infty}\{-r^*(i,p_n) - G^*(q_n,u)(i)\} \\ &= \lim_{n\to\infty} v_n^*(i) = T^*u(i) \end{aligned}$$

が成立する．以上の理由から次の事が成立している

$$-r^*(i,p^*) - G^*(q^*,u)(i) \geq T^*u(i),\ p^* + H_i^* q^* = \theta$$

かくて，各 $i \in S$ に対して $f^*(i) = q^* \in B^*$ とおくと，全ての $i \in S$ に対して

$$T_{f^*}^* u(i) \geq T^*u(i)$$

を満たす $f^* : S \to B^*$ が得られる．一方 $T^*u(i)$ の構成より全ての $q \in B^*$ に対して

$$T^*u(i) \geq T_{f^*}^* u(i)$$

が成り立つので，結論が得られ定理の証明は終わる． □

次の重要な補助定理を与えるために，定義された写像 Ψ_i に凸錐 Q

$$Q = [0,\infty) \times \{\theta\} \subset R \times A^*$$

を導入する．

Lemma 3 全ての $i \in S$ と $u \in B(S)$ に対して

$$\theta \in \mathrm{int}(H_i A_i - B_i(u)),\ \theta \in H_i^* B_i^*(u) + A_i^*$$

の条件の下で，全ての $i \in S$ に対して

$$(Tu(i), \theta) \in \Psi_i(A^*(i), B^*(i)) - Q$$

を満たすと仮定する．この時，全ての $i \in S$ に対して

$$Tu(i) = T^*u(i)$$

Proof Lemma 2 とこの補助定理の条件を適用すると, 全ての $i \in S$ に対して $p^* + H_i^* q^* = \theta$ を満たす $p^* \in A^*, q^* \in B^*$ が存在して

$$\begin{aligned} Tu(i) &\leq -r^*(i,p^*) - G^*(q^*,u)(i) \\ &= T_{q^*}^* u(i) \\ &\leq T^* u(i) \end{aligned}$$

が得られる. また Lemma 1 より

$$Tu(i) \geq T^* u(i)$$

が成立する事より, 補助定理の結論が得られ証明は終わる. □

本論文の主定理を成立させるために, 次の仮定を導入する. この仮定が成立するためのいろいろな充分条件が研究されて来ているが, ここでは省略する.

Assumption 最適方程式 : 全ての $i \in S$ に対して,

$$\begin{aligned} g + w(i) &= Tw(i) \\ &= \inf_{a \in A(i)} T_a w(i) \\ &= \inf_{a \in A(i)} \{r(i,a) + G(H_i a, w)(i)\} \end{aligned}$$

を満たす定数 g と実関数 $w : S \to R$ が存在する

Theorem 1 *D.P* モデルが上の仮定のもとで, 次の条件を満たしている. 全ての $i \in S$ に対して,

1. $\theta \in H_i^* B_i^*(u) + A_i^*,\ \theta \in \mathrm{int}(H_i A_i - B_i(u))$
2. $(Tu(i), \theta) \in \Psi_i(A_i^*, B_i^*(u)) - Q,\ \forall u \in B(S)$
3. $$\lim_{n \to \infty} \frac{E_\pi[w(s_{n+1}) | s_1 = i]}{n} = 0, \forall \pi \in \Pi$$

この時, 全ての $i \in S$ に対して,

$$\inf_{\pi \in \Pi} I(\pi)(i) = I^*(\pi^*)(i)$$

を満たすある双対定常戦略 $\pi^* = (f^*, f^*, \cdots, f^*, \cdots),\ f^* : S \to B^*$ が存在する. ただし,

$$I^*(\pi^*)(i) = \lim_{n \to \infty} \frac{1}{n} T^{*(n)}(\pi^*) w(i)$$

ここで, $T^{*(n)}(\pi^*)w(i) = T_{f^*}^* T_{f^*}^* \ldots T_{f^*}^* w(i)$ である.

Proof 最適方程式についての仮定より, 全ての $i \in S$ と任意の $f : S \to A$ に対して

$$g + w(i) \leq T_f w(i)$$

即ち,

$$w(i) \leq T_f w(i) - g$$

が得られ, 任意の $\pi = (f_1, f_2, \cdots) \in \Pi$ に対して, 上の不等式を繰り返し適用し, $T_\cdot$ の単調性を用いて

$$\begin{aligned} w(i) &\leq T_{f_1}(T_{f_2} w(i) - g) - g \\ &= T_{f_1} T_{f_2} w(i) - 2g \\ &\vdots \\ &\leq T_{f_1} T_{f_2} \cdots T_{f_n} w(i) - ng \end{aligned}$$

上の不等式を変形して

$$g \leq \frac{1}{n} \sum_{t=1}^{n} E_\pi[r(s_t, f_t(s_t)) | s_1 = i] + \frac{E_\pi[w(s_{t+1}) | s_1 = i]}{n} - \frac{w(i)}{n}$$

が得られ, 定理の条件 3 より $n \to \infty$ とすると

$$g \leq I(\pi)(i)$$

また、最適方程式を繰り返し適用して

$$g = \lim_{n \to \infty} \frac{1}{n} T^n w(i)$$

が得られ, よって

$$\lim_{n \to \infty} \frac{1}{n} T^n w(i) = \inf_\pi I(\pi)(i)$$

次に, 補助定理 2, 3 と定理の条件 1, 2 より

$$g + w(i) = Tw(i) = T^*_{f^*} w(i)$$

を満たす $f^* : S \to B^*$ が存在する. そこで, この式を繰り返し適用して

$$w(i) = T^{*(n)}_{\pi^*} w(i) - ng$$

が得られ, $n \to \infty$ とすると

$$g = \lim_{n \to \infty} \frac{1}{n} T^{*(n)}_{\pi^*} w(i)$$

かくて, 定理の結果が得られ証明は終わる. □

References

[1] J.P.Aubin, Optima and Equilibria – An Introduction to Nonlinear Analysis, Springer-Verlag, New York, 1993.

[2] J.P.Aubin, Mathematical Methods of Game and Economic Theory, Revised Edition, North-Holland, Amsterdam, 1982.

[3] J.P.Aubin, & I.Ekeland, Applied Nonlinear Analysis, Wiley-Interscience, 1984.

[4] J.P.Aubin, & H.Frankowska, Set-Valued Analysis, Birkhäuser, Boston, 1990.

[5] D.P.Bertsekas and S.E.Shreve, Stochastic Optimal Control: The Discrete Time Case, Academic Press, New York, 1978.

[6] D.Blackwell, Discrete dynamic programming, Ann. Math. Statist. 33 (1962) 719-726.

[7] D.Blackwell, Discounted dynamic programming, Ann. Math. Statist. 36 (1965) 226-235.

[8] R.M.Dudley, Real Analysis and Probability, Wadsworth & Brooks, 1989.

[9] E.B.Dynkin and A.A.Yushkevich, Controlled Markov Processes, Springer-Verlag, Berlin, 1979.

[10] I.Ekeland, On the variational principle, J.Math.Anal.Appl., 47 (1974) 324-353.

[11] I.Ekeland, Nonconvex minimization problems, Bull. Amer. Math., 47 (1979) 443-474.

[12] K.Hinderer, Foundations of non-stationary dynamic programming with discrete time parameter, Lecture Notes on Operations Research and Mathematical Systems 33, Springer-Verlag, Berlin, 1970.

[13] S.Iwamoto, Reverse function, reverse program, and reverse theorem in mathematical programming, J. Math. Anal. Appl. 95 (1983) 1–19.

[14] S.Iwamoto, A dynamic Inversion of the classical variational problems, J. Math. Anal. Appl. 100 (1984) 354–374.

[15] D.G.Luenberger, Optimization by Vector Space Methods, John Wiley & Sons, inc., 1969.

[16] R.T.Rockafellar, Extension of Fenchel's duality theorem for convex functions. Duke Math. J. 33 (1966) 81–89.

[17] K.Tanaka, On discounted dynamic programming with constraints, J. Math. Anal. Appl. 155 (1991) 264–277.

[18] K.Tanaka, On a perturbation of dynamic programming, Lecture Notes in Economic and Mathematical Systems, Springer-Verlag, Berlin, 419 (1995) 275–287.

[19] K.Tanaka, M.Hoshino, ans D.Kuroiwa, On a perturbation of continuous time Markov decision processes, Proceedings of APORS'94, edited by M.Fushimi and K.Tone, World Scientific, (1995) 320–329.

[20] K.Tanaka, M.Hoshino, and D.Kuroiwa, On an ε-optimal policy of discrete time stochastic control processes, Bulletin of Informatics and Cybernetics, 27 (1995) 107–119.