数理解析研究所講究録
第358巻 1979年 116-124

区分的動的計画法について

南山大学経営学部
沢木 勝茂

1. はじめに

Blackwell[1], Denardo[2], Strauch[6] らは単調かつ縮小写像の仮定の下で一般的な動的計画法を考察している。本稿では、Denardo[2]の設定の下における動的計画法の特別なクラスについて議論する。このクラスの動的計画法は次のような区分的な性質を有する。もし最終費用が区分的性質を有するならば、費用関数もまた区分的性質を有する。ここで区分的性質とは、関数が分割の部分集合上では、同一の性質を有することを意味する。例えば、部分観測可能なマルコフ決定過程[5],[7] は区分的線形の性質を有し、ある種の確率的在庫モデルは区分的凹の性質をもっている。その他、経済学や経営管理モデルなどにおける費用関数が区分的性質を有する場合が多い。Larson[3]やSawaki[4]は区分的動的計画法の特別な場合である。

本稿では、区分的動的計画法は一般的動的計画法に対して ε－最適な費用と ε－最適な政策が常に存在することを示す。更に、そのような ε－最適費用関数と ε－最適政策を次々と生成するためのアルゴリズムを提供する。

2. 区分的動的計画法

最初にDenardo[2]の設定による一般的な動的計画法を定義し、次に区分的動的計画法を定義する。次のような記号を使用する。

Ω = 状態空間（任意の集合）

x = Ωの要素，$x \in \Omega$

V = Ω上の有界な実数値関数の全体集合

v = Vの要素で費用関数と呼ぶ．$v \in V$

Vはノルム $\|v\| = \sup_{x\in\Omega}|v(x)|$ についてバナッハ空間である。$u, v \in V$ についてもしすべての $x\in\Omega$ に対して $u(x) \leq v(x)$ ならば $u \leq v$ と書く．もし $0 \leq u \leq v$ ならば $\|u\| \leq \|v\|$ である。すなわち、ノルムは非負なる u, v について単調性を有する。各 $x\in\Omega$ について行動の集合 A_x が存在する。デカルト積 $\displaystyle\times_{x\in\Omega} A_x$ を Δ と置く。Δ の要素 δ を政策と呼ぶ。$\displaystyle\bigcup_{x\in\Omega}\{x\}\times A_x \times V$ から実数への写像 h を損失関数として定義する。通常の動的計画法[1],[6]やマルコフ決定過程[5],[7]において損失関数 h は次のようになっている。

$$h(x,a,v) = c(x,a) + \beta\int_\Omega v(y)\, q(dy|x,a),\quad x\in\Omega,\ a\in A_x,\ v\in V.$$

ここで $c(x,a)$ は一期間の費用，β は割引因子，$q(\cdot|x,a)$ は Ω 上での推移確率である．

<u>単調性の仮定</u>　各 $x\in\Omega$ と $a\in A_x$ について，もし $u\le v$, $u, v\in V$, ならば $h(x,a,u)\le h(x,a,v)$ である。

<u>縮少性の仮定</u>　$\beta\in[0,1)$ について

$$|h(x,a,u)-h(x,a,v)|\le\beta\|u-v\|,\quad x\in\Omega,\ a\in A_x,\ u, v\in V,$$

である。

各 $\delta\in\Delta$ について V から V への写像 H_δ を次のように定義する．

$$(H_\delta v)(x) = h(x,\delta(x),v),\quad v\in V,\ x\in\Omega.$$

各 $v\in V$ について $H_{\bar\delta} v = \inf_{\delta\in\Delta} H_\delta v$ なる $\bar\delta\in\Delta$ が存在すると仮定する。この仮定が成立する十分条件を Denardo[2] は与えている。更に、V から V への写像 H_* を次のように定義する．

$$(H_* v)(x) = \inf_{\delta\in\Delta}(H_\delta v)(x),\quad x\in\Omega,\ v\in V.$$

単調性と縮少性の仮定の下で H_δ と H_* は明らかに単調縮少写像である。動的計画法の最適方程式は，$v^* = \sup_{a\in A_x} h(x,a,v^*)$ と表現される。$v^\delta = H_\delta v^\delta$ を満足する v^δ を政策 δ の費用と呼ぶ。$v^*=H_* v^*$ を満足する v^* を最適費用と呼ぶ。もし $\|v^\delta - v^*\|<\varepsilon$ ならば，δ は ε-最適政策である。もし $\|v-v^*\|<\varepsilon$ ならば，

v は ε－最適費用関数である。以上のような性質を有する動的計画法を単調縮少動的計画法と総称する。

仮定　これより以下では，すべての x について $A_x = A = \{a_1, a_2, \cdots, a_p\}$ なる有限な行動空間を仮定する．

定義1．(1) $\delta \in \Delta$ とする．もしすべての $x \in B_i$, $i = 1, 2, \cdots, m$, について $\delta(x) = a_i$ となるような Ω の分割 $\mathcal{P} = \{B_1, B_2, \cdots, B_m\}$ が存在するならば，政策 δ は区分的であると言う．

(2) $v \in V$ とする。もしすべての $x \in B_i$, $i = 1, 2, \cdots, l$, について $v(x) = v_i(x) \in V'$ となるような Ω の分割 $\mathcal{P} = \{B_1, B_2, \cdots, B_m\}$ と部分集合 $V' \subset V$ が存在するならば、費用関数 v は区分的であると言う。

定義2．次の2つの条件が成立するならば，単調縮少動的計画法は区分的であると言う：

(1)　もし δ と v が区分的ならば，$H_\delta v$ は区分的である．

(2)　もし v が区分的ならば、$H_\delta v = H_* v$ なる区分的政策 δ が存在する．

注意　Δ と V と比べて"十分小さい"部分集合 $\Delta' \subset \Delta$, $V' \subset V$ を考える。上の定義2は次のことを意味している。(1) もし $\delta \in \Delta'$ で $v \in V'$ ならば，$H_\delta v \in V'$ である。(2) もし $v \in V'$ ならば．$H_\delta v = H_* v$ なる $\delta \in \Delta'$ が存在する。

補助定理1．$a \in A$ で $v \in V$ とする。もし δ が分割 $\mathcal{P}^\delta$ に関して区

分的であって $h(\cdot, a, v)$ が分割 $\mathcal{P}^h$ に関して区分的ならば，$H_\delta v$ は積の分割 $\mathcal{P}^\delta \otimes \mathcal{P}^h$ に関して区分的である。

証明： $\mathcal{P}^\delta = \{B_1, B_2, \cdots, B_l\}$，$\mathcal{P}^h = \{D_1, D_2, \cdots, D_m\}$ とする。仮定により

$$\delta(x) = a_i \;,\; x \in B_i \;,\; i = 1, 2, \cdots, l$$

$$h(x, a, v) = h_j(x, a, v),\; x \in D_j \;,\; j = 1, 2, \cdots, m$$

となる。従って，

$$\begin{aligned}(H_\delta v)(x) &= h(x, \delta(x), v), & & x \in \Omega \\ &= h(x, a_i, v), & & x \in B_i \;,\; i = 1, 2, \cdots, l \\ &= h_j(x, a_i, v), & & x \in B_i \cap D_j \;,\; i = 1, 2, \cdots, l, \\ & & & j = 1, 2, \cdots, m.\end{aligned}$$

$\mathcal{P}^\delta$ と $\mathcal{P}^h$ の積分割 $\mathcal{P}^\delta \otimes \mathcal{P}' = \{B_i \cap D_j' : i = 1, 2, \cdots, l,\; j = 1, 2, \cdots, m\}$ に関して $H_\delta v$ は区分的である。$\boxtimes$

<u>定理1</u>．もし δ と $h(\cdot, a, v)$ が区分的ならば，動的計画法は区分的である。

証明：δ と $h(\cdot, a, v)$ が区分的ならば，補助定理より $H_\delta v$ もまた区分的である。従って区分的動的計画法の定義2の(2)が成立することのみを証明すればよい。$H_\delta v$ は分割 $\mathcal{P} = \{B_i, i = 1, 2, \cdots, m\}$ に関して区分的であるとする。すべての $x \in B_i$ について $(H_\delta v)(x) = h_i(x, a, v)$ である。B_i の部分集合 G_{ij} を次のように定義する。

$$G_{ij}=\{x\in B_i: h_i(x,a_j,v)\le h_i(x,a_k,v),\ \forall k\ne j\}.$$

$P_i=\{G_{ij}: j=1,2,\cdots,p\}$ は B_i の分割であり、$P=\bigcup_{i=1}^{m}P_i$ は Ω の分割となる。この部分 P に関して

$$(H_*v)(x)=h_i(x,a_j,v),\quad \forall x\in G_{ij},$$

が成立する。すべての $x\in G_{ij}$ に対して $\delta(x)=a_j$ として定義される δ は $H_\delta v=H_*v$ なる関係を満足する。$\boxtimes$

系. 2つの部分集合を $V'\subset V$ と $\Delta'\subset\Delta$ とする。もし $v\in V'$ ならば各 n について $H_*^n v\in V'$ である。もし $v\in V'$ で $\delta\in\Delta'$ ならば各 n について $H_\delta^n v\in V'$ である.

3. アルゴリズム

Δ と V の任意の部分集合を Δ' と V' とし，$v_0\ge H_*v_0$ なる $v_0\in V'$ が与えられたとする．$n=0$ とする．ε-最適費用と ε-最適政策を発見するためのアルゴリズムは次のようになる。β は縮少写像の係数である.

ステップ1　$H_{\delta_n}v_n=H_*v_n$ なる $\delta_n\in\Delta'$ を見つけよ.

ステップ2　もし $\|v_n-H_{\delta_n}v_n\|\le(1-\beta)\varepsilon$ ならばステップ4に進め.

ステップ3　もし $\|v_n-H_{\delta_n}v_n\|>(1-\beta)\varepsilon$ ならば、ある整数 k_n を選び $v_{n+1}=H_{\delta_n}^{k_n}v_n$ を計算せよ。n を $n+1$ と置いてステップ1に戻れ.

ステップ4　δ_n は ε-最適政策で，v_n と H_*v_n は ε-最適費用

である。

定理1とその系およびアルゴリズムの内容を整理すれば、次の定理2が得られる。

定理2．もし $v_0 \geq H_* v_0$ なる $v_0 \in V'$ が存在するならば、区分的動的計画法は ε－最適政策と ε－最適費用を持つ．

系．v_nをアルゴリズムによって生成された費用関数の系列とする．

(i) v_nはvに点収束する。

(ii) $v = H_* v$．すなわち，vは最適費用である。

縮少写像は連続であるから系の(ii)が成立する。換言すれば，(i)，(ii)は、アルゴリズムの収束を保証する．アルゴリズムと定理における $v_0 \geq H_* v_0$ なるv_0は、次のような場合存在する．もし、各々のaとvについて $h(\cdot, a, v) \leq 0$ ならば，$v_0 = 0$ は $H_* v_0 = \max_a h(x, a, 0) \leq 0 = v_0$ を満足する。もし $h(x, a, v) = c(x, a) + \beta \int_\Omega v(y) q(dy | x, a)$ ならば，$v_0 = \frac{M}{1-\beta}$ は $H_* v_0 \leq v_0$ を満足する．但し、$M = \max\{|c(x, a)| : x \in \Omega, a \in A\}$ である．

4．まとめ

一般的な単調縮少動的計画法の最適政策と最適費用の正確な関数形を計算することは不可能であるけれども，それらは区分的動的計画法によって近似できることを証明した．前節で述べたわれわれのアルゴリズムは、V'とΔ'に属する費用関

数と政策のみを取り扱えばよいことを示した。V'とΔ'がVとΔと比較して“大変小さい”部分集合（例えば、線形，二次形式，多項式の集合など）であるとき，アルゴリズムは一般的動的計画法の最適費用と最適政策に対してV'とΔ'の要素である費用関数と政策によって近似することを可能にしてくれる。ここで、最適政策と最適費用は必ずしもΔ'とV'の要素ではない。

アルゴリズムのステップ3において、$k_n=1$ならばアルゴリズムは逐次近似法に、$k_n=\infty$ならば政策改良法にそれぞれ相当する。従って、われわれのアルゴリズムはこの両方を特殊な場合として含んでいる。しかしながら、このk_nの最良の選択方法については、明らかではない。多分、繰返計算の初期の段階では小さいk_nを選択し、最適政策に十分近づいた段階では、大きいk_nを選ぶ方が、合理的であろう。

参考文献

[1] Blackwell, D.: Discounted Dynamic Programming. *Annals of Mathematical Statistics*, 36(1965), 226-235.

[2] Denardo, E.V.: Contraction Mappings in the Theory Underlying Dynamic Programming, *SIAM Review* 9(1967), 165-177.

[3] Larson, R.E.: *State Increment Dynamic Programming*, Elsevier, New York

(1968).

[4] Sawaki, K.: Piecewise Linear Markov Decision Processes with An Application into Partially Observable Models, presented at the 1978 International Conference on Markov Decision Processes, University of Manchester (1978).

[5] Sawaki, K. and Ichikawa, A: Optimal Control for Partially Observable Markov Decision Processes Over an Infinite Horizon, *Journal of Operations Research Society of Japan*, Vol. 21, No. 1(1978).

[6] Strauch, R.E.: Negative Dynamic Programming, *Annals of Mathematical Statistics*, 37(1966), 871-890.

[7] Sondik, E.J.: The Optimal Control of Partially Observable Markov Processes Over the Infinite Horizon: Discounted Costs, *Operations Research*, 26(1978), 282-304.