分岐・合流型マルコフ決定過程上での最小型評価について

九州工業大学・大学院工学研究院 * 藤田 敏治
Toshiharu Fujita
Graduate School of Engineering,
Kyushu Institute of Technology

1 はじめに

本稿では，分岐構造および合流構造をもつ決定過程において，最小型評価問題を考える．ここで扱っている推移構造は，ノンシリアル動的計画 ([2, 3, 4, 15]) におけるフィードフォワードループ型に相当する．これまでに，我々は分岐型推移システムに分類される非決定性動的計画 ([13, 6]) や相互依存型決定過程のモデル ([5, 7, 12]) をはじめ，合流型推移システムをもつ決定過程問題への動的計画法 ([1]) によるアプローチを試みてきた ([8, 9, 10, 11]).

ここでは，分岐構造と合流構造の両方を含む推移システム上での確定的決定過程を扱い，評価関数としては，利得関数の最小値をとる最小型評価を採用する．なお，複数に分岐した先の状態の推移においても，各状態の発生順序，言いかえれば決定を決める順序はあらかじめわかっている状況を仮定する．動的計画法による再帰的解法を与えるにあたって，最小型評価を考えた場合は，加法型評価と同様に扱うことはできず，埋め込み法を用いて対処する．

* 〒804-8550 北九州市戸畑区仙水町１－１

2　分岐・合流型推移をもつ決定過程

ここで扱う分岐・合流型推移をもつ決定過程とは，1 つの初期状態から始まり，状態推移を経る過程で分岐と合流を繰り返し，最後は 1 つの終端状態に達して終了するものである．分岐とは 1 つの状態から 2 つ以上の状態へ枝分かれする推移構造をあらわし，逆に合流とは 2 つ以上の状態から 1 つの状態への推移をあらわす．一般には，複数の初期状態から開始される問題が考えられるが，この場合，ダミーの状態を 1 つ追加し，そこから各初期状態へ常に確率 1 で推移する推移を追加することにより，1 つの初期状態を持つ問題への変換が可能である．複数の終端状態がある場合も同様である．

ここで，X を有限状態空間とし，状態変数を $x_1, x_2, \ldots, x_N \in X$ であらわす．x_1 が初期状態，x_N を終端状態とする．U は有限決定空間で $U_n(x_n) \subset U$ は状態 $x_n \in X$ に対し選択可能な決定全体を表し，$r_n : \mathrm{Gr}(U_n) \to \mathbf{R}$, $k : X \to \mathbf{R}$ はぞれぞれ利得関数，終端利得関数とする．なお，$\mathrm{Gr}(U_n)$ は U_n のグラフを表す：

$$\mathrm{Gr}(U_n) = \{(x, u) \in X \times U \mid u \in U_n(x)\}.$$

また，状態 x_i から状態 x_j へ直接推移するとき $e_{ij} = 1$，そうでないときは $e_{ij} = 0$ とおき，行列 $E = (e_{ij})$ を定め，x_j へ推移する状態のインデックス集合を $I_j = \{i \mid e_{ij} = 1\}$ とおく．そして，一般に添え字集合 $I = \{m_1, m_2, \ldots, m_M\}$ $(m_1 < m_2 < \cdots < m_M)$ が与えられた際，簡略化のため，状態・決定の交互列：

$$x_{m_1}, u_{m_1}, x_{m_2}, u_{m_2}, \ldots, x_{m_M}, u_{m_M}$$

を $\langle x_m, u_m \rangle_{m \in I}$ と表すこととする．このとき，状態 x_n へのマルコフ推移確率は，$I_n = \{m_1, m_2, \ldots, m_{M_n}\}$ $(m_1 < m_2 < \cdots < m_{M_n})$ に対し

$$p_n(\cdot \mid x_{m_1}, u_{m_1}, x_{m_2}, u_{m_2}, \ldots, x_{m_M}, u_{m_{M_n}})$$

で与えられ，この確率的推移を $x_n \sim p_n(\cdot \mid \langle x_m, u_m \rangle_{m \in I_n})$ と表す．

さらに，任意の $n \geq 2$ に対し，

$$i \leq n \quad (i \in I_n)$$

すなわち，推移元状態の添え字の値より推移先状態の添え字の値が大きいことを仮定し，状態が確定する順序は添え字の順とする．もしそうでない場合には，適当に添え字を付け替えることで対応可能であるため，この過程は一般性を失うものではない．

一般に決定過程問題における政策クラスはマルコフ政策全体を考える場合が多いが，ここで扱う問題はマルコフ政策の中に最適政策が存在するとは限らないため，一般政策の概念を導入する．一般政策 $\sigma = \{\sigma_1,\ \sigma_2, \ldots,\ \sigma_{N-1}\}$：

$$\sigma_1 : X \to U, \quad \sigma_2 : X \times X \to U, \ldots,\ \sigma_{N-1} : X^{N-1} \to U$$

は,

$$u_1 = \sigma_1(x_1),\ \ u_2 = \sigma_2(x_1,\ x_2), \ldots,\ u_{N-1} = \sigma_{N-1}(x_1,\ x_2, \ldots,\ x_{N-1})$$

のように状態履歴に依存して決定を与える決定関数 σ_n からなる政策である．

このとき，初期状態 x_1 に対する次の最小型評価最大化問題を考える：

$$\begin{array}{lll} \text{(P)} & \text{Maximize} & E\big[r_1(x_1, u_1) \wedge r_2(x_2, u_2) \wedge \cdots \wedge r_{N-1}(x_{N-1}, u_{N-1}) \wedge k(x_N)\big] \\ & \text{subject to} & x_{n+1} \sim p_n(\,\cdot \,|\, \langle x_m, u_m\rangle_{m \in I_n}) \qquad n = 1,\ 2, \ldots,\ N \\ & & \sigma = \{\sigma_1,\ \sigma_2, \ldots,\ \sigma_{N-1}\} \in \Sigma. \end{array}$$

ただし，2 項演算子 $\wedge$ は $a \wedge b = \min(a,\, b)$ $(a,\, b \in \mathbf{R})$ で定義され，Σ は一般政策全体を表す.

例 2.1　$N = 5$ とし $e_{12} = e_{14} = e_{23} = e_{35} = e_{45} = 1$ (それ以外の $(i,\, j)$ に対しては $e_{ij} = 0$) とする．初期状態 x_1 が与えられたとき，x_2, x_3, x_4, x_5 は次のマルコフ推移法則により確率的に定まる：

$$\begin{aligned} & x_2 \sim p_2(\,\cdot \,|\, x_1,\, u_1), \ \ x_1 \in X,\ u_1 \in U_1(x_1) \\ & x_3 \sim p_3(\,\cdot \,|\, x_2,\, u_2), \ \ x_2 \in X,\ u_2 \in U_2(x_2) \\ & x_4 \sim p_4(\,\cdot \,|\, x_1,\, u_1), \ \ x_1 \in X,\ u_1 \in U_1(x_1) \\ & x_5 \sim p_5(\,\cdot \,|\, x_3,\, u_3,\, x_4,\, u_4), \ \ x_3 \in X,\ x_4 \in X,\ u_3 \in U_3(x_3),\ u_4 \in U_4(x_4). \end{aligned}$$

図 1 は，この分岐合流型推移をもつ決定過程の状態推移図を示したものである．このとき，最小型評価最大化問題は次で与えられる：

$$\begin{array}{ll} \text{Maximize} & E\big[r_1(x_1, u_1) \wedge r_2(x_2, u_2) \wedge r_3(x_3, u_3) \wedge r_4(x_4, u_4) \wedge k(x_5)\big] \\ \text{subject to} & x_2 \sim p_2(\,\cdot \,|\, x_1, u_1), \ \ x_3 \sim p_3(\,\cdot \,|\, x_2, u_2), \\ & x_4 \sim p_4(\,\cdot \,|\, x_1, u_1), \ \ x_5 \sim p_5(\,\cdot \,|\, x_3, u_3, x_4, u_4) \\ & \sigma = \{\sigma_1, \sigma_2, \sigma_3, \sigma_4\} \in \Sigma. \end{array}$$

ただし，各決定は

$$\begin{aligned} & u_1 = \sigma_1(x_1) \in U_1(x_1),\ u_2 = \sigma_2(x_1, x_2) \in U_2(x_2), \\ & u_3 = \sigma_3(x_1, x_2, x_3) \in U_3(x_3),\ u_4 = \sigma_1(x_1, x_2, x_3, x_4) \in U_4(x_4) \end{aligned}$$

により与えられる.

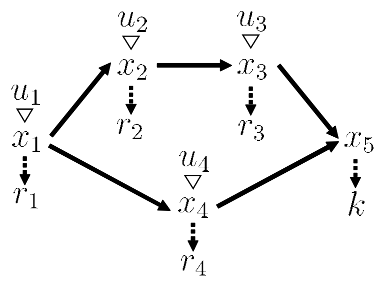

図 1　状態推移図

3　埋め込み問題

問題 (P) に対し，実パラメータ $\lambda \in \mathbf{R}$ を埋め込んだ次の問題を考える：

$$
\begin{array}{lll}
\text{(IP)} & \text{Maximize} & E\big[r_1(x_1, u_1) \wedge r_2(x_2, u_2) \wedge \cdots \wedge r_{N-1}(x_{N-1}, u_{N-1}) \wedge k(x_N)\big] \\
& \text{subject to} & x_{n+1} \sim p_n(\,\cdot \mid \langle x_m, u_m \rangle_{m \in I_n}) \qquad n = 1,\ 2, \ldots,\ N \\
& & \sigma = \{\sigma_1,\ \sigma_2, \ldots,\ \sigma_{N-1}\} \in \Sigma.
\end{array}
$$

問題 (IP) において $\lambda = \infty$ とおくと問題 (P) に一致することは明らかで（実用上は，十分大きな値を λ に代入すればよい，問題 (IP) を埋め込み問題と呼ぶ.

4　部分問題群

埋め込み問題 (IP) に対し，x_n から開始する部分問題群を次のように構成し，最適値関数を v^n であらわす：

$$
\begin{aligned}
& v^N(x_N, \lambda) \;=\; \lambda \wedge k(x_N), \qquad\qquad x_N \in X,\ \ \lambda \in \mathbf{R}, \\
& v^n(x_n, \lambda; \langle x_m, u_m \rangle_{m \in J_n}) \\
& = \max_{\{\sigma_n, \ldots, \sigma_{N-1}\} \in \Sigma_n} E\big[\lambda \wedge r_n(x_n, u_n) \wedge r_{n+1}(x_{n+1}, u_{n+1}) \wedge \\
& \qquad\qquad\qquad\qquad\qquad\qquad \cdots \wedge r_{N-1}(x_{N-1}, u_{N-1}) \wedge k(x_N)\big], \\
& \qquad\qquad\qquad x_n \in X,\ \ \lambda \in \mathbf{R}, \quad x_m \in X,\ u_m \in U_m(x_m)\ (m \in J_n).
\end{aligned}
$$

ただし，

$$
J_n \;=\; \bigcup_{l=n+1}^{N} \{j \in I_l \mid j < n\}
$$

であり，以後，$J_n = \phi$ のときは，$\langle x_m, u_m \rangle_{m \in J_n} = \phi$ と解釈し，

$$
v^n(x_n, \lambda; \langle x_m, u_m \rangle_{m \in J_n}) = v^n(x_n, \lambda; \phi) = v^n(x_n, \lambda)
$$

とみなす（3 種の表現は同一のものとみなす）．なお，Proposition 3.1 ([10]) により J_n を再帰的に構成する方法が与えられている．

上記の部分問題における一般政策について補足する．一般政策全体を表す Σ_n は，状態列 $(x_n,\ x_{n+1},\dots,\ x_N)$ に対応する問題についてのもので, $\sigma=\{\sigma_n,\ \sigma_{n+1},\dots,\ \sigma_{N-1}\}\in\Sigma_n$ は次の決定関数からなる：

$$\sigma_n : X\to U,\quad \sigma_{n+1} : X\times X\to U,\dots,\quad \sigma_{N-1} : X^{N-n}\to U.$$

すなわち，決定は

$$u_n=\sigma_n(x_n),\quad u_{n+1}=\sigma_{n+1}(x_n,\ x_{n+1}),\dots,\ u_{N-1}=\sigma_{N-1}(x_n,\ x_{n+1},\dots,\ x_{N-1})$$

と定まる．

5 再帰式

値関数 $v^n(\cdot)$ に対し，次の再帰式を得る：

定理 5.1

$$\begin{aligned}
v^N(x_N,\lambda)\ &=\ \lambda\wedge k(x_N), \qquad\quad x_N\in X,\quad \lambda\in\mathbf{R},\\
v^n(x_n,\lambda;\,&\langle x_m,u_m\rangle_{m\in J_n})\\
=\ \max_{u_n}\sum_{x_{n+1}} &v^{n+1}(x_{n+1},\lambda\wedge r_n(x_n,u_n);\,\langle x_m,u_m\rangle_{m\in J_{n+1}})p_{n+1}(x_{n+1}|\,\langle x_m,u_m\rangle_{m\in I_{n+1}}),\\
&x_n\in X,\quad \lambda\in\mathbf{R},\quad x_m\in X,\ u_m\in U(x_m)\ (m\in J_n),\\
&\qquad\qquad n=1,\ 2,\dots,\ N-1.
\end{aligned}$$

証明　値関数の定義より

$$\begin{aligned}
&v^n(x_n,\lambda;\,\langle x_m,u_m\rangle_{m\in J_n}) \qquad (1)\\
&=\max_{\{\sigma_n,\dots,\sigma_{N-1}\}\in\Sigma_n}\sum_{x_{n+1},\dots,x_N}\big[\lambda\wedge r_n(x_n,u_n)\wedge r_{n+1}(x_{n+1},u_{n+1})\wedge\\
&\qquad\cdots\wedge r_{N-1}(x_{N-1},u_{N-1})\wedge k(x_N)\big]p_{n+1}(x_{n+1}|\,\langle x_m,u_m\rangle_{m\in I_{n+1}})\\
&\qquad\qquad\cdots p_N(x_N|\,\langle x_m,u_m\rangle_{m\in I_N}).
\end{aligned}$$

ただし,

$$u_l=\sigma_l(x_n,\ x_{n+1},\dots,\ x_l)\qquad (l=n,\ n+1,\dots,\ N-1).$$

式 (1) 右辺の最大値を与える最適一般政策を $\{\hat{\sigma}_n,\ \hat{\sigma}_{n+1},\ldots,\ \hat{\sigma}_{N-1}\} \in \Sigma_n$ とする．式 (1) においては x_n が固定されているので，この最適一般政策により定まる決定は

$$\hat{u}_n = \hat{\sigma}_n(x_n),\ \hat{u}_{n+1}(x_{n+1}) = \hat{\sigma}_{n+1}(x_n, x_{n+1}),\\ \ldots,\ \hat{u}_{N-1}(x_{n+1}, \ldots x_{N-1}) = \hat{\sigma}_{N-1}(x_n, \ldots, x_{N-1})$$

と表すことができる．よって

$$\begin{aligned}
&(\text{ 式 (1) 右辺 }) \qquad (2)\\
&= \sum_{x_{n+1},\ldots,x_N} \big[\lambda \wedge r_n(x_n, \hat{u}_n) \wedge r_{n+1}(x_{n+1}, \hat{u}_{n+1}(x_{n+1})) \wedge \\
&\qquad \cdots \wedge r_{N-1}(x_{N-1}, \hat{u}_{N-1}(x_{n+1}, \ldots x_{N-1})) \wedge k(x_N)\big] p_{n+1}(x_{n+1} |\, \langle x_m, u_m \rangle_{m \in I_{n+1}}) \\
&\qquad \cdots p_N(x_N |\, \langle x_m, u_m \rangle_{m \in I_N}) \\
&= \sum_{x_{n+1}} \Big\{ \sum_{x_{n+2},\ldots,x_N} \big[(\lambda \wedge r_n(x_n, \hat{u}_n)) \wedge r_{n+1}(x_{n+1}, \hat{u}_{n+1}(x_{n+1})) \wedge \\
&\qquad \cdots \wedge r_{N-1}(x_{N-1}, \hat{u}_{N-1}(x_{n+1}, \ldots x_{N-1})) \wedge k(x_N)\big] p_{n+2}(x_{n+2} |\, \langle x_m, u_m \rangle_{m \in I_{n+2}}) \\
&\qquad \cdots p_N(x_N |\, \langle x_m, u_m \rangle_{m \in I_N}) \Big\} p_{n+1}(x_{n+1} |\, \langle x_m, u_m \rangle_{m \in I_{n+1}}) \\
&\leq \sum_{x_{n+1}} \Big\{ \max_{\{\sigma_{n+2},\ldots,\sigma_{N-1}\} \in \Sigma_{n+1}} \sum_{x_{n+1},\ldots,x_N} \big[(\lambda \wedge r_n(x_n, \hat{u}_n)) \wedge r_{n+1}(x_{n+1}, u_{n+1}) \wedge \\
&\qquad \cdots \wedge r_{N-1}(x_{N-1}, u_{N-1}) \wedge k(x_N)\big] p_{n+2}(x_{n+2} |\, \langle x_m, u_m \rangle_{m \in I_{n+2}}) \\
&\qquad \cdots p_N(x_N |\, \langle x_m, u_m \rangle_{m \in I_N}) \Big\} p_{n+1}(x_{n+1} |\, \langle x_m, u_m \rangle_{m \in I_{n+1}}) \\
&\qquad \big(u_{n+1} = \sigma_{n+1}(x_{n+1}), \ldots,\ u_{N-1} = \sigma_{N-1}(x_{n+1}, \ldots,\ x_{N-1})\big) \\
&= \sum_{x_{n+1}} v^{n+1}(x_{n+1}, \lambda \wedge r_n(x_n, \hat{u}_n);\ \langle x_m, u_m \rangle_{m \in J_{n+1}}) p_{n+1}(x_{n+1} |\, \langle x_m, u_m \rangle_{m \in I_{n+1}}).
\end{aligned}$$

上記の式変形においては，$n \in J_{n+1}$ のとき，$\langle x_m, u_m \rangle_{m \in J_{n+1}}$ 内の u_n は $\hat{u}_n$ と置き換えるものとし，$n \in I_l$ $(l = n+1,\ n+2, \cdots,\ N)$ のときも同様に置き換える．

式 (1), (2) より，

$$\begin{aligned}
&v^n(x_n, \lambda;\ \langle x_m, u_m \rangle_{m \in J_n}) \qquad (3)\\
&\leq \sum_{x_{n+1}} v^{n+1}(x_{n+1}, \lambda \wedge r_n(x_n, \hat{u}_n);\ \langle x_m, u_m \rangle_{m \in J_{n+1}}) p_{n+1}(x_{n+1} |\, \langle x_m, u_m \rangle_{m \in I_{n+1}}) \\
&\leq \max_{u_n} \sum_{x_{n+1}} v^{n+1}(x_{n+1}, \lambda \wedge r_n(x_n, u_n); \langle x_m, u_m \rangle_{m \in J_{n+1}}) p_{n+1}(x_{n+1} |\, \langle x_m, u_m \rangle_{m \in I_{n+1}}).
\end{aligned}$$

次に逆向きの不等式を示す．$\overline{u}_n \in U(x_n)$ が存在し次をみたす．

$$\max_{u_n} \sum_{x_{n+1}} v^{n+1}(x_{n+1}, \lambda \wedge r_n(x_n, u_n); \langle x_m, u_m \rangle_{m \in J_{n+1}}) p_{n+1}(x_{n+1} |\, \langle x_m, u_m \rangle_{m \in I_{n+1}}) \qquad (4)$$

$$= \sum_{x_{n+1}} v^{n+1}(x_{n+1}, \lambda \wedge r_n(x_n, \overline{u}_n);\ \langle x_m, u_m\rangle_{m\in J_{n+1}})p_{n+1}(x_{n+1}|\ \langle x_m, u_m\rangle_{m\in I_{n+1}}).$$

ただし，以後の式変形も含めて，右辺に含まれる u_n は必要に応じて $\overline{u}_n$ とみなす.

部分問題の定義より

$$\begin{aligned}
&v^{n+1}(x_{n+1}, \lambda \wedge r_n(x_n, \overline{u}_n);\ \langle x_m, u_m\rangle_{m\in J_{n+1}}) \qquad (5)\\
&= \max_{\{\sigma_{n+1},\ldots,\sigma_{N-1}\}\in\Sigma_{n+1}} \sum_{x_{n+2},\ldots,x_N} \big[(\lambda \wedge r_n(x_n, \overline{u}_n)) \wedge r_{n+1}(x_{n+1}, u_{n+1}) \wedge \\
&\qquad \cdots \wedge r_{N-1}(x_{N-1}, u_{N-1}) \wedge k(x_N)\big] p_{n+2}(x_{n+2}|\ \langle x_m, u_m\rangle_{m\in I_{n+2}}) \\
&\qquad \cdots p_N(x_N|\ \langle x_m, u_m\rangle_{m\in I_N}).
\end{aligned}$$

$\{\overline{\sigma}_{n+1}, \ldots,\ \overline{\sigma}_{N-1}\} \in \Sigma_{n+1}$ を上式の最適政策とするとき

$$\begin{aligned}
&(\text{ 式 (5) 右辺 }) \qquad (6)\\
&= \sum_{x_{n+2},\ldots,x_N} \big[(\lambda \wedge r_n(x_n, \overline{u}_n)) \wedge r_{n+1}(x_{n+1}, \overline{\sigma}_{n+1}(x_{n+1})) \wedge \\
&\qquad \cdots \wedge r_{N-1}(x_{N-1}, \overline{\sigma}_{N-1}(x_{n+1}, \ldots x_{N-1})) \wedge k(x_N)\big] p_{n+2}(x_{n+2}|\ \langle x_m, u_m\rangle_{m\in I_{n+2}}) \\
&\qquad \cdots p_N(x_N|\ \langle x_m, u_m\rangle_{m\in I_N}).
\end{aligned}$$

式 (4), (5), (6) より

$$\begin{aligned}
&\max_{u_n} \sum_{x_{n+1}} v^{n+1}(x_{n+1}, \lambda \wedge r_n(x_n, u_n);\ \langle x_m, u_m\rangle_{m\in J_{n+1}})p_{n+1}(x_{n+1}|\ \langle x_m, u_m\rangle_{m\in I_{n+1}}) \quad (7)\\
&= \sum_{x_{n+1}} \Big\{ \sum_{x_{n+2},\ldots,x_N} \big[(\lambda \wedge r_n(x_n, \overline{u}_n)) \wedge r_{n+1}(x_{n+1}, \overline{\sigma}_{n+1}(x_{n+1})) \wedge \\
&\qquad \cdots \wedge r_{N-1}(x_{N-1}, \overline{\sigma}_{N-1}(x_{n+1}, \ldots x_{N-1})) \wedge k(x_N)\big] p_{n+2}(x_{n+2}|\ \langle x_m, u_m\rangle_{m\in I_{n+2}}) \\
&\qquad \cdots p_N(x_N|\ \langle x_m, u_m\rangle_{m\in I_N})\Big\} p_{n+1}(x_{n+1}|\ \langle x_m, u_m\rangle_{m\in I_{n+1}}) \\
&= \sum_{x_{n+1},\ldots,x_N} (\lambda \wedge r_n(x_n, \overline{u}_n) \wedge r_{n+1}(x_{n+1}, \overline{\sigma}_{n+1}(x_{n+1})) \wedge \\
&\qquad \cdots \wedge r_{N-1}(x_{N-1}, \overline{\sigma}_{N-1}(x_{n+1}, \ldots, x_{N-1})) \wedge k(x_N)\big] p_{n+1}(x_{n+1}|\ \langle x_m, u_m\rangle_{m\in I_{n+1}}) \\
&\qquad \cdots p_N(x_N|\ \langle x_m, u_m\rangle_{m\in I_N}) \\
&\le \max_{\{\sigma_n,\ldots,\sigma_{N-1}\}\in\Sigma_n} \sum_{x_{n+1},\ldots,x_N} \big[(\lambda \wedge r_n(x_n, \sigma_n(x_n))) \wedge r_{n+1}(x_{n+1}, \sigma_{n+1}(x_n, x_{n+1})) \wedge \\
&\qquad \cdots \wedge r_{N-1}(x_{N-1}, \sigma_{N-1}(x_n, \ldots, x_{N-1})) \wedge k(x_N)\big] p_{n+1}(x_{n+1}|\ \langle x_m, u_m\rangle_{m\in I_{n+1}}) \\
&\qquad \cdots p_N(x_N|\ \langle x_m, u_m\rangle_{m\in I_N}) \\
&= v^n(x_n, \lambda;\ \langle x_m, u_m\rangle_{m\in J_n}).
\end{aligned}$$

よって，式 (3) と (7) により

$$v^n(x_n,\lambda;\,\langle x_m,u_m\rangle_{m\in J_n})$$
$$= \max_{u_n}\sum_{x_{n+1}} v^{n+1}(x_{n+1},\lambda\wedge r_n(x_n,u_n);\,\langle x_m,u_m\rangle_{m\in J_{n+1}})p_{n+1}(x_{n+1}|\,\langle x_m,u_m\rangle_{m\in I_{n+1}})$$

が成り立つ． □

例 5.1　例 2.1 の問題に対応する埋め込み問題を考え，部分問題を構成した結果は以下で与えられる：

$$\begin{aligned}
v^5(x_5,\lambda) &= \lambda\wedge k(x_5)\\
v^4(x_4,\lambda;\,x_3,u_3) &= \max_{\{\sigma_4\}\in\Sigma_4} E[\lambda\wedge r_4(x_4,u_4)\wedge k(x_5)]\\
v^3(x_3,\lambda;\,x_1,u_1) &= \max_{\{\sigma_3,\sigma_4\}\in\Sigma_3} E[\lambda\wedge r_3(x_3,u_3)\wedge r_4(x_4,u_4)\wedge k(x_5)]\\
v^2(x_2,\lambda;\,x_1,u_1) &= \max_{\{\sigma_2,\sigma_3,\sigma_4\}\in\Sigma_2} E[\lambda\wedge r_2(x_2,u_2)\wedge r_3(x_3,u_3)\wedge r_4(x_4,u_4)\wedge k(x_5)]\\
v^1(x_1,\lambda) &= \max_{\{\sigma_1,\sigma_2,\sigma_3,\sigma_4\}\in\Sigma} E[\lambda\wedge r_1(x_1,u_1)\wedge r_2(x_2,u_2)\wedge r_3(x_3,u_3)\\
&\qquad\qquad \wedge r_4(x_4,u_4)\wedge k(x_5)].
\end{aligned}$$

この値関数 v^n について以下の再帰式が成り立つ：

$$\begin{aligned}
v^5(x_5,\lambda) &= \lambda\wedge k(x_5)\\
v^4(x_4,\lambda;\,x_3,u_3) &= \max_{u_4}\sum_{x_5} v^5(x_5,\lambda\wedge r_4(x_4,u_4))p_5(x_5\,|\,x_3,u_3,x_4,u_4)\\
v^3(x_3,\lambda;\,x_1,u_1) &= \max_{u_3}\sum_{x_4} v^4(x_4,\lambda\wedge r_3(x_3,u_3);\,x_3,u_3)p_4(x_4\,|\,x_1,u_1)\\
v^2(x_2,\lambda;\,x_1,u_1) &= \max_{u_2}\sum_{x_3} v^3(x_3,\lambda\wedge r_2(x_2,u_2);\,x_1,u_1)p_3(x_3\,|\,x_2,u_2)\\
v^1(x_1,\lambda) &= \max_{u_1}\sum_{x_2} v^2(x_2,\lambda\wedge r_1(x_1,u_1);\,x_1,u_1)p_2(x_2\,|\,x_1,u_1).
\end{aligned}$$

6　最適政策

ここでは，埋め込み問題 (IP) に対する結果から，どのようにして元の問題 (P) に対する最適一般政策を構成するかについて考える．

まず，埋め込み問題 (IP) に対するパラメータ付き最適マルコフ決定関数 $\pi^*_n(x_n,\lambda;\langle x_m,u_m\rangle_{m\in J_n})$ $(n=1,\,2,\ldots,\,N-1)$ の値を，最大値 $v^n(x_n,\lambda;\langle x_m,u_m\rangle_{m\in J_n})$

を与える最適決定と定義する．すなわち

$$
\begin{aligned}
&\pi_n^*(x_n, \lambda; \langle x_m, u_m\rangle_{m\in J_n}) \\
&\in \ \underset{u_n}{\operatorname{argmax}} \sum_{x_{n+1}} v^{n+1}(x_{n+1}, \lambda \wedge r_n(x_n, u_n);\ \langle x_m, u_m\rangle_{m\in J_{n+1}}) \\
&\qquad\qquad\qquad\qquad\qquad\qquad \times p_{n+1}(x_{n+1} |\ \langle x_m, u_m\rangle_{m\in I_{n+1}}).
\end{aligned}
$$

このとき，与えられた初期状態 x_1 に対し，最適一般政策 $\sigma^* = \{\sigma_1^*,\ \sigma_2^*,\ \ldots,\ \sigma_{N-1}^*\}$ は次で与えられる．

$$
\begin{aligned}
&\lambda_1 = \infty, \\
&\qquad \sigma_1^*(x_1) = \pi_1^*(x_1, \lambda_1), \\
&\lambda_2 = \lambda_1 \wedge r_1(x_1, u_1^*),\ u_1^* = \sigma_1^*(x_1), \\
&\qquad \sigma_2^*(x_1, x_2) = \pi_2^*(x_2, \lambda_2;\ \langle x_m, u_m^*\rangle_{m\in J_2}), \\
&\quad \vdots \\
&\lambda_n = \lambda_{n-1} \wedge r_{n-1}(x_{n-1}, u_{n-1}^*),\ u_{n-1}^* = \sigma_{n-1}^*(x_1, x_2, \ldots,\ x_{n-1}), \\
&\qquad \sigma_n^*(x_1, x_2, , \ldots, x_n) = \pi_n^*(x_n, \lambda_n;\ \langle x_m, u_m^*\rangle_{m\in J_n}), \\
&\quad \vdots \\
&\lambda_{N-1} = \lambda_{N-2} \wedge r_{N-2}(x_{N-2}, u_{N-2}^*),\ u_{N-2}^* = \sigma_{N-2}^*(x_1, x_2, \ldots, x_{N-2}), \\
&\qquad \sigma_{N-1}^*(x_1, x_2, \ldots, x_{N-1}) = \pi_{N-1}^*(x_{N-1}, \lambda_{N-1};\ \langle x_m, u_m^*\rangle_{m\in J_{N-1}}).
\end{aligned}
$$

参考文献

[1] R.E. Bellman, Dynamic Programming, Princeton Univ. Press, NJ, 1957.

[2] U. Bertelé and F. Brioschi, Nonserial Dynamic Programming, Academic Press, New York, 1972.

[3] A.O. Esogbue and N.A. Warsi, A High-level Computing Algorithm for Diverging and Converging Branch Nonserial Dynamic Programming Systems, Computers and Mathematics with Applications, 12 (1986), 719–732.

[4] A.O. Esogbue, A High Level Dynamic Programming Algorithm for Processing Nonserial Looped System, Computers and Mathematics with Applications, 16 (1988), 801–813.

[5] T. Fujita, Mutually Dependent Markov Decision Processes, Journal of Advanced Computational Intelligence and Intelligent Informatics, 18 (2014), 992-998.

[6] T. Fujita, Sure Ways to win a Game — Nondeterministic Dynamic Programming Approach —, Bulletin of the Kyushu Institute of Technology, 62 (2015), 1–14.

[7] T. Fujita, Mutually Dependent Decision Processes Models, Bulletin of the Kyushu Institute of Technology, 63 (2016), 15–26.

[8] T. Fujita, Three Recursive Approaches for Decision Processes with a Converging Branch System, Bulletin of the Kyushu Institute of Technology, 65 (2018), 1–21.

[9] T. Fujita, On Markov Decision Process with Converging Branch System (in Japanese), RIMS Kokyuroku (Kyoto Univ.) 2158 (2020), 184–194.

[10] T. Fujita, Converging Decision Processes with Multiplicative Reward System, Bulletin of the Kyushu Institute of Technology, 68 (2021), 9–26.

[11] T. Fujita, Converging Markov Decision Processes with Multiplicative Reward System, Bulletin of the Kyushu Institute of Technology, 70 (2023), 33–41.

[12] T. Fujita and A. Kira, Associative Criteria in Mutually Dependent Markov Decision Processes, Proceedings of IIAI International Conference on Advanced Applied Informatics (2014), 147-150.

[13] T. Fujita, T. Ueno, and S. Iwamoto, A Nondeterministic Dynamic Programming Model, Lecture Notes in Artificial Intelligence 3214 (2004), 1208–1214.

[14] S. Iwamoto, Associative Dynamic Programs. Journal of Mathematical Analysis and Applications, 201 (1996), 195-211.

[15] G.L. Nemhauser, Introduction to Dynamic Programming, Wiley, New York, 1966.