# 再帰的方法による複合評価系の最適化

岩本, 誠一
九州大学経済学部 : 教授

https://doi.org/10.15017/4363004

出版情報：經濟學研究. 66 (1), pp.51-66, 1999-06-30. 九州大学経済学会

# 再帰的方法による複合評価系の最適化

岩　本　誠　一

## 1．はじめに

本論文では，多段確定的システムにおいて複数の単一評価から合成された複合評価系の最適化を考える。いわゆる動的計画法は利得系の再帰性と単調性の下で適用されるが，ここでは単調性を仮定しない。かわりに結合性を用いて確定的最適化の再帰的方法を考える。

再帰性は終端状態モデル化の過程で結合性から導かれることになる。複合評価系は加法型，最小型など，単一評価系を含む汎用性のある基準である。通常の加法型基準では暗黙裡にマルコフ政策クラスが用いられている。しかし，複合評価系ではマルコフクラスに最適政策が存在するとは限らない。したがって，本論文では，より広い「一般政策クラス」において再帰的方法で最適政策を導けることを示す。その基本原理はいわゆる不変埋没原理に基づく。ここでは，この原理をできるだけ小さな範囲で適用するという意味で新たに「極小不変埋没原理」を提唱している。この極小不変埋没方法は計算量を可能な限り少なくしているので，極めて有用である。第２節では複合最適化問題を一般クラスにおいて定式化する。この定式化の妥当性を第３節で，既存の単一評価型のマルコフクラスでの定式化と対比しながら述べる。

第４節では一般クラスに極小不変埋没原理を用いて複合最適化問題を終端型評価系をもつ拡大問題に変換する。ここで最適化問題を考察する上で，複合問題に対する一般クラスと終端問題に対するマルコフクラスの同値関係を導く。

第５節では両問題に対する再帰式を導き，同値性に基づいて複合最適化問題の最適一般政策を再帰的な方法で求める。

## 2．複合評価系

本論文では有限**状態空間** $X$ および有限**決定空間** $U$ 上の有限段逐次決定過程 $S$ を考える。決定過程 $S$ は以下の要素で構成されている：

$r, R : X \times U \to R^1$ **利得関数**

$k, K : X \to R^1$ **終端関数**

$T : X \times U \to X$ **状態変換**

$$\circ, \diamond : R^1 \times R^1 \to R^1 \text{ それぞれ左単位元 } \tilde{\lambda}, \tilde{\mu} \text{ をもつ2項結合演算} \tag{1}$$
$$\psi : R^1 \times R^1 \to R^1 \text{ 複合関数。}$$

また，以下の簡略記号を用いる：

$$r_n := r(x_n, u_n),\ R_n := R(x_n, u_n),\ k := k(x_{N+1}),\ K := K(x_{N+1})$$
$$H^{n+1} := X \times U \times X \times U \times \cdots \times X \times U \times X \ (2n+1)\text{個} \tag{2}$$
$$h_{N+1} := (x_1, u_1, x_2, u_2, \ldots, x_{N+1}) \in H^{N+1}$$
$$X^n := X \times X \times \cdots \times X \ (n\text{回})$$

さて，政策を導入しよう。**一般政策**$\sigma$とは**第n一般決定関数**$\sigma_n : X^n \to U$から成る列 $\sigma = \{\sigma_1, \sigma_2, \ldots, \sigma_N\}$をいい，一般政策全体を$\Pi_g$で表わす。特に，一般決定関数が最後(現在)の状態のみに依存して**第nマルコフ決定関数**$\pi_n : X \to U$になるとき，列$\pi = \{\pi_1, \pi_2, \ldots, \pi_N\}$を**マルコフ政策**といい，マルコフ政策全体を$\Pi(\subset \Pi_g)$で表わす。

このとき，次のような複合評価系の確定的最適化問題を一般政策のクラス$\Pi_g$の中で考える：

$$\mathrm{P}_1(x_1): \quad \begin{array}{ll} \text{Optimize} & \psi(r_1 \circ r_2 \circ \cdots \circ r_N \circ k,\ R_1 \diamond R_2 \diamond \cdots \diamond R_N \diamond K) \\ \text{subject to} & (\mathrm{i})_n\ x_{n+1} = T(x_n, u_n) \quad\quad 1 \le n \le N. \\ & (\mathrm{ii})_n\ u_n \in U \end{array} \tag{3}$$

問題$\mathrm{P}_1(x_1)$の評価は一般政策$\sigma = \{\sigma_1, \sigma_2, \ldots, \sigma_N\}$および初期状態$x_1$に依存して定まる値である。この評価値を$I_1(\sigma; x_1)$で表わす：

$$I_1(\sigma; x_1) = \psi(r_1 \circ r_2 \circ \cdots \circ r_N \circ k,\ R_1 \diamond R_2 \diamond \cdots \diamond R_N \diamond K) \tag{4}$$

ここに第$n$段利得$r_n = r(x_n, u_n), R_n = R(x_n, u_n)$ $(1 \le n \le N)$の決定$u_n$はそれまでの状態列$x_1, \ldots, x_n$に依存して第$n$一般決定関数$\sigma_n$によって決まっている：

$$u_1 = \sigma_1(x_1),\ u_2 = \sigma_2(x_1, x_2),\ \ldots,\ u_N = \sigma_N(x_1, x_2, \ldots, x_N). \tag{5}$$

問題$\mathrm{P}_1(x_1)$では逐次条件

$$(\mathrm{i})_n\ x_{n+1} = T(x_n, u_n),\ (\mathrm{ii})_n\ u_n \in U \quad 1 \le n \le N$$

を明示してダイナミックに（システムの状態推移を動的にして）最適化問題を表現しているが，1時点で単に一般政策を選ぶ静的な(時間的経過を圧縮した)最適化問題として

$$\text{Optimize} \quad I_1(\sigma; x_1) \quad \text{subject to} \quad (\mathrm{i})^*\ \sigma \in \Pi_g \tag{6}$$

で表わすこともできる。

## 3. マルコフ政策クラス

この節では複合型評価問題を一般政策クラスで考えることの妥当性を議論する。このため，いわゆる加法型，最小型，終端型などの「単調性をもつ結合型評価問題」ではマルコフ政策クラスの中で最適解が再帰的な方法で求められることと対比する。

### 3.1 複合型評価問題

さて，マルコフ政策クラス$\Pi$の中で複合型評価問題：

$$\mathrm{P}_1(x_1):\quad \begin{array}{ll} \text{Optimize} & \psi(r_1\circ r_2\circ\cdots\circ r_N\circ k, R_1\diamond R_2\diamond\cdots\diamond R_N\diamond K) \\ \text{subject to} & (\mathrm{i})_\mathrm{n}\ x_{n+1}=T(x_n,u_n) \\ & (\mathrm{ii})_\mathrm{n}\ u_n\in U \end{array} \quad 1\le n\le N \tag{7}$$

を考えてみよう。この節での問題 $\mathrm{P}_1(x_1)$の評価値 $I_1(\pi;x_1)$はマルコフ政策$\pi=\{\pi_1,\pi_2,\ldots,\pi_N\}$に依存している。すなわち

$$I_1(\pi;x_1)=\psi(r_1\circ r_2\circ\cdots\circ r_N\circ k,\ R_1\diamond R_2\diamond\cdots\diamond R_N\diamond K) \tag{8}$$

の決定列 $\{u_n\}$ は直前の状態のみで定まっている：

$$u_1=\pi_1(x_1),\ u_2=\pi_2(x_2),\ \ldots,\ u_N=\pi_N(x_N). \tag{9}$$

この問題は静的最適化問題で表わすと

$$\text{Optimize}\quad I_1(\pi;x_1)\quad \text{subject to}\quad (\mathrm{i})\ \pi\in\Pi \tag{10}$$

になる。

特に，複合関数$\psi$が**単一関数**$\psi=\psi(r,R)=r$になると，問題(7)の目的関数は結合型

$$r(x_1,u_1)\circ\cdots\circ r(x_N,u_N)\circ k(x_{N+1}) \tag{11}$$

になる。さらに，結合演算$\circ$の二つの場合；(i)最小演算：$\circ=\wedge$，(ii)加法演算：$\circ=+$を考えよう。

(i)最小型：$\circ=\wedge$ のとき，(11)は最小型評価値

$$r(x_1,u_1)\wedge\cdots\wedge r(x_N,u_N)\wedge k(x_{N+1}) \tag{12}$$

になる。最小型評価はファジィ理論で現れる基準である([2][4][16][17][18][21][23][24][32])。

(ii)加法型：$\circ=+$ では，(11)はいわゆる総和

$$r(x_1,u_1)+\cdots+r(x_N,u_N)+k(x_{N+1}) \tag{13}$$

である。言うまでもなく大多数の最適化問題は加法型である（参考文献省略)。さらに再帰加法型，一般再帰型の最適解も研究されている([5][6][7][8][11][19][27][28][29][30][34])。

### 3.2 加法型問題

さて，複合型評価問題の特別な場合として，マルコフ政策クラス$\Pi$の中で加法型問題

$$\mathrm{A}_1(x_1):\quad \begin{array}{ll} \text{Optimize} & r(x_1,u_1)+\cdots+r(x_N,u_N)+k(x_{N+1}) \\ \text{subject to} & (\mathrm{i})_\mathrm{n}\ x_{n+1}=T(x_n,u_n) \\ & (\mathrm{ii})_\mathrm{n}\ u_n\in U \end{array} \quad 1\le n\le N \tag{14}$$

を考えよう。

いわゆる動的計画法では所与の問題(14)を，**解き易い小さな**問題からなる問題群$A=\{A_n(x_n)\}$：

$$\mathrm{A}_n(x_n):\quad \begin{array}{ll} \text{Optimize} & r(x_n,u_n)+\cdots+r(x_N,u_N)+k(x_{N+1}) \\ \text{subject to} & (\mathrm{i})_\mathrm{m},\ (\mathrm{ii})_\mathrm{m}\quad n\le m\le N \end{array} \tag{15}$$

$$x_n\in X,\ 1\le n\le N+1$$

の一つと見なす。これがいわゆる**自然な埋め込み**(natural imbedding)である([19][20][21][25])。こ

のとき，解くべき問題が一番大きく，そして多分に一番解き難くくて最後に解かれる構成になっている場合がほとんどである。

さて，第 $n$ 段の状態 $x_n(\in X)$ から始まる $(N-n)$ 段部分決定過程 $A_n(x_n)$ を第１段での初期状態 $x_1(\in X)$ からの全 $N$ 段決定過程 $A_1(x_1)$ と同様に考えて，部分問題全体の間の最適構造関係を導こう。これがいわゆる動的計画法である([1][3][9][10][12][19][31][33][35][36][38][40])。これは最適化を含む再帰的方法である([41])。まず，第 $n$ 段から始まるマルコフ政策 $\pi=\{\pi_n, \pi_{n+1}, \ldots, \pi_N\}$ の全体を $\Pi(n)$ で表わしておこう。特に $n=1$ のとき，$\Pi(n)$ は $\Pi$ に一致する：$\Pi(1)=\Pi$. 次に部分問題 $A_n(x_n)$ の最適化をマルコフ政策クラス $\Pi(n)$ において考える。問題 $A_n(x_n)$ の評価値 $I_n(\pi; x_n)$ は初期状態 $x_n$ およびマルコフ政策 $\pi=\{\pi_n, \pi_{n+1}, \ldots, \pi_N\}$ に依存して定まる：

$$I_n(\pi; x_n) := r(x_n, u_n) + \cdots + r(x_N, u_N) + k(x_{N+1});$$
$$u_n = \pi_n(x_n),\ u_{n+1} = \pi_{n+1}(x_{n+1}),\ \ldots,\ u_N = \pi_N(x_N). \tag{16}$$

問題 $A_n(x_n)$ の $\Pi(n)$ にわたる最適値を $w_n(x_n)$ とする。すなわち，$w_n(x_n)$ は第 $n$ 段から始まるマルコフ政策全体にわたる最適値である：

$$w_n(x_n) := \mathop{\mathrm{Opt}}_{\pi\in\Pi(n)} I_n(\pi; x_n) \quad x_n\in X, \quad n = 1, 2, \ldots, N \tag{17}$$

$$w_{N+1}(x_{N+1}) := k(x_{N+1}) \quad x_{N+1} \in X. \tag{18}$$

このとき，次の最適値関数 $w_{n+1}(\cdot)$ との間には次の再帰式が成り立つ：

**定理3.1**

$$w_n(x) = \mathop{\mathrm{Opt}}_{u\in U} [\, r(x, u) + w_{n+1}(T(x, u))\,] \quad x \in X, \quad n = 1, 2, \ldots, N \tag{19}$$

$$w_{N+1}(x) = k(x) \qquad x\in X. \tag{20}$$

また，最適政策は次のように求められる。

**定理3.2**　再帰式(19)の最適子(最適値を与える $u$ の値)を $\pi^*_n(x)$ とすると，第 $n$ 決定関数 $\pi^*_n$ が得られる。このとき，決定関数列 $\pi^*=\{\pi^*_1, \pi^*_2, \ldots, \pi^*_N\}$ はマルコフ政策クラスの最適政策である：

$$I_1(\pi^*; x_1) \geq I_1(\pi; x_1) \quad \forall x_1 \in X,\ \forall \pi \in \Pi. \tag{21}$$

### 3.3　終端型問題

ここでは，加法型評価の特別な場合として，終端型評価を考えよう。これは加法型評価関数(14)における利得関数が零

$$r(x, u) = 0 \quad \forall (x, u) \in X\times U$$

の場合である：

$$\mathrm{T}_1(x_1):\quad \begin{array}{ll} \text{Optimize } k(x_{N+1}) & \\ \text{subject to (i)}_{\mathrm{n}}\ x_{n+1} = T(x_n, u_n) & \\ \qquad\qquad\ \text{(ii)}_{\mathrm{n}}\ u_n \in U & \end{array} \quad 1 \leq n \leq N \tag{22}$$

これに対する自然な埋め込みによって，問題 $T_1(x_1)$ を部分問題群 $\boldsymbol{T} = \{T_n(x_n)\}$：

$$\mathrm{T}_n(x_n):\quad \begin{array}{l} \text{Optimize } k(x_{N+1}) \\ \text{subject to (i)}_{\mathrm{m}},\ \text{(ii)}_{\mathrm{m}}\ \ n \le m \le N \\ \ x_n \in X,\ \ 1 \le n \le N+1 \end{array} \tag{23}$$

の一問題と考える。このとき部分問題 $T_n(x_n)$ の最適値 $t_n(x_n)$ は次の再帰式を満たす：

**系 3.1**

$$t_n(x) = \mathop{\mathrm{Opt}}_{u\in U}\ t_{n+1}(T(x,u)) \qquad x\in X,\ \ 1 \le n \le N \tag{24}$$

$$t_{N+1}(x) = k(x) \qquad x\in X. \tag{25}$$

**系 3.2** 再帰式(24)の最適子を $\pi_n^*(x)$ とする。このとき，決定関数列 $\pi^* = \{\pi_1^*, \pi_2^*, \ldots, \pi_N^*\}$ はマルコフ政策クラスの最適政策である：

$$I_1(\pi^*; x_1) \ge I_1(\pi; x_1) \quad \forall x_1 \in X,\ \ \forall \pi \in \Pi \tag{26}$$

ただしここでは，$I_1(\pi; x_1)$ は，初期状態 $x_1$ と政策 $\pi$ から一意に定まる履歴 $h_{N+1} = (x_1, u_1, x_2, u_2, \ldots, x_{N+1})$ の終端状態 $x_{N+1}$ に付随する終端値である：

$$\begin{aligned} &I_1(\pi; x_1) = k(x_{N+1}); \\ &u_1 = \pi_1(x_1),\ \ x_2 = T(x_1, u_1),\ \ u_2 = \pi_2(x_2),\ \ x_3 = T(x_2, u_2), \\ &\qquad\qquad \ldots,\ \ u_N = \pi_N(x_N),\ \ x_{N+1} = T(x_N, u_N). \end{aligned} \tag{27}$$

以上のように本節では簡単のため加法型問題を中心に最適政策が求められるまでを述べた。しかし，加法型問題に対する定理3.1, 定理3.2は**単調性をもつ結合型評価**(11)に対しても成り立つことに注意しよう。また，本節では，状態空間，決定空間，利得関数，状態変換がすべて時刻 $n$ に依存しないという意味で「定常モデル」で議論してきた。しかし，この節での結果は，これらの構成要素が時刻に依存する「非定常モデル」に対してもまったく同様にして成立することにも注意しよう。

## 4．複合型問題の埋め込み

前節でみたように，単調性をもつ結合型問題は加法型問題と同様に自然な埋め込みによって再帰的に解けることが分かった。さて，式(3)で与えられた問題 $\mathrm{P}_1(x_1)$ に戻ろう。この複合型問題は自然な埋め込みによって，（再帰式が導けて最適解を求めることができるという意味で）再帰的に解けるであろうか？

この複合型問題の自然な埋め込みによって，第 $n$ 段で $x_n(\in X)$ から始まる $(N-n)$ 段部分決定過程からなる問題群 $\boldsymbol{P} = \{\mathrm{P}_n(x_n)\}$：

$$\mathrm{P}_n(x_n):\quad \begin{array}{l} \text{Optimize } \psi(r_n\circ\cdots\circ r_N\circ k,\ R_n\diamond\cdots\diamond R_N\diamond K) \\ \text{subject to (i)}_{\mathrm{m}}\ x_{m+1}=T(x_m,u_m) \\ \qquad\qquad \text{(ii)}_{\mathrm{m}}\ u_m\in U \end{array}\quad n\le m\le N \tag{28}$$
$$x_n\in X,\ \ 1\le n\le N+1$$

が生成されることになる。ここでも，第 $n$ 段から始まる部分過程 $\mathrm{P}_n(x_n)$ を初期状態 $x_1(\in X)$ からの過程 $\mathrm{P}_1(x_1)$ と同様に考えよう。第 $n$ 段から始まる一般政策 $\sigma=\{\sigma_n,\ \ldots,\ \sigma_N\}$ の全体を $\Pi_g(n)$ とする。特に $\Pi_g(1)=\Pi_g$ である。このとき，部分問題 $\mathrm{P}_n(x_n)$ を一般政策クラス $\Pi_g(n)$ において最適化することを考える。したがって，問題 $\mathrm{P}_n(x_n)$ の評価値 $I_n(\sigma;x_n)$ は初期状態 $x_n$ および一般政策 $\sigma=\{\sigma_n,\ \sigma_{n+1},\ldots,\ \sigma_N\}$ に依存している：

$$I_n(\sigma;x_n):=\psi(r_n\circ\cdots\circ r_N\circ k,\ R_n\diamond\cdots\diamond R_N\diamond K);$$
$$u_n=\sigma_n(x_n),\ u_{n+1}=\sigma_{n+1}(x_n,x_{n+1}),\ \ldots,\ u_N=\sigma_N(x_n,x_{n+1},\ldots,x_N). \tag{29}$$

さて，部分問題 $\mathrm{P}_n(x_n)$ の一般政策クラス $\Pi_g(n)$ 上での最適値を $v_n(x_n)$ としよう：

$$v_n(x_n):=\mathop{\mathrm{Opt}}_{\sigma\in\Pi_g(n)} I_n(\sigma;x_{\mathrm{n}})\qquad x_n\in X,\quad n=1,2,\ldots,N \tag{30}$$

$$v_{N+1}(x_{N+1}):=\psi(k(x_{N+1}),K(x_{N+1}))\qquad x_{N+1}\in X. \tag{31}$$

しかし，このとき次期の最適値関数 $v_{n+1}(\cdot)$ との間には一般に再帰式は成立しない([16][17])。すなわち，複合問題群 $\boldsymbol{P}$ は上記のように導入できるが，**自然な埋め込みによる再帰的方法では**最適政策を一般政策クラス $\Pi_g$ に求めることは**できない**。それではどのような方法で最適政策を一般政策クラスの中に求めることができるであろうか？ この問題に対してポジティブに応えるために，新たにパラメータを追加して埋め込もう。以下では，原問題にパラメータを導入して等価な終端型問題に変換し，そこで自然な埋め込みをおこなって，原問題の最適政策を求める。このため，(1)状態空間の拡大，と(2)終端型問題への変換，の最適化問題としての二つの等価な書き換え(two equivalent transliterations as optimization problem)を行う。

### 4.1 状態空間の拡大

まず，2項演算の左単位元 $\tilde{\lambda}$ がどのような役割をはたしているか，確認しておこう。実際，左単位元は等式

$$\tilde{\lambda}\circ r_1\circ\cdots\circ r_N\circ k=r_1\circ\cdots\circ r_N\circ k \tag{32}$$

を成立せしめている。したがって，

$$\psi(\tilde{\lambda}\circ r_1\circ\cdots\circ r_N\circ k,\ \tilde{\mu}\diamond R_1\diamond\cdots\diamond R_N\diamond K)$$
$$=\psi(r_1\circ\cdots\circ r_N\circ k,\ R_1\diamond\cdots\diamond R_N\diamond K)$$

が成り立っている。ここで，**過去値集合**(past-value set)と**過去値関数**(past-value function) を導入しよう。第1段の状態 $x_1$ までの過去値集合は左単位元を用いて

$$\Omega_1(x_1):=\{(\tilde{\lambda},\ \tilde{\mu})\}$$

で定義する。第2の状態 $x_2$ までの過去値関数は

$$\lambda_2(x_1, u_1) := \tilde{\lambda} \circ r(x_1, u_1)$$
$$\mu_2(x_1, u_1) := \tilde{\mu} \diamond R(x_1, u_1)$$

で定義し，そこまでの過去値集合をその実行可能な値域

$$\Omega_2(x_2) := \{(\tilde{\lambda} \circ r(x_1, u_1),\ \tilde{\mu} \diamond R(x_1, u_1)) \mid (x_1, u_1) \in X \times U,\ T(x_1, u_1) = x_2\}$$

で定義する。このとき，条件式

$$\lambda_2 = \tilde{\lambda} \circ r(x_1, u_1)$$
$$\mu_2 = \tilde{\mu} \diamond R(x_1, u_1)$$

より

$$\lambda_2 \circ r_2 \circ \cdots \circ r_N \circ k = \tilde{\lambda} \circ r_1 \circ r_2 \circ \cdots \circ r_N \circ k$$
$$\mu_2 \diamond R_2 \diamond \cdots \diamond R_N \diamond K = \tilde{\mu} \diamond R_1 \diamond R_2 \diamond \cdots \diamond R_N \diamond K$$

が成り立つことに注意しよう。

一般に，第 $n$ 段の状態 $x_n$ までの過去値関数を

$$\lambda_n(x_1, u_1, \ldots, x_{n-1}, u_{n-1}) := \tilde{\lambda} \circ r(x_1, u_1) \circ \cdots \circ r(x_{n-1}, u_{n-1})$$
$$\mu_n(x_1, u_1, \ldots, x_{n-1}, u_{n-1}) := \tilde{\mu} \diamond R(x_1, u_1) \diamond \cdots \diamond R(x_{n-1}, u_{n-1}) \tag{33}$$

で定義し，その実行可能な値域

$$\Omega_n(x_n) := \{(\tilde{\lambda} \circ r(x_1, u_1) \circ \cdots \circ r(x_{n-1}, u_{n-1}),\ \tilde{\mu} \diamond R(x_1, u_1) \diamond \cdots \diamond R(x_{n-1}, u_{n-1}))$$
$$\mid \text{実効可能な}(u_1, x_2, \ldots, x_{n-1}, u_{n-1}) \in U \times X \times \cdots \times X \times U\}. \tag{34}$$

で，段と状態の対 $(n, x_n)$ までの過去値集合を定義する。ただし，「実行可能な」とは制約条件

$$\begin{array}{ll} \text{(i)}_{\mathrm{m}}\ x_{m+1} = T(x_m, u_m) & \\ \text{(ii)}_{\mathrm{m}}\ u_m \in U & 1 \leq m \leq n-1 \end{array}$$

を満たすことを意味する。このとき，過去値集合の族 $\{\Omega_n\}$ の間には次のように前向きの再帰式が成立する：

**補題 4.1**

$$\Omega_1(x) = \{(\tilde{\lambda},\ \tilde{\mu})\}$$
$$\Omega_{n+1}(y) = \{(\lambda \circ r(x, u),\ \mu \diamond R(x, u)) \mid (\lambda,\ \mu) \in \Omega_n(x),\ (x, u) \in X \times U,\ T(x, u) = y\}. \tag{35}$$

**証明** 自明。

さて，２つのパラメータ $\lambda_n, \mu_n$ に対する**第 n 条件**をそれぞれ

$$(\lambda)_{\mathrm{n}} \overset{def}{\Leftrightarrow} \lambda_n = \tilde{\lambda} \circ r(x_1, u_1) \circ \cdots \circ r(x_{n-1}, u_{n-1}) \tag{36}$$
$$(\mu)_{\mathrm{n}} \overset{def}{\Leftrightarrow} \mu_n = \tilde{\mu} \diamond R(x_1, u_1) \diamond \cdots \diamond R(x_{n-1}, u_{n-1}) \tag{37}$$

で定義する。このとき，条件(36), (37)より，

$$\lambda_n \circ r_n \circ \cdots \circ r_N \circ k = \tilde{\lambda} \circ r_1 \circ \cdots \circ r_N \circ k, \tag{38}$$
$$\mu_n \diamond R_n \diamond \cdots \diamond R_N \diamond K = \tilde{\mu} \diamond R_1 \diamond \cdots \diamond R_N \diamond K. \tag{39}$$

したがって

$$\psi(\lambda_n \circ r_n \circ \cdots \circ r_N \circ k,\ \mu_n \diamond R_n \diamond \cdots \diamond R_N \diamond K)$$
$$= \psi(\tilde{\lambda} \circ r_1 \circ \cdots \circ r_N \circ k,\ \tilde{\mu} \diamond R_1 \diamond \cdots \diamond R_N \diamond K) \tag{40}$$

が成り立つことに注意しよう。さらに，２つのパラメータ列$\{\lambda_1, \lambda_2, \ldots, \lambda_n\}$, $\{\mu_1, \mu_2, \ldots, \mu_n\}$に対する**第n累積条件**をそれぞれ

$$(\Lambda)_{\mathrm{n}} \overset{def}{\Leftrightarrow} (\lambda)_{\mathrm{m}}, \ m = 1, 2, \ldots, n \tag{41}$$

$$(\Gamma)_{\mathrm{n}} \overset{def}{\Leftrightarrow} (\mu)_{\mathrm{m}}, \ m = 1, 2, \ldots, n \tag{42}$$

で定義しておく。(41), (42)において，特に$n := N+1$とすると，この累積条件は

$$(\Lambda)_{\mathrm{N+1}} \Leftrightarrow \begin{cases} \lambda_1 = \tilde{\lambda} \\ \lambda_2 = \tilde{\lambda} \circ r(x_1, u_1) \\ \quad \vdots \\ \lambda_{N+1} = \tilde{\lambda} \circ r(x_1, u_1) \circ \cdots \circ r(x_N, u_N) \end{cases} \tag{43}$$

$$(\Gamma)_{\mathrm{N+1}} \Leftrightarrow \begin{cases} \mu_1 = \tilde{\mu} \\ \mu_2 = \tilde{\mu} \diamond R(x_1, u_1) \\ \quad \vdots \\ \mu_{N+1} = \tilde{\mu} \diamond R(x_1, u_1) \diamond \cdots \diamond R(x_N, u_N) \end{cases} \tag{44}$$

になる。この２つは逐次条件：

$$(\Lambda)_{\mathrm{N+1}} \Leftrightarrow \begin{cases} \lambda_1 = \tilde{\lambda} \\ \lambda_2 = \lambda_1 \circ r(x_1, u_1) \\ \quad \vdots \\ \lambda_{N+1} = \tilde{\lambda}_N \circ r(x_N, u_N) \end{cases} \tag{45}$$

$$(\Gamma)_{\mathrm{N+1}} \Leftrightarrow \begin{cases} \mu_1 = \tilde{\mu} \\ \mu_2 = \mu_1 \diamond R(x_1, u_1) \\ \quad \vdots \\ \mu_{N+1} = \tilde{\mu}_N \diamond R(x_N, u_N) \end{cases} \tag{46}$$

にそれぞれ同値である。しかも，式(45),(46)で示されているように，$(\lambda_n, \mu_n)$から$(\lambda_{n+1}, \mu_{n+1})$への制御$(x_n, u_n) \in X \times U$の下での変換は確定的である：

$$\lambda_{n+1} = \lambda_n \circ r(x_n, u_n), \quad \mu_{n+1} = \mu_n \diamond R(x_n, u_n). \tag{47}$$

### 4.2　終端型評価の導入

これまで，式(33)によって２つのパラメータ列$\{\lambda_1, \lambda_2, \ldots, \lambda_{N+1}\}$, $\{\mu_1, \mu_2, \ldots, \mu_{N+1}\}$を導入して制約条件(36), (37)を課した。この２パラメータは変換(47)のもとで推移している。ここではこの推移を本来の状態変換に組み込んで新しい状態空間上に新たな変換を構成しよう。

さて，本来の状態空間$X$に過去値空間$\Omega_n(x_n)$を貼り付けて，新しい状態空間列$\{Y_n\}$を点対集合値写像$\Omega_n(\cdot)$のグラフ

$$Y_n := \{(x; \lambda, \mu) \mid x \in X, (\lambda, \mu) \in \Omega_n(x)\} \quad 1 \leq n \leq N+1 \tag{48}$$

によって定義する。この空間上の確定的な**新しい状態変換列**$\tilde{T} = \{\tilde{T}_n\}$を

$$\tilde{T}_n((x;\lambda,\mu),u) = (y;\lambda',\mu') \overset{def}{\Leftrightarrow} \begin{cases} T(x,u) = y \\ \lambda \circ r(x,u) = \lambda' \\ \mu \diamond R(x,u) = \mu' \end{cases} \tag{49}$$

で定義する。すなわち，本来の状態変換 $T$ と2パラメータ $(\lambda_n, \mu_n)$ 間の変換は拡大状態空間上の変換に合成されて次のようになる：1つの拡大状態 $\tilde{x}=(x;\lambda,\mu)\in Y_n$ が決定 $u\in U$ による新しい変換

$$\tilde{y} = \tilde{T}_n(\tilde{x},u) \tag{50}$$

によって $\tilde{y}=(y;\lambda',\mu')\in Y_{n+1}$ に推移するということは

$$y = T(x,u),\quad \lambda' = \lambda\circ r(x,u),\quad \mu' = \mu\diamond R(x,u) \tag{51}$$

に他ならない。

次に**終端関数** $\Psi: Y_{N+1}\to R^1$ を

$$\Psi(y_{N+1}) = \Psi(x_{N+1};\lambda_{N+1},\mu_{N+1}) := \psi(\lambda_{N+1}\circ k(x_{N+1}),\ \mu_{N+1}\diamond K(x_{N+1})) \tag{52}$$

で定義する。

さて，冒頭の複合型評価問題 $\mathrm{P}_1(x_1)$ に対して拡大状態空間上の終端型評価問題 $Q_1(x_1;\tilde{\lambda},\tilde{\mu})$:

$$Q_1(x_1;\tilde{\lambda},\tilde{\mu}): \quad \begin{array}{ll} \text{Optimize } \Psi(y_{N+1}) & \\ \text{subject to (i)}'_{\mathrm{n}}\ y_{n+1} = \tilde{T}_n(y_n,u_n) & \\ \qquad\qquad \text{(ii)}_{\mathrm{n}}\ u_n\in U & 1\le n\le N \end{array} \tag{53}$$

を考えよう。この過程は初期状態 $y_1=(x_1;\tilde{\lambda},\tilde{\mu})\ (\in Y_1)$ から出発してある終端状態 $y_{N+1}=(x_{N+1};\lambda_{N+1},\mu_{N+1})\ (\in Y_{N+1})$ で終わる。このとき過程全体の評価値は終端状態のみに依存した $\Psi(y_{N+1})=\psi(\lambda_{N+1}\circ k(x_{N+1}),\ \mu_{N+1}\diamond K(x_{N+1}))$ である。さて，拡大状態空間上での政策を導入しよう。第nマルコフ決定関数が

$$\gamma_n: Y_n\to U,\quad (n=1,2,\ldots,N)$$

となる決定列 $\gamma=\{\gamma_1,\gamma_2,\ldots,\gamma_N\}$ を**マルコフ政策**という。この拡大空間上でのマルコフ政策の全体を $\tilde{\Pi}$ で表し，本来の状態空間上のそれ $\Pi$ と区別する。

ここで，拡大問題 $Q_1(x_1;\tilde{\lambda},\tilde{\mu})$ を「自然な意味」で，部分問題群 $\boldsymbol{Q}=\{Q_n(x_n;\lambda_n,\mu_n)\}$:

$$Q_n(x_n;\lambda_n,\mu_n): \quad \begin{array}{ll} \text{Optimize } \Psi(y_{N+1}) & \\ \text{subject to (i)}'_{\mathrm{m}}\ y_{m+1} = \tilde{T}_m(y_m,u_m) & \\ \qquad\qquad \text{(ii)}_{\mathrm{m}}\ u_m\in U & n\le m\le N \end{array} \tag{54}$$

$$(x_n;\lambda_n,\mu_n)\in Y_n,\quad 1\le n\le N+1$$

に埋め込もう。部分問題 $Q_n(x_n;\lambda_n,\mu_n)$ は第 $n$ 段の始発状態 $y_n=(x_n;\lambda_n,\mu_n)\ (\in Y_n)$ からある終端状態 $y_{N+1}$ までの $(N-n)$ 段過程である。ここでまず，第 $n$ 段から始まるマルコフ政策 $\gamma=\{\gamma_n,\gamma_{n+1},\ldots,\gamma_N\}$ の全体を $\tilde{\Pi}(n)$ で表わそう。特に $n=1$ のとき，$\tilde{\Pi}(n)$ は $\tilde{\Pi}$ に一致する：$\tilde{\Pi}(1)=\tilde{\Pi}$. 次に，部分問題 $Q_n(x_n;\lambda_n,\mu_n)$ の最適化をマルコフ政策クラス $\tilde{\Pi}(n)$ において考える。問題 $Q_n(x_n;\lambda_n,\mu_n)$ の評価値 $J_n(\gamma;x_n;\lambda_n,\mu_n)$ は初期状態 $(x_n;\lambda_n,\mu_n)$ およびマルコフ政策 $\gamma=\{\gamma_n,\gamma_{n+1},\ldots,\gamma_N\}$ に依存して定まる：

$$J_n(\gamma;x_n;\lambda_n,\mu_n) := \Psi(x_{N+1};\lambda_{N+1},\mu_{N+1})$$

$$u_n=\gamma_n(y_n),\ u_{n+1}=\gamma_{n+1}(y_{n+1}),\ \ldots,\ u_N=\gamma_N(y_N). \tag{55}$$

問題 $Q_n(x_n;\lambda_n,\mu_n)$ の $\tilde{\Pi}(n)$ にわたる最適値を $u^n(x_n;\lambda_n,\mu_n)$ とする。すなわち，$u^n(x_n;\lambda_n,\mu_n)$ は第 $n$ 段から始まるマルコフ政策全体にわたる最適値である：

$$u^n(x_n;\lambda_n,\mu_n) := \operatorname*{Opt}_{\gamma\in\tilde{\Pi}(n)} J_n(\gamma;x_n;\lambda_n,\mu_n) \qquad (x_n;\lambda_n,\mu_n)\in Y_n \quad n=1,2,\ \ldots,\ N \tag{56}$$

$$u^{N+1}(x_{N+1};\lambda_{N+1},\mu_{N+1}) := \psi(\lambda_{N+1}\circ k(x_{N+1}),\ \mu_{N+1}\diamond K(x_{N+1})) \tag{57}$$

## 5．同値性から最適性へ

この節では本来の状態空間上の一般政策クラスと拡大状態空間上のマルコフ政策クラスが最適値を考察する上で1対1に対応していること(同値性)を示し，後者でのマルコフ最適政策の存在・構成が前者の一般最適政策を導くことを示す。

**補題 5.1**　(同値性)

(i) 拡大問題 $Q_1(x_1;\tilde{\lambda},\tilde{\mu})$ の任意のマルコフ政策 $\gamma$ は原問題 $\mathrm{P}_1(x_1)$ の一般政策 $\sigma$ を生成して

$$I_1(\sigma;x_1) = J_1(\gamma;x_1,\tilde{\lambda},\tilde{\mu}) \qquad \forall x_1\in X \tag{58}$$

になる。

(ii) 原問題 $\mathrm{P}_1(x_1)$ の一般政策 $\sigma$ が任意に与えられたとき，拡大問題 $Q_1(x_1;\tilde{\lambda},\tilde{\mu})$ のマルコフ政策 $\gamma$ を構成して式(58)が成り立つようにできる。

証明　(i) まず，$\gamma=\{\gamma_1,\ \gamma_2,\ \ldots,\ \gamma_N\}$ を拡大問題の任意のマルコフ政策としよう。このとき，各 $n$ $(1\le n\le N)$ に対して第 $n$ 一般決定関数 $\sigma_n$ を定めて $\gamma$ と等価な一般政策 $\sigma$ を構成しよう。さて，$(x_1,\ x_2,\ \ldots,x_n)$ を第 $n(1\le n\le N+1)$ 段までの任意の状態列としよう。このとき，$\{\gamma_1,\ \gamma_2,\ \ldots,\ \gamma_N\}$ から生成された3つの中間列 $\{\lambda_m\}_{1\le m\le n}$, $\{\mu_m\}_{1\le m\le n}$, $\{y_m\}_{1\le m\le n}$ を用いて，次のように $\sigma_n(x_1,x_2,\ \ldots,x_n)$ を定める：

$$\begin{aligned}
&\lambda_1 := \tilde{\lambda},\ \ \mu_1 := \tilde{\mu}\\
&y_1 := (x_1;\ \lambda_1,\mu_1)\\
&u_1 := \gamma_1(y_1),\ \ \lambda_2 := \lambda_1\circ r(x_1,u_1),\ \ \mu_2 := \mu_1\diamond R(x_1,u_1)\\
&y_2 := (x_2;\ \lambda_2,\mu_2)\\
&u_2 := \gamma_2(y_2),\ \ \lambda_3 := \lambda_2\circ r(x_2,u_2),\ \ \mu_3 := \mu_2\diamond R(x_2,u_2)\\
&y_3 := (x_3;\ \lambda_3,\mu_3)\\
&\qquad\qquad \vdots\\
&u_{n-1} := \lambda_{n-1}(y_{n-1}),\\
&\qquad\qquad \lambda_n := \lambda_{n-1}\circ r(x_{n-1},u_{n-1}),\quad \mu_n := \mu_{n-1}\diamond R(x_{n-1},u_{n-1})\\
&y_n := (x_n;\lambda_n,\mu_n)\\
&\sigma_n(x_1,x_2,\ \ldots,x_n) := \gamma_n(y_n).
\end{aligned} \tag{59}$$

(この $\sigma$ の定義は一義的ではないが，実行可能な唯一の行動（状態と決定の交互列）は一意に定まる。)

さて，このように定義された $\sigma=\{\sigma_1,\ \sigma_2,\ \ldots,\ \sigma_N\}$ が $\gamma$ と等価であることを示そう。ここで拡大状

態列（$y_1, y_2, \ldots, y_{N-1}$）が政策$\gamma$で生成されたとしよう。このとき実はある状態列（$x_1, x_2, \ldots, x_{N-1}$）が政策$\sigma$で生成され，前者の終端評価値と後者の複合評価値が一致することを示す。さて，拡大状態の定義より，列（$y_1, y_2, \ldots, y_{N-1}$）は次を満たしながら定まっている：

$$
\begin{aligned}
y_1 &= (x_1; \lambda_1, \mu_1) \quad \text{ここに } \lambda_1 = \tilde{\lambda},\ \mu_1 = \tilde{\mu} \\
y_2 &= (x_2; \lambda_2, \mu_2) \\
&\qquad \text{ここに } \lambda_2 = \lambda_1 \circ r(x_1, u_1),\ \mu_2 = \mu_1 \diamond R(x_1, u_1) \text{ ただし } u_1 = \gamma_1(y_1) \\
y_3 &= (x_3; \lambda_3, \mu_3) \\
&\qquad \text{ここに } \lambda_3 = \lambda_2 \circ r(x_2, u_2),\ \mu_3 = \mu_2 \diamond R(x_2, u_2) \text{ ただし } u_2 = \gamma_2(y_2) \\
&\qquad \vdots \\
u_N &= (x_N; \lambda_N, \mu_N) \\
&\qquad \text{ここに } \lambda_N = \lambda_{N-1} \circ r(x_{N-1}, u_{N-1}),\ \mu_N = \mu_{N-1} \diamond R(x_{N-1}, u_{N-1}) \\
&\qquad\qquad \text{ただし } u_{N-1} = \gamma_{N-1}(y_{N-1}) \\
u_{N+1} &= (x_{N+1}; \lambda_{N+1}, \mu_{N+1}) \\
&\qquad \text{ここに } \lambda_{N+1} = \lambda_N \circ r(x_N, u_N),\ \mu_{N+1} = \mu_N \diamond R(x_N, u_N) \\
&\qquad\qquad \text{ただし } u_N = \gamma_N(y_N).
\end{aligned} \tag{60}
$$

したがって，評価値は次の等式を満たす：

$$
\begin{aligned}
&\Psi(x_{N+1}; \lambda_{N+1}, \mu_{N+1}) \\
&= \psi(\lambda_{N+1} \circ k(x_{N+1}),\ \mu_{N+1} \diamond R(x_{N+1})) \\
&= \psi(\lambda_N \circ r(x_N, u_N) \circ k(x_{N+1}),\ \mu_N \diamond R(x_N, u_N) \diamond K(x_{N+1})) \\
&\qquad \vdots \\
&= \psi(\lambda_n \circ r_n \circ \cdots \circ r_N \circ k,\ \mu_n \diamond R_n \diamond \cdots \diamond R_N \diamond K) \\
&\qquad \vdots \\
&= \psi(\lambda_1 \circ r_1 \cdots \circ r_N \circ k,\ \mu_1 \diamond R_1 \diamond \cdots \diamond R_N \diamond K) \\
&= \psi(\tilde{\lambda} \circ r_1 \circ \cdots \circ r_N \circ k,\ \tilde{\mu} \diamond R_1 \diamond \cdots \diamond R_N \diamond K) \\
&= \psi(r_1 \circ \cdots \circ r_N \circ k,\ R_1 \diamond \cdots \diamond R_N \diamond K).
\end{aligned} \tag{61}
$$

すなわち

$$J_1(\gamma; x_1, \tilde{\lambda}, \tilde{\mu}) = I_1(\sigma; x_1) \tag{62}$$

が成り立つ。

(ii) 次に $P_1(x_1)$の一般政策$\sigma = \{\sigma_1, \sigma_2, \ldots, \sigma_N\}$を任意に与えて，$Q_1(x_1; \tilde{\lambda}, \tilde{\mu})$ のある等価なマルコフ政策$\gamma$を構成しよう。

このとき初期状態$x_1$と政策$\sigma$から，５つの列 $\{\lambda_n\}_{1 \leq n \leq N+1}$，$\{\mu_n\}_{1 \leq n \leq N+1}$，$\{x_n\}_{1 \leq n \leq N+1}$，$\{u_n\}_{1 \leq n \leq N}$，$\{y_n\}_{1 \leq n \leq N+1}$ が次のように定まる：

$$
\begin{aligned}
\lambda_1 &:= \tilde{\lambda},\ \mu_1 := \tilde{\mu} \\
y_1 &:= (x_1; \lambda_1, \mu_1) \\
u_1 &:= \sigma_1(x_1),\ \lambda_2 := \lambda_1 \circ r(x_1, u_1),\ \mu_2 := \mu_1 \diamond R(x_1, u_1)
\end{aligned}
$$

$$x_2 := T(x_1, u_1),\ y_2 := (x_2; \lambda_2, \mu_2)$$
$$u_2 := \sigma_2(x_1, x_2),\ \lambda_3 := \lambda_2 \circ r(x_2, u_2),\ \mu_3 := \mu_2 \diamond R(x_2, u_2)$$
$$x_3 := T(x_2, u_2),\ y_3 := (x_3; \lambda_3, \mu_3)$$
$$\vdots$$
$$u_{n-1} := \sigma_{n-1}(x_1, x_2, \ldots, x_{n-1}),$$
$$\lambda_n := \lambda_{n-1} \circ r(x_{n-1}, u_{n-1}),\ \mu_n := \mu_{n-1} \diamond R(x_{n-1}, u_{n-1})$$
$$x_n := T(x_{n-1}, u_{n-1}),\ y_n := (x_n; \lambda_n, \mu_n)$$
$$u_n := \sigma_n(x_1, x_2, \ldots, x_n) \tag{63}$$
$$\vdots$$
$$u_N := \sigma_N(x_1, x_2, \ldots, x_N)$$
$$\lambda_{N+1} := \lambda_N \circ r(x_N, u_N),\ u_{N+1} := u_N \diamond R(x_N, u_N)$$
$$x_{N+1} := T(x_N, u_N),\ y_{N+1} := (x_{N+1}; \lambda_{N+1}, \mu_{N+1}).$$

式(63)を満たす $y_n = (x_n; \lambda_n, \mu_n)$ における決定 $\gamma_n(y_n)$ は状態列 $\{x_1, x_2, \ldots, x_{n-1}, x_n\}$ を用いて

$$\gamma_n(y_n) := \sigma_n(x_1, x_2, \ldots, x_{n-1}, x_n) \qquad (1 \le n \le N) \tag{64}$$

で定義する。((63)を満たさない $y_n$ については適宜定める。こうしても以下の議論に影響しない。)このとき，$(x_1; \tilde{\lambda}, \tilde{\mu})$ からのマルコフ政策 $\gamma$ による終端評価値は $x_1$ からの一般政策 $\sigma$ による複合評価値に等しいことを見よう。実際,

$$
\begin{aligned}
&\psi(r_1 \circ \cdots \circ r_N \circ k,\ R_1 \diamond \cdots \diamond R_N \diamond K). \\
&= \psi(\tilde{\lambda} \circ r_1 \circ \cdots \circ r_N \circ k,\ \tilde{\mu} \diamond R_1 \diamond \cdots \diamond R_N \diamond K) \\
&= \psi(\lambda_1 \circ r_1 \cdots \circ r_N \circ k,\ \mu_1 \diamond R_1 \diamond \cdots \diamond R_N \diamond K) \\
&\qquad \vdots \\
&= \psi(\lambda_n \circ r_n \circ \cdots \circ r_N \circ k,\ \mu_n \diamond R_n \diamond \cdots \diamond R_N \diamond K) \qquad (65) \\
&\qquad \vdots \\
&= \psi(\lambda_N \circ r(x_N, u_N) \circ k(x_{N+1}),\ \mu_N \diamond R(x_N, u_N) \diamond K(x_{N+1})) \\
&= \psi(\lambda_{N+1} \circ k(x_{N+1}),\ \mu_{N+1} \diamond R(x_{N+1})) \\
&= \Psi(x_{N+1}; \lambda_{N+1}, \mu_{N+1})
\end{aligned}
$$

となり

$$I_1(\sigma; x_1) = J_1(\gamma; x_1, \tilde{\lambda}, \tilde{\mu}) \tag{66}$$

が成立する。これで同値性（補題5.1）の証明を終える。

さて，この同値性(補題5.1)を第 $n$ 段からの過程に用いれば，部分問題 $Q_n(x_n; \tilde{\lambda}, \tilde{\mu})$ の $\tilde{\Pi}(n)$ にわたる最適値 $u^n(x_n; \tilde{\lambda}, \tilde{\mu})$ と部分問題 $\mathrm{P}_n(x_n)$ の一般政策クラス $\Pi_g(n)$ 上の最適値を $v_n(x_n)$ は一致する：

**系 5.1**

$$u^n(x_n; \tilde{\lambda}, \tilde{\mu}) = v_n(x_n) \qquad x \in X,\ n = 1, 2, \ldots, N+1. \tag{67}$$

証明　最適値はそれぞれ

$$v_n(x_n) = \underset{\sigma \in \Pi_g(n)}{\mathrm{Opt}}\ I_n(\sigma; x_n) \qquad x_n \in X \tag{68}$$

$$u^n(x_n; \lambda_n, \mu_n) = \underset{\gamma \in \tilde{\Pi}(n)}{\mathrm{Opt}}\ J_n(\gamma; x_n, \lambda_n, \mu_n) \qquad (x_n; \lambda_n, \mu_n) \in Y_n \tag{69}$$

であることに注意すれば，$(x_n; \tilde{\lambda}, \tilde{\mu})$からの同値性より等式(67)が成立する。

### 5.1　拡大問題の表現

ここでは拡大問題(53)に対する部分問題 $Q_n(x_n; \lambda_n, \mu_n)$の状態変換 $(\mathrm{i})'_{\mathrm{m}}$ ： $\tilde{T}_m(y_m, u_m) = y_{m+1}$ を本来の状態変換 $T$ とパラメータ間の変換 $(\lambda_m, \mu_m) \xmapsto{(x_m,\, u_m)} (\lambda_{m+1}, \mu_{m+1})$ に還元することによって幾つかの同値表現が得られることを述べよう。まず，補題5.1より，拡大問題 $Q_1(x_1; \tilde{\lambda}, \tilde{\mu})$ は次のように表わされることに注意しよう：

$$Q_1(x_1; \tilde{\lambda}, \tilde{\mu}): \quad \begin{array}{ll} \text{Optimize} & \psi(\tilde{\lambda} \circ r_1 \circ \cdots \circ r_N \circ k,\ \tilde{\mu} \diamond R_1 \diamond \cdots \diamond R_N \diamond K) \\ \text{subject to} & (\mathrm{i})_{\mathrm{n}}\ x_{n+1} = T(x_n, u_n) \\ & (\mathrm{ii})_{\mathrm{n}}\ u_n \in U \qquad\qquad 1 \le n \le N \\ & (\mathrm{iii})_{\mathrm{n}}\ \ \lambda_{n+1} = \lambda_n \circ r(x_n, u_n),\ \ \mu_{n+1} = \mu_n \diamond R(x_n, u_n) \end{array} \tag{70}$$

また，逐次条件 $\{(\mathrm{iii})_{\mathrm{n}}\}$ より

$$\begin{aligned} &\psi(\tilde{\lambda} \circ r_1 \circ \cdots \circ r_N \circ k,\ \tilde{\mu} \diamond R_1 \diamond \cdots \diamond R_N \diamond K) \\ &= \psi(\lambda_{N+1} \circ k,\ \mu_{N+1} \diamond K) \end{aligned} \tag{71}$$

が成立するから，問題 $Q_1(x_1; \tilde{\lambda}, \tilde{\mu})$ は終端評価問題：

$$Q_1(x_1; \tilde{\lambda}, \tilde{\mu}): \quad \begin{array}{ll} \text{Optimize} & \psi(\lambda_{N+1} \circ k,\ \mu_{N+1} \diamond K) \\ \text{subject to} & (\mathrm{i})_{\mathrm{n}}\ x_{n+1} = T(x_n, u_n) \\ & (\mathrm{ii})_{\mathrm{n}}\ u_n \in U \qquad\qquad 1 \le n \le N \\ & (\mathrm{iii})_{\mathrm{n}}\ \ \lambda_{n+1} = \lambda_n \circ r(x_n, u_n),\ \ u_{n+1} = u_n \diamond R(x_n, u_n) \end{array} \tag{72}$$

に表わされる。ただし，ここでは三つ組 $(x_n; \lambda_n, \mu_n)$ が状態を表わしている。このように表わすと，問題 $Q_1(x_1; \tilde{\lambda}, \tilde{\mu})$は部分問題群 $\boldsymbol{Q} = \{Q_n(x_n; \lambda_n, \mu_n)\}$：

$$Q_n(x_n; \lambda_n, \mu_n): \quad \begin{array}{ll} \text{Optimize} & \psi(\lambda_{N+1} \circ k,\ \mu_{N+1} \diamond K) \\ \text{subject to} & (\mathrm{i})_{\mathrm{m}},\ \ (\mathrm{ii})_{\mathrm{m}} \qquad n \le m \le N \\ & (\mathrm{iii})_{\mathrm{m}}\ \ \lambda_{m+1} = \lambda_m \circ r(x_m, u_m),\ \ \mu_{m+1} = \mu_m \diamond R(x_m, u_m) \end{array} \tag{73}$$

すなわち

$$Q_n(x_n; \lambda_n, \mu_n): \quad \begin{array}{ll} \text{Optimize} & \psi(\lambda_n \circ r_n \circ \cdots \circ r_N \circ k,\ \mu_n \diamond R_n \diamond \cdots \diamond R_N \diamond K) \\ \text{subject to} & (\mathrm{i})_{\mathrm{m}}\ x_{m+1} = T(x_m, u_m) \\ & (\mathrm{ii})_{\mathrm{m}}\ u_m \in U \qquad\qquad n \le m \le N \\ & (\mathrm{iii})_{\mathrm{m}}\ \ \lambda_{m+1} = \lambda_m \circ r(x_m, u_m),\ \ \mu_{m+1} = \mu_m \diamond R(x_m, u_m). \end{array} \tag{74}$$

に「自然に」埋め込まれたと解釈できる。

## 5.2 最適政策

さて，$u^n(x_n;\lambda_n,\mu_n)$は部分問題 $Q_n(x_n;\lambda_n,\mu_n)$に対するマルコフ政策全体$\tilde{\Pi}(n)$ にわたる最適であることに注意しよう。このとき，最適値 $u^n(x_n;\lambda_n,\mu_n)$と次の最適値関数 $u^{n+1}(\cdot;\cdot,\cdot)$との間には次の後向きの再帰式が成り立つ：

**定理 5.1**

$$u^n(x;\lambda,\mu) = \mathop{\mathrm{Opt}}_{u\in U} u^{n+1}(T(x,u);\ \lambda\circ r(x,u),\ \mu\diamond R(x,u)) \tag{75}$$

$$x\in X,\ (\lambda,\mu)\in\Omega_n(x),\ n=1,2,\ldots,N$$

$$u^{N+1}(x;\lambda,\mu) = \psi(\lambda\circ k(x),\ \mu\diamond K(x)) \qquad x\in X,\ (\lambda,\mu)\in\Omega_{N+1}(x) \tag{76}$$

証明　問題族$\boldsymbol{Q} = \{Q_n(x_n;\lambda_n,\mu_n)\}$ が新状態空間 $\{Y_n\}$ 上の終端型評価問題の部分族であり，終端型評価系は加法型系であることに注意すれば，系3.1より再帰式(75),(76)が成立する。

**定理 5.2**　式(75)の最適子(最適値を与える$u(\in U)$の値)を$\gamma_n^*(x;\lambda,\mu)$とすると，第$n$ 決定関数$\gamma_n^*$: $Y_n\to U$ が定まる。列 $\gamma^*=\{\gamma_1^*,\gamma_2^*,\ldots,\gamma_N^*\}$はマルコフ政策クラスで最適である：

$$J_1(\gamma^*;x_1,\tilde{\lambda},\tilde{\mu}) \geq J_1(\gamma;x_1,\tilde{\lambda},\tilde{\mu}) \quad \forall\gamma\in\tilde{\Pi},\ (x_1,\tilde{\lambda},\tilde{\mu})\in Y_1. \tag{77}$$

証明　これは系3.2からただちに従う。

従って，原問題 $\mathrm{P}_1(x_1)$の**最適値** $v_1(x_1)$は拡大問題 $Q_1(x_1;\tilde{\lambda},\tilde{\mu})$の**最適値** $u^1(x_1;\tilde{\lambda},\tilde{\mu})$で与えられる：

$$v_1(x_1) = u^1(x_1;\tilde{\lambda},\tilde{\mu}). \tag{78}$$

さらに，最適マルコフ政策$\gamma^*=\{\gamma_1^*,\gamma_2^*,\ldots,\gamma_N^*\}$は(3)の一般政策$\sigma^*=\{\sigma_1^*,\sigma_2^*,\ldots,\sigma_N^*\}$を以下のように生成する。ここに$\sigma_n^*(x_1,x_2,\cdots,x_n)$は式(59)と同様に

$$\begin{aligned}
&\lambda_1 := \tilde{\lambda},\ \mu_1 := \tilde{\mu}\\
&y_1 := (x_1;\lambda_1,\mu_1)\\
&u_1 := \gamma_1^*(y_1),\ \lambda_2 := \lambda_1\circ r(x_1,u_1),\ \mu_2 := \mu_1\diamond R(x_1,u_1)\\
&y_2 := (x_2;\lambda_2,\mu_2)\\
&u_2 := \gamma_2^*(y_2),\ \lambda_3 := \lambda_2\circ r(x_2,u_2),\ \mu_3 := \mu_2\diamond R(x_2,u_2)\\
&y_3 := (x_3;\lambda_3,\mu_3)\\
&\quad\vdots\\
&u_{n-1} := \gamma_{n-1}^*(y_{n-1}),\\
&\qquad\qquad \lambda_n := \lambda_{n-1}\circ r(x_{n-1},u_{n-1}),\ \mu_n := \mu_{n-1}\diamond R(x_{n-1},u_{n-1})\\
&y_n := (x_n;\lambda_n,\mu_n)\\
&\sigma_n^*(x_1,x_2,\ldots,x_n) := \gamma_n^*(y_n).
\end{aligned} \tag{79}$$

で定義する。このとき

**定理 5.3**　政策$\sigma^*$は一般クラスにおいて最適である：

$$I_1(\sigma^*;x_1) \geq I_1(\sigma;x_1) \qquad \forall\sigma\in\Pi_g,\ x_1\in X. \tag{80}$$

証明　系5.1，定理5.2より従う。

## 参 考 文 献


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

[2] R.E. Bellman and L.A. Zadeh, Decision-making in a fuzzy environment, Management Sci. **17**(1970), B141-B164.

[3] E.V. Denardo, *Dynamic Programming: Models and Applications,* Prentice-Hall, NJ, 1982.

[4] A.O. Esogbue and R.E. Bellman, Fuzzy dynamic programming and its extensions, TIMS/Studies in the Management Sciences **20**(1984), 147-167.

[5] N. Furukawa and S. Iwamoto, Markovian decision processes with recursive reward functions, *Bull. Math. Statist.* **15** (1973), 79-91.

[6] N. Furukawa and S. Iwamoto, Dynamic programming on recursive reward systems, *Bull. Math. Statist.* **17**(1976), 103-126.

[7] S.W. Golmb, E2118, Amer. Math. Monthly **75**(1968), 878.

[8] S.W. Golmb, Iterated binomial coefficients, Amer. Math. Monthly **87**(1980), 719-727.

[9] K. Hinderer, *Foundations of Non-Stationary Dynamic Programming with Discrete-Time Parameter,* Lect. Notes in Operation Research and Mathematical Systems, Vol.33, Springer-Verlag, Berlin, 1970.

[10] R. A. Howard, *Dynamic Programming and Markov Processes,* MIT Press, Cambridge, Mass., 1960.

[11] S. Iwamoto, Discrete dynamic programming with recursive additive system, *Bull. Math. Statist.* **15**(1974), 49-66.

[12] 岩本誠一，動的計画の理論と応用，数学 **31** 巻４号，1979年, 331-348.

[13] 岩本誠一，「動的計画論」，九大出版会，1987.

[14] S. Iwamoto, From dynamic programming to bynamic programming, *J. Math. Anal. Appl.* **177**(1993), 56-74.

[15] S. Iwamoto, On bidecision processes, *J. Math. Anal. Appl.* **187**(1994), 676-699.

[16] 岩本誠一，不確実性の下での最適危機管理について，経済学研究 **61** 巻5/6号，1996年，1-18.

[17] 岩本誠一，ファジィ環境下の確率的意志決定過程 — ベルマン・ザデーのアプローチとの比較 —，経済学研究 **62** 巻1/6号，1996年，349-365.

[18] 岩本誠一，ファジィ動的計画法，日本ファジィ学会編「ソフトコンピューティング用語集」，朝倉書店，1996年，pp. 42.

[19] S. Iwamoto, Associative dynamic programs, *J. Math. Anal. Appl.* **201**(1996), 195-211.

[20] S. Iwamoto, On expected values of Markov statistics, *Bull. Informatics and Cybernetics* **30**(1998), 1-24.

[21] S. Iwamoto, Decision-making in fuzzy environment: a survey from stochastic decision process, Ed. L.C. Jain and R.K. Jain, Proceedings of The Second International Conference on Knowledge-based Intelligent Electronics Systems (KES '98), Adelaide, 1998, pp.542-546.

[22] S. Iwamoto, Conditional decision processes with recursive reward function, *J. Math. Anal. Appl.* **230**(1999), 193-210.

[23] S. Iwamoto and T. Fujita, Stochastic decision-making in a fuzzy environment, *J. Operations Res. Soc. Japan* **38**(1995), 467-482.

[24] S. Iwamoto and M. Sniedovich, Sequential decision making in fuzzy environment, *J. Math. Anal. Appl.* **222**(1998), 208-224.

[25] E. S. Lee, *Quasilinearization and Invariant Imbedding,* Academic Press, New York, 1968.

[26] J. Kacprzyk, Decision-making in a fuzzy environment with fuzzy termination time, *Fuzzy Sets and Systems* **1**(1978), 169-179.

[27] D.M. Kreps, Decision problems with expected utility criteria I, *Math. Oper. Res.* **2**(1977), 45-53.

[28] D.M. Kreps, Decision problems with expected utility criteria II; stationarity, *Math. Oper.* Res. **2**(1977), 266-274.

[29] D.M. Kreps, *A Course in Microeconomics Theory,* Harvester Wheatsheaf, New York, 1990.

[30] W. Lipfert, Über ein stochastisches dynamiches Entscheidungsmodell mit allegemeinen Ertragsfunktionalen, *Optimization* **16**(1985), 313-328.

[31] L.G. Mitten, Composition principles for sysnthesis of optimal multi-stage processes, *Operations Res.* **12**(1964), 610-619.

[32] 水本雅晴，「ファジィ理論とその応用」，サイエンス社，1988.

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

[34] A. Nowak, On a general dynamic programming problem, *Colloquium Mathematicum* **37**(1977), Fasc.1, 131-138.

[35] E. Porteus, An informal look at the principle of optimality, *Management Sci.* **21**(1975), 1346-1348.

[36] E. Porteus, Conditions for characterizing the structure of optimal strategies in infinite-horizon dynamic programs, *J. Opt. Theo. Anal.* **36**(1982), 419-432.

[37] M.L. Puterman, *Markov Decision Processes:* Stochastic Models, Chap. VIII, D.P. Heyman and M.J. Sobel (Ed's),

Handbooks in Operations Research and Management Science Vol. 2, Elsevier, Amsterdam, 1990.
[38] M.L. Puterman, *Markov Decision Processes: discrete stochastic dynamic programming,* Wiley & Sons, New York, 1994.
[39] M. Sniedovich, A class of variance constrained problems, *Operations Research* **31** (1983), 338-353.
[40] M. Sniedovich, *Dynamic Programming,* Marcel Dekker, Inc. NY, 1992.
[41] N.L. Stokey and R.E. Lucas Jr., *Recursive Methods in Economic Dynamics,* Harvard Univ. Press, Cambridge, MA, 1989.

[九州大学経済学部教授]