数理解析研究所講究録
第 1682 巻 2010 年 55-61

全期間依存制約つき決定過程
ー確率的推移システム上での結合型制約ー

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

全期間依存制約をもつ確率システム上の決定過程問題に対し，動的計画法による再帰式を導く．この問題は，制約中に，通常の目的関数と同型あるいは類似した形の関数が現れるものである．一般に 2 種類の評価 — 例えばリスクとリターン — を考えた場合，リスクを一定以下に抑えつつリターンを最大化したり，逆にリターンを一定以上に保ちつつリスクを最小にしたりといった基準が求められる．決定過程問題に全期間依存制約を導入することで，この種の問題に対応することが可能となる。これまでの研究により，確定システム上で考えた場合 ([1])，および確率システム上で加法型の制約を扱った場合 ([2]) についてはすでに結果が得られた．本報告の目的は，確率システム上の決定過程において，より一般的な 2 種類の評価を扱うことである．すなわち，目的関数および制約部にあらわれる関数をともに加法型に限定せず，より一般の結合律を満たす演算子 ([3]) を導入することによって，多様な評価へ対応させることである．以下，本文中で用いる記号と用語を述べる．ただし，$\mathbf{R}$ は実数全体を表すものとし，$D, D' \subset \mathbf{R}$ とする．

(1) $N \geq 2$ は段の総数を表す正整数

(2) $X = \{s_1, s_2, \ldots, s_l\}$ は有限状態空間

(3) $U = \{a_1, a_2, \ldots, a_k\}$ は有限決定空間

(4) p はマルコフ推移法則;

$$p(y|x,u) \geq 0 \quad \forall (x,u,y) \in X \times U \times X,$$
$$\sum_{y \in X} p(y|x,u) = 1 \quad \forall (x,u) \in X \times U;$$

$p(y|x,u)$ は，状態 x で決定 u をとったとき，次の状態が y になる条件つき確率を表す．

このpにより表される確率的推移を $y \sim p(\cdot|x,u)$ と表現する．

(5) $r_n : X \times U \to D$ は第 n 利得関数 $(n = 0, 1, 2, \ldots, N-1)$;

$r_n(x,u)$ は第 n 期に状態 x で決定 u をとったとき，システムが得られる利得（リターン）を表す．

(6) $r_G : X \to D$ は終端利得関数；

$r_G(x)$ は最終 N 期に状態 x になったとき，システムが得られる利得を表す．

(7) $q_n : X \times U \to D'$ は第 n 損失関数 $(n = 0, 1, 2, \ldots, N-1)$;

$q_n(x,u)$ は第 n 期に状態 x で決定 u をとったとき，システムが被る損失（リスク）を表す．

(8) $q_G : X \to D'$ は終端損失関数；

$q_G(x)$ は最終 N 期に状態 x になったとき，システムが被る損失を表す．

2 全期間依存制約つき決定過程

2.1 定式化

初期状態 x_0 は与えられるものとする．決定 $u_0 \in U$ をとるとき，次期の状態 x_1 は確率的推移：

$$x_1 \sim p(\cdot|x_0, u_0)$$

に従う．同様に繰り返し，終了期までの状態決定列：

$$u_1 \in U,\ x_2 \sim p(\cdot|x_1, u_1),\ u_2 \in U,\ x_3 \sim p(\cdot|x_2, u_2), \ldots,\ u_{N-1} \in U,\ x_N \sim p(\cdot|x_{N-1}, u_{N-1})$$

を得る．ここで，$\circ$ は D 上で定義された結合律を満たす 2 項演算子で単位元 $\tilde{\lambda}$ をもつものとし，次の結合型評価の期待値：

$$E[r_0(x_0, u_0) \circ r_1(x_1, u_1) \circ \cdots \circ r_{N-1}(x_{N-1}, u_{N-1}) \circ r_G(x_N)]$$

の最大化を考える．そして，これと同時に，$\bullet$ を D' 上で定義された結合律を満たす 2 項演算子で単位元 $\tilde{\mu}$ をもつものとし，次の結合型制約を課す．

$$E[q_0(x_0, u_0) \bullet q_1(x_1, u_1) \bullet \cdots \bullet q_{N-1}(x_{N-1}, u_{N-1}) \bullet q_G(x_N)] \ \leq\ T$$

ただし，T は損失の最大許容値であり，定数として与えられるものとする．このとき，問題は次のように定式化される．

$$\begin{array}{ll} \text{Maximize} & E[r_0(x_0, u_0) \circ r_1(x_1, u_1) \circ \cdots \circ r_{N-1}(x_{N-1}, u_{N-1}) \circ r_G(x_N)] \\ \text{subject to} & \left\{ \begin{array}{l} x_{n+1} \sim p(\,\cdot\,|x_n, u_n) \qquad n = 0, 1, 2, \ldots, N-1 \\ \sigma = \{\sigma_0, \sigma_1, \cdots, \sigma_{N-1}\} \\ E[q_0(x_0, u_0) \bullet q_1(x_1, u_1) \bullet \cdots \bullet q_{N-1}(x_{N-1}, u_{N-1}) \bullet q_G(x_N)] \ \leq\ T \end{array} \right. \end{array}$$

これを全期間依存制約付問題と呼ぶ．ここで $\sigma = \{\sigma_0, \sigma_1, \cdots, \sigma_{N-1}\}$ は一般政策 ([6]) をあらわす．一般政策とは, 各期において，その時点までのすべての状態に依存して決定を定める一般決定関数の列：

$$\sigma_n \ :\ X^n \to U \qquad n = 0, 1, 2, \ldots, N-1$$

からなる政策である．すなわち，全期間依存制約付問題における決定は，一般政策 σ により

$$u_0 = \sigma_0(x_0),\ u_1 = \sigma_1(x_0, x_1),\ u_2 = \sigma_2(x_0, x_1, x_2),\ \ldots,\ u_{N-1} = \sigma_{N-1}(x_0, x_1, \ldots, x_{N-1})$$

と定まり，最大化 (Maximize) は，あらゆる一般政策に関してとられる．

2.2 再帰式

不変埋没原理 ([1]) を用いて再帰式を導く。全期間依存制約付問題において，新たなパラメータ $\lambda \in D$, $\mu \in D'$ を導入し，最大許容値 T をパラメータ $t \in \mathbf{R}$ に置き換えた次の問題を考える．

$$\begin{array}{ll} \text{Maximize} & E[\lambda \circ r_0(x_0, u_0) \circ r_1(x_1, u_1) \circ \cdots \circ r_{N-1}(x_{N-1}, u_{N-1}) \circ r_G(x_N)] \\ \text{subject to} & \left\{ \begin{array}{l} x_{n+1} \sim p(\,\cdot\,|x_n, u_n) \qquad n = 0, 1, 2, \ldots, N-1 \\ \sigma = \{\sigma_0, \sigma_1, \cdots, \sigma_{N-1}\} \\ E[\mu \bullet q_0(x_0, u_0) \bullet q_1(x_1, u_1) \bullet \cdots \bullet q_{N-1}(x_{N-1}, u_{N-1}) \bullet q_G(x_N)] \ \leq\ t \end{array} \right. \end{array}$$

この問題は，$\lambda=\tilde{\lambda}$, $\mu=\tilde{\mu}$, $t=T$ とおくとき，与問題と等価になる．すなわち，全期間依存制約付問題はこの問題に埋め込まれている．この意味で，この問題を埋め込み問題と呼ぶ．

次に，埋め込み問題の部分問題群を考える．これは，開始段 n と始発状態 x_n を変化させて得られる一連の部分問題で，その最適値を最適値関数 v^n とおく．

$$v^n(x_n,\,t,\,\lambda,\,\mu) \;=\; \underset{\substack{(\mathrm{i})_n,(\mathrm{ii})_n;\\ E[\mu\bullet q_n(x_n,u_n)\bullet\cdots\bullet q_G(x_N)]\le t}}{\mathrm{Max}} E[\lambda\circ r_n(x_n,u_n)\circ\cdots\circ r_{N-1}(r_{N-1},u_{N-1})\circ r_G(x_N)]$$

$$x_n\in X,\ t\in\mathbf{R},\ \lambda\in\mathbf{R},\ \mu\in D,\quad n=0,1,2,\ldots,N-1$$

ここで，

$$(\mathrm{i})_n \qquad x_{m+1}\sim p(\,\cdot\,|x_m,u_m) \qquad m=n,n+1,\ldots,N-1$$

$$(\mathrm{ii})_n \qquad \sigma=\{\sigma_n,\sigma_{n+1},\cdots,\sigma_{N-1}\}$$

である．ただし，このときの一般政策 σ は，開始段が n であるので

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

からなる．また，終了期 $(n=N)$ に対する値関数は

$$v^N(x_N,\,t,\,\lambda,\,\mu) \;=\; \begin{cases} \lambda\circ r_G(x_N) & \mu\bullet q_G(x_N)\le t \\ \text{—} & \mu\bullet q_G(x_N)>t \end{cases} \qquad x_N\in X,\ t\in\mathbf{R},\ \lambda\in\mathbf{R},\ \mu\in D$$

である．開始段 N に対する部分問題の制約は $\mu\bullet q_G(x_N)\le t$ であり，$\mu\bullet q_G(x_N)>t$ のときは常に制約を満たさないので，値なしという意味で，"—" とおく．同様に，開始段 n $(n=0,1,2,\ldots,N-1)$ に対する部分問題において，最大化の条件を満たす決定列 $u_n,u_{n+1},\cdots,u_{N-1}$ が存在しない場合も

$$v^n(x_n,\,t,\,\lambda,\,\mu) \;=\; \text{—}$$

と表す．

"—" に関する各演算については次のように定義する．$a\in\mathbf{R}$ に対して，

$$\begin{array}{lcl} a+\text{—}\;=\;\text{—} & \qquad & \text{—}+\text{—}\;=\;\text{—} \\ a\times\text{—}\;=\;\text{—} & & \text{—}\times\text{—}\;=\;\text{—} \\ \mathrm{Max}(a,\,\text{—})\;=\;a & & \mathrm{Max}(\text{—},\,\text{—})\;=\;\text{—} \end{array}$$

とする．

このとき次の再帰式が成り立つ

定理 1

$$v^N(x,\,t,\,\lambda,\,\mu) \;=\; \begin{cases} \lambda\circ r_G(x) & \mu\bullet q_G(x)\le t \\ \text{—} & \mu\bullet q_G(x)>t \end{cases} \qquad x\in X,\ t\in\mathbf{R},\ \lambda\in\mathbf{R},\ \mu\in D$$

$$v^n(x,\,t,\,\lambda,\,\mu) \;=\; \underset{u\in U}{\mathrm{Max}}\left[\underset{\substack{t_1,t_2,\cdots,t_l;\\ \sum_{i=1}^{l} t_i p(s_i|x,u)=t}}{\mathrm{Max}} \left\{\sum_{j=1}^{l} v^{n+1}(s_j,\ t_j,\ \lambda\circ r_n(x,u),\ \mu\bullet q_n(x,u))p(s_j|x,u)\right\}\right]$$

$$x\in X,\ t\in\mathbf{R},\ \lambda\in\mathbf{R},\ \mu\in D,\quad n=0,1,2,\ldots,N-1$$

この再帰式の等号は，両辺ともに値なしの場合を含む．

2.3 最適政策の構成

各部分問題の最適値 $v^n(x,\ t,\ \lambda,\ \mu)$ を与える決定 u^* を $\pi_n^*(x,\ t,\ \lambda,\ \mu)$，決定 u^* を取るとき，最適値を与える各 t_i^* を $\tau_{in}^*(x,\ u^*,\ t,\ \lambda,\ \mu)$ で表すとき，与問題の最適政策 $\sigma^* = \{\sigma_0^*, \sigma_1^*, \ldots, \sigma_{N-1}^*\}$ は次のように構成される．

$$
\begin{aligned}
\sigma_0^*(x_0) &= \pi_0^*(x_0, t_0, \lambda_0, \mu_0)\ , \quad t_0 = T,\ \lambda_0 = \hat{\lambda},\ \mu_0 = \hat{\mu} \\
\sigma_1^*(x_0,\ s_i) &= \pi_1^*(s_i, t_{i1}, \lambda_1, \mu_1)\ , \\
&\quad t_{i1} = \tau_{i1}^*(x_0,\ \sigma_0^*(x_0), t_0, \lambda_0, \mu_0), \\
&\quad \lambda_1 = \lambda_0 \circ r_0(x_0,\ \sigma_0^*(x_0)),\ \mu_1 = \mu_0 \bullet q_0(x_0,\ \sigma_0^*(x_0)) \\
&\ \vdots
\end{aligned}
$$

一般に

$$
\begin{aligned}
\sigma_n^*(x_0,\ \cdots,\ x_{n-1},\ s_i) &= \pi_n^*(s_i,\ t_{in}, \lambda_n, \mu_n)\ , \\
&\quad t_{in} = \tau_{in}^*(x_{n-1}, \sigma_{n-1}^*(x_0, \cdots, x_{n-1}), t_{n-1}, \lambda_{n-1}, \mu_{n-1}), \\
&\quad \lambda_n = \lambda_{n-1} \circ r_{n-1}(x_{n-1},\ \sigma_{n-1}^*(x_0, \cdots, x_{n-1})), \\
&\quad \mu_n = \mu_{n-1} \bullet q_{n-1}(x_{n-1},\ \sigma_{n-1}^*(x_0, \cdots, x_{n-1})) \\
&\qquad\qquad n = 1, 2, \ldots, N-1
\end{aligned}
$$

である．

2.4 定理１の証明

簡単のため期数 2 ($N = 2$)，状態数 2 ($X = \{s_1, s_2\}$) とし，各状態への推移確率は正とする (一般の場合も同様)．このとき

$$
v^1(x_1, t, \lambda, \mu) = \operatorname*{Max}_{u_1 \in U}\left[\operatorname*{Max}_{t_1, t_2; (*^1)}\left\{\sum_{j=1}^{2} v^2(s_j,\ t_j,\ \lambda \circ r_1(x_1, u_1),\ \mu \bullet q_1(x_1, u_1)) p(s_j|x_1, u_1)\right\}\right] \quad (1)
$$

$$
\left((*^1) : \sum_{i=1}^{2} t_i p(s_i|x_1, u_1) = t\right)
$$

$$
v^0(x_0, t, \lambda, \mu) = \operatorname*{Max}_{u_0 \in U}\left[\operatorname*{Max}_{t_1, t_2; (*^0)}\left\{\sum_{j=1}^{2} v^1(s_j,\ t_j,\ \lambda \circ r_0(x_0, u_0),\ \mu \bullet q_0(x_0, u_0)) p(s_j|x_0, u_0)\right\}\right] \quad (2)
$$

$$
\left((*^0) : \sum_{i=1}^{2} t_i p(s_i|x_0, u_0) = t\right)
$$

を示せばよいが，式 (1) については v^2 を具体的に代入することで比較的容易に示すことができるので，式 (2) について考える．

定義より，左辺 $v^0(x_0, t, \lambda, \mu)$ は

$$
\begin{aligned}
v^0(x_0, t, \lambda, \mu) &= \operatorname*{Max}_{\sigma = \{\sigma_0, \sigma_1\}; C_0(x_0, t, \mu)} \sum\sum_{x_1, x_2} \{\lambda \circ r_0(x_0, u_0) \circ r_1(x_1, u_1) \circ r_G(x_2)\} \\
&\qquad \times p_0(x_1|x_0, u_0) p_1(x_2|x_1, u_1)
\end{aligned}
$$

ただし

$$C_0(x_0,t,\mu) : E[\mu \bullet q_0(x_0,u_0) \bullet q_1(x_1,u_1) \bullet q_G(x_2)] \leq t$$

となるので

$$\begin{aligned}
&\operatorname*{Max}_{\sigma=\{\sigma_0,\sigma_1\};C_0(x_0,t,\mu)} \sum_{x_1,x_2}\sum \{\lambda \circ r_0(x_0,u_0) \circ r_1(x_1,u_1) \circ r_G(x_2)\} p_0(x_1|x_0,u_0) p_1(x_2|x_1,u_1) \\
&= \operatorname*{Max}_{u_0 \in U}\left[\operatorname*{Max}_{t_1,t_2;(*^0)}\left\{\sum_{j=1}^{2} v^1(s_j,\ t_j,\ \lambda \circ r_0(x_0,u_0),\ \mu \bullet q_0(x_0,u_0)) p(s_j|x_0,u_0)\right\}\right] \qquad (3)
\end{aligned}$$

を示せばよい．ここで $\sigma = \{\sigma_0,\ \sigma_1\}$ に対し

$$u_0 = \sigma_0(x_0)\ , \qquad \begin{cases} u_1^1 = \sigma_1(x_0,\ s_1) \\ u_1^2 = \sigma_1(x_0,\ s_2) \end{cases}$$

とおくと，左辺は

$$\begin{aligned}
\operatorname*{Max}_{u_0,u_1^1,u_1^2;C_0(x_0,t,\mu)} \Bigg[& \sum_{x_2} \{\lambda \circ r_0(x_0,u_0) \circ r_1(s_1,u_1^1) \circ r_G(x_2)\} p(s_1|x_0,u_0) p(x_2|s_1,u_1^1) \\
& + \sum_{x_2} \{\lambda \circ r_0(x_0,u_0) \circ r_1(s_2,u_1^2) \circ r_G(x_2)\} p(s_2|x_0,u_0) p(x_2|s_2,u_1^2) \Bigg] \qquad (4)
\end{aligned}$$

となり，決定に課されている制約 $C_0(x_0,t,\mu)$ は

$$\begin{aligned}
C_0(x_0,t,\mu) \quad : \quad & \sum_{x_2} \{\mu \bullet q_0(x_0,u_0) \bullet q_1(s_1,u_1^1) \bullet q_G(x_2)\} p(s_1|x_0,u_0) p(x_2|s_1,u_1^1) \\
& + \sum_{x_2} \{\mu \bullet q_0(x_0,u_0) \bullet q_1(s_2,u_1^2) \bullet q_G(x_2)\} p(s_2|x_0,u_0) p(x_2|s_2,u_1^2) \leq t
\end{aligned}$$

である．一方，$C_1(x_1,t,\mu)$ で制約：

$$C_1(x_1,t,\mu) \quad : \quad (\mu \bullet q_1(x_1,u_1) \bullet q_G(s_1)) p(s_1|x_1,u_1) + (\mu \bullet q_1(x_1,u_1) \bullet q_G(s_2)) p(s_2|x_1,u_1) \leq t$$

を表すとき，式 (3) の右辺は $v^1(x_1,t,\lambda,\mu)$ の定義から

$$\begin{aligned}
&\operatorname*{Max}_{u_0 \in U}\Bigg[\operatorname*{Max}_{t_1,t_2;(*^0)}\Bigg\{ \\
&\quad \operatorname*{Max}_{u_1^1;C_1(s_1,t_1,\mu \bullet q_0(x_0,u_0))} \sum_{x_2} \{\lambda \circ r_0(x_0,u_0) \circ r_1(s_1,u_1^1) \circ r_G(x_2)\} p(x_2|s_1,u_1^1) \times p(s_1|x_0,u_0) \\
&\quad + \operatorname*{Max}_{u_1^2;C_1(s_2,t_2,\mu \bullet q_0(x_0,u_0))} \sum_{x_2} \{\lambda \circ r_0(x_0,u_0) \circ r_1(s_2,u_1^2) \circ r_G(x_2)\} p(x_2|s_2,u_1^2) \times p(s_2|x_0,u_0) \Bigg\}\Bigg] \qquad (5)
\end{aligned}$$

と書ける．ただし右辺においては

$$u_1^1 \ = \ \sigma_1(s_1), \quad u_1^2 \ = \ \sigma_1(s_2)$$

とおいた．ここで $C_1(s_1,t_1,\mu \bullet q_0(x_0,u_0)), C_1(s_2,t_2,\mu \bullet q_0(x_0,u_0))$ はそれぞれ

$$\begin{aligned}
&(\mu \bullet q_0(x_0,u_0) \bullet q_1(s_1,u_1^1) \bullet q_G(s_1)) p(s_1|s_1,u_1^1) \\
&\qquad\qquad + (\mu \bullet q_0(x_0,u_0) \bullet q_1(s_1,u_1^1) \bullet q_G(s_2)) p(s_2|s_1,u_1^1) \leq t_1, \\
&(\mu \bullet q_0(x_0,u_0) \bullet q_1(s_2,u_1^2) \bullet q_G(s_1)) p(s_1|s_2,u_1^2) \\
&\qquad\qquad + (\mu \bullet q_0(x_0,u_0) \bullet q_1(s_2,u_1^2) \bullet q_G(s_2)) p(s_2|s_2,u_1^2) \leq t_2
\end{aligned}$$

を表している．このとき，第1式に $p(s_1|x_0,u_0)$，第2式に $p(s_2|x_0,u_0)$ をそれぞれ乗じて，さらに両辺をそれぞれ加えると

$$
\begin{aligned}
&(\mu \bullet q_0(x_0,u_0) \bullet q_1(s_1,u_1^1) \bullet q_G(s_1)) \times p(s_1|x_0,u_0)p(s_1|s_1,u_1^1)\\
&\quad +(\mu \bullet q_0(x_0,u_0) \bullet q_1(s_1,u_1^1) \bullet q_G(s_2)) \times p(s_1|x_0,u_0)p(s_2|s_1,u_1^1)\\
&\quad\quad +(\mu \bullet q_0(x_0,u_0) \bullet q_1(s_2,u_1^2) \bullet q_G(s_1)) \times p(s_2|x_0,u_0)p(s_1|s_2,u_1^2)\\
&\quad\quad\quad +(\mu \bullet q_0(x_0,u_0) \bullet q_1(s_2,u_1^2) \bullet q_G(s_2)) \times p(s_2|x_0,u_0)p(s_2|s_2,u_1^2)\\
&\le t_1 p(s_1|x_0,u_0) + t_2 p(s_2|x_0,u_0) \;=\; t
\end{aligned}
\tag{6}
$$

となり，これは制約 $C_0(x_0,t,\mu)$ に一致する．逆に制約 $C_0(x_0,t,\mu)$ が満たされるとき，式(6)より，$(*^0)$ をみたす t_1,t_2 が存在し，$C_1(s_1,t_1,\mu \bullet q_0(x_0,u_0)), C_1(s_2,t_2,\mu \bullet q_0(x_0,u_0))$ が成り立つ．たとえば，

$$
\begin{aligned}
t_1 &= (\mu \bullet q_0(x_0,u_0) \bullet q_1(s_1,u_1^1) \bullet q_G(s_1))p(s_1|s_1,u_1^1)\\
&\qquad\qquad +(\mu \bullet q_0(x_0,u_0) \bullet q_1(s_1,u_1^1) \bullet q_G(s_2))p(s_2|s_1,u_1^1)\\
t_2 &= \frac{t - t_1 p(s_1|x_0,u_0)}{p(s_2|x_0,u_0)}
\end{aligned}
$$

とおけばよい．したがって，左辺において $C_0(x_0,t,\mu)$ を満たす u_0,u_1^1,u_1^2 が存在する条件と，右辺において u_0,u_1^1,u_1^2 が存在する条件が一致する．すなわち両辺において，実行可能解が存在しない条件は一致し，この場合，式(2)は両辺とも値なしという意味で成り立つ．

後は両辺とも最大値が存在する場合について，それらが等しくなることを示せばよい．まず，式(4)において $C_0(x_0,t,\mu)$ を満たす決定の組のなかで，最大値を与えるものを $\hat{u}_0,\hat{u}_1^1,\hat{u}_1^2$ とおくと

$$
\begin{aligned}
(\text{左辺}) &= \sum_{x_2}\Big\{\lambda \circ r_0(x_0,\hat{u}_0) \circ r_1(s_1,\hat{u}_1^1) \circ r_G(x_2)\Big\}p(s_1|x_0,\hat{u}_0)p(x_2|s_1,\hat{u}_1^1)\\
&\qquad +\sum_{x_2}\Big\{\lambda \circ r_0(x_0,\hat{u}_0) \circ r_1(s_2,\hat{u}_1^2) \circ r_G(x_2)\Big\}p(s_2|x_0,\hat{u}_0)p(x_2|s_2,\hat{u}_1^2)\\
&= \left[\sum_{x_2}\Big\{\lambda \circ r_0(x_0,\hat{u}_0) \circ r_1(s_1,\hat{u}_1^1) \circ r_G(x_2)\Big\}p(x_2|s_1,\hat{u}_1^1)\right] p(s_1|x_0,\hat{u}_0)\\
&\qquad + \left[\sum_{x_2}\Big\{\lambda \circ r_0(x_0,\hat{u}_0) \circ r_1(s_2,\hat{u}_1^2) \circ r_G(x_2)\Big\}p(x_2|s_2,\hat{u}_1^2)\right] p(s_2|x_0,\hat{u}_0)
\end{aligned}
$$

と変形できる．先の議論より，$\hat{u}_1^1,\hat{u}_1^2$ に対しては，$(*^0)$ をみたす t_1,t_2 が存在して $C_1(s_1,t_1,\mu \bullet q_0(x_0,\hat{u}_0)), C_1(s_2,t_2,\mu \bullet q_0(x_0,\hat{u}_0))$ が成り立つので

$$
\begin{aligned}
&(\text{左辺})\\
&\le \operatorname*{Max}_{u_1^1;C_1(s_1,t_1,\mu\bullet q_0(x_0,\hat{u}_0))} \left[\sum_{x_2}\Big\{\lambda \circ r_0(x_0,\hat{u}_0) \circ r_1(s_1,u_1^1) \circ r_G(x_2)\Big\}p(x_2|s_1,u_1^1)\right] p(s_1|x_0,\hat{u}_0)\\
&\quad + \operatorname*{Max}_{u_1^2;C_1(s_2,t_2,\mu\bullet q_0(x_0,\hat{u}_0))} \left[\sum_{x_2}\Big\{\lambda \circ r_0(x_0,\hat{u}_0) \circ r_1(s_2,u_1^2) \circ r_G(x_2)\Big\}p(x_2|s_2,u_1^2)\right] p(s_2|x_0,\hat{u}_0)\\
&= \sum_{j=1}^{2} v^1(s_j,\ t_j,\ \lambda \circ r_0(x_0,\hat{u}_0),\ \mu \bullet q_0(x_0,u_0))p(s_j|x_0,\hat{u}_0)\\
&\le \operatorname*{Max}_{t_1,t_2;(*^0)} \left[\sum_{j=1}^{2} v^1(s_j,\ t_j,\ \lambda \circ r_0(x_0,\hat{u}_0),\ \mu \bullet q_0(x_0,\hat{u}_0))p(s_j|x_0,\hat{u}_0)\right]
\end{aligned}
$$

$$\leq \operatorname*{Max}_{u_0 \in U}\left[\operatorname*{Max}_{t_1,t_2;(*^0)}\left[\sum_{j=1}^{2} v^1(s_j,\ t_j,\ \lambda \circ r_0(x_0,u_0),\ \mu \bullet q_0(x_0,u_0))p(s_j|x_0,\hat{u}_0)\right]\right]$$
$$= (右辺) \tag{7}$$

次に，右辺を変形した式 (5) において $\tilde{u}_0, \tilde{t}_1, \tilde{t}_2$ が存在し次を満たす.

$$\begin{aligned}
&(右辺)\\
&= \operatorname*{Max}_{u_1^1;C_1(s_1,\tilde{t}_1,\mu\bullet q_0(x_0,\tilde{u}_0))}\sum_{x_2}\left\{\lambda \circ r_0(x_0,\tilde{u}_0) \circ r_1(s_1,u_1^1) \circ r_G(x_2)\right\}p(x_2|s_1,u_1^1) \times p(s_1|x_0,\tilde{u}_0)\\
&\quad + \operatorname*{Max}_{u_1^2;C_1(s_2,\tilde{t}_2,\mu\bullet q_0(x_0,\tilde{u}_0))}\sum_{x_2}\left\{\lambda \circ r_0(x_0,\tilde{u}_0) \circ r_1(s_2,u_1^2) \circ r_G(x_2)\right\}p(x_2|s_2,u_1^2) \times p(s_2|x_0,\tilde{u}_0)
\end{aligned}$$

さらに，$C_1(s_1,t_1,\mu \bullet q_0(x_0,\hat{u}_0)), C_1(s_2,t_2,\mu \bullet q_0(x_0,\hat{u}_0))$ をみたす $\tilde{u}_1^1,\ \tilde{u}_1^2$ がそれぞれ存在し

$$\begin{aligned}
(右辺) &= \sum_{x_2}\left\{\lambda \circ r_0(x_0,\tilde{u}_0) \circ r_1(s_1,\tilde{u}_1^1) \circ r_G(x_2)\right\}p(x_2|s_1,\tilde{u}_1^1) \times p(s_1|x_0,\tilde{u}_0)\\
&\qquad + \sum_{x_2}\left\{\lambda \circ r_0(x_0,\tilde{u}_0) \circ r_1(s_2,\tilde{u}_1^2) \circ r_G(x_2)\right\}p(x_2|s_2,\tilde{u}_1^2) \times p(s_2|x_0,\tilde{u}_0)
\end{aligned}$$

を満たす. ここで，$\tilde{u}_0, \tilde{u}_1^1,\ \tilde{u}_1^2$ はその定め方により，$C_0(x_0,t,\mu)$ を満たすので,

$$\begin{aligned}
(右辺) &\leq \operatorname*{Max}_{u_0,u_1^1,u_1^2;C_0(x_0,t,\mu)}\left[\sum_{x_2}\{\lambda \circ r_0(x_0,u_0) \circ r_1(s_1,u_1^1) \circ r_G(x_2)\}p(x_2|s_1,\tilde{u}_1^1) \times p(s_1|x_0,u_0)\right.\\
&\qquad \left. + \sum_{x_2}\{\lambda \circ r_0(x_0,u_0) \circ r_1(s_2,u_1^2) \circ r_G(x_2)\}p(x_2|s_2,\tilde{u}_1^2) \times p(s_2|x_0,u_0)\right]\\
&= (左辺)
\end{aligned} \tag{8}$$

よって式 (7) と式 (8) より式 (2) が示された.

参考文献

[1] R.E. Bellman and E.D. Denman, "Invariant Imbedding," Lect. Notes in Operation Research and Mathematical Systems, Vol. **52**, Springer-Verlag, Berlin, 1971.

[2] T. Fujita, Deterministic Decision Process under Range Constraint Through all Stages, Proceedings of MDAI 2008(CD-ROM), 60-70

[3] T. Fujita, Stochastic decision process under additive constraint through all stages, RIMS Kokyuroku **1636**, Kyoto Univ. (2009), 9-16

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

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

[6] S. Iwamoto, K. Tsurusaki and T. Fujita, On Markov Policies for Minimax Decision Processes, J. Math. Anal. Appl., **253**(2001), 58-78.