数理解析研究所講究録
1108 巻 1999 年 29-43

A Priori and a Posteriori Conditional Decision Processes with Nonadditively Recursive Utility

Seiichi Iwamoto
Department of Economic Engineering, Faculty of Economics
Kyushu University 27, Fukuoka 812-8581, Japan
Tel&Fax. +81(92)642-2488 Email: iwamoto@en.kyushu-u.ac.jp

Abstract

We consider stochastic optimization of not necessarily additive but recursive utilities over multi-stage decision processes. Without assuming any monotonicity, we optimize a regular process through a direct dynamic programming approach. On the regular decision process, we propose two related conditional decision processes: an a posteriori conditional decision process and an a priori. When the Markov transition law degenerates into a deterministic dynamics, the two conditional processes reduce to the same deterministic decision process. The conditional processes with monotonicity are optimized by the usual backward dynamic programming. We show that under additional convexity the regular process dominates the a priori in maximum value function and the priori does the a posteriori. We show that the a posteriori process illustrates Kreps and Porteus's dynamic choice problem. The numerical example also verifies the dominace relation in three optimal value functions.

1 Introduction

In this paper we are concerned with a broad class of multi-stage stochastic decision processes with recursive utility system. It is well known that a dynamic programming recursive equation is valid under both separability and monotonicity in criterion function ([1],[5],[6],[22],[23],[26],[27]). The criterion for stochastic optimization problem is the expected value, which is a multiple summation in discrete process ([4],[7],[10],[11],[18],[19], [25]). The expected value of additive or multiplicative utility is easily decomposed into the current (immediate) return and the resulting remaining. Thus the expected value is separable, because of lineality of expectation operator. Of course, it is monotone. We forcus on the decomposition (separability and monotonicity) of the *expected value* of *nonadditively* recursive utility, which is generated by the recursive system.

In Section 2, we optimize a regular process without an explicit monotonicity through a direct dynamic programming approach.

In Section 3, in contrast to the regular decision process, we propose two related conditional decision processes, which admit separability and monotonicity. One is an a posteriori conditional decision process. The other is an a priori. The conditional processes are optimized by the usual backward dynamic programming. We show that for a deterministic dynamics the two conditional decision processes reduce to a common deterministic decision process.

In Section 4, we cmpare three decision processes under convexity or concavity. It is shown that, for convex utility system, the maximum value function of regular process dominates that of a priori, which in turn dominates that of a posteriori and that the dominance is reversed for concave system.

In Section 5, we illustrate three decision processes through a three-state, two-decision and two-stage model. We show that Kreps and Porteus's dynamic choice problem is nothing but an a posteriori decision precess.

2 Regular Decision Process

Throughout the paper, the following data is given :

$$\begin{array}{l} N \geq 2 \text{ is an integer; the } \textit{total number of stages} \\ X = \{s_1,\, s_2, \ldots, s_p\} \text{ is a finite } \textit{state space} \\ U = \{a_1,\, a_2, \ldots, a_q\} \text{ is a finite } \textit{action space} \\ g_n : X \times U \times R^1 \to R^1 \text{ is an } \textit{n-th utility function} \quad (1 \leq n \leq N) \\ k : X \to R^1 \text{ is a } \textit{terminal utility function} \\ p \text{ is a } \textit{Markov transition law} \\ \quad : \; p(y|x,u) \geq 0 \;\; \forall (x,u,y) \in X \times U \times X, \quad \sum_{y \in X} p(y|x,u) = 1 \;\; \forall (x,u) \in X \times U \end{array} \tag{1}$$

$y \sim p(\cdot\,|x,u)$ denotes that next state y conditioned on state x and action u appears with probability $p(y|x,u)$.

We use the following notations :

$$\begin{aligned} & X^n := X \times X \cdots \times X \text{ (n-times)} \\ & H_n := X \times U \times X \times U \times \cdots \times X \quad ((2n-1)\text{-factors}) \\ & h_n := (x_1, u_1, x_2, u_2, \ldots, x_n) \\ & g_n(g_{n+1}(\cdots g_N(k)\cdots)) \\ & \quad := g_n(x_n, u_n; g_{n+1}(x_{n+1}, u_{n+1}; \ldots; g_N(x_N, u_N; k(x_{N+1}))\cdots)) \\ & E^{\sigma}_{x_1} g_1(g_2(\cdots g_N(k)\cdots)) \\ & \quad := \sum\sum\cdots\sum_{(x_2,\ldots,x_{N+1}) \in X^N} g_1(x_1, u_1; g_2(x_2, u_2; \cdots; g_N(x_N, u_N; k(x_{N+1}))\cdots)) \\ & \qquad\qquad \times p(x_2|x_1, u_1) p(x_3|x_2, u_2) \cdots p(x_{N+1}|x_N, u_N)\} \\ & \qquad\qquad (u_n = \sigma_n(x_1, \ldots, x_n) \quad 1 \leq n \leq N) \\ & E^u_x l := \sum_{y \in X} l(y) p(y|x,u) \quad \text{for } l = l(\cdot) \end{aligned}$$

As a *regular* decision process, we consider the following optimization problem subject to a succesive constraint :

$$\begin{array}{lll} \text{Maximize} & E^{\sigma}_{x_1} g_1(g_2(\cdots g_N(k)\cdots)) & \\ \text{subject to} & (\mathrm{i})_{\mathrm{n}} \quad x_{n+1} \sim p(\cdot|\, x_n, u_n), \quad u_n \in U \quad 1 \le n \le N & \end{array} \tag{2}$$

where $E^{\sigma}_{x_1}$ denotes the (regular) expectation operator on X^N induced from the conditional probability functions $p(x_{n+1}|x_n, u_n)$, a *general* policy $\sigma = \{\sigma_1, \sigma_2, ..., \sigma_N\}$ and an initial state x_1.

We derive directly a recursive formula for this process. Let us consider for any given n $(1 \le n \le N+1)$, $h_n = (x_1, u_1, x_2, u_2, \ldots, x_n) \in H_n$ the maximization problem :

$$v_n(h_n) = \operatorname*{Max}_{\mu} \; E^{\mu}_{h_n}[g_1(\cdots g_N(k)\cdots)|\,(\mathrm{i})_{\mathrm{m}} \;\; n \le m \le N] \quad h_n \in H_n, \;\; 1 \le n \le N \tag{3}$$

$$v_{N+1}(h_{N+1}) = g_1(x_1, u_1; \cdots; g_N(x_N, u_N; k(x_{N+1}))\cdots) \quad h_{N+1} \in H_{N+1} \tag{4}$$

where the sequence of action and state $(u_n, x_{n+1}, u_{n+1}, \ldots, u_N, x_{N+1})$ after starting state h_n is governed stochastically by a *primitive* policy $\mu = \{\mu_n, \mu_{n+1}, ..., \mu_N\}$ consisting of decision functions

$$\mu_m : H_m \to U \quad n \le m \le N \tag{5}$$

as follows :

$$\begin{aligned} &\mu_n(h_n) = u_n \;\to\; p(\cdot\,|x_n, u_n) \sim x_{n+1} \\ &\to\; \mu_{n+1}(h_{n+1}) = u_{n+1} \;\to\; p(\cdot\,|x_{n+1}, u_{n+1}) \sim x_{n+2} \\ &\to\; \cdots \;\to\; \mu_N(h_N) = u_N \;\to\; p(\cdot\,|x_N, u_N) \sim x_{N+1}. \end{aligned} \tag{6}$$

The maximization is taken for all *primitive* policies μ for the subprocess starting from state $h_n \in H_n$ at stage n and terminating at state $h_{N+1} \in H_{N+1}$. Note that any primitive policy $\mu = \{\mu_n, \mu_{n+1}, ..., \mu_N\}$ for the subprocess yields the expected value in (3) defined by the multiple summation :

$$\begin{aligned} &E^{\mu}_{h_n}[\,g_1(\cdots g_N(k)\cdots)|\,(\mathrm{i})_{\mathrm{m}} \;\; n \le m \le N\,] \\ &= \sum_{(x_{n+1},\ldots,x_{N+1}) \in X^{N-n+1}} \!\!\!\! \sum\cdots\sum \; g_1(x_1, u_1; \cdots; g_N(x_N, u_N; k(x_{N+1}))\cdots) \\ &\qquad\qquad \times p(x_{n+1}|x_n, u_n) \cdots p(x_{N+1}|x_N, u_N). \end{aligned} \tag{7}$$

Then we have the recursive equation between value $v_n(h)$ and two-variable function $v_{n+1}(h, \cdot, \cdot)$:

Theorem 1

$$v_n(h) = \operatorname*{Max}_{u \in U} E^u_x v_{n+1}(h, u, \cdot) \qquad h \in H_n, \quad n = 1, 2, \ldots, N \tag{8}$$

$$v_{N+1}(h) = g_1(x_1, u_1; \cdots; g_N(x_N, u_N; k(x_{N+1}))\cdots) \quad h \in H_{N+1}. \tag{9}$$

Proof The addition $a+b : R^1 \times R^1 \to R^1$ is commutative, associative, and distributive over multiplication $\times$. These properties imply the validity of recursive formula (8). □

Solving the recursive equation (8) yields an n-th optimal decision function $\mu_n^* : H_n \to U$. As a whole, we have a *primitive* optimal policy

$$\mu^* = \{\mu_1^*, \mu_2^*, ..., \mu_N^*\}.$$

By successively projecting the optimal decision function $\mu_n^* : H_n \to U$ onto the original state space X^n, we obtain a *general* optimal policy

$$\sigma^* = \{\sigma_1^*, \sigma_2^*, ..., \sigma_N^*\}$$

as follows :

$$\begin{aligned}
&\sigma_1^*(x_1) := \mu_1^*(h_1) \quad (\, h_1 = x_1 \,) \\
&\sigma_2^*(x_1, x_2) := \mu_2^*(h_2) \quad (\, h_2 = (x_1, u_1, x_2), \; u_1 = \mu_1^*(h_1) \,) \\
&\sigma_3^*(x_1, x_2, x_3) := \mu_3^*(h_3) \quad (\, h_3 = (h_2, u_2, x_3), \; u_2 = \mu_2^*(h_2) \,) \\
&\quad \cdots \\
&\sigma_N^*(x_1, x_2, \ldots, x_N) := \mu_N^*(h_N) \quad (\, h_N = (h_{N-1}, u_{N-1}, x_N), \; u_{N-1} = \mu_{N-1}^*(h_{N-1}) \,).
\end{aligned} \tag{10}$$

3 Conditional Decision Processes

In this section, we propose two conditional optimization problems subject to the successive constraint; one is an a posteriori conditional decision process and the other an a priori.

Throughout this section, we consider the class of all Markov policies on the original state space X. Note that any Markov policy $\pi = \{\pi_1, \pi_2, ..., \pi_N\}$ is specified by a sequence of Markov decision functions :

$$\pi_n : X \to U \quad 1 \le n \le N. \tag{11}$$

We assume that the utility system $\{g_n(x, u; \cdot)\}_1^N$ is *monotone* with respect to the third-variable :

$$a < b \implies g_n(x, u; a) \le g_n(x, u; b). \tag{12}$$

Then we are concerned with optimization of expected value of the backward accumulated returns :

$$\begin{aligned}
&E_{x_1}^{\pi} g_1(g_2(\cdots g_N(k) \cdots)) \\
= &\sum_{(x_2,\ldots,x_{N+1}) \in X^N} \!\!\!\sum \cdots \sum g_1(x_1, u_1; g_2(x_2, u_2; \cdots; g_N(x_N, u_N; k(x_{N+1})) \cdots)) \\
&\qquad \times p(x_2|x_1, u_1) p(x_3|x_2, u_2) \cdots p(x_{N+1}|x_N, u_N)
\end{aligned} \tag{13}$$

where the sequence of actions are determined through Markov policy π :

$$u_n = \pi_n(x_n) \quad 1 \le n \le N.$$

The multiple summation (13) is not necessarily decomposed into iterative (or repeated) summation. We present two types of decomposition by taking backward expectations. In the following subsections, we optimize such decomposed forms in the class of Markov policies.

3.1 A Posteriori Conditional Decision Process

First, we take at each stage backward conditional expectation of remaining process *after* performing take-action for regular decision process. This generates an *a posteriori* conditional decision process as follows:

$$\begin{array}{ll} \text{Maximize} & g_1(x_1, u_1; E_{x_1}^{u_1} g_2(x_2, u_2; \cdots; E_{x_{N-1}}^{u_{N-1}} g_N(x_N, u_N; E_{x_N}^{u_N} k) \cdots)) \\ \text{subject to} & (\mathrm{i})_\mathrm{n} \quad x_{n+1} \sim p(\cdot | x_n, u_n), \quad u_n \in U \quad 1 \le n \le N. \end{array} \tag{14}$$

Here we note that

$$E_x^u l = \sum_{y \in X} l(y) p(y|x, u) \qquad \text{for } l = l(\cdot). \tag{15}$$

For the sake of simplicity we use the following short notations :

$$E^n l \; := \; E_{x_n}^{u_n} l \tag{16}$$

$$g_n(E^n l) \; := \; g_n(x_n, u_n; E^n l) \quad 1 \le n \le N. \tag{17}$$

Thus the objective function in (14) is written as follows :

$$\begin{aligned} & g_1(E^1 g_2(\cdots E^{N-1} g_N(E^N k) \cdots)) \\ := \; & g_1(x_1, u_1; E_{x_1}^{u_1} g_2(x_2, u_2; \cdots; E_{x_{N-1}}^{u_{N-1}} g_N(x_N, u_N; E_{x_N}^{u_N} k) \cdots)). \end{aligned} \tag{18}$$

We should remark that Markov policy π is implicit in the notation E^n in (18). That is,

$$E^n l = E_{x_n}^{u_n} l, \qquad u_n = \pi_n(x_n) \;\; 1 \le n \le N. \tag{19}$$

Thus the a posteriori conditional expected value in (14) is not always equal to the called expected value (13). That is, in general, the equality

$$E_{x_1}^{\pi} g_1(g_2(\cdots g_N(k) \cdots)) \;=\; g_1(E^1 g_2(\cdots E^{N-1} g_N(E^N k) \cdots)) \tag{20}$$

does not hold. However, two typical processes admit the equality (20). One is the additive process : $g_n(x, u; h) = g_n(x, u) + h$. The other is the multiplicative process with nonnegative stage-wise return : $g_n(x, u; h) = g_n(x, u) \times h \;\; (g_n(x, u) \ge 0)$. Throughout the remainder, we are mainly concerned with the class of processes which do not admit the equality (20).

Let us consider for any given n $(1 \le n \le N+1)$, $x_n \in X$ the maximization problem :

$$w_n(x_n) = \operatorname*{Max}_{\pi} \; [\, g_n(E^n g_{n+1}(\cdots E^{N-1} g_N(E^N k) \cdots)) \mid (\mathrm{i})_\mathrm{m} \;\; n \le m \le N \,] \tag{21}$$

$$w_{N+1}(x_{N+1}) = k(x_{N+1}). \tag{22}$$

Then we have the recursive equation between value $w_n(x)$ and one-variable function $w_{n+1} = w_{n+1}(\cdot)$:

Theorem 2

$$w_n(x) = \operatorname*{Max}_{u \in U} g_n(x, u; E_x^u w_{n+1}) \quad x \in X, \;\; n = 1, 2, \ldots, N \tag{23}$$

$$w_{N+1}(x) = k(x) \qquad x \in X. \tag{24}$$

Proof The monotonicity in utility systems implies the validity of the recursive formula. □

The validity of recursive formula (23),(24) is equivalent to the validity of equality

$$\begin{aligned}
& \operatorname*{Max}_{\pi} g_1(E^1 g_2 \cdots E^{N-1} g_N(E^N k)\cdots)) \\
= \ & \operatorname*{Max}_{\pi_1} g_1(E^1 \operatorname*{Max}_{\pi_2} g_2(\cdots E^{N-1} \operatorname*{Max}_{\pi_N} g_N(E^N k)\cdots)) \qquad (25) \\
& \qquad (u_n = \pi_n(x_n) \ \ 1 \le n \le N).
\end{aligned}$$

We remark that the a posteriori cdp (14) is expressed in the following problem with backward aggregated return-variables $\{m_n(\cdot)\}_{n=1}^{N+1}$:

$$\begin{aligned}
\text{Maximize} \ \ & m_1(x_1) \\
\text{subject to} \ \ & (\mathrm{i})_n \ \ x_{n+1} \sim p(\cdot | x_n, u_n), \ \ u_n \in U \quad 1 \le n \le N \qquad (26) \\
& (\mathrm{ii}) \ \ m_{N+1}(x_{N+1}) = k(x_{N+1}) \\
& (\mathrm{iii})_n \ \ m_n = g_n(E^n m_{n+1}) \quad N \ge n \ge 1.
\end{aligned}$$

3.2 A Priori Conditional Decision Process

Second, *before* in turn performing take-action for regular decision process, we take at each stage backward conditional expectation of remaining process. This generates the following *a priori* conditional decision process:

$$\begin{aligned}
\text{Maximize} \ \ & E_{x_1}^{u_1} g_1(x_1, u_1; E_{x_2}^{u_2} g_2(x_2, u_2; \cdots ; E_{x_N}^{u_N} g_N(x_N, u_N; k)\cdots)) \qquad (27) \\
\text{subject to} \ \ & (\mathrm{i})_n \ \ x_{n+1} \sim p(\cdot | x_n, u_n), \ \ u_n \in U \quad 1 \le n \le N.
\end{aligned}$$

Here we note that

$$E_x^u g_n(x, u; l) = \sum_{y \in X} g_n(x, u; l(y)) p(y|x, u) \quad \text{for } l = l(\cdot). \qquad (28)$$

We use the following short notations :

$$E^n g_n(l) := E^n g_n(x_n, u_n; l) := E_{x_n}^{u_n} g_n(x_n, u_n; l) \quad 1 \le n \le N. \qquad (29)$$

Henceforth, the objective function in (27) is written as follows :

$$\begin{aligned}
& E^1 g_1(E^2 g_2(\cdots E^N g_N(k)\cdots)) \\
:= \ & E_{x_1}^{u_1} g_1(x_1, u_1; E_{x_2}^{u_2} g_2(x_2, u_2; \cdots ; E_{x_N}^{u_N} g_N(x_N, u_N; k)\cdots)). \qquad (30)
\end{aligned}$$

In the above notation E^n, the relevant Markov policy π is also implicit :

$$E^n g_n(l) = E_{x_n}^{u_n} g_n(l), \qquad u_n = \pi_n(x_n) \ \ 1 \le n \le N. \qquad (31)$$

We remark that the a priori conditonal expected value in (27) is not always identical with the a posteriori in (14). It may also different from the so-called expected value (13). However, the three expected values are identical both for the additive process and for the multiplicative process. The reason is nothing but the linearity of the expectation operator.

Let us consider for any given n $(1 \leq n \leq N+1)$, $x_n \in X$ the maximization problem :

$$W_n(x_n) = \underset{\pi}{\text{Max}} \; [E^n g_n(E^{n+1} g_{n+1}(\cdots E^N g_N(k)\cdots)) \mid (\text{i})_\text{m} \;\; n \leq m \leq N\,] \quad (32)$$
$$W_{N+1}(x_{N+1}) = k(x_{N+1}). \quad (33)$$

Then we have the recursive equation between value $W_n(x)$ and one-variable function $W_{n+1} = W_{n+1}(\cdot)$:

Theorem 3

$$W_n(x) = \underset{u \in U}{\text{Max}}\, E^u_x g_n(x, u; W_{n+1}) \quad x \in X, \;\; n = 1, 2, \ldots, N \quad (34)$$
$$W_{N+1}(x) = k(x) \qquad x \in X. \quad (35)$$

Proof The monotonicity also implies the validity of the recursive formula. □
The recursive formula (34),(35) states the equality

$$\begin{aligned} & \underset{\pi}{\text{Max}}\, E^1 g_1(E^2 g_2(\cdots E^N g_N(k)\cdots)) \\ = \; & \underset{\pi_1}{\text{Max}}\, E^1 g_1(\underset{\pi_2}{\text{Max}}\, E^2 g_2(\cdots \underset{\pi_N}{\text{Max}}\, E^N g_N(k)\cdots)). \end{aligned} \quad (36)$$

We remark that the a priori cdp (27) is stated in the following problem with backward aggregated return-variables $\{m_n(\cdot)\}_{n=1}^{N+1}$:

$$\begin{array}{lll} \text{Maximize} & m_1(x_1) & \\ \text{subject to} & (\text{i})_\text{n} \;\; x_{n+1} \sim p(\cdot \mid x_n, u_n), \;\; u_n \in U \quad 1 \leq n \leq N & \quad (37) \\ & (\text{ii}) \;\; m_{N+1}(x_{N+1}) = k(x_{N+1}) & \\ & (\text{iii})_\text{n} \;\; m_n = E^n g_n(m_{n+1}) \quad N \geq n \geq 1 & \end{array}$$

3.3 Deterministic Decision Process

We consider the special dynamics where the Markov transition law $p = p(y|x, u)$ degenerates into a deterministic dynamics :

$$f = f(x, u) \text{ represents the successor state of } x \text{ for action } u. \quad (38)$$

(See also [5],[6],[8]). Then we have no difference between the a posterior conditional decision process and the a priori process :

$$\begin{array}{ll} \text{Maximize} & g_1(x_1, u_1; g_2(x_2, u_2; \cdots; g_N(x_N, u_N; k(x_{N+1}))\cdots)) \quad (39) \\ \text{subject to} & (\text{i})_\text{n} \;\; f(x_n, u_n) = x_{n+1}, \;\; u_n \in U \quad 1 \leq n \leq N \end{array}$$

Then the correponding optimal value functions $\{v_n(\cdot)\}$ satisfy the following :

Corollary 1

(i) $$W_n(x) = w_n(x) = v_n(x) \qquad x \in X \quad (40)$$

(ii) $$v_n(x) = \underset{u \in U}{\text{Max}}\, g_n(x, u\,; v_{n+1}(f(x, u)) \quad x \in X, \;\; n = 1, 2, \ldots, N \quad (41)$$
$$v_{N+1}(x) = k(x) \qquad x \in X. \quad (42)$$

4 Convexity/Concavity

In this section, we compare the optimal value functions of regular, a priori and a posteriori decision processes under an additional convexity or concavity. We say that utility system $\{g_n\}$ is *convex* (resp. *concave*) if $g_n(x,u;\cdot) : R^1 \to R^1$ is convex (resp. concave) for $(x,u) \in X \times U,\ 1 \le n \le N$.

Theorem 4 *Let utility system $\{g_n\}$ be convex. Then we have*

$$\begin{aligned} & E^{\pi}_{x_n} g_n(g_{n+1}(\cdots g_N(k)\cdots)) \\ \geq\ & E^n g_n(E^{n+1} g_{n+1}(\cdots E^N g_N(k)\cdots)) \\ \geq\ & g_n(E^n g_{n+1}(\cdots E^{N-1} g_N(E^N k)\cdots)) \end{aligned} \tag{43}$$

for any Markov policy $\pi = \{\pi_n, \pi_{n+1}, ..., \pi_N\}$. The inequalities are reversed under concavity.

Proof We prove the inequalities for two-stage convex processes, because the inequalities for N-stage processes are similarly proved. First we note that the convexity implies

$$E^{u_1}_{x_1} g_1(x_1, u_1; l) \ \geq\ g_1(x_1, u_1; E^{u_1}_{x_1} l)$$

and

$$E^{u_2}_{x_2} g_1(x_1, u_1; g_2(x_2, u_2; k)) \ \geq\ g_1(x_1, u_1; E^{u_2}_{x_2} g_2(x_2, u_2; k))$$

where

$$\begin{aligned} E^{u_2}_{x_2} g_1(x_1, u_1; g_2(x_2, u_2; k)) &= \sum_{x_3 \in X} g_1(x_1, u_1; g_2(x_2, u_2; k(x_3)))p(x_3|x_2, u_2) \\ g_1(x_1, u_1; E^{u_2}_{x_2} g_2(x_2, u_2; k)) &= g_1(x_1, u_1; \sum_{x_3 \in X} g_2(x_2, u_2; k(x_3))p(x_3|x_2, u_2)). \end{aligned}$$

Since the expectation operator $E^{u_1}_{x_1}$ is monotone, we have

$$\begin{aligned} & E^{\pi}_{x_1} g_1(x_1, u_1; g_2(x_2, u_2; k)) \\ =\ & E^{u_1}_{x_1} E^{u_2}_{x_2} g_1(x_1, u_1; g_2(x_2, u_2; k)) \\ \geq\ & E^{u_1}_{x_1} g_1(x_1, u_1; E^{u_2}_{x_2} g_2(x_2, u_2; k)). \end{aligned}$$

This implies

$$E^{\pi}_{x_1} g_1(g_2(k)) \geq E^1 g_1(E^2 g_2(k)). \tag{44}$$

Second, we have

$$E^{u_2}_{x_2} g_2(x_2, u_2; k) \geq g_2(x_2, u_2; E^{u_2}_{x_2} k).$$

This together with monotonicity of $g_1(x_1, u_1; \cdot)$ implies

$$g_1(x_1, u_1; E^{u_2}_{x_2} g_2(x_2, u_2; k)) \geq g_1(x_1, u_1; g_2(x_2, u_2; E^{u_2}_{x_2} k)) \tag{45}$$

Thus by taking expectation operator $E^{u_1}_{x_1}$ on both hand-sides, we get

$$\begin{aligned} E^{u_1}_{x_1} g_1(x_1, u_1; E^{u_2}_{x_2} g_2(x_2, u_2; k)) &\geq E^{u_1}_{x_1} g_1(x_1, u_1; g_2(x_2, u_2; E^{u_2}_{x_2} k)) \\ &\geq g_1(x_1, u_1; E^{u_1}_{x_1} g_2(x_2, u_2; E^{u_2}_{x_2} k)) \end{aligned}$$

which implies

$$E^1 g_1(E^2 g_2(k)) \geq g_1(E^1 g_2(E^2 k)). \tag{46}$$

This completes the proof. □

Theorem 5 *Let utility system $\{g_n\}$ be convex. Then we have for any given n $(1 \leq n \leq N)$*

$$\begin{aligned}
\text{(i)} \quad & v_n(h_n) \geq g_1(x_1, u_1; \cdots; g_{n-1}(x_{n-1}, u_{n-1}; W_n(x_n)) \cdots) \\
& \qquad \geq g_1(x_1, u_1; \cdots; g_{n-1}(x_{n-1}, u_{n-1}; w_n(x_n)) \cdots) \\
& \qquad\qquad h_n = (x_1, u_1, x_2, u_2, \ldots, x_n) \in H_n \\
\text{(ii)} \quad & W_n(x_n) \geq w_n(x_n) \qquad x_n \in S_n.
\end{aligned} \tag{47}$$

The inequalities are reversed for minimization problems under concavity.

Proof We show the inequalities for two-stage convex processes, because the inequalities for N-stage processes are similarly shown. First we note

$$v_3(x_1, u_1, x_2, u_2, x_3) = g_1(x_1, u_1; g_2(x_2, u_2; W_3(x_3))), \qquad W_3(x_3) = w_3(x_3) = k(x_3).$$

Let $\pi_2^* : X \to U$ be an optimal (Markov) decision function for the remaining one-stage a priori process. Then we have

$$E_{x_2}^{u_2^*} g_2(x_2, u_2; k) = W_2(x_2) \qquad u_2^* = \pi_2^*(x_2).$$

From the definition of $v_2(\cdot)$ and convexity of $g_1(x_1, u_1; \cdot)$ we have

$$\begin{aligned}
v_2(x_1, u_1, x_2) &\geq E_{x_2}^{u_2^*} g_1(x_1, u_1; g_2(x_2, u_2^*; k)) \\
&\geq g_1(x_1, u_1; E_{x_2}^{u_2^*} g_2(x_2, u_2^*; k)) \\
&= g_1(x_1, u_1; W_2(x_2)).
\end{aligned}$$

Further we get

$$\begin{aligned}
v_1(x_1) &= \operatorname*{Max}_{\pi_1} v_2(x_1, u_1, x_2) \\
&\geq \operatorname*{Max}_{\pi_1} g_1(x_1, u_1; W_2) \\
&= W_1(x_1).
\end{aligned}$$

Second, taking maximum operator $\operatorname*{Max}_{\pi_2}$ on both hand-sides

$$E^2 g_2(k) \geq g_2(E^2 k)$$

we have

$$W_2(x_2) \geq w_2(x_2).$$

Further by succesive operation of $\operatorname*{Max}_{\pi_2}$, $\operatorname*{Max}_{\pi_1}$ for

$$E^1 g_1(E^2 g_2(k)) \geq g_1(E^1 g_2(E^2 k))$$

we get

$$W_1(x_1) \geq w_1(x_1).$$

This completes the proof. $\square$

Example 1. (Non-additive/Additive Utility System [11]) Let

$$g_n(x, u; h) = [g_n(x, u) + h^\alpha]^{1/\alpha} \qquad g_n(x, u) \geq 0, \quad \alpha > 0. \tag{48}$$

Then $g_n(x, u; \cdot)$ is increasing on $[0, \infty)$. If $0 < \alpha < 1$ (resp. $\alpha > 1$), it is concave (resp. convex). Then the utility function (48) generates the recursive utility :

$$(g_1(x_1, u_1) + g_2(x_2, u_2) + \cdots + g_N(x_N, u_N) + k^\alpha(x_{N+1}))^{1/\alpha}.$$

If $\alpha = 1$, it is linear. The resulting utility is additive :

$$g_1(x_1, u_1) + g_2(x_2, u_2) + \cdots + g_N(x_N, u_N) + k(x_{N+1}).$$

Example 2. (Maximum/Minimum Utility System [9],[12],[13],[14],[15],[16],[17]) Let

$$g_n(x, u; h) = g_n(x, u) \vee h \quad (\text{resp. } g_n(x, u) \wedge h) \qquad -\infty < g_n(x, u) < \infty. \tag{49}$$

Then $g_n(x, u; \cdot)$ is nondecreasing and convex (resp. concave) on R^1. The utility system yields the maximum (resp. minimum) utility :

$$g_1(x_1, u_1) \vee g_2(x_2, u_2) \vee \cdots \vee g_N(x_N, u_N) \vee k(x_{N+1})$$

$$(\text{resp. } g_1(x_1, u_1) \wedge g_2(x_2, u_2) \wedge \cdots \wedge g_N(x_N, u_N) \wedge k(x_{N+1}).)$$

5 Examples

In this section, we illustrate three decision processes; a regular decision process and two conditional decision processes. One conditional decision process is the dynamic choice theory which has been originally introduced by Kreps and Porteus ([20],[21]). The other is its a priori process. In this section, we consider the utility system $\{g_n(x, u; \cdot)\}_1^N$ as follows :

$$g_n(x, u; h) = [g_n(u) + h^\alpha]^{1/\alpha} \qquad (g_n(u) \geq 0, \quad \alpha > 0). \tag{50}$$

(see also Epstein and Zin [3] and Ozaki and Streufert [24]). For the sake of simplicity we take the case $N = 2$, $\alpha = 2$ over Bellman and Zadeh's data [2, pp. B154]:

$$k(s_1) = 0.3 \qquad k(s_2) = 1.0 \qquad k(s_3) = 0.8 \tag{51}$$

$$g_2(a_1) = 1.0 \qquad g_2(a_2) = 0.6 \tag{52}$$

$$g_1(a_1) = 0.7 \qquad g_1(a_2) = 1.0 \tag{53}$$

$u_t = a_1$

$x_t \diagdown x_{t+1}$	s_1	s_2	s_3
s_1	0.8	0.1	0.1
s_2	0.0	0.1	0.9
s_3	0.8	0.1	0.1

$u_t = a_2$

$x_t \diagdown x_{t+1}$	s_1	s_2	s_3
s_1	0.1	0.9	0.0
s_2	0.8	0.1	0.1
s_3	0.1	0.0	0.9

5.1 Regular Decision Process

First we note that

$$\begin{aligned} & g_1(x_1, u_1; g_2(x_2, u_2; k(x_3))) \\ = \ & [g_1(u_1) + ([g_2(u_2) + k^2(x_3)]^{1/2})^2]^{1/2} \\ = \ & (g_1(u_1) + g_2(u_2) + k^2(x_3))^{1/2}. \end{aligned} \tag{54}$$

Then the resulting optimal equations (8),(9) reduce to the recursive equations :

$$\begin{aligned} v_3(h_3) &= (g_1(u_1) + g_2(u_2) + k^2(x_3))^{1/2} \\ v_2(h_2) &= \operatorname*{Max}_{u_2} \sum_{x_3} v_3(h_2, u_2, x_3) p(x_3 | x_2, u_2) \\ v_1(x_1) &= \operatorname*{Max}_{u_1} \sum_{x_2} v_2(x_1, u_1, x_2) p(x_2 | x_1, u_1) \end{aligned} \tag{55}$$

where

$$\operatorname*{Max}_{u_n} = \operatorname*{Max}_{u_n \in \{a_1, a_2\}}, \qquad \sum_{x_n} = \sum_{x_n \in \{s_1, s_2, s_3\}}. \tag{56}$$

First, we have $v_3(h_3); h_3 = (x_1, u_1, x_2, u_2, x_3)$:

$v_3(\cdot, \underline{a_1}, \cdot, u_2, x_3)$

$u_2 \setminus x_3$	s_1	s_2	s_3
a_1	1.3379	1.6432	1.5297
a_2	1.1790	1.5166	1.3928

$v_3(\cdot, \underline{a_2}, \cdot, u_2, x_3)$

$u_2 \setminus x_3$	s_1	s_2	s_3
a_1	1.4457	1.7321	1.6248
a_2	1.3000	1.6125	1.4967

Second we calculate $v_2(h_2); h_2 = (x_1, u_1, x_2)$:

$v_2(\cdot, u_1, x_2), \ \mu_2^*(\cdot, u_1, x_2)$

$u_1 \setminus x_2$	s_1	s_2	s_3
a_1	1.4828, a_2	1.5411, a_1	1.3876, a_1
a_2	1.5813, a_2	1.6355, a_1	1.4923, a_1

Here we note that

$$v_3(x_1, u_1, x_2, u_2, x_3) = v_3(x_1', u_1, x_2', u_2, x_3) \quad {}^\forall x_1, x_2, x_1', x_2' \in X$$

$$v_2(x_1, u_1, x_2) = v_2(x_1', u_1, x_2) \quad {}^\forall x_1, x_1' \in X.$$

Finally, we get

$$v_1(s_1) = 1.6301, \quad v_1(s_2) = 1.5778, \quad v_1(s_3) = 1.5012$$

$$\mu_1^*(s_1) = a_2, \quad \mu_1^*(s_2) = a_2, \quad \mu_1^*(s_3) = a_2.$$

The optimal primitive policy $\mu^* = \{\mu_1^*, \mu_2^*\}$ yields an optimal general policy $\sigma^* = \{\sigma_1^*, \sigma_2^*\}$:

$$\sigma_1^*(s_1) = a_2, \quad \sigma_1^*(s_2) = a_2, \quad \sigma_1^*(s_3) = a_2 \tag{57}$$

$$\begin{aligned} \sigma_2^*(s_1, s_1) = a_2, \quad \sigma_2^*(s_2, s_1) = a_2, \quad \sigma_2^*(s_3, s_1) = a_2 \\ \sigma_2^*(s_1, s_2) = a_1, \quad \sigma_2^*(s_2, s_2) = a_1, \quad \sigma_2^*(s_3, s_2) = a_1 \\ \sigma_2^*(s_1, s_3) = a_1 \quad \sigma_2^*(s_2, s_3) = a_1, \quad \sigma_2^*(s_3, s_3) = a_1. \end{aligned} \tag{58}$$

Note that this optimal general policy σ^* is Markov.

5.2 Dynamic Choice Process

We consider the following conditional optimization problem :

$$\begin{array}{ll} \text{Maximize} & [g_1(u_1) + (E^{u_1}_{x_1}[g_2(u_2) + (E^{u_2}_{x_2}k)^2]^{1/2})^2]^{1/2} \\ \text{subject to} & (\text{i})_n \quad x_{n+1} \sim p(\cdot \mid x_n, u_n), \quad u_n \in \{a_1, a_2\} \quad n = 1, 2 \end{array} \tag{59}$$

Then, the "deterministic" dynamic programming technique, as we have already pointed in §3.1, yields the identity

$$\begin{aligned} & \operatorname*{Max}_{\pi} \; [g_1(u_1) + (E^{u_1}_{x_1}[g_2(u_2) + (E^{u_2}_{x_2}k)^2]^{1/2})^2]^{1/2} \\ = \; & \operatorname*{Max}_{\pi_1} [g_1(u_1) + (E^{u_1}_{x_1} \operatorname*{Max}_{\pi_2} [g_2(u_2) + (E^{u_2}_{x_2}k)^2]^{1/2})^2]^{1/2} \\ & \qquad (u_n = \pi_n(x_n) \quad n = 1, 2). \end{aligned} \tag{60}$$

This reduces to the recurrence equations :

$$\begin{aligned} w_3(x_3) &= k(x_3) \\ w_2(x_2) &= \operatorname*{Max}_{u_2}[g_2(u_2) + (\sum_{x_3} w_3(x_3)p(x_3|x_2, u_2))^2]^{1/2} \\ w_1(x_1) &= \operatorname*{Max}_{u_1}[g_1(u_1) + (\sum_{x_2} w_2(x_2))p(x_2|x_1, u_1))^2]^{1/2}. \end{aligned} \tag{61}$$

We have the following optimal solution for the a posteriori conditional process :

$$w_3(s_1) = 0.3, \quad w_3(s_2) = 1.0, \quad w_3(s_3) = 0.8 \tag{62}$$

$$w_2(s_1) = 1.2103, \quad w_2(s_2) = 1.2932, \quad w_2(s_3) = 1.0846 \tag{63}$$

$$\pi_2(s_1) = a_2, \quad \pi_2(s_2) = a_1, \quad \pi_2(s_3) = a_1, \tag{64}$$

$$w_1(s_1) = 1.6282, \quad w_1(s_2) = 1.5667, \quad w_1(s_3) = 1.4845 \tag{65}$$

$$\pi_1(s_1) = a_2, \quad \pi_1(s_2) = a_2, \quad \pi_1(s_3) = a_2. \tag{66}$$

Now we evaluate $g_1(x_1, u_1; w_2(x_2))$:

$g_1(\cdot, u_1; w_2(x_2))$

$u_1 \setminus x_2$	s_1	s_2	s_3
a_1	1.4713	1.5403	1.3698
a_2	1.5700	1.6347	1.4753

where

$$g_1(\cdot, u_1; w_2(x_2)) = [g_1(u_1) + w_2^2(x_2)]^{1/2}.$$

5.3 A Priori Process

As an a aprior procees for the Kreps and Porteus's process ([20],[21]), we consider the following problem :

$$\begin{array}{ll} \text{Maximize} & E^{u_1}_{x_1}[g_1(u_1) + (E^{u_2}_{x_2}[g_2(u_2) + k^2]^{1/2})^2]^{1/2} \\ \text{subject to} & (\text{i})_n \;\; x_{n+1} \sim p(\cdot \mid x_n, u_n), \;\; u_n \in \{a_1, a_2\} \quad n = 1, 2. \end{array} \tag{67}$$

Then, for the preceding data, the corresponding recursive equation

$$\begin{aligned} W_3(x_3) &= k(x_3) \\ W_2(x_2) &= \operatorname*{Max}_{u_2} \sum_{x_3} [\, g_2(u_2) + W_3^2(x_3)\,]^{1/2} p(x_3|x_2, u_2) \\ W_1(x_1) &= \operatorname*{Max}_{u_1} \sum_{x_2} [\, g_1(u_1) + W_2^2(x_2)\,]^{1/2} p(x_2|x_1, u_1) \end{aligned} \tag{68}$$

yields in turn

$$W_3(s_1) = 0.3, \quad W_3(s_2) = 1.0, \quad W_3(s_3) = 0.8, \tag{69}$$

$$W_2(s_1) = 1.2215, \quad W_2(s_2) = 1.2940, \quad W_2(s_3) = 1.1047 \tag{70}$$

$$\pi_2^*(s_1) = a_2, \quad \pi_2^*(s_2) = a_1, \quad \pi_2^*(s_3) = a_1, \tag{71}$$

$$W_1(s_1) = 1.6297, \quad W_1(s_2) = 1.5754, \quad W_1(s_3) = 1.4990 \tag{72}$$

$$\pi_1^*(s_1) = a_2, \quad \pi_1^*(s_2) = a_2, \quad \pi_1^*(s_3) = a_2. \tag{73}$$

Now we evaluate $g_1(x_1, u_1; W_2(x_2))$:

$g_1(\,\cdot\,, u_1; W_2(x_2))$

$u_1 \setminus x_2$	s_1	s_2	s_3
a_1	1.4806	1.5409	1.3858
a_2	1.5786	1.6354	1.4901

where

$$g_1(\,\cdot\,, u_1; W_2(x_2)) = [g_1(u_1) + W_2^2(x_2)]^{1/2}.$$

Finally we observe that

$$\begin{aligned} v_3(x_1, u_1, x_2, u_2, x_3) = g_1(x_1, u_1; g_2(x_2, u_2; W_3(x_3))) &= g_1(x_1, u_1; g_2(x_2, u_2; w_3(x_3))) \\ v_2(x_1, u_1, x_2) \geq g_1(x_1, u_1; W_2(x_2)) &\geq g_1(x_1, u_1; w_2(x_2)) \\ v_1(x_1) \geq W_1(x_1) &\geq w_1(x_1) \end{aligned} \tag{74}$$

and

$$\begin{aligned} k(x_3) = W_3(x_3) &= w_3(x_3) \\ W_2(x_2) &\geq w_2(x_2). \end{aligned} \tag{75}$$

References

[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] L.G. Epstein and S.E. Zin, Substitution, risk aversion, and the temporal behavior of consumption and asset returns: a theoretical framework, Econometrica **57**(1989), 937-969.

[4] T. Fujita and K. Tsurusaki, Stochastic optimization of multiplicative functions with negative value, J. Operations Res. Soc. Japan,**41**(1998), No.3, 351-373.

[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] 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.

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

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

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

[11] S. Iwamoto, Conditional decision processes with recursive reward function, J. Math. Anal. Appl., to appear.

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

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

[14] S. Iwamoto, Y. Tsurusaki and T. Fujita, On Markov policies for minimax decision processes, under preparation.

[15] S. Iwamoto, K. Tsurusaki and T. Fujita, Conditional decision-making in a fuzzy environment, J. Operations Res. Soc. Japan, to appear.

[16] J. Kacprzyk and A.O. Esogbue, Fuzzy dynamic programming: Main developments and applications, Fuzzy Sets and Systems **81**(1996), 31-45.

[17] J. Kacprzyk and P. Staniewski, A new approach to the control of stochastic systems in a fuzzy environment, Archiwum Automatyki i Telemechaniki **XXV**(1980), 443-444.

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

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

[20] D.M. Kreps and E. Porteus, Temporal resolution of uncertainty and dynamic choice theory, Econometrica **46**(1978), 185-200.

[21] D.M. Kreps and E. Porteus, Dynamic choice theory and dynamic programming, Econometrica **47**(1979), 91-100.

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

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

[24] H. Ozaki and P. E. Streufert, Dynamic programming and non-additive stochastic objects, J. Math. Econo. **25**(1996), 391-442.

[25] M. L. Puterman, *Markov Decision Processes : discrete stochastic dynamic programming*, Wiley & Sons, New York, 1994.

[26] M. Sniedovich, A class of nonseparable dynamic programming problems, J. Opt. Theo. Anal. **52**(1987), 111-121.

[27] M. Sniedovich, *Dynamic Programming*, Marcel Dekker, Inc. NY, 1992.

Sets and Systems. **3**(1980), 253-259.