Journal of the Operations Research
Society of Japan
Vol. 38, No. 4, December 1995

STOCHASTIC DECISION-MAKING IN A FUZZY ENVIRONMENT

Seiichi Iwamoto
Kyushu University

Toshiharu Fujita
Kyushu University

(Received January 13, 1994; Revised September 19, 1994)

Abstract In this paper we consider multistage decision processes in Bellman and Zadeh's paper "Decision-making in a fuzzy environment" from a mathematical view-point. We propose another recursive equation which solves both stochastic and deterministic multistage decision processes in a fuzzy environment in their sense. Our result for deterministic processes coincides with their result. However, our result for stochastic processes is more or less different from theirs. Our "stochastic" recursive equation is derived through invariant imbedding. On the other hand, their "stochastic" recursive equation is a direct analogy for their deterministic one. As an example, we illustrate their numerical data, which verify the equality between simultaneous and sequential optimizations.

1 Introduction

It has been well known that dynamic programming is an iterative optimization technique which assures that in a sequential deterministic or stochastic system the simultaneous optimization is attained and calculated through a sequential optimization [1], [4]. The sequential optimization — optimization of expected value for stochastic problem — reduces to a recursive formula, which is sometimes called "recursive equation" or "Bellman equation".

In this paper, from such a dynamic programming viewpoint — sequential optimization assures simultaneous one —, we consider the stochastic decision-making problem in a fuzzy environment in Bellman and Zadeh [2].

Throughout the paper we use the same notations and problems as in [2]. However, we introduce a different notion and analysis from theirs. In Section 2, we consider the stochastic decision processes in a rigorous way. In Section 3, we treat the deterministic decision process as a special case of stochastic ones. On the other hand, Bellman and Zadeh have first derived mathematically a recursive equation for deterministic process. Then for stochastic process they have just replaced formally the recursive equation with a stochastic version through a straightforward analogy. This is a main difference between their approach and ours. Section 4 illustrates a numerical example.

2 Stochastic Multistage Decision Processes

We use the notations in §4 (Deterministic) Multistage Decision Processes and §5 Stochastic Systems in a Fuzzy Environment in Bellman and Zadeh [2, pp. B151–B155]. In this section, we focus our attention on [2, §5].

First, let us cite their stochastic multistage decision processes [2, pp. B153] in the following style.

As in the preceding (deterministic) problem, assume that the termination time N is fixed and that an initial state x_0 is specified. The system is assumed to be characterized by a conditional probability function $p(x_{t+1} \mid x_t, u_t)$. The problem is to maximize the probability

of attainment of the fuzzy goal at time N, subject to the fuzzy constraints $C^0, \cdots, C^{N-1}$.

If the fuzzy goal G^N is regarded as a fuzzy event in X, then the conditional probability of this event given x_{N-1} and u_{N-1} is expressed by

$$\mathrm{Prob}(G^N \mid x_{N-1}, u_{N-1}) = E\mu_{G^N}(x_N) = \sum_{x_N} p(x_N \mid x_{N-1}, u_{N-1})\mu_{G^N}(x_N) \tag{1}$$

Where E denotes the conditional expectation and μ_{G^N} is the membership function of the given fuzzy goal.

From a notational viewpoint, $E\mu_{G^N}(x_N)$ may be replaced with the adequate notation $E\mu_{G^N}(\bullet \mid x_{N-1}, u_{N-1})$. Thus Eq (1) had better be replaced with

$$\mathrm{Cond.Exp}(G^N \mid x_{N-1}, u_{N-1}) = E\mu_{G^N}(\bullet \mid x_{N-1}, u_{N-1}) = \sum_{x_N} p(x_N \mid x_{N-1}, u_{N-1})\mu_{G^N}(x_N). \tag{2}$$

We observe that Eq (1) expresses $\mathrm{Prob}(G^N \mid x_{N-1}, u_{N-1})$ or, equivalently, $E\mu_{G^N}(x_N)$, as a function of x_{N-1} and u_{N-1}, just as in the preceding (deterministic) problem $\mu_{G^N}(x_N)$ was expressed as a function of x_{N-1} and u_{N-1} via the deterministic dynamics

$$x_{t+1} = f(x_t, u_t), \quad t = 0, 1, 2, \cdots \tag{3}$$

This implies that $E\mu_{G^N}(x_N)$ can be treated in the same way as $\mu_{G^N}(x_N)$ was treated in the nonstochastic case, thus making it possible to reduce the solution of the problem under consideration to that of the preceding problem.

More specifically, the deterministic recurrence equations

$$\mu_{G^{N-\nu}}(x_{N-\nu}) = \mathrm{Max}_{u_{N-\nu}}(\mu_{N-\nu}(u_{N-\nu}) \wedge \mu_{G^{N-\nu+1}}(x_{N-\nu+1})) \tag{4}$$

$$x_{N-\nu+1} = f(x_{N-\nu}, u_{N-\nu}), \quad \nu = 1, \cdots, N, \tag{5}$$

are replaced by the stochastic ones

$$\mu_{G^{N-\nu}}(x_{N-\nu}) = \mathrm{Max}_{u_{N-\nu}}(\mu_{N-\nu}(u_{N-\nu}), E\mu_{G^{N-\nu+1}}(x_{N-\nu+1})) \tag{6}$$

$$E\mu_{G^{N-\nu+1}}(x_{N-\nu+1}) = \sum_{x_{N-\nu+1}} p(x_{N-\nu+1} \mid x_{N-\nu}, u_{N-\nu})\mu_{G^{N-\nu+1}}(x_{N-\nu+1})) \tag{7}$$

where $\mu_{G^{N-\nu}}(x_{N-\nu})$ denotes the membership of the fuzzy goal at $t = N - \nu$ induced by the fuzzy goal at $t = N - \nu + 1$, $\nu = 1, \cdots, N$.

The Eqs (6), (7) may be replaced with the following equations:

$$\mu_{G^{N-\nu}}(x_{N-\nu}) = \mathrm{Max}_{u_{N-\nu}}[\mu_{N-\nu}(u_{N-\nu}) \wedge E\mu_{G^{N-\nu+1}}(\bullet \mid x_{N-\nu}, u_{N-\nu})] \tag{8}$$

$$E\mu_{G^{N-\nu+1}}(\bullet \mid x_{N-\nu}, u_{N-\nu}) = \sum_{x_{N-\nu+1}} p(x_{N-\nu+1} \mid x_{N-\nu}, u_{N-\nu})\mu_{G^{N-\nu+1}}(x_{N-\nu+1}). \tag{9}$$

In fact, their Example in [2, pp. B154–B155] is calculated through Eqs (8), (9) (See also [3, pp. 153], [5, pp. 172]).

Second, let us consider the conditional optimization problem subject to a successive constraint as follows:

$$\begin{array}{ll} \text{Maximize} & E[\mu_0(u_0) \wedge \mu_1(u_1) \wedge \cdots \wedge \mu_{N-1}(u_{N-1}) \wedge \mu_{G^N}(x_N)] \\ \text{subject to} & (i)_n \ x_{n+1} \sim p(\bullet \mid x_n, u_n) \ 0 \le n \le N-1 \\ & (ii)_n \ u_n \in U \ 0 \le n \le N-1 \end{array} \tag{10}$$

where E denotes the expectation (integral) operator on $U \times X \times U \times X \cdots \times U \times X$ induced from the conditional probability functions $p(x_{n+1} \mid x_n, u_n)$, a policy $\pi = \{\pi_0, \pi_1, \cdots, \pi_{N-1}\}$ and an initial state x_0. Now, let us define for any given $x_{N-\nu}$ the subproblem:

$$\mu_{G^{N-\nu}}(x_{N-\nu}) = \text{Max } E[\mu_{N-\nu}(u_{N-\nu}) \wedge \cdots \wedge \mu_{N-1}(u_{N-1}) \wedge \mu_{G^N}(x_N) \mid (i)_m, (ii)_m \; N-\nu \leq m \leq N-1]. \tag{11}$$

Then we want to find a recursive equation between value $\mu_{G^{N-\nu}}(x_{N-\nu})$ and function $\{\mu_{G^{N-\nu+1}}(x_{N-\nu+1})\}$. However, it is somewhat difficult to get such an equation [4]. Thus, we imbed the problem into the following family of parameterized problems. Let us consider for any given $x_{N-\nu}$ and λ the maximization problem:

$$\begin{aligned}\mu_{G^{N-\nu}}(x_{N-\nu};\lambda) = \text{Max } E[&\lambda \wedge \mu_{N-\nu}(u_{N-\nu}) \wedge \cdots \wedge \mu_{N-1}(u_{N-1}) \wedge \mu_{G^N}(x_N) \\ &\mid (i)_m, \; (ii)_m \; N-\nu \leq m \leq N-1] \\ &1 \leq \nu \leq N\end{aligned} \tag{12}$$

$$\mu_{G^N}(x_N;\lambda) = \lambda \wedge \mu_{G^N}(x_N) \quad 0 \leq \lambda \leq 1. \tag{13}$$

Then we have the recursive equation between value $\mu_{G^{N-\nu}}(x_{N-\nu};\lambda)$ and two-variable function $\{\mu_{G^{N-\nu+1}}(x_{N-\nu+1};\lambda)\}$:

Theorem 1

$$\begin{aligned}\mu_{G^{N-\nu}}(x_{N-\nu};\lambda) = \text{Max}_{u_{N-\nu}} &\sum_{x_{N-\nu+1}} \mu_{G^{N-\nu+1}}(x_{N-\nu+1};\lambda \wedge \mu_{N-\nu}(u_{N-\nu})) \\ &\times p(x_{N-\nu+1} \mid x_{N-\nu}, u_{N-\nu}) \\ &x_{N-\nu} \in X, \quad 0 \leq \lambda \leq 1 \quad \nu = 1, 2, \cdots, N\end{aligned} \tag{14}$$

$$\mu_{G^N}(x_N;\lambda) = \lambda \wedge \mu_{G^N}(x_N) \qquad x_N \in X, \quad 0 \leq \lambda \leq 1. \tag{15}$$

Proof We have the identity

$$\begin{aligned}&\lambda \wedge [(\mu_{N-\nu}(u_{N-\nu}) \wedge \cdots \wedge \mu_{N-1}(u_{N-1}) \wedge \mu_{G^N}(x_N)] \\ =&(\lambda \wedge \mu_{N-\nu}(u_{N-\nu})) \wedge [\mu_{N-\nu+1}(u_{N-\nu+1}) \wedge \cdots \wedge \mu_{N-1}(u_{N-1}) \wedge \mu_{G^N}(x_N)].\end{aligned} \tag{16}$$

We note that the common value is denote by

$$\lambda \wedge \mu_{N-\nu}(u_{N-\nu}) \wedge \cdots \wedge \mu_{N-1}(u_{N-1}) \wedge \mu_{G^N}(x_N).$$

This completes the proof.□

Let $\tilde{\pi}_{N-\nu}(x_{N-\nu};\lambda)$ be any value of $u_{N-\nu}$ which attains the maximum in Eq (14). We call the sequence $\tilde{\pi} = \{\tilde{\pi}_0, \tilde{\pi}_1, \cdots, \tilde{\pi}_{N-1}\}$ an *optimal policy* for parametrized problems (12),(13). In the following, we should discriminate one-variable function $\mu_{G^{N-\nu}}(x_{N-\nu})$ from two-variable function $\mu_{G^{N-\nu}}(x_{N-\nu};\lambda)$. In general, we have the inequality

$$\mu_{G^{N-\nu}}(x_{N-\nu};\lambda) \neq \lambda \wedge \mu_{G^{N-\nu}}(x_{N-\nu}) \qquad \nu = 1, 2, \cdots, N-1. \tag{17}$$

(See [4]). However, we have for a sufficiently large value $\hat{\lambda}$ of λ

$$\mu_{G^{N-\nu}}(x_{N-\nu}) = \mu_{G^{N-\nu}}(x_{N-\nu};\hat{\lambda}). \tag{18}$$

For instance, choose $\hat{\lambda}$ satisfying

$$\begin{aligned}&\hat{\lambda} \geq \mu_m(u_m) \quad u_m \in U \quad N-\nu \leq m \leq N-1 \\ &\hat{\lambda} \geq \mu_{G^N}(x_N) \quad x_N \in X\end{aligned} \tag{19}$$

or more rigorously

$$\hat{\lambda} \geq \mathrm{Max}[\mu_{N-\nu}(u_{N-\nu}) \wedge \cdots \wedge \mu_{N-1}(u_{N-1}) \wedge \mu_{G^N}(x_N) \mid (i)_m,\ (ii)_m \quad N-\nu \leq m \leq N-1]. \tag{20}$$

Then we have the equality (18). Thus the desired maximum expected value $\mu_{G^0}(x_0)$ is given by $\mu_{G^0}(x_0;\hat{\lambda})$ for a sufficiently large value $\hat{\lambda}(\leq 1)$ of λ:

$$\mu_{G^0}(x_0) = \mu_{G^0}(x_0;\hat{\lambda}). \tag{21}$$

Here, of course, $\hat{\lambda} = 1$ is available, because of $0 \leq \mu_A(x) \leq 1$.

3 Deterministic Multistage Decision Processes

In this section, we focus our attention on [2, §4]. Let us reconsider the deterministic dynamics

$$x_{t+1} = f(x_t, u_t), \quad t = 0, 1, 2, \cdots, N-1 \tag{22}$$

in the following. This deterministic system is a special case of stochastic system:

$$p(x_{t+1} \mid x_t, u_t) = \delta_{f(x_t,u_t)}(x_{t+1}) \tag{23}$$

where $\delta_a(\bullet)$ is a Dirac's measure concentrated on a with probability one:

$$\delta_a(x) = \begin{cases} 1 & \text{for } x = a \\ 0 & \text{for } x \neq a. \end{cases}$$

Then the stochastic recursive equations (14),(15) reduce to the following ones:

$$\mu_{G^{N-\nu}}(x_{N-\nu};\lambda) = \mathrm{Max}_{u_{N-\nu}} \mu_{G^{N-\nu+1}}(f(x_{N-\nu}, u_{N-\nu}); \lambda \wedge \mu_{N-\nu}(u_{N-\nu})) \quad \nu = 1, 2, \cdots, N \tag{24}$$

$$\mu_{G^N}(x_N;\lambda) = \lambda \wedge \mu_{G^N}(x_N) \qquad 0 \leq \lambda \leq 1. \tag{25}$$

On the other hand, taking account of the deterministic system, we see that Eqs (11), (12) become as follows, respectively:

$$\mu_{G^{N-\nu}}(x_{N-\nu}) = \mathrm{Max}[\mu_{N-\nu}(u_{N-\nu}) \wedge \cdots \wedge \mu_{N-1}(u_{N-1}) \wedge \mu_{G^N}(x_N) \mid (i)_m,\ (ii)_m \quad N-\nu \leq m \leq N-1] \tag{26}$$

$$\mu_{G^{N-\nu}}(x_{N-\nu};\lambda) = \mathrm{Max}[\lambda \wedge \mu_{N-\nu}(u_{N-\nu}) \wedge \cdots \wedge \mu_{N-1}(u_{N-1}) \wedge \mu_{G^N}(x_N) \mid (i)_m,\ (ii)_m \quad N-\nu \leq m \leq N-1] \tag{27}$$

where, in the deterministic system, the constraints $(i)_n$, $(ii)_n$ mean

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

respectively. Now, if λ is a constant and g is any function, we have the identity

$$\mathrm{Max}_u[\lambda \wedge g(u)] = \lambda \wedge \mathrm{Max}_u g(u).$$

Consequently, we have the relation

$$\mu_{G^{N-\nu}}(x_{N-\nu};\lambda) = \mu_{G^{N-\nu}}(x_{N-\nu}) \wedge \lambda \qquad 0 \le \lambda \le 1, \quad x_{N-\nu} \in X, \quad 0 \le \nu \le N. \tag{28}$$

Substituting Eq (28) into Eq (24), we have

$$\mu_{G^{N-\nu}}(x_{N-\nu}) \wedge \lambda = \mathrm{Max}_{u_{N-\nu}}[\mu_{G^{N-\nu+1}}(f(x_{N-\nu}, u_{N-\nu})) \wedge \lambda \wedge \mu_{N-\nu}(u_{N-\nu})]. \tag{29}$$

This implies

$$\lambda \wedge \mu_{G^{N-\nu}}(x_{N-\nu}) = \lambda \wedge \{\mathrm{Max}_{u_{N-\nu}}[\mu_{N-\nu}(u_{N-\nu}) \wedge \mu_{G^{N-\nu+1}}(f(x_{N-\nu}, u_{N-\nu}))]\}.$$

Since λ is arbitrary in the interval [0,1], we finally obtain the desired "deterministic" recursive equation:

Theorem 2

$$\mu_{G^{N-\nu}}(x_{N-\nu}) = \mathrm{Max}_{u_{N-\nu}}[\mu_{N-\nu}(u_{N-\nu}) \wedge \mu_{G^{N-\nu+1}}(f(x_{N-\nu}, u_{N-\nu}))] \qquad x_{N-\nu} \in X, \qquad \nu = 1, 2, \cdots, N. \tag{30}$$

This equation coincides with Bellman and Zadeh's deterministic recurrence Eqs (4), (5):

$$\mu_{G^{N-\nu}}(x_{N-\nu}) = \mathrm{Max}_{u_{N-\nu}}[\mu_{N-\nu}(u_{N-\nu}) \wedge \mu_{G^{N-\nu+1}}(x_{N-\nu+1}))] \tag{31}$$

$$x_{N-\nu+1} = f(x_{N-\nu}, u_{N-\nu}), \qquad \nu = 1, \cdots, N. \tag{32}$$

4 Bellman and Zadeh's Example

Throughout this section, we use Bellman and Zadeh's example in [2, pp. B154] to verify that the sequential optimization assures the simultaneous optimization. Their numerical data are as follows:

$$\mu_{G^2}(\sigma_1) = 0.3, \qquad \mu_{G^2}(\sigma_2) = 1.0, \qquad \mu_{G^2}(\sigma_3) = 0.8 \tag{33}$$

$$\mu_1(\alpha_1) = 1.0, \qquad \mu_1(\alpha_2) = 0.6 \tag{34}$$

$$\mu_0(\alpha_1) = 0.7, \qquad \mu_0(\alpha_2) = 1.0 \tag{35}$$

$u_t = \alpha_1$

$x_t \backslash x_{t+1}$	σ_1	σ_2	σ_3
σ_1	0.8	0.1	0.1
σ_2	0.0	0.1	0.9
σ_3	0.8	0.1	0.1

$u_t = \alpha_2$

$x_t \backslash x_{t+1}$	σ_1	σ_2	σ_3
σ_1	0.1	0.9	0.0
σ_2	0.8	0.1	0.1
σ_3	0.1	0.0	0.9

4.1 Recursive Equations for Imbedded Problem

In this subsection, we apply the preceding recursive equations with parameter λ:

$$\mu_{G^{N-\nu}}(x_{N-\nu};\lambda) = \mathrm{Max}_{u_{N-\nu}} \sum_{x_{N-\nu+1}} \mu_{G^{N-\nu+1}}(x_{N-\nu+1}; \lambda \wedge \mu_{N-\nu}(u_{N-\nu})) \times p(x_{N-\nu+1} \mid x_{N-\nu}, u_{N-\nu}) \qquad \nu = 1, 2, \cdots, N \tag{36}$$

$$\mu_{G^N}(x_N;\lambda) = \lambda \wedge \mu_{G^N}(x_N) \qquad 0 \le \lambda \le 1. \tag{37}$$

First, letting

$$N = 2, \qquad \mu_{G^2}(\sigma_1) = 0.3, \qquad \mu_{G^2}(\sigma_2) = 1, \qquad \mu_{G^2}(\sigma_3) = 0.8,$$

we have

$$\begin{aligned} \mu_{G^2}(\sigma_1;\lambda) &= \lambda \wedge 0.3 \\ \mu_{G^2}(\sigma_2;\lambda) &= \lambda \wedge 1 \\ \mu_{G^2}(\sigma_3;\lambda) &= \lambda \wedge 0.8. \end{aligned} \tag{38}$$

Second, the equation

$$\mu_{G^1}(x_1;\lambda) = \mathrm{Max}_{u_1\in\{\alpha_1,\alpha_2\}} \sum_{x_2\in\{\sigma_1,\sigma_2,\sigma_3\}} \mu_{G^2}(x_2;\lambda\wedge\mu_1(u_1))p(x_2 \mid x_1,u_1) \tag{39}$$

for $x_1 = \sigma_1$ becomes as follows:

$$\begin{aligned} \mu_{G^1}(\sigma_1;\lambda) =& [((\lambda\wedge 1)\wedge 0.3)0.8 + ((\lambda\wedge 1)\wedge 1)0.1 + ((\lambda\wedge 1)\wedge 0.8)0.1] \\ & \vee [((\lambda\wedge 0.6)\wedge 0.3)0.1 + ((\lambda\wedge 0.6)\wedge 1)0.9 + ((\lambda\wedge 0.6)\wedge 0.8)0.0] \\ =& [(\lambda\wedge 0.3)0.8 + (\lambda\wedge 1)0.1 + (\lambda\wedge 0.8)0.1] \\ & \vee [(\lambda\wedge 0.3)0.1 + (\lambda\wedge 0.6)0.9 + (\lambda\wedge 0.6)0.0]. \end{aligned} \tag{40}$$

A simple calculation yields

$$\mu_{G^1}(\sigma_1;\lambda) = \begin{cases} \lambda & \text{for } 0\le\lambda\le 0.3 \\ 0.9\lambda + 0.03 & \text{for } 0.3\le\lambda\le 0.6 \\ 0.57 & \text{for } 0.6\le\lambda\le 1 \end{cases}$$

$$\tilde{\pi}_1(\sigma_1;\lambda) = \begin{cases} \alpha_1 \text{ or } \alpha_2 & \text{for } 0\le\lambda\le 0.3 \\ \alpha_2 & \text{for } 0.3\le\lambda\le 0.6 \\ \alpha_2 & \text{for } 0.6\le\lambda\le 1. \end{cases}$$

Similarly, we have

$$\mu_{G^1}(\sigma_2;\lambda) = \begin{cases} \lambda & \text{for } 0\le\lambda\le 0.3 \\ \lambda & \text{for } 0.3\le\lambda\le 0.8 \\ 0.1\lambda + 0.72 & \text{for } 0.8\le\lambda\le 1 \end{cases}$$

$$\tilde{\pi}_1(\sigma_2;\lambda) = \begin{cases} \alpha_1 \text{ or } \alpha_2 & \text{for } 0\le\lambda\le 0.3 \\ \alpha_1 & \text{for } 0.3\le\lambda\le 0.8 \\ \alpha_1 & \text{for } 0.8\le\lambda\le 1 \end{cases}$$

and

$$\mu_{G^1}(\sigma_3;\lambda) = \begin{cases} \lambda & \text{for } 0\le\lambda\le 0.3 \\ 0.9\lambda + 0.3 & \text{for } 0.3\le\lambda\le 0.6 \\ 0.57 & \text{for } 0.6\le\lambda\le 1 \end{cases}$$

$$\tilde{\pi}_1(\sigma_3;\lambda) = \begin{cases} \alpha_1 \text{ or } \alpha_2 & \text{for } 0\le\lambda\le 0.3 \\ \alpha_2 & \text{for } 0.3\le\lambda\le 0.6 \\ \alpha_2 & \text{for } 0.6\le\lambda\le 1. \end{cases}$$

Third, the equation

$$\mu_{G^0}(x_0;\lambda) = \mathrm{Max}_{u_0\in\{\alpha_1,\alpha_2\}} \sum_{x_1\in\{\sigma_1,\sigma_2,\sigma_3\}} \mu_{G^1}(x_1;\lambda\wedge\mu_0(u_0))p(x_1 \mid x_0,u_0) \tag{41}$$

yields

$$\mu_{G^0}(\sigma_1;\lambda) = \begin{cases} \lambda & \text{for } 0\le\lambda\le 0.3 \\ 0.99\lambda + 0.003 & \text{for } 0.3\le\lambda\le 0.6 \\ 0.9\lambda + 0.057 & \text{for } 0.6\le\lambda\le 0.8 \\ 0.09\lambda + 0.705 & \text{for } 0.8\le\lambda\le 1 \end{cases}$$

$$\tilde{\pi}_0(\sigma_1;\lambda) = \begin{cases} \alpha_1 \text{ or } \alpha_2 & \text{for } 0 \le \lambda \le 0.3 \\ \alpha_2 & \text{for } 0.3 \le \lambda \le 0.6 \\ \alpha_2 & \text{for } 0.6 \le \lambda \le 0.8 \\ \alpha_2 & \text{for } 0.8 \le \lambda \le 1 \end{cases}$$

$$\mu_{G^0}(\sigma_2;\lambda) = \begin{cases} \lambda & \text{for } 0 \le \lambda \le 0.3 \\ 0.91\lambda + 0.027 & \text{for } 0.3 \le \lambda \le 0.6 \\ 0.1\lambda + 0.513 & \text{for } 0.6 \le \lambda \le 0.7 \\ 0.1\lambda + 0.513 & \text{for } 0.7 \le \lambda \le 0.8 \\ 0.01\lambda + 0.585 & \text{for } 0.8 \le \lambda \le 1 \end{cases}$$

$$\tilde{\pi}_0(\sigma_2;\lambda) = \begin{cases} \alpha_1 \text{ or } \alpha_2 & \text{for } 0 \le \lambda \le 0.3 \\ \alpha_1 \text{ or } \alpha_2 & \text{for } 0.3 \le \lambda \le 0.6 \\ \alpha_1 \text{ or } \alpha_2 & \text{for } 0.6 \le \lambda \le 0.7 \\ \alpha_2 & \text{for } 0.7 \le \lambda \le 0.8 \\ \alpha_2 & \text{for } 0.8 \le \lambda \le 1 \end{cases}$$

and

$$\mu_{G^0}(\sigma_3;\lambda) = \begin{cases} \lambda & \text{for } 0 \le \lambda \le 0.3 \\ 0.91\lambda + 0.027 & \text{for } 0.3 \le \lambda \le 0.6 \\ 0.1\lambda + 0.513 & \text{for } 0.6 \le \lambda \le 0.7 \\ 0.583 & \text{for } 0.7 \le \lambda \le 1 \end{cases}$$

$$\tilde{\pi}_0(\sigma_3;\lambda) = \begin{cases} \alpha_1 \text{ or } \alpha_2 & \text{for } 0 \le \lambda \le 0.3 \\ \alpha_1 & \text{for } 0.3 \le \lambda \le 0.6 \\ \alpha_1 & \text{for } 0.6 \le \lambda \le 0.7 \\ \alpha_1 & \text{for } 0.7 \le \lambda \le 1. \end{cases}$$

Therefore, the conditional optimization problem:

$$\begin{aligned} &\text{Maximize } E[\mu_0(u_0) \wedge \mu_1(u_1) \wedge \mu_{G^2}(x_2)] \\ &\text{subject to } (i)_n \ x_{n+1} \sim p(\bullet \mid x_n, u_n) \quad n = 0,1 \\ &\qquad (ii)_n \ u_n \in \{\alpha_1, \alpha_2\} \quad n = 0,1 \end{aligned} \tag{42}$$

has the following maximum expected values:

$$\begin{aligned} \mu_{G^0}(\sigma_1) &= \mu_{G^0}(\sigma_1;1) = 0.795 \\ \mu_{G^0}(\sigma_2) &= \mu_{G^0}(\sigma_2;1) = 0.595 \\ \mu_{G^0}(\sigma_3) &= \mu_{G^0}(\sigma_3;1) = 0.583. \end{aligned} \tag{43}$$

These maximum expected values are yielded by optimal policy $\tilde{\pi} = \{\tilde{\pi}_0, \tilde{\pi}_1\}$ from initial state $(x_0;1)$. Figures 1, 2 and 3 give not only the optimal behaviors resulting from optimal policy $\tilde{\pi}$ but also the corresponding maximum expected values. Here, of course, a behavior is a cyclic sequence of state, action, stage-wise reward and one-step transition probability.

We use the following notation in Figures 1, 2 and 3.

$$u_0 = \tilde{\pi}_0(x_0;1),\ \mu_0 = \mu_0(u_0),\ p_0 = p(x_1 \mid x_0, u_0),\ x_1 \sim p(\bullet \mid x_0, u_0),\ \lambda_1 = 1 \wedge u_0$$

$$u_1 = \tilde{\pi}_1(x_1;\lambda_1),\ \mu_1 = \mu_1(u_1),\ p_1 = p(x_2 \mid x_1, u_1),\ x_2 \sim p(\bullet \mid x_1, u_1),\ \lambda_2 = \lambda_1 \wedge \mu_1$$

$$\mu_2 = \mu_{G^2}(x_2),\ \text{min} = \mu_0 \wedge \mu_1 \wedge \mu_2,\ \text{prob} = p_0, p_1,\ \text{multi.} = \text{prob} \times \text{min}$$

$$\text{max. ttl.} = \text{maximum total expected value.}$$

These maximum expected values are also obtained through the direct enumeration method as are shown in Tables 1, 2 and 3, respectively.

We use the following notations in Tables 1, 2 and 3.

$$\text{history} = x_0\ u_0\ \mu_0(u_0)\ p(x_1 \mid x_0, u_0)\ x_1\ u_1\ \mu_1(u_1)\ p(x_2 \mid x_1, u_1)\ x_2$$

$(x_0;1.0)$	u_0	μ_0	p_0	$(x_1;\lambda_1)$	u_1	μ_1	p_1	$(x_2;\lambda_2)$	μ_2	prob	min	multi.
							0.1	$(\sigma_1;0.6)$	0.3			
										0.01	0.3	0.003
			0.1	$(\sigma_1;1.0)$	α_2	0.6	*0.9*	$(\sigma_2;0.6)$	1.0			
										0.09	0.6	0.054
							0.0	$(\sigma_3;0.6)$	0.8			
										0.00	0.6	0.000
							0.0	$(\sigma_1;1.0)$	0.3			
										0.00	0.3	0.000
$(\sigma_1;1.0)$	α_2	1.0	*0.9*	$(\sigma_2;1.0)$	α_1	1.0	*0.1*	$(\sigma_2;1.0)$	1.0			
										0.09	1.0	0.090
							0.9	$(\sigma_3;1.0)$	0.8			
										0.81	0.8	0.648
							0.1	$(\sigma_1;0.6)$	0.3			
										0.00	0.3	0.000
			0.0	$(\sigma_3;1.0)$	α_2	0.6	*0.0*	$(\sigma_2;0.6)$	1.0			
										0.00	0.6	0.000
							0.9	$(\sigma_3;0.6)$	0.8			
										0.00	0.6	0.000
										max.	ttl.	**0.795**

Figure 1: optimal behavoir and maximum expected value from $(\sigma_1;1)$

$(x_0;1.0)$	u_0	μ_0	p_0	$(x_1;\lambda_1)$	u_1	μ_1	p_1	$(x_2;\lambda_2)$	μ_2	
									prob min	multi.
							0.1	$(\sigma_1;0.6)$	0.3	
									0.08 0.3	0.024
			0.8	$(\sigma_1;1.0)$	α_2	0.6	*0.9*	$(\sigma_2;0.6)$	1.0	
									0.72 0.6	0.432
							0.0	$(\sigma_3;0.6)$	0.8	
									0.00 0.6	0.000
							0.0	$(\sigma_1;1.0)$	0.3	
									0.00 0.3	0.000
$(\sigma_2;1.0)$	α_2	1.0	*0.1*	$(\sigma_2;1.0)$	α_1	1.0	*0.1*	$(\sigma_2;1.0)$	1.0	
									0.01 1.0	0.010
							0.9	$(\sigma_3;1.0)$	0.8	
									0.09 0.8	0.072
							0.1	$(\sigma_1;0.6)$	0.3	
									0.01 0.3	0.003
			0.1	$(\sigma_3;1.0)$	α_2	0.6	*0.0*	$(\sigma_2;0.6)$	1.0	
									0.00 0.6	0.000
							0.9	$(\sigma_3;0.6)$	0.8	
									0.09 0.6	0.054
									max. ttl.	**0.595**

Figure 2: optimal behavoir and maximum expected value from $(\sigma_2;1)$

$(x_0;1.0)$	u_0	μ_0	p_0	$(x_1;\lambda_1)$	u_1	μ_1	p_1	$(x_2;\lambda_2)$	μ_2 min	prob	multi.
							0.1	$(\sigma_1;0.6)$	0.3		
										0.08	0.024
			0.8	$(\sigma_1;0.7)$	α_2	0.6	*0.9*	$(\sigma_2;0.6)$	1.0		
										0.72	0.432
							0.0	$(\sigma_3;0.6)$	0.8		
										0.00	0.000
							0.0	$(\sigma_1;0.7)$	0.3		
										0.00	0.000
$(\sigma_3;1.0)$	α_1	0.7	*0.1*	$(\sigma_2;0.7)$	α_1	1.0	*0.1*	$(\sigma_2;0.7)$	1.0		
										0.01	0.007
							0.9	$(\sigma_3;0.7)$	0.8		
										0.09	0.063
							0.1	$(\sigma_1;0.6)$	0.3		
										0.01	0.003
			0.1	$(\sigma_3;0.7)$	α_2	0.6	*0.0*	$(\sigma_2;0.6)$	1.0		
										0.00	0.000
							0.9	$(\sigma_3;0.6)$	0.8		
										0.09	0.054
										max. ttl.	**0.583**

Figure 3: optimal behavoir and maximum expected value from $(\sigma_3;1)$

Table 1 : all behaviors from σ_1 and selection of maximum branch

history									ter.	path	min.	mult.	sub.	ttl.
σ_1	α_1	0.7	*0.8*	σ_1	α_1	1.0	*0.8*	σ_1	0.3	*0.64*	0.3	0.192		
σ_1	α_1	0.7	*0.8*	σ_1	α_1	1.0	*0.1*	σ_2	1.0	*0.08*	0.7	0.056	0.304	
σ_1	α_1	0.7	*0.8*	σ_1	α_1	1.0	*0.1*	σ_3	0.8	*0.08*	0.7	0.056		
σ_1	α_1	0.7	*0.8*	σ_1	α_2	0.6	*0.1*	σ_1	0.3	*0.08*	0.3	0.024		
σ_1	α_1	0.7	*0.8*	σ_1	α_2	0.6	*0.9*	σ_2	1.0	*0.72*	0.6	0.432	**0.456**	
σ_1	α_1	0.7	*0.8*	σ_1	α_2	0.6	*0.0*	σ_3	0.8	*0.0*	0.6	0		
σ_1	α_1	0.7	*0.1*	σ_2	α_1	1.0	*0.0*	σ_1	0.3	*0.0*	0.3	0		
σ_1	α_1	0.7	*0.1*	σ_2	α_1	1.0	*0.1*	σ_2	1.0	*0.01*	0.7	0.007	**0.070**	
σ_1	α_1	0.7	*0.1*	σ_2	α_1	1.0	*0.9*	σ_3	0.8	*0.09*	0.7	0.063		0.583
σ_1	α_1	0.7	*0.1*	σ_2	α_2	0.6	*0.8*	σ_1	0.3	*0.08*	0.3	0.024		
σ_1	α_1	0.7	*0.1*	σ_2	α_2	0.6	*0.1*	σ_2	1.0	*0.01*	0.6	0.006	0.036	
σ_1	α_1	0.7	*0.1*	σ_2	α_2	0.6	*0.1*	σ_3	0.8	*0.01*	0.6	0.006		
σ_1	α_1	0.7	*0.1*	σ_3	α_1	1.0	*0.8*	σ_1	0.3	*0.08*	0.3	0.024		
σ_1	α_1	0.7	*0.1*	σ_3	α_1	1.0	*0.1*	σ_2	1.0	*0.01*	0.7	0.007	0.038	
σ_1	α_1	0.7	*0.1*	σ_3	α_1	1.0	*0.1*	σ_3	0.8	*0.01*	0.7	0.007		
σ_1	α_1	0.7	*0.1*	σ_3	α_2	0.6	*0.1*	σ_1	0.3	*0.01*	0.3	0.003		
σ_1	α_1	0.7	*0.1*	σ_3	α_2	0.6	*0.0*	σ_2	1.0	*0.0*	0.6	0	**0.057**	
σ_1	α_1	0.7	*0.1*	σ_3	α_2	0.6	*0.9*	σ_3	0.8	*0.09*	0.6	0.054		
σ_1	α_2	1.0	*0.1*	σ_1	α_1	1.0	*0.8*	σ_1	0.3	*0.08*	0.3	0.024		
σ_1	α_2	1.0	*0.1*	σ_1	α_1	1.0	*0.1*	σ_2	1.0	*0.01*	1.0	0.01	0.042	
σ_1	α_2	1.0	*0.1*	σ_1	α_1	1.0	*0.1*	σ_3	0.8	*0.01*	0.8	0.008		
σ_1	α_2	1.0	*0.1*	σ_1	α_2	0.6	*0.1*	σ_1	0.3	*0.01*	0.3	0.003		
σ_1	α_2	1.0	*0.1*	σ_1	α_2	0.6	*0.9*	σ_2	1.0	*0.09*	0.6	0.054	**0.057**	
σ_1	α_2	1.0	*0.1*	σ_1	α_2	0.6	*0.0*	σ_3	0.8	*0.0*	0.6	0		
σ_1	α_2	1.0	*0.9*	σ_2	α_1	1.0	*0.0*	σ_1	0.3	*0.0*	0.3	0		
σ_1	α_2	1.0	*0.9*	σ_2	α_1	1.0	*0.1*	σ_2	1.0	*0.09*	1.0	0.09	**0.738**	
σ_1	α_2	1.0	*0.9*	σ_2	α_1	1.0	*0.9*	σ_3	0.8	*0.81*	0.8	0.648		**0.795**
σ_1	α_2	1.0	*0.9*	σ_2	α_2	0.6	*0.8*	σ_1	0.3	*0.72*	0.3	0.216		
σ_1	α_2	1.0	*0.9*	σ_2	α_2	0.6	*0.1*	σ_2	1.0	*0.09*	0.6	0.054	0.324	
σ_1	α_2	1.0	*0.9*	σ_2	α_2	0.6	*0.1*	σ_3	0.8	*0.09*	0.6	0.054		
σ_1	α_2	1.0	*0.0*	σ_3	α_1	1.0	*0.8*	σ_1	0.3	*0.0*	0.3	0		
σ_1	α_2	1.0	*0.0*	σ_3	α_1	1.0	*0.1*	σ_2	1.0	*0.0*	1.0	0	**0**	
σ_1	α_2	1.0	*0.0*	σ_3	α_1	1.0	*0.1*	σ_3	0.8	*0.0*	0.8	0		
σ_1	α_2	1.0	*0.0*	σ_3	α_2	0.6	*0.1*	σ_1	0.3	*0.0*	0.3	0		
σ_1	α_2	1.0	*0.0*	σ_3	α_2	0.6	*0.0*	σ_2	1.0	*0.0*	0.6	0	**0**	
σ_1	α_2	1.0	*0.0*	σ_3	α_2	0.6	*0.9*	σ_3	0.8	*0.0*	0.6	0		

Table 2 : all behaviors from σ_2 and selection of maximum branch

history	ter.	path	min.	mult.	sub.	ttl.
σ_2 α_1 0.7 *0.0* σ_1 α_1 1.0 *0.8* σ_1	0.3	*0.0*	0.3	0		
σ_2 α_1 0.7 *0.0* σ_1 α_1 1.0 *0.1* σ_2	1.0	*0.0*	0.7	0	**0**	
σ_2 α_1 0.7 *0.0* σ_1 α_1 1.0 *0.1* σ_3	0.8	*0.0*	0.7	0		
σ_2 α_1 0.7 *0.0* σ_1 α_2 0.6 *0.1* σ_1	0.3	*0.0*	0.3	0		
σ_2 α_1 0.7 *0.0* σ_1 α_2 0.6 *0.9* σ_2	1.0	*0.0*	0.6	0	**0**	
σ_2 α_1 0.7 *0.0* σ_1 α_2 0.6 *0.0* σ_3	0.8	*0.0*	0.6	0		
σ_2 α_1 0.7 *0.1* σ_2 α_1 1.0 *0.0* σ_1	0.3	*0.0*	0.3	0		
σ_2 α_1 0.7 *0.1* σ_2 α_1 1.0 *0.1* σ_2	1.0	*0.01*	0.7	0.007	**0.070**	
σ_2 α_1 0.7 *0.1* σ_2 α_1 1.0 *0.9* σ_3	0.8	*0.09*	0.7	0.063		0.583
σ_2 α_1 0.7 *0.1* σ_2 α_2 0.6 *0.8* σ_1	0.3	*0.08*	0.3	0.024		
σ_2 α_1 0.7 *0.1* σ_2 α_2 0.6 *0.1* σ_2	1.0	*0.01*	0.6	0.006	0.036	
σ_2 α_1 0.7 *0.1* σ_2 α_2 0.6 *0.1* σ_3	0.8	*0.01*	0.6	0.006		
σ_2 α_1 0.7 *0.9* σ_3 α_1 1.0 *0.8* σ_1	0.3	*0.72*	0.3	0.216		
σ_2 α_1 0.7 *0.9* σ_3 α_1 1.0 *0.1* σ_2	1.0	*0.09*	0.7	0.063	0.342	
σ_2 α_1 0.7 *0.9* σ_3 α_1 1.0 *0.1* σ_3	0.8	*0.09*	0.7	0.063		
σ_2 α_1 0.7 *0.9* σ_3 α_2 0.6 *0.1* σ_1	0.3	*0.09*	0.3	0.027		
σ_2 α_1 0.7 *0.9* σ_3 α_2 0.6 *0.0* σ_2	1.0	*0.0*	0.6	0	**0.513**	
σ_2 α_1 0.7 *0.9* σ_3 α_2 0.6 *0.9* σ_3	0.8	*0.81*	0.6	0.486		
σ_2 α_2 1.0 *0.8* σ_1 α_1 1.0 *0.8* σ_1	0.3	*0.64*	0.3	0.192		
σ_2 α_2 1.0 *0.8* σ_1 α_1 1.0 *0.1* σ_2	1.0	*0.08*	1.0	0.08	0.336	
σ_2 α_2 1.0 *0.8* σ_1 α_1 1.0 *0.1* σ_3	0.8	*0.08*	0.8	0.064		
σ_2 α_2 1.0 *0.8* σ_1 α_2 0.6 *0.1* σ_1	0.3	*0.08*	0.3	0.024		
σ_2 α_2 1.0 *0.8* σ_1 α_2 0.6 *0.9* σ_2	1.0	*0.72*	0.6	0.432	**0.456**	
σ_2 α_2 1.0 *0.8* σ_1 α_2 0.6 *0.0* σ_3	0.8	*0.0*	0.6	0		
σ_2 α_2 1.0 *0.1* σ_2 α_1 1.0 *0.0* σ_1	0.3	*0.0*	0.3	0		
σ_2 α_2 1.0 *0.1* σ_2 α_1 1.0 *0.1* σ_2	1.0	*0.01*	1.0	0.01	**0.082**	
σ_2 α_2 1.0 *0.1* σ_2 α_1 1.0 *0.9* σ_3	0.8	*0.09*	0.8	0.072		**0.595**
σ_2 α_2 1.0 *0.1* σ_2 α_2 0.6 *0.8* σ_1	0.3	*0.08*	0.3	0.024		
σ_2 α_2 1.0 *0.1* σ_2 α_2 0.6 *0.1* σ_2	1.0	*0.01*	0.6	0.006	0.036	
σ_2 α_2 1.0 *0.1* σ_2 α_2 0.6 *0.1* σ_3	0.8	*0.01*	0.6	0.006		
σ_2 α_2 1.0 *0.1* σ_3 α_1 1.0 *0.8* σ_1	0.3	*0.08*	0.3	0.024		
σ_2 α_2 1.0 *0.1* σ_3 α_1 1.0 *0.1* σ_2	1.0	*0.01*	1.0	0.01	0.042	
σ_2 α_2 1.0 *0.1* σ_3 α_1 1.0 *0.1* σ_3	0.8	*0.01*	0.8	0.008		
σ_2 α_2 1.0 *0.1* σ_3 α_2 0.6 *0.1* σ_1	0.3	*0.01*	0.3	0.003		
σ_2 α_2 1.0 *0.1* σ_3 α_2 0.6 *0.0* σ_2	1.0	*0.0*	0.6	0	**0.057**	
σ_2 α_2 1.0 *0.1* σ_3 α_2 0.6 *0.9* σ_3	0.8	*0.09*	0.6	0.054		

Table 3 : all behaviors from σ_3 and selection of maximum branch

history	ter.	path	min.	mult.	sub.	ttl.
σ_3 α_1 0.7 *0.8* σ_1 α_1 1.0 *0.8* σ_1	0.3	*0.64*	0.3	0.192		
σ_3 α_1 0.7 *0.8* σ_1 α_1 1.0 *0.1* σ_2	1.0	*0.08*	0.7	0.056	0.304	
σ_3 α_1 0.7 *0.8* σ_1 α_1 1.0 *0.1* σ_3	0.8	*0.08*	0.7	0.056		
σ_3 α_1 0.7 *0.8* σ_1 α_2 0.6 *0.1* σ_1	0.3	*0.08*	0.3	0.024		
σ_3 α_1 0.7 *0.8* σ_1 α_2 0.6 *0.9* σ_2	1.0	*0.72*	0.6	0.432	**0.456**	
σ_3 α_1 0.7 *0.8* σ_1 α_2 0.6 *0.0* σ_3	0.8	*0.0*	0.6	0		
σ_3 α_1 0.7 *0.1* σ_2 α_1 1.0 *0.0* σ_1	0.3	*0.0*	0.3	0		
σ_3 α_1 0.7 *0.1* σ_2 α_1 1.0 *0.1* σ_2	1.0	*0.01*	0.7	0.007	**0.070**	
σ_3 α_1 0.7 *0.1* σ_2 α_1 1.0 *0.9* σ_3	0.8	*0.09*	0.7	0.063		**0.583**
σ_3 α_1 0.7 *0.1* σ_2 α_2 0.6 *0.8* σ_1	0.3	*0.08*	0.3	0.024		
σ_3 α_1 0.7 *0.1* σ_2 α_2 0.6 *0.1* σ_2	1.0	*0.01*	0.6	0.006	0.036	
σ_3 α_1 0.7 *0.1* σ_2 α_2 0.6 *0.1* σ_3	0.8	*0.01*	0.6	0.006		
σ_3 α_1 0.7 *0.1* σ_3 α_1 1.0 *0.8* σ_1	0.3	*0.08*	0.3	0.024		
σ_3 α_1 0.7 *0.1* σ_3 α_1 1.0 *0.1* σ_2	1.0	*0.01*	0.7	0.007	0.038	
σ_3 α_1 0.7 *0.1* σ_3 α_1 1.0 *0.1* σ_3	0.8	*0.01*	0.7	0.007		
σ_3 α_1 0.7 *0.1* σ_3 α_2 0.6 *0.1* σ_1	0.3	*0.01*	0.3	0.003		
σ_3 α_1 0.7 *0.1* σ_3 α_2 0.6 *0.0* σ_2	1.0	*0.0*	0.6	0	**0.057**	
σ_3 α_1 0.7 *0.1* σ_3 α_2 0.6 *0.9* σ_3	0.8	*0.09*	0.6	0.054		
σ_3 α_2 1.0 *0.1* σ_1 α_1 1.0 *0.8* σ_1	0.3	*0.08*	0.3	0.024		
σ_3 α_2 1.0 *0.1* σ_1 α_1 1.0 *0.1* σ_2	1.0	*0.01*	1.0	0.01	0.042	
σ_3 α_2 1.0 *0.1* σ_1 α_1 1.0 *0.1* σ_3	0.8	*0.01*	0.8	0.008		
σ_3 α_2 1.0 *0.1* σ_1 α_2 0.6 *0.1* σ_1	0.3	*0.01*	0.3	0.003		
σ_3 α_2 1.0 *0.1* σ_1 α_2 0.6 *0.9* σ_2	1.0	*0.09*	0.6	0.054	**0.057**	
σ_3 α_2 1.0 *0.1* σ_1 α_2 0.6 *0.0* σ_3	0.8	*0.0*	0.6	0		
σ_3 α_2 1.0 *0.0* σ_2 α_1 1.0 *0.0* σ_1	0.3	*0.0*	0.3	0		
σ_3 α_2 1.0 *0.0* σ_2 α_1 1.0 *0.1* σ_2	1.0	*0.0*	1.0	0	**0**	
σ_3 α_2 1.0 *0.0* σ_2 α_1 1.0 *0.9* σ_3	0.8	*0.0*	0.8	0		0.570
σ_3 α_2 1.0 *0.0* σ_2 α_2 0.6 *0.8* σ_1	0.3	*0.0*	0.3	0		
σ_3 α_2 1.0 *0.0* σ_2 α_2 0.6 *0.1* σ_2	1.0	*0.0*	0.6	0	**0**	
σ_3 α_2 1.0 *0.0* σ_2 α_2 0.6 *0.1* σ_3	0.8	*0.0*	0.6	0		
σ_3 α_2 1.0 *0.8* σ_3 α_1 1.0 *0.8* σ_1	0.3	*0.72*	0.3	0.216		
σ_3 α_2 1.0 *0.8* σ_3 α_1 1.0 *0.1* σ_2	1.0	*0.09*	1.0	0.09	0.378	
σ_3 α_2 1.0 *0.8* σ_3 α_1 1.0 *0.1* σ_3	0.8	*0.09*	0.8	0.072		
σ_3 α_2 1.0 *0.8* σ_3 α_2 0.6 *0.1* σ_1	0.3	*0.09*	0.3	0.027		
σ_3 α_2 1.0 *0.8* σ_3 α_2 0.6 *0.0* σ_2	1.0	*0.0*	0.6	0	**0.513**	
σ_3 α_2 1.0 *0.8* σ_3 α_2 0.6 *0.9* σ_3	0.8	*0.81*	0.6	0.486		

$$\text{ter.} = \text{terminal reward} = \mu_{G^2}(x_2)$$
$$\text{path} = \text{path probability} = p(x_1 \mid x_0, u_0) \times p(x_2 \mid x_1, u_1)$$
$$\text{min.} = \text{minimum} = \mu_0(u_0) \wedge \mu_1(u_1) \wedge \mu_{G^2}(x_2)$$
$$\text{mult.} = \text{multiplication} = \text{path} \times \text{min.}$$
$$\text{sub.} = \text{sub expected value}, \qquad \text{ttl.} = \text{total expected value.}$$

Furthermore, an *italic* number is a probability, and a **bold** number denotes a selection of the greater (maximum) value of the two up-and-down expected values.

4.2 Bellman and Zadeh's Recursive Equations

In this subsection, we consider Bellman and Zadeh's approach [2] with the preceding data. They have applied their recurrence equations (8), (9):

$$\mu_{G^{N-\nu}}(x_{N-\nu}) = \mathrm{Max}_{u_{N-\nu}}[\mu_{N-\nu}(u_{N-\nu}) \wedge E\mu_{G^{N-\nu+1}}(\bullet \mid x_{N-\nu}, u_{N-\nu})] \tag{44}$$

$$E\mu_{G^{N-\nu+1}}(\bullet \mid x_{N-\nu}, u_{N-\nu}) = \sum_{x_{N-\nu+1}} p(x_{N-\nu+1} \mid x_{N-\nu}, u_{N-\nu})\mu_{G^{N-\nu+1}}(x_{N-\nu+1}). \tag{45}$$

Their approach [2, pp. B154] solves the "deterministic" sequential optimization problem, which has been taken, as it were, the backward conditional expectations:

$$\begin{array}{ll} \text{Maximize} & [\mu_0(\pi_0) \wedge E^{\pi_0}[\mu_1(\pi_1) \wedge E^{\pi_1}\mu_{G^2}(x_2)]] \\ \text{subject to} & (i)_n \ x_{n+1} \sim p(\bullet \mid x_n, u_n) \quad n = 0, 1 \\ & (ii)_n \ \pi_n(x_n) \in \{\alpha_1, \alpha_2\} \quad n = 0, 1 \end{array} \tag{46}$$

where

$$E^{\pi_1}k(x_2) = \sum_{x_2} k(x_2)p(x_2 \mid x_1, \pi_1(x_1)) \quad \text{for} \quad k = k(x_2)$$
$$\mu_1(\pi_1) = \mu_1(\pi_1(x_1))$$

are functions of x_1, and

$$E^{\pi_0}l(x_1) = \sum_{x_1} l(x_1)p(x_1 \mid x_0, \pi_0(x_0)) \quad \text{for} \quad l = l(x_1)$$
$$\mu_0(\pi_0) = \mu_0(\pi_0(x_0))$$

are functions of x_0. Then, the usual dynamic programming technique yields the identity

$$\begin{aligned} &\mathrm{Max}_{\pi_0,\pi_1}[\mu_0(\pi_0) \wedge E^{\pi_0}[\mu_1(\pi_1) \wedge E^{\pi_1}\mu_{G^2}(x_2)]] \\ =&\mathrm{Max}_{\pi_0}[\mu_0(\pi_0) \wedge E^{\pi_0}\mathrm{Max}_{\pi_1}[\mu_1(\pi_1) \wedge E^{\pi_1}\mu_{G^2}(x_2)]]. \end{aligned} \tag{47}$$

This is equivalent to the recurrence equations:

$$\mu_{G^1}(x_1) = \mathrm{Max}_{u_1 \in \{\alpha_1,\alpha_2\}}[\mu_1(u_1) \wedge \sum_{x_2 \in \{\sigma_1,\sigma_2,\sigma_3\}} \mu_{G^2}(x_2)p(x_2 \mid x_1, u_1)] \tag{48}$$

$$\mu_{G^0}(x_0) = \mathrm{Max}_{u_0 \in \{\alpha_1,\alpha_2\}}[\mu_0(u_0) \wedge \sum_{x_1 \in \{\sigma_1,\sigma_2,\sigma_3\}} \mu_{G^1}(x_1)p(x_1 \mid x_0, u_0)]. \tag{49}$$

They give the following optimal solution through the backward equations:

$$\mu_{G^1}(\sigma_1) = 0.6, \quad \mu_{G^1}(\sigma_2) = 0.82, \quad \mu_{G^1}(\sigma_3) = 0.6 \tag{50}$$

$$\pi_1(\sigma_1) = \alpha_1, \quad \pi_1(\sigma_2) = \alpha_1, \quad \pi_1(\sigma_3) = \alpha_2, \tag{51}$$

$$\mu_{G^0}(\sigma_1) = 0.8, \quad \mu_{G^0}(\sigma_2) = 0.62, \quad \mu_{G^0}(\sigma_3) = 0.62 \tag{52}$$

$$\pi_0(\sigma_1) = \alpha_1, \quad \pi_0(\sigma_2) = \alpha_1 \text{ or } \alpha_2, \quad \pi_0(\sigma_3) = \alpha_1. \tag{53}$$

However, an exact expression of $\mu_{G^0}(x_0), \pi_0(x_0)$ becomes as follows:

$$\mu_{G^0}(\sigma_1) = 0.798, \quad \mu_{G^0}(\sigma_2) = 0.622, \quad \mu_{G^0}(\sigma_3) = 0.622 \tag{54}$$

$$\pi_0(\sigma_1) = \alpha_2, \quad \pi_0(\sigma_2) = \alpha_1 \text{ or } \alpha_2, \quad \pi_0(\sigma_3) = \alpha_1. \tag{55}$$

Now, let us compare optimal solution (43) of our stochastic problem (42) with optimal solution (54) of Bellman and Zadh's "deterministic" problem (46). Thus, we should remark that problems (42), (46) are not the same problems (see [4] for a detail):

$$\begin{aligned} &\mathrm{Max}_{\pi} E^{\pi}[\mu_0(u_0) \wedge \mu_1(u_1) \wedge \mu_{G^2}(x_2)] \\ &\neq \mathrm{Max}_{\pi_0}[\mu_0(\pi_0) \wedge E^{\pi_0} \mathrm{Max}_{\pi_1}[\mu_1(\pi_1) \wedge E^{\pi_1} \mu_{G^2}(x_2)]]. \end{aligned} \tag{56}$$

As is shown in the preceding section, the invariant imbedding technique with a parameter λ solves the former problem (42).

4.3 Another Recursive Equations

Finally, we have, as another candidate, the third recursive equations as follows:

$$\begin{aligned} \xi_{G^{N-\nu}}(x_{N-\nu}) \mathrm{Max}_{u_{N-\nu}} \sum_{x_{N-\nu+1}} &[\mu_{N-\nu}(u_{N-\nu}) \wedge \xi_{G^{N-\nu+1}}(x_{N-\nu+1})] \\ &\times p(x_{N-\nu+1} \mid x_{N-\nu}, u_{N-\nu}) \\ &\nu = 1, 2, \cdots, N \end{aligned} \tag{57}$$

$$\xi_{G^N}(x_N) = \mu_{G^N}(x_N). \tag{58}$$

Let $\pi^*_{N-\nu}(x_{N-\nu})$ be any value of $u_{N-\nu}$ which attains the maximum in Eq (57). However, there is no reason why we may call the sequence $\pi^* = \{\pi^*_0, \pi^*_1, \cdots, \pi^*_{N-1}\}$ an optimal policy for problems (57),(58). Then, for the preceding data, the corresponding recursive equations

$$\xi_{G^2}(x_2) = \mu_{G^2}(x_2) \tag{59}$$

$$\xi_{G^1}(x_1) = \mathrm{Max}_{u_1 \in \{\alpha_1, \alpha_2\}} \sum_{x_2 \in \{\sigma_1, \sigma_2, \sigma_3\}} [\mu_1(u_1) \wedge \xi_{G^2}(x_2)] p(x_2 \mid x_1, u_1) \tag{60}$$

$$\xi_{G^0}(x_0) = \mathrm{Max}_{u_0 \in \{\alpha_1, \alpha_2\}} \sum_{x_1 \in \{\sigma_1, \sigma_2, \sigma_3\}} [\mu_0(u_0) \wedge \xi_{G^1}(x_1)] p(x_1 \mid x_0, u_0) \tag{61}$$

yield in turn

$$\xi_{G^2}(\sigma_1) = 0.3, \quad \xi_{G^2}(\sigma_2) = 1.0, \quad \xi_{G^2}(\sigma_3) = 0.8, \tag{62}$$

$$\xi_{G^1}(\sigma_1) = 0.57, \quad \xi_{G^1}(\sigma_2) = 0.82, \quad \xi_{G_1}(\sigma_3) = 0.57 \tag{63}$$

$$\pi^*_1(\sigma_1) = \alpha_2, \quad \pi^*_1(\sigma_2) = \alpha_1, \quad \pi^*_1(\sigma_3) = \alpha_2, \tag{64}$$

$$\xi_{G^0}(\sigma_1) = 0.795, \quad \xi_{G^0}(\sigma_2) = 0.595, \quad \xi_{G^0}(\sigma_3) = 0.583 \tag{65}$$

$$\pi^*_0(\sigma_1) = \alpha_2, \quad \pi^*_0(\sigma_2) = \alpha_2, \quad \pi^*_0(\sigma_3) = \alpha_1. \tag{66}$$

Now, let us compare optimal solution (43) of our stochastic problem (42) with solution (62)-(66) of problems (59)-(61). Thus, it happens that the coincidences

$$\mu_{G^i}(x_i; 1) = \xi_{G^i}(x_i) \quad \text{for} \quad x_i = \sigma_1, \sigma_2, \sigma_3, \quad i = 0, 1$$

$$\tilde{\pi}_i(x_i; 1) = \pi_i^*(x_i) \quad \text{for} \quad x_i = \sigma_1, \sigma_2, \sigma_3, \quad i = 0, 1$$

hold. However, these two equalities do not remain in general true.

Acknowledgements

The authors wish to thank the anonymous referees for careful reading of this paper and helpful comments.

References

[1] Bellman, R. E.: *Dynamic Programming*, Princeton Univ. Press, NJ, 1957.
[2] Bellman, R. E. and Zadeh, L. A.: Decision-making in a fuzzy environment, *Management Science*, Vol. 17 (1970), B141–B164.
[3] Esogbue, A. O. and Bellman, R. E.: Fuzzy dynamic programming and its extensions, *TIMS/Studies in the Management Sciences*, Vol. 20 (1984), 147–167.
[4] Iwamoto, S.: Associative dynamic programs, under consideration.
[5] Kacprzyk, J.: Decision-making in a fuzzy environment with fuzzy termination time, *Fuzzy Sets and Systems*, Vol. 1 (1978), 169–179.

Seiichi IWAMOTO
Department of Economic Engineering
Faculty of Economics
Kyushu University 27
Fukuoka 812, Japan