Journal of the Operations Research
Society of Japan
Vol. 41, No. 3, September 1998

© 1998 The Operations Research Society of Japan

STOCHASTIC OPTIMIZATION OF MULTIPLICATIVE FUNCTIONS WITH NEGATIVE VALUE

Toshiharu Fujita
Kyushu Institute of Technology

Kazuyoshi Tsurusaki
Kyushu University

(Received February 6, 1997; Revised August 25, 1997)

Abstract In this paper we show three methods for solving optimization problems of expected value of multiplicative functions with negative values; multi-stage stochastic decision tree, Markov bidecision process and invariant imbedding approach.

1. Introduction

Since Bellman and Zadeh [3], a large amounts of efforts has been devoted to the study of stochastic optimization of minimum criterion in the field of "Decision-making in a fuzzy environment" (Esogbue and Bellman [5], Kacprzyk [11] and others). Recently Iwamoto and Fujita [9] have solved the optimal value function through invariant imbedding. Iwamoto, Tsurusaki and Fujita [10] give a detailed structure of optimal policy. Further, the regular dynamic programming is extended to a two-way programming under the name of bidecision process [7] or bynamic programming [6].

In this paper, we are concerned with stochastic maximization problems of *multiplicative* function with *negative* returns. We raise the question whether there exists an optimal policy for the stochastic maximum problem or not. Further, if it exists, we focus our attention on the question whether the optimal policy is Markov or not.

Stochastic optimization of multiplicative function has been studied under the restriction that return is nonnegative. In this paper we remove the nonnegativity. The multiplicative function with *negative* returns applies to a class of sequential decision processes in which the total reliability of an information system is accumulateled through the degree of stage-wise reliabilities taking both positive and negative values. The negativity means unreliabilty (or incredibility) and the positivity does reliability (or truth). We are concerned with two extreme behaviors of the system under uncertainty. One is a maximizing behavior. The other is a minimizing behavior. This leads to both maximum problem and minimum problem for such a multiplicative criterion function. We show three methods — bidecision process approach, invariant imbedding approach, and multi-stage stochastic decision tree approach — yield the common optimal solution. Section 2 discusses stochastic maximization of multiplicative function with nonnegative returns. The optimization problem with negative returns are discussed in Sections 3, 4 and 5. Section 3 solves it through bidecision process. Section 4 solves it through invariant imbedding. Section 5 solves an example through multi-stage stochastic decision tree approach.

Throughout the paper the following data is given :

$N \geq 2$ is an integer; the *total number of stages*

$X = \{s_1, s_2, \dots, s_p\}$ is a finite *state space*

$U = \{a_1, a_2, \dots, a_k\}$ is a finite *action space*

$r_n : X \times U \to R^1$ is an *n-th reward function* $(0 \le n \le N-1)$

$r_G : X \to R^1$ is a *terminal reward function*

p is a *Markov transition law*

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

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

2. Nonnegative Returns

In this section we consider the stochastic maximization of multiplicative function as follows :

$$\begin{array}{lll} \text{Maximize} & E[\, r_0(x_0,u_0)r_1(x_1,u_1)\cdots r_{N-1}(x_{N-1},u_{N-1})r_G(x_N)\,] & \\ \text{subject to} & \text{(i)} \ \ x_{n+1} \sim p(\cdot|x_n,u_n) & \\ & \text{(ii)} \ \ u_n \in U \qquad n = 0,1,\dots,N-1. & \end{array} \tag{2.1}$$

We treat the case for multiplicative process with nonnegative returns. Thus, we assume the nonnegativity of reward functions :

$$r_n(x,u) \ge 0 \qquad (x,u) \in X \times U, \ 0 \le n \le N-1. \tag{2.2}$$

2.1. General policies

In this subsection we consider the original problem (2.1) with the set of all general policies. We call this problem *general problem.* With any general policy $\sigma = \{\sigma_n, \dots, \sigma_{N-1}\}$ over the $(N-n)$-stage process starting on n-th stage and terminating at the last stage, we associate the expected value :

$$\begin{aligned} J^n(x_n;\sigma) \ = \ & \sum\sum_{(x_{n+1},\dots,x_N)\in X\times\cdots\times X}\cdots\sum\{[r_n(x_n,u_n)\cdots r_{N-1}(x_{N-1},u_{N-1})r_G(x_N)] \\ & \times p(x_{n+1}|x_n,u_n)\cdots p(x_N|x_{N-1},u_{N-1})\}. \end{aligned} \tag{2.3}$$

We define the family of the corresponding *general subproblems* as follows :

$$\begin{aligned} V^N(x_N) &= r_G(x_N) \qquad x_N \in X \\ V^n(x_n) &= \operatorname*{Max}_{\sigma} J^n(x_n;\sigma) \qquad x_n \in X, \ \ 0 \le n \le N-1. \end{aligned} \tag{2.4}$$

Then, we have the recursive formula for the general subproblems :

Theorem 2.1

$$\begin{aligned} V^N(x) &= r_G(x) \qquad x \in X \\ V^n(x) &= \operatorname*{Max}_{u \in U}\Big[r_n(x,u)\sum_{y \in X} V^{n+1}(y)p(y|x,u)\Big] \qquad x \in X, \ \ 0 \le n \le N-1. \end{aligned} \tag{2.5}$$

Copyright © by ORSJ. Unauthorized reproduction of this article is prohibited.

2.2. Markov policies

In this subsection we restrict the problem (2.1) to the set of all Markov policies. We call this problem *Markov problem.*

Any Markov policy $\pi = \{\pi_n, \ldots, \pi_{N-1}\}$ over the $(N-n)$-stage process is associated with its expected value $J^n(x_n;\pi)$ defined by (2.3). For the corresponding *Markov subproblems* :

$$\begin{aligned} v^N(x_N) &= r_G(x_N) \qquad x_N \in X \\ v^n(x_n) &= \operatorname*{Max}_{\pi} J^n(x_n;\pi) \qquad x_n \in X, \quad 0 \le n \le N-1, \end{aligned} \tag{2.6}$$

we have the recursive formula :

Theorem 2.2

$$\begin{aligned} v^N(x) &= r_G(x) \qquad x \in X \\ v^n(x) &= \operatorname*{Max}_{u \in U}\Big[r_n(x,u) \sum_{y \in X} v^{n+1}(y) p(y|x,u)\Big] \qquad x \in X, \quad 0 \le n \le N-1. \end{aligned} \tag{2.7}$$

Theorem 2.3 (i) *A Markov policy yields the optimal value function $V^0(\cdot)$ for the general problem. That is, there exists an optimal Markov policy π^* for the general problem* (2.1) :

$$J^0(x_0;\pi^*) = V^0(x_0) \quad \textit{for all} \ \ x_0 \in X.$$

In fact, letting $\pi_n^*(x)$ be a maximizer of (2.5) (*or* (2.7)) *for each $x \in X$, $0 \le n \le N-1$, we have the optimal Markov policy $\pi^* = \{\pi_0^*, \ldots, \pi_{N-1}^*\}$.*
(ii) *The optimal value functions for the Markov subproblems* (2.6) *are equal to the optimal value functions for the general problems* (2.4) :

$$v^n(x) = V^n(x) \qquad x \in X, \quad 0 \le n \le N.$$

3. Bidecision Processes

In this section we take away the nonnegativity assumption (2.2) for return functions. We rather assume that it takes at least a negative value :

$$r_n(x,u) < 0 \quad \text{for some} \ \ 0 \le n \le N-1, \ (x,u) \in X \times U. \tag{3.1}$$

Then, in general, neither recursive formula (2.5) nor (2.7) holds.

Nevertheless, we have the following positive result :

Theorem 3.1 *A general policy yields the optimal value function $V^0(\cdot)$ for the general problem. That is, there exists an optimal general policy σ^* for the general problem* (2.1) :

$$J^0(x_0;\sigma^*) = V^0(x_0) \quad \textit{for all} \ \ x_0 \in X.$$

The proofs of Theorem 3.1 and 3.3 are postponed to Subsection 3.3.

Theorem 3.2 *In general, Markov policy does not yield the optimal value function $V^0(\cdot)$ for the general problem. That is, there exists a stochastic decision process with multiplicative function such that for any Markov policy π*

$$V^0(x_0) > J^0(x_0;\pi) \quad \textit{for some} \ \ x_0 \in X.$$

Proof The proof will be completed by illustrating an example in §5. □

In the following we show two alternatives for the *negative* case, i.e., under assumption (3.1). One is a bidecision approach. The other is an invariant imbedding approach.

Copyright © by ORSJ. Unauthorized reproduction of this article is prohibited.

3.1. General policies

In this subsection we consider the problem (2.1) with the set of all general policies. We call this problem *general problem.* With any general policy $\sigma = \{\sigma_n, \ldots, \sigma_{N-1}\}$, we associate the corresponding expected value :

$$J^n(x_n;\sigma) = \sum_{(x_{n+1},\ldots,x_N)\in X\times\cdots\times X}\sum\cdots\sum\{[r_n(x_n,u_n)\cdots r_{N-1}(x_{N-1},u_{N-1})r_G(x_N)] \times p(x_{n+1}|x_n,u_n)\cdots p(x_N|x_{N-1},u_{N-1})\}.$$

We define both the *family of maximum subproblems* and the *family of minimum subproblems* as follows :

$$\begin{aligned} V^N(x_N) &= r_G(x_N) \qquad x_N\in X \\ V^n(x_n) &= \operatorname*{Max}_{\sigma} J^n(x_n;\sigma) \qquad x_n\in X,\quad 0\le n\le N-1 \end{aligned} \tag{3.2}$$

$$\begin{aligned} W^N(x_N) &= r_G(x_N) \qquad x_N\in X \\ W^n(x_n) &= \min_{\sigma} J^n(x_n;\sigma) \qquad x_n\in X,\quad 0\le n\le N-1. \end{aligned} \tag{3.3}$$

For each $n\,(0\le n\le N-1), x\in X$ we divide the control space U into two disjoint subsets :

$$U(n,x,-)=\{u\in U|r_n(x,u)<0\},\quad U(n,x,+)=\{u\in U|r_n(x,u)\ge 0\}. \tag{3.4}$$

Then, we have the ***bicursive formula*** (system of two recursive formulae) for the both subproblems :

Theorem 3.3 (Bicursive Formula [7, pp.685,*l*.13-22])

$$\begin{aligned} V^N(x) &= W^N(x)=r_G(x) \qquad x\in X \\ V^n(x) &= \operatorname*{Max}_{u\in U(n,x,-)}\Big[r_n(x,u)\sum_{y\in X}W^{n+1}(y)p(y|x,u)\Big] \\ &\qquad \vee \operatorname*{Max}_{u\in U(n,x,+)}\Big[r_n(x,u)\sum_{y\in X}V^{n+1}(y)p(y|x,u)\Big], \end{aligned} \tag{3.5}$$

$$\begin{aligned} W^n(x) &= \min_{u\in U(n,x,-)}\Big[r_n(x,u)\sum_{y\in X}V^{n+1}(y)p(y|x,u)\Big] \\ &\qquad \wedge \min_{u\in U(n,x,+)}\Big[r_n(x,u)\sum_{y\in X}W^{n+1}(y)p(y|x,u)\Big] \\ &\qquad x\in X,\quad 0\le n\le N-1. \end{aligned} \tag{3.6}$$

Let $\pi = \{\pi_0, \ldots, \pi_{N-1}\}$ be a Markov policy for maximum problem and $\sigma = \{\sigma_0, \ldots, \sigma_{N-1}\}$ be a Markov policy for minimum problem, respectively. Then, the ordered pair (π, σ) is called a *strategy* for *both maximum and minimum problem* (2.1).

Given any strategy (π, σ), we regenerate two policies, upper policy and lower policy, together with corresponding two stochastic processes. The *upper policy* $\mu = \{\mu_0, \ldots, \mu_{N-1}\}$, which governs the *upper process* $Y = \{Y_0, \ldots, Y_N\}$ on the state space $X = \{s_1, s_2, \ldots, s_p\}$ ([7, pp.683]) , is defined as follows :

$$\mu_0(x_0) := \pi_0(x_0) \tag{3.7}$$

Copyright © by ORSJ. Unauthorized reproduction of this article is prohibited.

$$\mu_1(x_0,x_1) := \begin{cases} \sigma_1(x_1) \\ \pi_1(x_1) \end{cases} \text{for } r_0(x_0,u_0) \begin{cases} <0 \\ \geq 0 \end{cases} \tag{3.8}$$

$$\mu_2(x_0,x_1,x_2) := \begin{cases} \pi_2(x_2) \\ \sigma_2(x_2) \\ \sigma_2(x_2) \\ \pi_2(x_2) \end{cases} \text{for } r_1(x_1,u_1) \begin{cases} <0 \\ <0 \\ \geq 0 \\ \geq 0 \end{cases} u_1 = \begin{cases} \sigma_1(x_1) \\ \pi_1(x_1) \\ \sigma_1(x_1) \\ \pi_1(x_1) \end{cases} \tag{3.9}$$

$$\vdots$$

$$\mu_n(x_0,\ldots,x_n) := \begin{cases} \pi_n(x_n) \\ \sigma_n(x_n) \\ \sigma_n(x_n) \\ \pi_n(x_n) \end{cases} \text{for } r_{n-1}(x_{n-1},u_{n-1}) \begin{cases} <0 \\ <0 \\ \geq 0 \\ \geq 0 \end{cases} u_{n-1} = \begin{cases} \sigma_{n-1}(x_{n-1}) \\ \pi_{n-1}(x_{n-1}) \\ \sigma_{n-1}(x_{n-1}) \\ \pi_{n-1}(x_{n-1}) \end{cases} \tag{3.10}$$

and so on, where

$$u_i = \mu_i(x_0,\ldots,x_i) \quad i = 0,1,\ldots,n-1.$$

On the other hand, the replacement of triplet $\{\mu,\ \sigma,\ \pi\}$ by $\{\nu,\ \pi,\ \sigma\}$ in the regeneration process above yields the *lower policy* $\nu = \{\nu_0,\ \ldots,\ \nu_{N-1}\}$, which in turn governs the *lower process* $Z = \{Z_0,\ldots,Z_N\}$ on the state space X ([7, pp.684]).

Now let us return to the problem of selecting an optimal policy for *maximum problem* (2.1) *with the set of all general policies.* We have obtained the bicursive formula (3.5),(3.6) for the general subproblems. Let for each $n(0 \leq n \leq N-1)$, $x \in X$ $\pi_n^*(x)$ and $\hat{\sigma}_n(x)$ be a maximizer for (3.5) and a minimizer for (3.6), respectively. Then, we have a pair of policies $\pi^* = \{\pi_0^*,\ \ldots,\ \pi_{N-1}^*\}$ and $\hat{\sigma} = \{\hat{\sigma}_0,\ \ldots,\ \hat{\sigma}_{N-1}\}$. Thus, the pair $(\pi^*,\ \hat{\sigma})$ is a strategy for problem (2.1). The preceding discussion for strategy $(\pi^*,\ \hat{\sigma})$ regenerates both upper policy $\mu^* = \{\mu_0^*,\ \ldots,\ \mu_{N-1}^*\}$ and lower policy $\hat{\nu} = \{\hat{\nu}_0,\ \ldots,\ \hat{\nu}_{N-1}\}$. From the construction (3.7)-(3.10) together with bicursive formula (3.5),(3.6), we see that *upper policy* $\mu^* = \{\mu_0^*,\ \ldots,\ \mu_{N-1}^*\}$ *is optimal policy for maximum problem* (2.1). Thus, the general policy μ^* yields the optimal value function $V^0(\cdot)$ in (3.2) for the general maximum problem.

Similarly, the lower policy $\hat{\nu} = \{\hat{\nu}_0,\ \ldots,\ \hat{\nu}_{N-1}\}$ is optimal for minimum problem (2.1). The general policy $\hat{\nu}$ yields the optimal value function $W^0(\cdot)$ in (3.3) for the general minimum problem.

3.2. Markov policies

Further, restricting the problem (2.1) to the *set of all Markov policies*, we have the *Markov problem.* However, the corresponding optimal value functions for Markov subproblems $\{v^n(\cdot), w^n(\cdot)\}$ do not satisfy the bicursive formula (3.5),(3.6). Further, the optimal value functions are not identical to the optimal value functions $\{V^n(\cdot), W^n(\cdot)\}$ in (3.2),(3.3), respectively. In general, we have inequalities :

$$V^n(x) \geq v^n(x), \qquad W^n(x) \leq w^n(x) \qquad x \in X \ \ 0 \leq n \leq N. \tag{3.11}$$

3.3. Proofs of Theorems 3.1 and 3.3

In this subsection we prove Theorems 3.1 and 3.3. It suffices to prove these two facts for the two-stage process, because those for the N-stage process are proved in a similar way.

We note that for $x_n \in X$

$$\begin{aligned} V^2(x_2) &= W^2(x_2) = r_G(x_2) \\ V^1(x_1) &= \operatorname*{Max}_{\sigma_1} \sum_{x_2 \in X} [\, r_1(x_1,u_1) r_G(x_2)\,] p(x_2|x_1,u_1) \end{aligned} \tag{3.12}$$

Copyright © by ORSJ. Unauthorized reproduction of this article is prohibited.

$$W^1(x_1) = \min_{\sigma_1} \sum_{x_2 \in X} [\, r_1(x_1,u_1) r_G(x_2)\,] p(x_2|x_1,u_1) \tag{3.13}$$

$$V^0(x_0) = \operatorname*{Max}_{\sigma_0,\sigma_1} \sum \sum_{(x_1,x_2)\in X\times X} \{[\, r_0(x_0,u_0) r_1(x_1,u_1) r_G(x_2)\,] \times p(x_1|x_0,u_0) p(x_2|x_1,u_1)\} \tag{3.14}$$

$$W^0(x_0) = \min_{\sigma_0,\sigma_1} \sum \sum_{(x_1,x_2)\in X\times X} \{[\, r_0(x_0,u_0) r_1(x_1,u_1) r_G(x_2)\,] \times p(x_1|x_0,u_0) p(x_2|x_1,u_1)\} \tag{3.15}$$

where $u_1 = \sigma_1(x_1)$ in (3.12),(3.13) and $u_0 = \sigma_0(x_0),\ u_1 = \sigma_1(x_0,x_1)$ in (3.14),(3.15), respectively.

Thus, the equalities

$$\begin{aligned} V^1(x_1) &= \operatorname*{Max}_{u_1\in U(1,x_1,-)} \left[r_1(x_1,u_1) \sum_{x_2\in X} W^2(x_2) p(x_2|x_1,u_1) \right] \\ &\qquad \vee \operatorname*{Max}_{u_1\in U(1,x_1,+)} \left[r_1(x_1,u_1) \sum_{x_2\in X} V^2(x_2) p(x_2|x_1,u_1) \right] \\ W^1(x_1) &= \min_{u_1\in U(1,x_1,-)} \left[r_1(x_1,u_1) \sum_{x_2\in X} V^2(x_2) p(x_2|x_1,u_1) \right] \\ &\qquad \wedge \min_{u_1\in U(1,x_1,+)} \left[r_1(x_1,u_1) \sum_{x_2\in X} W^2(x_2) p(x_2|x_1,u_1) \right] \\ &\qquad x_1 \in X \end{aligned}$$

are trivial. Therefore we must show the equalities

$$\begin{aligned} V^0(x_0) &= \operatorname*{Max}_{u_0\in U(0,x_0,-)} \left[r_0(x_0,u_0) \sum_{x_1\in X} W^1(x_1) p(x_1|x_0,u_0) \right] \\ &\qquad \vee \operatorname*{Max}_{u_0\in U(0,x_0,+)} \left[r_0(x_0,u_0) \sum_{x_1\in X} V^1(x_1) p(x_1|x_0,u_0) \right] \quad (3.16) \\ W^0(x_0) &= \min_{u_0\in U(0,x_0,-)} \left[r_0(x_0,u_0) \sum_{x_1\in X} V^1(x_1) p(x_1|x_0,u_0) \right] \\ &\qquad \wedge \min_{u_0\in U(0,x_0,+)} \left[r_0(x_0,u_0) \sum_{x_1\in X} W^1(x_1) p(x_1|x_0,u_0) \right] \quad (3.17) \\ &\qquad x_0 \in X. \end{aligned}$$

Since (3.17) is proved in a similar way, we prove (3.16) in the following.

Let us choose an optimal (Markov) policy π_1^* for the one-stage maximum process :

$$V^1(x_1) = \sum_{x_2\in X} \{[\, r_1(x_1,u_1) r_G(x_2)\,] p(x_2|x_1,u_1)\} \quad \forall x_1 \in X \tag{3.18}$$

where $u_1 = \pi_1^*(x_1)$ and choose an optimal (Markov) policy $\hat{\sigma}_1$ for the one-stage minimum process :

$$W^1(x_1) = \sum_{x_2\in X} \{[\, r_1(x_1,u_1) r_G(x_2)\,] p(x_2|x_1,u_1)\} \quad \forall x_1 \in X \tag{3.19}$$

where $u_1 = \hat{\sigma}_1(x_1)$. From the definition (3.14), we can for each $x_0 \in X$ choose an optimal (not necessarily Markov) policy $\tilde{\sigma} = \{\tilde{\sigma}_0, \tilde{\sigma}_1\}$ for the two-stage process :

$$V^0(x_0) = \sum \sum_{(x_1,x_2)\in X\times X} \{[r_0(x_0,u_0) r_1(x_1,u_1) r_G(x_2)] \times p(x_1|x_0,u_0) p(x_2|x_1,u_1)\} \tag{3.20}$$

Copyright © by ORSJ. Unauthorized reproduction of this article is prohibited.

where

$$u_0 = \tilde{\sigma}_0(x_0), \quad u_1 = \tilde{\sigma}_1(x_0, x_1).$$

We note that

$$\sum_{(x_1,x_2)\in X\times X} f(x_1,x_2) = \sum_{x_1\in X}\sum_{x_2\in X} f(x_1,x_2) \tag{3.21}$$

and

$$W^1(x_1) \le \sum_{x_2\in X} [r_1(x_1,u_1)r_G(x_2)]p(x_2|x_1,u_1) \le V^1(x_1) \quad \forall x_1 \in X. \tag{3.22}$$

From (3.20),(3.21) and (3.22) we have for $u_0 \in U$ satisfying $r_0(x_0,u_0) > 0$

$$\begin{aligned} V^0(x_0) &= \sum_{x_1\in X}\sum_{x_2\in X}\{[r_0(x_0,u_0)r_1(x_1,u_1)r_G(x_2)] \times p(x_1|x_0,u_0)p(x_2|x_1,u_1)\} \\ &= \sum_{x_1\in X} r_0(x_0,u_0)\Big\{\sum_{x_2\in X}[r_1(x_1,u_1)r_G(x_2)]p(x_2|x_1,u_1)\Big\}p(x_1|x_0,u_0) \\ &\le r_0(x_0,u_0)\sum_{x_1\in X} V^1(x_1)p(x_1|x_0,u_0). \end{aligned}$$

On the other hand, we have for $u_0 \in U$ satisfying $r_0(x_0,u_0) \le 0$

$$V^0(x_0) \le r_0(x_0,u_0)\sum_{x_1\in X} W^1(x_1)p(x_1|x_0,u_0).$$

Thus, taking maximum over $u_0 \in U(0,x_0,+)$ and once more over $u_0 \in U(0,x_0,-)$, we get

$$\begin{aligned} V^0(x_0) \le & \operatorname*{Max}_{u_0\in U(0,x_0,-)}\Big[r_0(x_0,u_0)\sum_{x_1\in X} W^1(x_1)p(x_1|x_0,u_0)\Big] \\ & \vee \operatorname*{Max}_{u_0\in U(0,x_0,+)}\Big[r_0(x_0,u_0)\sum_{x_1\in X} V^1(x_1)p(x_1|x_0,u_0)\Big] \quad \forall x_0 \in X. \end{aligned} \tag{3.23}$$

On the other hand, let for any $x_0 \in X$, $u^* = u^*(x_0) \in U$ be a maximizer of the right hand side of (3.23)(i.e., maximum of the two maxima). This defines a Markov decision function

$$\pi_0^* : X \to U \qquad \pi_0^*(x_0) = u^*(x_0).$$

First let us assume

$$r_0(x_0,u_0) > 0 \qquad u_0 = \pi_0^*(x_0).$$

Then, we have

$$\begin{aligned} & \operatorname*{Max}_{u_0\in U(0,x_0,-)}\Big[r_0(x_0,u_0)\sum_{x_1\in X} W^1(x_1)p(x_1|x_0,u_0)\Big] \\ & \qquad \vee \operatorname*{Max}_{u_0\in U(0,x_0,+)}\Big[r_0(x_0,u_0)\sum_{x_1\in X} V^1(x_1)p(x_1|x_0,u_0)\Big] \\ &= \operatorname*{Max}_{u_0\in U(0,x_0,+)}\Big[r_0(x_0,u_0)\sum_{x_1\in X} V^1(x_1)p(x_1|x_0,u_0)\Big] \\ &= r_0(x_0,u_0)\sum_{x_1\in X} V^1(x_1)p(x_1|x_0,u_0) \qquad (u_0 = \pi_0^*(x_0)). \end{aligned} \tag{3.24}$$

Copyright © by ORSJ. Unauthorized reproduction of this article is prohibited.

From (3.18) and (3.19), we get

$$V^1(x_1) = \sum_{x_2 \in X} \{[\, r_1(x_1,u_1) r_G(x_2)\,] p(x_2|x_1,u_1)\} \qquad u_1 = \pi_1^*(x_1) \tag{3.25}$$

and

$$W^1(x_1) = \sum_{x_2 \in X} \{[\, r_1(x_1,u_1) r_G(x_2)\,] p(x_2|x_1,u_1)\} \qquad u_1 = \hat{\sigma}_1(x_1), \tag{3.26}$$

respectively. Thus, we have from (3.24),(3.25)

$$\begin{aligned} & r_0(x_0,u_0) \sum_{x_1 \in X} V^1(x_1) p(x_1|x_0,u_0) \qquad (\, u_0 = \pi_0^*(x_0)\,) \\ = & \sum_{x_1 \in X} r_0(x_0,u_0) \Big\{ \sum_{x_2 \in X} \{[r_1(x_1,u_1) r_G(x_2)] p(x_2|x_1,u_1)\} p(x_1|x_0,u_0) \Big\} \\ = & \sum_{(x_1,x_2) \in X \times X} \{[r_0(x_0,u_0) r_1(x_1,u_1) r_G(x_2)] p(x_1|x_0,u_0) p(x_2|x_1,u_1)\}. \end{aligned} \tag{3.27}$$

Combining (3.24) and (3.27), we obtain

$$\begin{aligned} & \operatorname*{Max}_{u_0 \in U(0,x_0,+)} r_0(x_0,u_0) \sum_{x_1 \in X} V^1(x_1) p(x_1|x_0,u_0) \\ = & \sum_{(x_1,x_2) \in X \times X} \{[r_0(x_0,u_0) r_1(x_1,u_1) r_G(x_2)] p(x_1|x_0,u_0) p(x_2|x_1,u_1)\} \\ & \qquad (\, u_0 = \pi_0^*(x_0), \;\; u_1 = \pi_1^*(x_1)\,) \\ \le & \operatorname*{Max}_{\sigma_0,\sigma_1} \sum_{(x_1,x_2) \in X \times X} \{[\, r_0(x_0,u_0) r_1(x_1,u_1) r_G(x_2)\,] \times \; p(x_1|x_0,u_0) p(x_2|x_1,u_1)\,\} \\ = & V^0(x_0). \end{aligned} \tag{3.28}$$

Second let assume

$$r_0(x_0,u_0) \le 0 \qquad u_0 = \pi_0^*(x_0).$$

Similarly, for this case, we obtain through (3.26)

$$\operatorname*{Max}_{u_0 \in U(0,x_0,-)} r_0(x_0,u_0) \sum_{x_1 \in X} W^1(x_1) p(x_1|x_0,u_0) \le V^0(x_0). \tag{3.29}$$

From (3.28),(3.29), we have

$$\begin{aligned} & \operatorname*{Max}_{u_0 \in U(0,x_0,-)} \Big[r_0(x_0,u_0) \sum_{x_1 \in X} W^1(x_1) p(x_1|x_0,u_0) \Big] \\ & \qquad \vee \operatorname*{Max}_{u_0 \in U(0,x_0,+)} \Big[r_0(x_0,u_0) \sum_{x_1 \in X} V^1(x_1) p(x_1|x_0,u_0) \Big] \\ \le \; & V^0(x_0). \end{aligned} \tag{3.30}$$

Both equations (3.23) and (3.30) imply the desired equality (3.16). This completes the proof of Theorem 3.3.

Furthermore, from the Markov policy $\pi^* = \{\pi_0^*, \pi_1^*\}$ and the Markov decision function $\hat{\sigma}_1$ we construct a general policy $\mu^* = \{\mu_0^*, \mu_1^*\}$ through (3.7),(3.8). Then, the equality in (3.30) implies that the optimal value function $V^0(\cdot)$ is attained by this general policy μ^* :

$$V^0(x_0) = J^0(x_0; \mu^*\,) \qquad x_0 \in X.$$

Thus, Theorem 3.1 is proved. This completes the proofs.

Copyright © by ORSJ. Unauthorized reproduction of this article is prohibited.

4. Imbedded Processes

In this section we imbed the problem (2.1) into a family of ***terminal processes on one-dimensionally augmented state space.*** We note that the return, which may take negative values, is multiplicatively accumulating.

Let us return to the original stochastic maximization problem (2.1) with multiplicative function. Without loss of generality, we may assume that

$$\begin{aligned} -1 \le r_n(x,u) \le 1 &\quad (x,u) \in X \times U,\ 0 \le n \le N-1 \\ -1 \le r_G(x) \le 1 &\quad x \in X. \end{aligned} \tag{4.1}$$

Under the condition (4.1), we imbed the problem (2.1) into the family of parameterized problems as follows :

$$\begin{array}{ll} \text{Maximize} & E[\,\lambda_0 r_0(x_0,u_0) r_1(x_1,u_1) \cdots r_{N-1}(x_{N-1},u_{N-1}) r_G(x_N)\,] \\ \text{subject to} & \text{(i)}\ \ x_{n+1} \sim p(\cdot|x_n,u_n) \\ & \text{(ii)}\ \ u_n \in U \quad n = 0,1,\ldots,N-1 \end{array} \tag{4.2}$$

where the parameter ranges over $\lambda_0 \in [-1,1]$.

4.1. General policies

First we consider the imbedded problem (4.2) with the set of all general policies, called ***general problem.*** Here we note that any general policy :

$$\sigma = \{\sigma_0,\ \sigma_1,\ \ldots,\ \sigma_{N-1}\}$$

consists of the following decision functions

$$\begin{aligned} \sigma_0 &:\ X \times [-1,1] \to U \\ \sigma_1 &:\ (X \times [-1,1]) \times (X \times [-1,1]) \to U \\ &\cdots \\ \sigma_{N-1} &:\ (X \times [-1,1]) \times (X \times [-1,1]) \times \cdots \times (X \times [-1,1]) \to U. \end{aligned}$$

Thus, any general policy $\sigma = \{\sigma_n,\ \ldots,\ \sigma_{N-1}\}$ over the $(N-n)$-stage process yields its expected value :

$$\begin{aligned} K^n(x_n,\lambda_n\,;\sigma) \;=\; & \sum_{(x_{n+1},\ldots,x_N)\in X\times\cdots\times X}\!\!\sum\cdots\sum \{[\lambda_n r_n(x_n,u_n)\cdots r_{N-1}(x_{N-1},u_{N-1}) r_G(x_N)] \\ & \times p(x_{n+1}|x_n,u_n)\cdots p(x_N|x_{N-1},u_{N-1})\} \end{aligned} \tag{4.3}$$

where the alternating sequence of action and augmented state

$$\{u_n,\ (x_{n+1},\lambda_{n+1}),\ u_{n+1},\ (x_{n+2},\lambda_{n+2}),\ \ldots,\ u_{N-1},\ (x_N,\lambda_N)\,\}$$

is stochastically generated through the policy σ and the starting state (x_n,λ_n) as follows :

$$\begin{aligned} & \sigma_n(x_n,\lambda_n) = u_n \ \to \left\{ \begin{array}{l} p(\cdot|x_n,u_n) \sim x_{n+1} \\ \lambda_n r_n(x_n,u_n) = \lambda_{n+1} \end{array} \right. \\ & \to\ \sigma_{n+1}(x_n,\lambda_n,x_{n+1},\lambda_{n+1}) = u_{n+1} \ \to \left\{ \begin{array}{l} p(\cdot|x_{n+1},u_{n+1}) \sim x_{n+2} \\ \lambda_{n+1} r_{n+1}(x_{n+1},u_{n+1}) = \lambda_{n+2} \end{array} \right. \\ & \to\ \sigma_{n+2}(x_n,\lambda_n,x_{n+1},\lambda_{n+1},x_{n+2},\lambda_{n+2}) = u_{n+2} \end{aligned} \tag{4.4}$$

Copyright © by ORSJ. Unauthorized reproduction of this article is prohibited.

$$\to \left\{ \begin{array}{l} p(\cdot|x_{n+2}, u_{n+2}) \sim x_{n+3} \\ \lambda_{n+2} r_{n+2}(x_{n+2}, u_{n+2}) = \lambda_{n+3} \end{array} \right. \quad \to \quad \cdots$$
$$\to \sigma_{N-1}(x_n, \lambda_n, x_{n+1}, \lambda_{n+1}, \ldots, x_{N-1}, \lambda_{N-1}) = u_{N-1}$$
$$\to \left\{ \begin{array}{l} p(\cdot|x_{N-1}, u_{N-1}) \sim x_N \\ \lambda_{N-1} r_{N-1}(x_{N-1}, u_{N-1}) = \lambda_N. \end{array} \right.$$

However, note that the sequence of the latter halves of the states $\{\lambda_{n+1}, \lambda_{n+2}, \ldots, \lambda_N\}$ behaves deterministically.

We define the family of the corresponding *general subproblems* :

$$\begin{array}{ll} V^N(x_N, \lambda_N) = \lambda_N r_G(x_N) & x_N \in X, \quad -1 \leq \lambda_N \leq 1 \\ V^n(x_n, \lambda_n) = \operatorname*{Max}_{\sigma} K^n(x_n, \lambda_n\,;\sigma) & x_n \in X, \quad -1 \leq \lambda_n \leq 1, \quad 0 \leq n \leq N-1. \end{array} \tag{4.5}$$

Then, the general problem (4.2) is identical to (4.5) with $n = 0$. We have the recursive formula for the general subproblems :

Theorem 4.1

$$\begin{array}{rcll} V^N(x, \lambda) & = & \lambda r_G(x) & x \in X, \quad \lambda \in [-1, 1] \\ V^n(x, \lambda) & = & \operatorname*{Max}_{u \in U} \sum_{y \in X} V^{n+1}(y, \lambda r_n(x, u)) p(y|x, u) & \\ & & & x \in X, \quad \lambda \in [-1, 1], \quad 0 \leq n \leq N-1. \end{array} \tag{4.6}$$

4.2. Markov policies

Second we consider the *Markov problem.* That is, we restrict the imbedded problem (4.2) to the set of all Markov policies. Here Markov policy

$$\pi = \{\pi_0, \pi_1, \ldots, \pi_{N-1}\}$$

consists in turn of two-variable decision functions :

$$\pi_n : X \times [-1, 1] \to U \qquad 0 \leq n \leq N-1.$$

Note that any Markov policy $\pi = \{\pi_n, \ldots, \pi_{N-1}\}$ over the $(N-n)$-stage process yields its expected value $K^n(x_n, \lambda_n\,;\pi)$ through (4.3). The alternating sequence of action and augmented state

$$\{u_n, (x_{n+1}, \lambda_{n+1}), u_{n+1}, (x_{n+2}, \lambda_{n+2}), \ldots, u_{N-1}, (x_N, \lambda_N)\}$$

is similarly generated through the policy π and the state (x_n, λ_n) as in (4.4), where

$$\begin{array}{l} \pi_n(x_n, \lambda_n) = u_n \\ \pi_{n+1}(x_{n+1}, \lambda_{n+1}) = u_{n+1} \\ \cdots \\ \pi_{N-1}(x_{N-1}, \lambda_{N-1}) = u_{N-1}. \end{array}$$

Of course, the sequence of the latter halves of the states $\{\lambda_{n+1}, \lambda_{n+2}, \ldots, \lambda_N\}$ behaves deterministically.

We define the family of the corresponding *Markov subproblems* :

$$\begin{array}{ll} v^N(x_N, \lambda_N) = \lambda_N r_G(x_N) & x_N \in X, \quad -1 \leq \lambda_N \leq 1 \\ v^n(x_n, \lambda_n) = \operatorname*{Max}_{\pi} K^n(x_n, \lambda_n\,;\pi) & x_n \in X, \quad -1 \leq \lambda_n \leq 1, \quad 0 \leq n \leq N-1. \end{array} \tag{4.7}$$

Note that the Markov problem (4.2) is also (4.7) with $n = 0$. Then, we have the recursive formula for the Markov subproblems :

Copyright © by ORSJ. Unauthorized reproduction of this article is prohibited.

Theorem 4.2

$$\begin{aligned} v^N(x,\lambda) &= \lambda r_G(x) \qquad x \in X, \quad \lambda \in [-1,1] \\ v^n(x,\lambda) &= \operatorname*{Max}_{u \in U} \sum_{y \in X} v^{n+1}(y, \lambda r_n(x,u))p(y|x,u) \\ & \qquad\qquad x \in X, \quad \lambda \in [-1,1], \quad 0 \le n \le N-1. \end{aligned} \tag{4.8}$$

Theorem 4.3 (i) *A Markov policy yields the optimal value function* $V^0(\cdot)$ *for the general problem. That is, there exists an optimal Markov policy* π^* *for the general problem* (4.2) :

$$V^0(x_0,\lambda_0) = K^0(x_0,\lambda_0\,;\,\pi^*) \quad \textit{for all} \ (x_0,\lambda_0) \in X \times [-1,1].$$

In fact, letting $\pi_n^*(x,\lambda)$ *be a maximizer of* (4.8) (*or* (4.6)) *for each* $(x,\lambda) \in X \times [-1,1]$, $0 \le n \le N-1$, *we have the optimal Markov policy* $\pi^* = \{\pi_0^*, \ldots, \pi_{N-1}^*\}$.
(ii) *The optimal value functions for the Markov subproblems* (4.7) *are equal to the optimal value functions for the general problems* (4.5) :

$$v^n(x,\lambda) = V^n(x,\lambda) \qquad (x,\lambda) \in X \times [-1,1], \quad 0 \le n \le N.$$

4.3. Proofs of Theorems 4.1 - 4.3

In this subsection we prove only Theorems 4.1 and 4.3(i) because Theorems 4.2 and 4.3(ii) are the direct consequences of Theorems 4.1 and 4.3(i). We prove both theorems for the two-stage process, because the theorems for the N-stage process are proved similarly.

We note that for $(x_n,\lambda_n) \in X \times [-1,1]$

$$V^2(x_2,\lambda_2) = \lambda_2 r_G(x_2)$$

$$V^1(x_1,\lambda_1) = \operatorname*{Max}_{\sigma_1} \sum_{x_2 \in X} [\,\lambda_1 r_1(x_1,u_1) r_G(x_2)\,] p(x_2|x_1,u_1) \tag{4.9}$$

$$V^0(x_0,\lambda_0) = \operatorname*{Max}_{\sigma_0,\sigma_1} \sum \sum_{(x_1,x_2)\in X \times X} \{[\,\lambda_0 r_0(x_0,u_0) r_1(x_1,u_1) r_G(x_2)\,] \times p(x_1|x_0,u_0)p(x_2|x_1,u_1)\} \tag{4.10}$$

where $u_1 = \sigma_1(x_1,\lambda_1)$ in (4.9) and $u_0 = \sigma_0(x_0,\lambda_0)$, $\lambda_1 = \lambda_0 r_0(x_0,u_0)$, $u_1 = \sigma_1(x_0,\lambda_0,x_1,\lambda_1)$ in (4.10), respectively.

Thus, the equality

$$V^1(x_1,\lambda_1) = \operatorname*{Max}_{u_1 \in U} \sum_{x_2 \in X} V^2(x_2,\, \lambda_1 r_1(x_1,u_1))p(x_2|x_1,u_1) \quad x_1 \in X, \ \ \lambda_1 \in [-1,1]$$

is trivial. We prove

$$V^0(x_0,\lambda_0) = \operatorname*{Max}_{u_0 \in U} \sum_{x_1 \in X} V^1(x_1,\, \lambda_0 r_0(x_0,u_0))p(x_1|x_0,u_0) \quad x_0 \in X, \ \ \lambda_0 \in [-1,1]. \tag{4.11}$$

Let us choose an optimal (Markov) policy σ_1^* for the one-stage process :

$$V^1(x_1,\lambda_1) = K^1(x_1,\lambda_1;\sigma_1^*) \quad \forall (x_1,\lambda_1) \in X \times [-1,1]. \tag{4.12}$$

From the definition (4.5), we can for each $(x_0,\lambda_0) \in X \times [-1,1]$ choose an optimal (not necessarily Markov) policy $\tilde{\sigma} = \{\tilde{\sigma}_0, \tilde{\sigma}_1\}$ for the two-stage process :

$$V^0(x_0,\lambda_0) = K^0(x_0,\lambda_0;\tilde{\sigma}) \quad (x_0,\lambda_0) \in X \times [-1,1].$$

Copyright © by ORSJ. Unauthorized reproduction of this article is prohibited.

Thus, we see that

$$V^0(x_0,\lambda_0) = \sum_{(x_1,x_2)\in X\times X} \{[\lambda_0 r_0(x_0,u_0) r_1(x_1,u_1) r_G(x_2)] \times p(x_1|x_0,u_0)p(x_2|x_1,u_1)\} \quad (4.13)$$

where

$$u_0 = \tilde{\sigma}_0(x_0,\lambda_0), \quad \lambda_1 = \lambda_0 r_0(x_0,u_0), \quad u_1 = \tilde{\sigma}_1(x_0,\lambda_0,x_1,\lambda_1). \quad (4.14)$$

We note that

$$\begin{aligned}&\sum_{x_2\in X}[\lambda_1 r_1(x_1,u_1) r_G(x_2)] p(x_2|x_1,u_1)\\ \le\ & K^1(x_1,\lambda_1;\sigma_1^*) = V^1(x_1,\lambda_1) \qquad \forall(x_1,\lambda_1)\in X\times[-1,1].\end{aligned} \quad (4.15)$$

From (4.13) together with (4.14) and (4.15) we have

$$\begin{aligned} V^0(x_0,\lambda_0) &= \sum_{x_1\in X}\sum_{x_2\in X}\{[\lambda_0 r_0(x_0,u_0) r_1(x_1,u_1) r_G(x_2))] \times p(x_1|x_0,u_0)p(x_2|x_1,u_1)\} \\ &= \sum_{x_1\in X}\left\{\sum_{x_2\in X}\{[\lambda_1 r_1(x_1,u_1) r_G(x_2)]p(x_2|x_1,u_1)\}p(x_1|x_0,u_0)\right\} \\ &\qquad\qquad\qquad\qquad (\,\lambda_1 = \lambda_0 r_0(x_0,u_0)\,) \\ &\le \sum_{x_1\in X} V^1(x_1,\lambda_1)p(x_1|x_0,u_0) \qquad (\,\lambda_1 = \lambda_0 r_0(x_0,u_0)\,) \\ &= \sum_{x_1\in X} V^1(x_1,\ \lambda_0 r_0(x_0,u_0))p(x_1|x_0,u_0). \end{aligned}$$

Consequently, we have

$$V^0(x_0,\lambda_0) \le \sum_{x_1\in X} V^1(x_1,\ \lambda_0 r_0(x_0,u_0))p(x_1|x_0,u_0) \qquad \forall(x_0,\lambda_0)\in X\times[-1,1].$$

Thus, taking maximum over $u\in U$, we get

$$V^0(x_0,\lambda_0) \le \operatorname*{Max}_{u_0\in U}\sum_{x_1\in X} V^1(x_1,\ \lambda_0 r_0(x_0,u_0))p(x_1|x_0,u_0) \qquad \forall(x_0,\lambda_0)\in X\times[-1,1]. \quad (4.16)$$

On the other hand, let for any $(x_0,\lambda_0)\in X\times[-1,1]$, $u^* = u^*(x_0,\lambda_0)\in U$ be a maximizer of the right hand side of (4.16). This defines a Markov decision function

$$\pi_0^* : X\times[-1,1]\to U \qquad \pi_0^*(x_0,\lambda_0) = u^*(x_0,\lambda_0).$$

Then, we have

$$\begin{aligned}&\operatorname*{Max}_{u_0\in U}\sum_{x_1\in X} V^1(x_1,\ \lambda_0 r_0(x_0,u_0))p(x_1|x_0,u_0)\\ =\ & \sum_{x_1\in X} V^1(x_1,\ \lambda_0 r_0(x_0,u_0))p(x_1|x_0,u_0) \qquad (\,u_0 = \pi_0^*(x_0,\lambda_0)\,).\end{aligned} \quad (4.17)$$

From (4.12), we get

$$V^1(x_1,\lambda_1) = \sum_{x_2\in X}[\lambda_1 r_1(x_1,u_1) r_G(x_2)]p(x_2|x_1,u_1) \quad (\,u_1 = \sigma_1^*(x_1,\lambda_1)\,). \quad (4.18)$$

Copyright © by ORSJ. Unauthorized reproduction of this article is prohibited.

Thus, we have from (4.18)

$$
\begin{aligned}
&\sum_{x_1\in X} V^1(x_1,\ \lambda_0 r_0(x_0,u_0))p(x_1|x_0,u_0) \qquad (\,u_0=\pi_0^*(x_0,\lambda_0)\,)\\
=\ &\sum_{x_1\in X}\Big\{\sum_{x_2\in X}[\lambda_1 r_1(x_1,u_1)r_G(x_2)]p(x_2|x_1,u_1)\Big\}p(x_1|x_0,u_0) \quad (\text{for } \lambda_1=\lambda_0 r_0(x_0,u_0)\,)\\
=\ &\sum_{(x_1,x_2)\in X\times X}\sum \{[\lambda_0 r_0(x_0,u_0)r_1(x_1,u_1)r_G(x_2)]p(x_1|x_0,u_0)p(x_2|x_1,u_1)\}. \qquad (4.19)
\end{aligned}
$$

Combining (4.17) and (4.19), we obtain

$$
\begin{aligned}
&\operatorname*{Max}_{u_0\in U}\sum_{x_1\in X} V^1(x_1,\ \lambda_0 r_0(x_0,u_0))p(x_1|x_0,u_0)\\
=\ &\sum_{(x_1,x_2)\in X\times X}\sum \{[\lambda_0 r_0(x_0,u_0)r_1(x_1,u_1)r_G(x_2)]p(x_1|x_0,u_0)p(x_2|x_1,u_1)\}\\
&\qquad (\,u_0=\pi_0^*(x_0,\lambda_0),\ \ \lambda_2=\lambda_0 r_0(x_0,u_0),\ \ u_1=\sigma_1^*(x_1,\lambda_1)\,)\\
\le\ &\operatorname*{Max}_{\pi_0,\pi_1}\sum_{(x_1,x_2)\in X\times X}\sum \{[\,\lambda_0 r_0(x_0,u_0)r_1(x_1,u_1)r_G(x_2)\,]\times\ p(x_1|x_0,u_0)p(x_2|x_1,u_1)\,\}\\
=\ &V^0(x_0,\lambda_0). \qquad (4.20)
\end{aligned}
$$

Both equations (4.16) and (4.20) imply the desired equality (4.11). This completes the proof of Theorem 4.1.

Furthermore, the equalities in (4.20) imply that the optimal value function $V^0(\cdot)$ is attained by the Markov policy $\bar{\pi}=\{\pi_0^*,\sigma_1^*\}$:

$$V^0(x_0,\lambda_0)=K^0(x_0,\lambda_0;\ \bar{\pi}\,) \qquad (x_0,\lambda_0)\in X\times[-1,1].$$

This completes the proof of Theorem 4.3(i).

4.4. Proof of Theorem 3.1

Now, in this subsection, let us prove Theorem 3.1 by use of the result of Theorem 4.3.

First we note that any Markov policy for the imbedded problem (4.2) $\pi=\{\pi_0,\ \ldots,\ \pi_{N-1}\}$ together with a specified value of the parameter λ_0 induces the general policy for the problem (2.1) $\sigma=\{\sigma_0,\ \ldots,\ \sigma_{N-1}\}$ as follows :

$$
\begin{aligned}
&\sigma_0(x_0):=\pi_0(x_0,\lambda_0)\\
&\sigma_1(x_0,x_1):=\pi_1(x_1,\lambda_1)\\
&\qquad \text{where } \lambda_1=\lambda_0 r_0(x_0,u_0),\quad u_0=\pi_0(x_0,\lambda_0)\\
&\sigma_2(x_0,x_1,x_2):=\pi_2(x_2,\lambda_2)\\
&\qquad \text{where } \lambda_2=\lambda_1 r_1(x_1,u_1),\quad u_1=\pi_1(x_1,\lambda_1),\quad \lambda_1=\lambda_0 r_0(x_0,u_0),\\
&\qquad\qquad u_0=\pi_0(x_0,\lambda_0)\\
&\cdots \qquad (4.21)\\
&\sigma_{N-1}(x_0,x_1,\ldots,x_{N-1}):=\pi_{N-1}(x_{N-1},\lambda_{N-1})\\
&\qquad \text{where } \lambda_{N-1}=\lambda_{N-2}r_{N-2}(x_{N-2},u_{N-2}),\quad u_{N-2}=\pi_{N-2}(x_{N-2},\lambda_{N-2}),\\
&\qquad\qquad \lambda_{N-2}=\lambda_{N-3}r_{N-3}(x_{N-3},u_{N-3}),\quad u_{N-3}=\pi_{N-3}(x_{N-3},\lambda_{N-3}),\\
&\qquad\qquad ,\cdots,\quad \lambda_1=\lambda_0 r_0(x_0,u_0),\quad u_0=\pi_0(x_0,\lambda_0).
\end{aligned}
$$

Furthermore we see that both the Markov policy π with a specified value $\lambda_0=1$ and the resulting general policy σ yield the same value function :

$$K^0(x_0,\ 1\,;\,\pi)=J^0(x_0\,;\,\sigma) \qquad x_0\in X.$$

Copyright © by ORSJ. Unauthorized reproduction of this article is prohibited.

Second we note that Theorem 4.3 assures the existence of an optimal Markov policy for the imbedded problem (4.2) π^* which together with the value $\lambda_0 = 1$ induces the corresponding general policy for the problem (2.1) σ^*, as is shown by (4.21). Thus, we get

$$K^0(x_0,\, 1\,;\, \pi^*) = J^0(x_0\,;\, \sigma^*) \quad x_0 \in X. \tag{4.22}$$

On the other hand, for any general policy for the problem (2.1) $\sigma = \{\sigma_0, \ldots, \sigma_{N-1}\}$ we define a general policy for the imbedded problem (4.2) $\tilde{\sigma} = \{\tilde{\sigma}_0, \ldots, \tilde{\sigma}_{N-1}\}$ by

$$\tilde{\sigma}_n(x_0, \lambda_0, x_1, \lambda_1, \ldots, x_n, \lambda_n) := \sigma_n(x_0, x_1, \ldots, x_n)$$
$$\text{on } (X \times [-1,1]) \times (X \times [-1,1]) \times \cdots \times (X \times [-1,1]), \quad 0 \le n \le N-1.$$

Then, we have

$$K^0(x_0,\, 1\,;\, \tilde{\sigma}) = J^0(x_0\,;\, \sigma) \quad x_0 \in X. \tag{4.23}$$

Therefore, the optimality of the policy π^* implies

$$K^0(x_0,\, 1\,;\, \pi^*) \ge K^0(x_0,\, 1\,;\, \tilde{\sigma}) \quad x_0 \in X. \tag{4.24}$$

Combining (4.22),(4.24) and (4.23), we get for any general policy σ

$$J^0(x_0\,;\, \sigma^*) = K^0(x_0,\, 1\,;\, \pi^*) \ge K^0(x_0,\, 1\,;\, \tilde{\sigma}) = J^0(x_0\,;\, \sigma).$$

Thus, the policy σ^* is optimal for the general problem (2.1). This completes the proof of Theorem 3.1.

5. Example

In this section we illustrate a multiplicative decision process with negative returns which does not admit any optimal *Markov* policy. As was mentioned in §3, the illustration also proves Theorem 3.2. We show that bidecision process approach, invariant imbedding approach, and multi-stage stochastic decision tree approach yield a common pair of optimal value functions and optimal policy.

Let us consider the two-stage, three-state and two-action problem as follows :

$$\begin{array}{ll} \text{Maximize} & E[\, r_0(u_0) r_1(u_1) r_G(x_2)\,] \\ \text{subject to} & \text{(i)} \;\; x_{n+1} \sim p(\,\cdot\,|x_n, u_n) \qquad n = 0, 1 \\ & \text{(ii)} \;\; u_0 \in U, \;\; u_1 \in U \end{array} \tag{5.1}$$

where the data is given as follows (see also [3, pp.152, *l*.19-22, pp.B154],[9]) :

$$r_G(s_1) = 0.3 \qquad r_G(s_2) = 1.0 \qquad r_G(s_3) = -0.8$$
$$r_1(a_1) = -1.0 \quad r_1(a_2) = 0.6\,; \qquad r_0(a_1) = -0.7 \quad r_0(a_2) = 1.0$$

$u_t = a_1$

$x_t \diagdown x_{t+1}$	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. Bidecision processes

We note that Theorem 3.3 for Markov problem reduces to

$$\begin{aligned} V^2(x) &= W^2(x) = r_G(x) \qquad x \in X \\ V^1(x) &= \operatorname*{Max}_{u \in U(1,x,-)} \Big[r_1(u) \sum_{y \in X} W^2(y) p(y|x,u) \Big] \vee \operatorname*{Max}_{u \in U(1,x,+)} \Big[r_1(u) \sum_{y \in X} V^2(y) p(y|x,u) \Big] \end{aligned}$$

Copyright © by ORSJ. Unauthorized reproduction of this article is prohibited.

$$
\begin{aligned}
W^1(x) &= \min_{u\in U(1,x,-)}\Big[r_1(u)\sum_{y\in X}V^2(y)p(y|x,u)\Big]\wedge\min_{u\in U(1,x,+)}\Big[r_1(u)\sum_{y\in X}W^2(y)p(y|x,u)\Big]\\
V^0(x) &= \operatorname*{Max}_{u\in U(0,x,-)}\Big[r_0(u)\sum_{y\in X}W^1(y)p(y|x,u)\Big]\vee\operatorname*{Max}_{u\in U(0,x,+)}\Big[r_0(u)\sum_{y\in X}V^1(y)p(y|x,u)\Big]\\
W^0(x) &= \min_{u\in U(0,x,-)}\Big[r_0(u)\sum_{y\in X}V^1(y)p(y|x,u)\Big]\wedge\min_{u\in U(0,x,+)}\Big[r_0(u)\sum_{y\in X}W^1(y)p(y|x,u)\Big]
\end{aligned}
$$

The computation proceeds as follows. First

$$
\begin{aligned}
&V^2(s_1)=0.3 \quad V^2(s_2)=1.0 \quad V^2(s_3)=-0.8\\
&W^2(s_1)=0.3 \quad W^2(s_2)=1.0 \quad W^2(s_3)=-0.8.
\end{aligned}
$$

Second we have

$$
\begin{aligned}
V^1(s_1) &= [(-1.0)\times\{0.3\times 0.8+1.0\times 0.1+(-0.8)\times 0.1\}]\\
&\qquad \vee[0.6\times\{0.3\times 0.1+1.0\times 0.9+(-0.8)\times 0.0\}]\\
&= (-0.26)\vee 0.558=0.558 \qquad \pi_1^*(s_1)=a_2.
\end{aligned}
$$

Similarly, we have

$$
\begin{aligned}
& & & W^1(s_1)=-0.26 \quad \hat\sigma_1(s_1)=a_1\\
&V^1(s_2)=0.62 \quad &&\pi_1^*(s_2)=a_1 \qquad W^1(s_2)=0.156 \quad \hat\sigma_1(s_2)=a_2\\
&V^1(s_3)=-0.26 \quad &&\pi_1^*(s_3)=a_1 \qquad W^1(s_3)=-0.414 \quad \hat\sigma_1(s_3)=a_2.
\end{aligned}
$$

Third we have

$$
\begin{aligned}
&V^0(s_1)=0.6138 \quad &&\pi_0^*(s_1)=a_2 \qquad W^0(s_1)=-0.33768 \quad \hat\sigma_0(s_1)=a_1\\
&V^0(s_2)=0.4824 \quad &&\pi_0^*(s_2)=a_2 \qquad W^0(s_2)=-0.2338 \quad \hat\sigma_0(s_2)=a_2\\
&V^0(s_3)=0.16366 \quad &&\pi_0^*(s_3)=a_1 \qquad W^0(s_3)=-0.3986 \quad \hat\sigma_0(s_3)=a_2.
\end{aligned}
$$

Thus, we have obtained the Markov strategy $(\pi^*,\ \hat\sigma)$ as follows :

$$\pi^*=\{\pi_0^*,\pi_1^*\} \quad \hat\sigma=\{\hat\sigma_0,\hat\sigma_1\}$$

where

$$
\begin{aligned}
&\pi_0^*(s_1)=a_2, \quad \pi_0^*(s_2)=a_2, \quad \pi_0^*(s_3)=a_1\\
&\hat\sigma_0(s_1)=a_1, \quad \hat\sigma_0(s_2)=a_2, \quad \hat\sigma_0(s_3)=a_2\\
&\pi_1^*(s_1)=a_2, \quad \pi_1^*(s_2)=a_1, \quad \pi_1^*(s_3)=a_1\\
&\hat\sigma_1(s_1)=a_1, \quad \hat\sigma_1(s_2)=a_2, \quad \hat\sigma_1(s_3)=a_2.
\end{aligned}
$$

Now let us construct an optimal policy $\mu^*=\{\mu_0^*,\mu_1^*\}$ for ***maximum problem*** from the Markov strategy $(\pi^*,\ \hat\sigma)$. The ***upper policy*** $\mu^*=\{\mu_0^*,\mu_1^*\}$ is defined as follows :

$$\mu_0^*(x_0):=\pi_0^*(x_0)$$

$$\mu_1^*(x_0,x_1):=\begin{cases}\hat\sigma_1(x_1)\\ \pi_1^*(x_1)\end{cases} \text{for } r_0(u_0)\begin{cases}<0\\ \ge 0\end{cases} \text{where } u_0=\pi_0^*(x_0).$$

First we have

$$\mu_0^*(s_1)=a_2, \quad \mu_0^*(s_2)=a_2, \quad \mu_0^*(s_3)=a_1.$$

Second we have the following components of $\mu_1^*(x_0,x_1)$.
Since $r_0(\pi_0^*(s_1))=r_0(a_2)=1.0>0$, we have

Copyright © by ORSJ. Unauthorized reproduction of this article is prohibited.

$$\mu_1^*(s_1, s_1) = \pi_1^*(s_1) = a_2 \qquad \mu_1^*(s_1, s_2) = \pi_1^*(s_2) = a_1 \qquad \mu_1^*(s_1, s_3) = \pi_1^*(s_3) = a_1.$$

Similarly $r_0(\pi_0^*(s_2)) = r_0(a_2) = 1.0 > 0$ yields

$$\mu_1^*(s_2, s_1) = \pi_1^*(s_1) = a_2 \qquad \mu_1^*(s_2, s_2) = \pi_1^*(s_2) = a_1 \qquad \mu_1^*(s_2, s_3) = \pi_1^*(s_3) = a_1.$$

Further $r_0(\pi_0^*(s_3)) = r_0(a_1) = -0.7 < 0$ does

$$\mu_1^*(s_3, s_1) = \hat{\sigma}_1(s_1) = a_1 \qquad \mu_1^*(s_3, s_2) = \hat{\sigma}_1(s_2) = a_2 \qquad \mu_1^*(s_3, s_3) = \hat{\sigma}_1(s_3) = a_2.$$

5.2. Imbedded processes

In this subsection we solve the following parametric recursive formula :

$$\begin{aligned}
v^2(x_2; \lambda_2) &= \lambda_2 \times r_G(x_2) \quad x_2 \in X, \ \lambda_2 \in [-1, 1] \\
v^1(x_1; \lambda_1) &= \operatorname*{Max}_{u_1 \in U} \sum_{x_2 \in X} v^2(x_2;\ \lambda_1 \times r_1(x_1, u_1))p(x_2|x_1, u_1) \ \ x_1 \in X, \ \lambda_1 \in [-1, 1] \\
v^0(x_0; \lambda_0) &= \operatorname*{Max}_{u_0 \in U} \sum_{x_1 \in X} v^1(x_1;\ \lambda_0 \times r_0(x_0, u_0))p(x_1|x_0, u_0) \ \ x_0 \in X, \ \lambda_0 \in [-1, 1].
\end{aligned}$$

The computation proceeds as follows :

$$v^2(s_1; \lambda_2) = \lambda_2 \times 0.3 \qquad v^2(s_2) = \lambda_2 \times 1.0 \qquad v^2(s_3) = \lambda_2 \times (-0.8).$$

$$\begin{aligned}
v^1(s_1; \lambda_1) &= \sum_{x_2 \in X} v^2(x_2; \lambda_1 \times r_1(a_1))p(x_2 \,|\, s_1, a_1) \vee \sum_{x_2 \in X} v^2(x_2; \lambda_1 \times r_1(a_2))p(x_2 \,|\, s_1, a_2) \\
&= [\lambda_1 \times (-1.0) \times 0.3 \times 0.8 + \lambda_1 \times (-1.0) \times 1.0 \times 0.1 \\
&\qquad + \lambda_1 \times (-1.0) \times (-0.8) \times 0.1] \vee [\lambda_1 \times 0.6 \times 0.3 \times 0.1 \\
&\qquad\quad + \lambda_1 \times 0.6 \times 1.0 \times 0.9 + \lambda_1 \times 0.6 \times (-0.8) \times 0.0] \\
&= [\lambda_1 \times (-0.26)] \vee [\lambda_1 \times 0.558] \\
&= \begin{cases} \lambda_1 \times (-0.26) \\ \lambda_1 \times 0.558 \end{cases}, \quad \pi_1^*(s_1; \lambda_1) = \begin{cases} a_1 \\ a_2 \end{cases} \text{for} \begin{cases} -1 \le \lambda_1 \le 0 \\ 0 \le \lambda_1 \le 1 \end{cases}
\end{aligned}$$

$$v^1(s_2; \lambda_1) = \begin{cases} \lambda_1 \times 0.156 \\ \lambda_1 \times 0.62 \end{cases}, \quad \pi_1^*(s_2; \lambda_1) = \begin{cases} a_2 \\ a_1 \end{cases} \text{for} \begin{cases} -1 \le \lambda_1 \le 0 \\ 0 \le \lambda_1 \le 1 \end{cases}$$

$$v^1(s_3; \lambda_1) = \begin{cases} \lambda_1 \times (-0.414) \\ \lambda_1 \times (-0.26) \end{cases}, \quad \pi_1^*(s_3; \lambda_1) = \begin{cases} a_2 \\ a_1 \end{cases} \text{for} \begin{cases} -1 \le \lambda_1 \le 0 \\ 0 \le \lambda_1 \le 1. \end{cases}$$

Thus, we have optimal value function v^1 and optimal second decision function π_1^*:

	$-1 \le \lambda_1 \le 0$		$0 \le \lambda_1 \le 1$	
$v^1(s_1; \lambda_1)$, $\pi_1^*(s_1; \lambda_1)$	$\lambda_1 \times (-0.26)$,	a_1	$\lambda_1 \times 0.558$,	a_2
$v^1(s_2; \lambda_1)$, $\pi_1^*(s_2; \lambda_1)$	$\lambda_1 \times 0.156$,	a_2	$\lambda_1 \times 0.62$,	a_1
$v^1(s_3; \lambda_1)$, $\pi_1^*(s_3; \lambda_1)$	$\lambda_1 \times (-0.414)$,	a_2	$\lambda_1 \times (-0.26)$,	a_1

$$\begin{aligned}
v^0(s_1; \lambda_0) &= \sum_{x_1 \in X} v^1(x_1; \lambda_0 \times r_0(a_1))p(x_1 \,|\, s_1, a_1) \vee \sum_{x_1 \in X} v^1(x_1; \lambda_0 \times r_0(a_2))p(x_1 \,|\, s_1, a_2) \\
&= [v^1(s_1; \lambda_0 \times (-0.7)) \times 0.8 + v^1(s_2; \lambda_0 \times (-0.7)) \times 0.1 \\
&\qquad + v^1(s_3; \lambda_0 \times (-0.7)) \times 0.1] \vee [v^1(s_1; \lambda_0 \times 1.0) \times 0.1 \\
&\qquad\quad + v^1(s_2; \lambda_0 \times 1.0) \times 0.9 + v^1(s_3; \lambda_0 \times 1.0) \times 0.0].
\end{aligned}$$

Copyright © by ORSJ. Unauthorized reproduction of this article is prohibited.

For $-1 \leq \lambda_0 \leq 0$, we have

$$\begin{aligned} v^0(s_1;\lambda_0) &= [\lambda_0 \times (-0.7) \times 0.558 \times 0.8 + \lambda_0 \times (-0.7) \times 0.62 \times 0.1 \\ &\quad +\lambda_0 \times (-0.7) \times (-0.26) \times 0.1] \vee [\lambda_0 \times 1.0 \times (-0.26) \times 0.1 \\ &\quad +\lambda_0 \times 1.0 \times 0.156 \times 0.9 + \lambda_0 \times 1.0 \times (-0.414) \times 0.0] \\ &= [\lambda_0 \times (-0.33768)] \vee [\lambda_0 \times 0.144] = \lambda_0 \times (-0.33768), \qquad \pi_0^*(s_1;\lambda_0) = a_1, \end{aligned}$$

and for $0 \leq \lambda_0 \leq 1$, we have

$$\begin{aligned} v^0(s_1;\lambda_0) &= [\lambda_0 \times (-0.7) \times (-0.26) \times 0.8 + \lambda_0 \times (-0.7) \times 0.156 \times 0.1 \\ &\quad +\lambda_0 \times (-0.7) \times (-0.414) \times 0.1] \vee [\lambda_0 \times 1.0 \times 0.558 \times 0.1 \\ &\quad +\lambda_0 \times 1.0 \times 0.62 \times 0.9 + \lambda_0 \times 1.0 \times (-0.26) \times 0.0] \\ &= [\lambda_0 \times 0.16366] \vee [\lambda_0 \times 0.6138] = \lambda_0 \times 0.6138, \qquad \pi_0^*(s_1;\lambda_0) = a_2. \end{aligned}$$

Similar computation yields

$$v^0(s_2;\lambda_0) = \begin{cases} \lambda_0 \times (-0.2338) \\ \lambda_0 \times 0.4824 \end{cases} \text{for} \begin{cases} -1 \leq \lambda_0 \leq 0 \\ 0 \leq \lambda_0 \leq 1 \end{cases} \pi^*(s_2;\lambda_0) = a_2,$$

$$v^0(s_3;\lambda_0) = \begin{cases} \lambda_0 \times (-0.3986) \\ \lambda_0 \times 0.16366 \end{cases} \pi^*(s_0;\lambda_0) = \begin{cases} a_2 \\ a_1 \end{cases} \text{for} \begin{cases} -1 \leq \lambda_0 \leq 0 \\ 0 \leq \lambda_0 \leq 1. \end{cases}$$

Thus, we have optimal value function v^0 and optimal first decision function π_0^*:

	$-1 \leq \lambda_0 \leq 0$	$0 \leq \lambda_0 \leq 1$
$v^0(s_1;\lambda_0),\ \pi_0^*(s_1;\lambda_0)$	$\lambda_0 \times (-0.33768),\quad a_1$	$\lambda_0 \times 0.6138,\quad a_2$
$v^0(s_2;\lambda_0),\ \pi_0^*(s_2;\lambda_0)$	$\lambda_0 \times (-0.2338),\quad a_2$	$\lambda_0 \times 0.4824,\quad a_2$
$v^0(s_3;\lambda_0),\ \pi_0^*(s_3;\lambda_0)$	$\lambda_0 \times (-0.3986),\quad a_2$	$\lambda_0 \times 0.16366,\quad a_1$

Hence, substituting $\lambda_0 = 1$, we have

$$v^0(s_1;1) = 0.6138 \qquad v^0(s_2;1) = 0.4824 \qquad v^0(s_3;1) = 0.16366.$$

Of course, these optimal values obtained by solving parametric recursive formula are identical to those by bicursive formula:

$$V^0(s_1) = 0.6138 \qquad V^0(s_2) = 0.4824 \qquad V^0(s_3) = 0.16366.$$

At the same time, we have obtained the Markov policy $\pi^* = \{\pi_0^*, \pi_1^*\}$ for the imbedded process, where

$$\pi_0^*(s_1;\lambda_0) = \begin{cases} a_1 \\ a_2 \end{cases} \quad \pi_0^*(s_2;\lambda_0) = a_2 \qquad \pi_0^*(s_3;\lambda_0) = \begin{cases} a_2 \\ a_1 \end{cases} \text{for} \begin{cases} -1 \leq \lambda_0 \leq 0 \\ 0 \leq \lambda_0 \leq 1 \end{cases}$$

$$\pi_1^*(s_1;\lambda_1) = \begin{cases} a_1 \\ a_2 \end{cases} \quad \pi_1^*(s_2;\lambda_1) = \begin{cases} a_2 \\ a_1 \end{cases} \quad \pi_1^*(s_3;\lambda_1) = \begin{cases} a_2 \\ a_1 \end{cases} \quad \text{for} \begin{cases} -1 \leq \lambda_0 \leq 0 \\ 0 \leq \lambda_0 \leq 1. \end{cases}$$

Now let us from the Markov policy π^* construct an optimal general policy $\tilde{\gamma} = \{\tilde{\gamma}_0, \tilde{\gamma}_1\}$. The first decision function is

$$\tilde{\gamma}_0(s_1) = \pi_0^*(s_1, 1) = a_2 \qquad \tilde{\gamma}_0(s_2) = \pi_0^*(s_2, 1) = a_2 \qquad \tilde{\gamma}_0(s_3) = \pi_0^*(s_3, 1) = a_1.$$

The second decision function

$$\tilde{\gamma}_1(x_0, x_1) = \pi_1^*(x_1, \lambda_1) = \pi_1^*(x_1, \lambda_0 \times r_0(u_0)), \quad u_0 = \pi_0^*(x_0, \lambda_0), \quad \lambda_0 = 1.0$$

reduces in our data to

$$\begin{aligned} \tilde{\gamma}_1(s_1, x_1) &= \pi_1^*(x_1, r_0(a_2)) = \pi_1^*(x_1, 1.0) \\ \tilde{\gamma}_1(s_2, x_1) &= \pi_1^*(x_1, r_0(a_2)) = \pi_1^*(x_1, 1.0) \end{aligned}$$

Copyright © by ORSJ. Unauthorized reproduction of this article is prohibited.

$$\tilde{\gamma}_1(s_3, x_1) = \pi_1^*(x_1, r_0(a_1)) = \pi_1^*(x_1, -0.7).$$

This yields

$$\begin{array}{lll} \tilde{\gamma}_1(s_1, s_1) = a_2, & \tilde{\gamma}_1(s_2, s_1) = a_2, & \tilde{\gamma}_1(s_3, s_1) = a_1 \\ \tilde{\gamma}_1(s_1, s_2) = a_1, & \tilde{\gamma}_1(s_2, s_2) = a_1, & \tilde{\gamma}_1(s_3, s_2) = a_2 \\ \tilde{\gamma}_1(s_1, s_3) = a_1, & \tilde{\gamma}_1(s_2, s_3) = a_1, & \tilde{\gamma}_1(s_3, s_3) = a_2. \end{array}$$

Thus, we have through invariant imbedding obtained an optimal policy $\tilde{\gamma}$, which is not Markov but general. The optimal policy $\tilde{\gamma}$ is completely coincident with μ^* obtained through the bidecision process in §5.2.

5.3. Stochastic decision tree

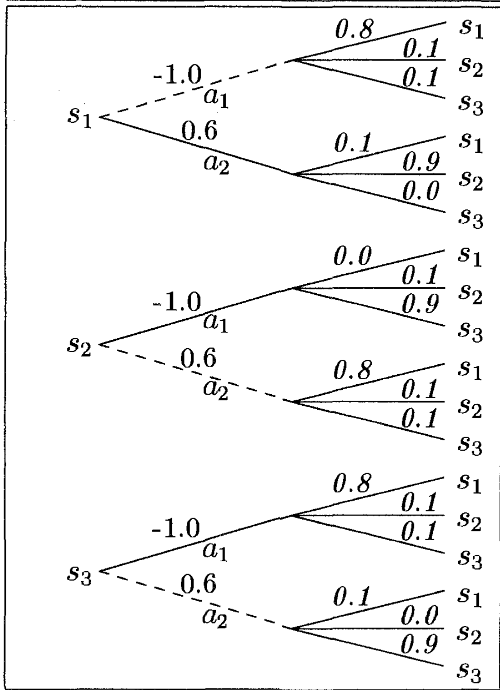
In this subsection we solve directly the problem (5.1) by generating two-stage stochastic decision trees and enumerating all the possible histories together with the related expected values.

We remark that the size yields $2^3 = 8$ first decision functions $\sigma_0 = \begin{pmatrix} \sigma_0(s_1) \\ \sigma_0(s_2) \\ \sigma_0(s_3) \end{pmatrix}$ and $2^9 = 512$ second decision functions

$$\sigma_1 = \begin{pmatrix} \sigma_1(s_1, s_1) & \sigma_1(s_2, s_1) & \sigma_1(s_3, s_1) \\ \sigma_1(s_1, s_2) & \sigma_1(s_2, s_2) & \sigma_1(s_3, s_2) \\ \sigma_1(s_1, s_3) & \sigma_1(s_2, s_3) & \sigma_1(s_3, s_3) \end{pmatrix}.$$

As a total, there are $8 \times 512 = 4096$ general policies $\sigma = \{\sigma_0, \sigma_1\}$ for the problem (5.1).

First, the decision tree method in Figure 2 shows $V^0(s_1) = 0.6138$. Similarly, the method



history	ter.	path	mult.	times	total
	0.3	*0.8*	-0.3	-0.24	
	1.0	*0.1*	-1.0	-0.1	-0.26
	-0.8	*0.1*	0.8	0.08	
	0.3	*0.1*	0.18	0.018	
	1.0	*0.9*	0.6	0.54	**0.558**
	-0.8	*0.0*	-0.48	-0.0	
	0.3	*0.0*	-0.3	-0.0	
	1.0	*0.1*	-1.0	-0.1	**0.62**
	-0.8	*0.9*	0.8	0.72	
	0.3	*0.8*	0.18	0.144	
	1.0	*0.1*	0.6	0.06	0.156
	-0.8	*0.1*	-0.48	-0.048	
	0.3	*0.8*	-0.3	-0.24	
	1.0	*0.1*	-1.0	-0.1	**-0.26**
	-0.8	*0.1*	0.8	0.08	
	0.3	*0.1*	0.18	0.018	
	1.0	*0.0*	0.6	0.0	-0.414
	-0.8	*0.9*	-0.48	-0.432	

Figure 1 : One-stage stochastic decision tree from s_1, s_2 and s_3

Copyright © by ORSJ. Unauthorized reproduction of this article is prohibited.

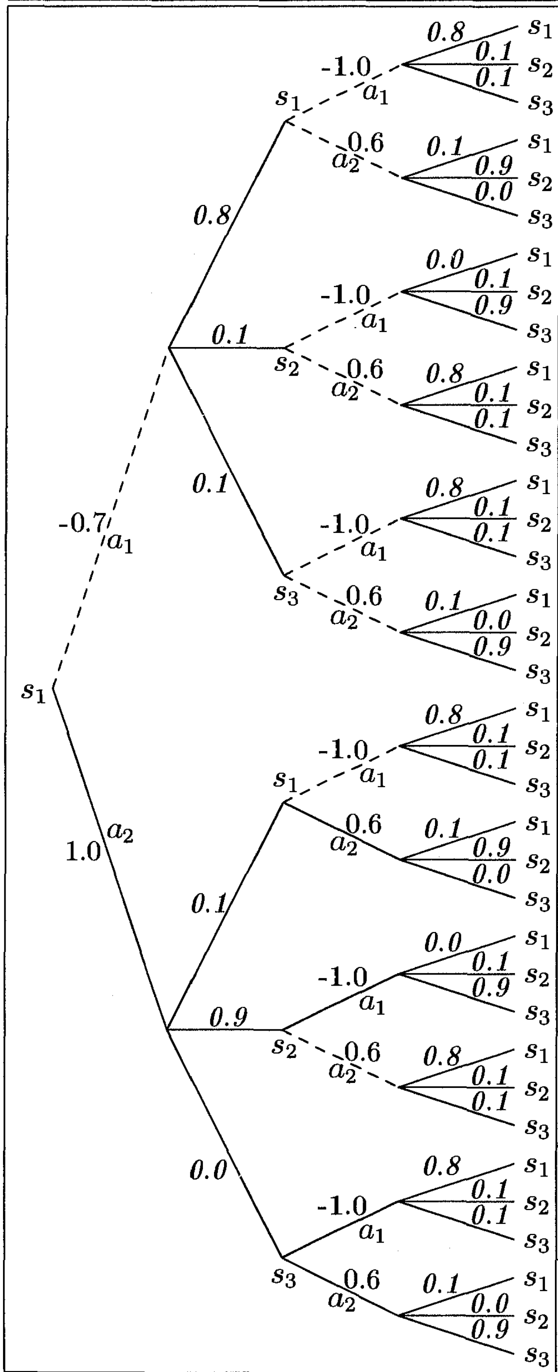
calculates the maximum expected values $V^0(s_2), V^0(s_3)$ on Figures 3,4, respectively. Then, we have

$$V^0(s_1) = 0.6138, \quad V^0(s_2) = 0.4824, \quad V^0(s_3) = 0.16366.$$

The calculation yields, at the same time, the optimal policy $\sigma^* = \{\sigma_0^*(x_0),\ \sigma_1^*(x_0, x_1)\}$:

$$\sigma_0^*(s_1) = a_2, \quad \sigma_0^*(s_2) = a_2, \quad \sigma_0^*(s_3) = a_1$$

$$\begin{array}{lll} \sigma_1^*(s_1, s_1) = a_2, & \sigma_1^*(s_2, s_1) = a_2, & \sigma_1^*(s_3, s_1) = a_1 \\ \sigma_1^*(s_1, s_2) = a_1, & \sigma_1^*(s_2, s_2) = a_1, & \sigma_1^*(s_3, s_2) = a_2 \\ \sigma_1^*(s_1, s_3) = a_1 \text{ or } a_2, & \sigma_1^*(s_2, s_3) = a_1, & \sigma_1^*(s_3, s_3) = a_2. \end{array}$$



history	ter.	path	mult.	times	sub.	total
	0.3	*0.64*	0.21	0.1344		
	1.0	*0.08*	0.7	0.056	**0.1456**	
	-0.8	*0.08*	-0.56	-0.0448		
	0.3	*0.08*	-0.126	-0.01008		
	1.0	*0.72*	-0.42	-0.3024	-0.31248	
	-0.8	*0.0*	0.336	0.0		
	0.3	*0.0*	0.21	0.0		
	1.0	*0.01*	0.7	0.007	-0.0434	
	-0.8	*0.09*	-0.56	-0.0504		0.16366
	0.3	*0.08*	-0.126	-0.01008		
	1.0	*0.01*	-0.42	-0.0042	**-0.01092**	
	-0.8	*0.01*	0.336	0.00336		
	0.3	*0.08*	0.21	0.0168		
	1.0	*0.01*	0.7	0.007	0.0182	
	-0.8	*0.01*	-0.56	-0.0056		
	0.3	*0.01*	-0.126	-0.00126		
	1.0	*0.0*	-0.42	-0.0	**0.02898**	
	-0.8	*0.09*	0.336	0.03024		
	0.3	*0.08*	-0.3	-0.024		
	1.0	*0.01*	-1.0	-0.01	-0.026	
	-0.8	*0.01*	0.8	0.008		
	0.3	*0.01*	0.18	0.0018		
	1.0	*0.09*	0.6	0.054	**0.0558**	
	-0.8	*0.0*	-0.48	-0.0		
	0.3	*0.0*	-0.3	-0.0		
	1.0	*0.09*	-1.0	-0.09	**0.558**	
	-0.8	*0.81*	0.8	0.648		**0.6138**
	0.3	*0.72*	0.18	0.1296		
	1.0	*0.09*	0.6	0.054	0.1404	
	-0.8	*0.09*	-0.48	-0.0432		
	0.3	*0.0*	-0.3	-0.0		
	1.0	*0.0*	-1.0	-0.0	**0.0**	
	-0.8	*0.0*	0.8	0.0		
	0.3	*0.0*	0.18	0.0		
	1.0	*0.0*	0.6	0.0	**0.0**	
	-0.8	*0.0*	-0.48	-0.0		

Figure 2 : Two-stage stochastic decision tree from s_1

Copyright © by ORSJ. Unauthorized reproduction of this article is prohibited.

Note that

$$\sigma_1^*(s_1, s_1) \neq \sigma_1^*(s_3, s_1).$$

Thus, the optimal policy σ^* is not Markov (but general).

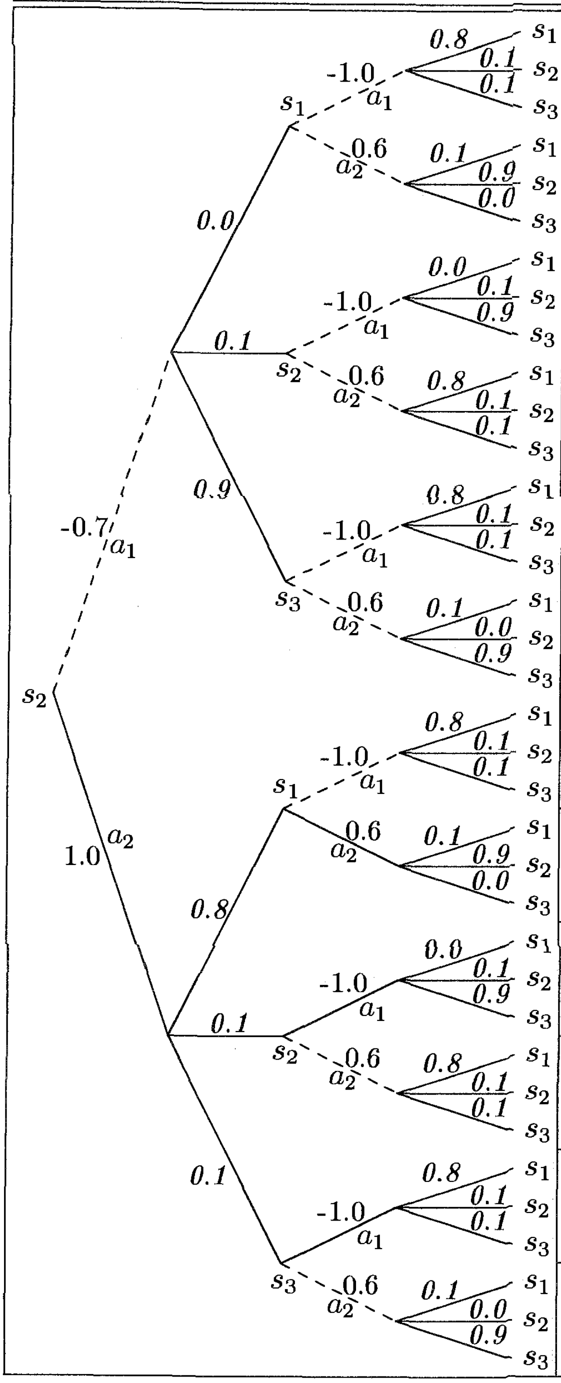
In Figure 1 (resp. Figures 2, 3 and 4) we use the following notations :

history $= x_1 \;\; r_1(u_1) \,/\, u_1 \;\; p(x_2 \mid x_1, u_1) \;\; x_2$

(resp. history $= x_0 \;\; r_0(u_0) \,/\, u_0 \;\; p(x_1 \mid x_0, u_0) \;\; x_1 \;\; r_1(u_1) \,/\, u_1 \;\; p(x_2 \mid x_1, u_1) \;\; x_2$)

ter. = terminal value $= r_G(x_2)$

history



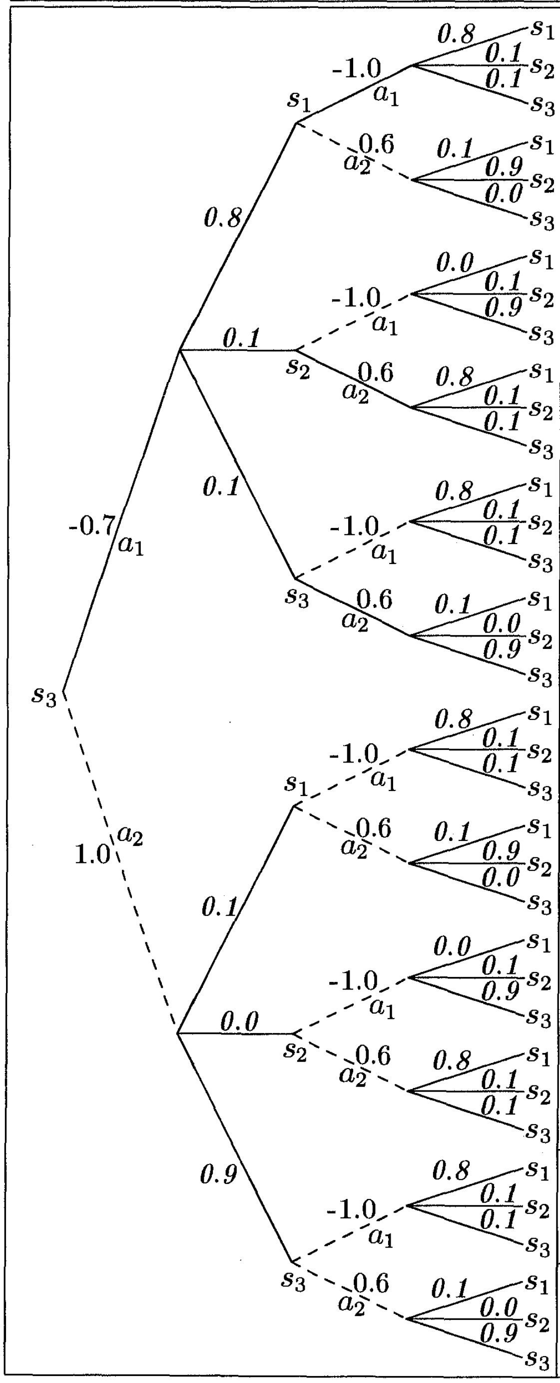
ter.	path	mult.	times	sub.	total
0.3	*0.0*	0.21	0.0		
1.0	*0.0*	0.7	0.0	**0.0**	
-0.8	*0.0*	-0.56	-0.0		
0.3	*0.0*	-0.126	-0.0		
1.0	*0.0*	-0.42	-0.0	**0.0**	
-0.8	*0.0*	0.336	0.0		
0.3	*0.0*	0.21	0.0		
1.0	*0.01*	0.7	0.007	-0.0434	
-0.8	*0.09*	-0.56	-0.0504		0.2499
0.3	*0.08*	-0.126	-0.01008		
1.0	*0.01*	-0.42	-0.0042	**-0.01092**	
-0.8	*0.01*	0.336	0.00336		
0.3	*0.72*	0.21	0.1512		
1.0	*0.09*	0.7	0.063	0.1638	
-0.8	*0.09*	-0.56	-0.0504		
0.3	*0.09*	-0.126	-0.01134		
1.0	*0.0*	-0.42	-0.0	**0.26082**	
-0.8	*0.81*	0.336	0.27216		
0.3	*0.64*	-0.3	-0.192		
1.0	*0.08*	-1.0	-0.08	-0.208	
-0.8	*0.08*	0.8	0.064		
0.3	*0.08*	0.18	0.0144		
1.0	*0.72*	0.6	0.432	**0.4464**	
-0.8	*0.0*	-0.48	-0.0		
0.3	*0.0*	-0.3	-0.0		
1.0	*0.01*	-1.0	-0.01	**0.062**	
-0.8	*0.09*	0.8	0.072		**0.4824**
0.3	*0.08*	0.18	0.0144		
1.0	*0.01*	0.6	0.006	0.0156	
-0.8	*0.01*	-0.48	-0.0048		
0.3	*0.08*	-0.3	-0.024		
1.0	*0.01*	-1.0	-0.01	**-0.026**	
-0.8	*0.01*	0.8	0.008		
0.3	*0.01*	0.18	0.0018		
1.0	*0.0*	0.6	0.0	-0.0414	
-0.8	*0.09*	-0.48	-0.0432		

Figure 3 : Two-stage stochastic decision tree from s_2

Copyright © by ORSJ. Unauthorized reproduction of this article is prohibited.

path = path probability = $p(x_1 \mid x_0, u_0)p(x_2 \mid x_1, u_1)$
mult. = multiplication of the two = $r_1(u_1) \times r_G(x_2)$
(resp. mult. = multiplication of the three = $r_0(u_0) \times r_1(u_1) \times r_G(x_2)$)
times = path × mult.
sub. = subtotal expected value
total = total expected value.

history



ter.	path	mult.	times	sub.	total
0.3	*0.64*	0.21	0.1344		
1.0	*0.08*	0.7	0.056	**0.1456**	
-0.8	*0.08*	-0.56	-0.0448		
0.3	*0.08*	-0.126	-0.01008		
1.0	*0.72*	-0.42	-0.3024	-0.31248	
-0.8	*0.0*	0.336	0.0		
0.3	*0.0*	0.21	0.0		
1.0	*0.01*	0.7	0.007	-0.0434	
-0.8	*0.09*	-0.56	-0.0504		**0.16366**
0.3	*0.08*	-0.126	-0.01008		
1.0	*0.01*	-0.42	-0.0042	**-0.01092**	
-0.8	*0.01*	0.336	0.00336		
0.3	*0.08*	0.21	0.0168		
1.0	*0.01*	0.7	0.007	0.0182	
-0.8	*0.01*	-0.56	-0.0056		
0.3	*0.01*	-0.126	-0.00126		
1.0	*0.0*	-0.42	-0.0	**0.02898**	
-0.8	*0.09*	0.336	0.03024		
0.3	*0.08*	-0.3	-0.024		
1.0	*0.01*	-1.0	-0.01	-0.026	
-0.8	*0.01*	0.8	0.008		
0.3	*0.01*	0.18	0.0018		
1.0	*0.09*	0.6	0.054	**0.0558**	
-0.8	*0.0*	-0.48	-0.0		
0.3	*0.0*	-0.3	-0.0		
1.0	*0.0*	-1.0	-0.0	**0.0**	
-0.8	*0.0*	0.8	0.0		-0.1782
0.3	*0.0*	0.18	0.0		
1.0	*0.0*	0.6	0.0	**0.0**	
-0.8	*0.0*	-0.48	-0.0		
0.3	*0.72*	-0.3	-0.216		
1.0	*0.09*	-1.0	-0.09	**-0.234**	
-0.8	*0.09*	0.8	0.072		
0.3	*0.09*	0.18	0.0162		
1.0	*0.0*	0.6	0.0	-0.3726	
-0.8	*0.81*	-0.48	-0.3888		

Figure 4 : Two-stage stochastic decision tree from s_3

Copyright © by ORSJ. Unauthorized reproduction of this article is prohibited.

Table 1 : all expected value vectors $J^0(\pi)$, where $\pi = \{\pi_0, \pi_1\}$ is Markov

$\pi_0 \setminus \pi_1$	$\begin{pmatrix} a_1 \\ a_1 \\ a_1 \end{pmatrix}$	$\begin{pmatrix} a_1 \\ a_1 \\ a_2 \end{pmatrix}$	$\begin{pmatrix} a_1 \\ a_2 \\ a_1 \end{pmatrix}$	$\begin{pmatrix} a_1 \\ a_2 \\ a_2 \end{pmatrix}$	$\begin{pmatrix} a_2 \\ a_1 \\ a_1 \end{pmatrix}$	$\begin{pmatrix} a_2 \\ a_1 \\ a_2 \end{pmatrix}$	$\begin{pmatrix} a_2 \\ a_2 \\ a_1 \end{pmatrix}$	$\begin{pmatrix} a_2 \\ a_2 \\ a_2 \end{pmatrix}$
$\begin{pmatrix} a_1 \\ a_1 \\ a_1 \end{pmatrix}$	$\begin{pmatrix} 0.1204 \\ 0.1204 \\ 0.1204 \end{pmatrix}$	$\begin{pmatrix} 0.13118 \\ 0.21742 \\ 0.13118 \end{pmatrix}$	$\begin{pmatrix} 0.15288 \\ 0.15288 \\ 0.15288 \end{pmatrix}$	$\begin{pmatrix} 0.16366 \\ 0.2499 \\ 0.16366 \end{pmatrix}$	$\begin{pmatrix} -0.33768 \\ 0.1204 \\ -0.33768 \end{pmatrix}$	$\begin{pmatrix} -0.3269 \\ 0.21742 \\ -0.3269 \end{pmatrix}$	$\begin{pmatrix} -0.3052 \\ 0.15288 \\ -0.3052 \end{pmatrix}$	$\begin{pmatrix} -0.29442 \\ 0.2499 \\ -0.29442 \end{pmatrix}$
$\begin{pmatrix} a_1 \\ a_1 \\ a_2 \end{pmatrix}$	$\begin{pmatrix} 0.1204 \\ 0.1204 \\ -0.26 \end{pmatrix}$	$\begin{pmatrix} 0.13118 \\ 0.21742 \\ -0.3986 \end{pmatrix}$	$\begin{pmatrix} 0.15288 \\ 0.15288 \\ -0.26 \end{pmatrix}$	$\begin{pmatrix} 0.16366 \\ 0.2499 \\ -0.3986 \end{pmatrix}$	$\begin{pmatrix} -0.33768 \\ 0.1204 \\ -0.1782 \end{pmatrix}$	$\begin{pmatrix} -0.3269 \\ 0.21742 \\ -0.3168 \end{pmatrix}$	$\begin{pmatrix} -0.3052 \\ 0.15288 \\ -0.1782 \end{pmatrix}$	$\begin{pmatrix} -0.29442 \\ 0.2499 \\ -0.3168 \end{pmatrix}$
$\begin{pmatrix} a_1 \\ a_2 \\ a_1 \end{pmatrix}$	$\begin{pmatrix} 0.1204 \\ -0.172 \\ 0.1204 \end{pmatrix}$	$\begin{pmatrix} 0.13118 \\ -0.1874 \\ 0.13118 \end{pmatrix}$	$\begin{pmatrix} 0.15288 \\ -0.2184 \\ 0.15288 \end{pmatrix}$	$\begin{pmatrix} 0.16366 \\ -0.2338 \\ 0.16366 \end{pmatrix}$	$\begin{pmatrix} -0.33768 \\ 0.4824 \\ -0.33768 \end{pmatrix}$	$\begin{pmatrix} -0.3269 \\ 0.467 \\ -0.3269 \end{pmatrix}$	$\begin{pmatrix} -0.3052 \\ 0.436 \\ -0.3052 \end{pmatrix}$	$\begin{pmatrix} -0.29442 \\ 0.4206 \\ -0.29442 \end{pmatrix}$
$\begin{pmatrix} a_1 \\ a_2 \\ a_2 \end{pmatrix}$	$\begin{pmatrix} 0.1204 \\ -0.172 \\ -0.26 \end{pmatrix}$	$\begin{pmatrix} 0.13118 \\ -0.1874 \\ -0.1986 \end{pmatrix}$	$\begin{pmatrix} 0.15288 \\ -0.2184 \\ -0.26 \end{pmatrix}$	$\begin{pmatrix} 0.16366 \\ -0.2338 \\ -0.3986 \end{pmatrix}$	$\begin{pmatrix} -0.33768 \\ 0.4824 \\ -0.1782 \end{pmatrix}$	$\begin{pmatrix} -0.3269 \\ 0.467 \\ -0.3168 \end{pmatrix}$	$\begin{pmatrix} -0.3052 \\ 0.436 \\ -0.1782 \end{pmatrix}$	$\begin{pmatrix} -0.29442 \\ 0.4206 \\ -0.3168 \end{pmatrix}$
$\begin{pmatrix} a_2 \\ a_1 \\ a_1 \end{pmatrix}$	$\begin{pmatrix} 0.532 \\ 0.1204 \\ 0.1204 \end{pmatrix}$	$\begin{pmatrix} 0.532 \\ 0.21742 \\ 0.13118 \end{pmatrix}$	$\begin{pmatrix} 0.1144 \\ 0.15288 \\ 0.15288 \end{pmatrix}$	$\begin{pmatrix} 0.1144 \\ 0.2499 \\ 0.16366 \end{pmatrix}$	$\begin{pmatrix} 0.6138 \\ 0.1204 \\ -0.33768 \end{pmatrix}$	$\begin{pmatrix} 0.6138 \\ 0.21742 \\ -0.3269 \end{pmatrix}$	$\begin{pmatrix} 0.1962 \\ 0.15288 \\ -0.3052 \end{pmatrix}$	$\begin{pmatrix} 0.1962 \\ 0.2499 \\ -0.29442 \end{pmatrix}$
$\begin{pmatrix} a_2 \\ a_1 \\ a_2 \end{pmatrix}$	$\begin{pmatrix} 0.532 \\ 0.1204 \\ -0.26 \end{pmatrix}$	$\begin{pmatrix} 0.532 \\ 0.21742 \\ -0.3986 \end{pmatrix}$	$\begin{pmatrix} 0.1144 \\ 0.15288 \\ -0.26 \end{pmatrix}$	$\begin{pmatrix} 0.1144 \\ 0.2499 \\ -0.3986 \end{pmatrix}$	$\begin{pmatrix} 0.6138 \\ 0.1204 \\ -0.1782 \end{pmatrix}$	$\begin{pmatrix} 0.6138 \\ 0.21742 \\ -0.3168 \end{pmatrix}$	$\begin{pmatrix} 0.1962 \\ 0.15288 \\ -0.1782 \end{pmatrix}$	$\begin{pmatrix} 0.1962 \\ 0.2499 \\ -0.3168 \end{pmatrix}$
$\begin{pmatrix} a_2 \\ a_2 \\ a_1 \end{pmatrix}$	$\begin{pmatrix} 0.532 \\ -0.172 \\ 0.1204 \end{pmatrix}$	$\begin{pmatrix} 0.532 \\ -0.1874 \\ 0.13118 \end{pmatrix}$	$\begin{pmatrix} 0.1144 \\ -0.2184 \\ 0.15288 \end{pmatrix}$	$\begin{pmatrix} 0.1144 \\ -0.2338 \\ 0.16366 \end{pmatrix}$	$\begin{pmatrix} 0.6138 \\ 0.4824 \\ -0.33768 \end{pmatrix}$	$\begin{pmatrix} 0.6138 \\ 0.467 \\ -0.3269 \end{pmatrix}$	$\begin{pmatrix} 0.1962 \\ 0.436 \\ -0.3052 \end{pmatrix}$	$\begin{pmatrix} 0.1962 \\ 0.4206 \\ -0.29442 \end{pmatrix}$
$\begin{pmatrix} a_2 \\ a_2 \\ a_2 \end{pmatrix}$	$\begin{pmatrix} 0.532 \\ -0.172 \\ -0.26 \end{pmatrix}$	$\begin{pmatrix} 0.532 \\ -0.1874 \\ -0.3986 \end{pmatrix}$	$\begin{pmatrix} 0.1144 \\ -0.2184 \\ -0.26 \end{pmatrix}$	$\begin{pmatrix} 0.1144 \\ -0.2338 \\ -0.3986 \end{pmatrix}$	$\begin{pmatrix} 0.6138 \\ 0.4824 \\ -0.1782 \end{pmatrix}$	$\begin{pmatrix} 0.6138 \\ 0.467 \\ -0.3168 \end{pmatrix}$	$\begin{pmatrix} 0.1962 \\ 0.436 \\ -0.1782 \end{pmatrix}$	$\begin{pmatrix} 0.1962 \\ 0.4206 \\ -0.3168 \end{pmatrix}$

Further, the *italic* face means probability, and the **bold** number denotes a selection of maximum of up expected value or down.

Second, Table 1 is an arrangement of Figures 2, 3 and 4 by selecting all ($8 \times 8 = 64$) Markov policies $\pi = \{\pi_0, \pi_1\}$. The table lists up the corresponding expected value vectors

$$J^0(\pi) = \begin{pmatrix} J^0(s_1; \pi) \\ J^0(s_2; \pi) \\ J^0(s_3; \pi) \end{pmatrix}$$

where

$$J^0(x_0; \pi) = \sum_{(x_1, x_2) \in X \times X} \{[\, r_0(u_0) r_1(u_1) r_G(x_2)\,] p(x_1|x_0, u_0) p(x_2|x_1, u_1)\}$$

$$u_0 = \pi_0(x_0), \quad u_1 = \pi_1(x_1), \qquad x_0 = s_1,\, s_2,\, s_3$$

$$\pi_0 = \begin{pmatrix} \pi_0(s_1) \\ \pi_0(s_2) \\ \pi_0(s_3) \end{pmatrix} \qquad \pi_1 = \begin{pmatrix} \pi_1(s_1) \\ \pi_1(s_2) \\ \pi_1(s_3) \end{pmatrix}.$$

Copyright © by ORSJ. Unauthorized reproduction of this article is prohibited.

We see that the optimal value vector $V^0 = \begin{pmatrix} V^0(s_1) \\ V^0(s_2) \\ V^0(s_3) \end{pmatrix}$ becomes $V^0 = \begin{pmatrix} 0.6138 \\ 0.4824 \\ 0.16366 \end{pmatrix}$. Thus, Table 1 shows that for any Markov policy π

$$V^0(x_0) > J^0(x_0; \pi) \quad \text{for some} \quad x_0 \in \{s_1, s_2, s_3\},$$

which completes the proof of Theorem 3.2.

Acknowledgments

The authors wishe to thank Professor Seiichi Iwamoto for valuable advice on this investigation. The authors would like to thank the anonymous referees for careful reading of this paper and helpful comments.

References

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

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

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

[4] D.P. Bertsekas and S.E. Shreve: *Stochastic Optimal Control* (Academic Press, New York, 1978).

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

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

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

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

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

[10] S. Iwamoto, K. Tsurusaki and T. Fujita: On Markov policies for minimax decision process. submitted.

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

[12] E. S. Lee: *Quasilinearization and Invariant Imbedding* (Academic Press, New York, 1968).

[13] M. L. Puterman: *Markov Decision Processes : Discrete Stochastic Dynamic Programming* (Wiley & Sons, New York, 1994).

[14] M. Sniedovich: *Dynamic Programming* (Marcel Dekker, Inc. NY, 1992).

[15] N.L. Stokey and R.E. Lucas Jr.: *Recursive Methods in Economic Dynamics* (Harvard Univ. Press, Cambridge, MA, 1989).

Toshiharu Fujita
Department of Electric, Electoronic and Computer Engineering
Kyushu Institute of Technology,
Tobata, Kitakyushu 804-8550, Japan
E-mail: fujita@comp.kyutech.ac.jp

Copyright © by ORSJ. Unauthorized reproduction of this article is prohibited.