数理解析研究所講究録
987巻 1997年 143-163

Optimal Policies for Optimization of Associative Functionals

岩本 誠一 (Seiichi IWAMOTO)

九州大学 経済学部 経済工学科

1 Introduction

Since Bellman [1], an enormous amounts of efforts has been devoted to the study of "dynamic programming". There are two types – deterministic dynamic programming and stochastic dynamic programming – . Since any deterministic system is considered as a special (degenerate) case of stochastic system, we in this paper are mainly concerned with stochastic dynamic programming, which is frequently called "Markov decision process". In this field, there are many research monographs (Howard [13], White [38][39], Nemhauser [28], Denardo [8], Hinderer [12], Bertsekas [4], Bertsekas and Shreve [5], Whittle [40], Hartley, Thomas and White[11], Sniedovich [36], Puterman [31][32] and others) as well as research papers (Blackwell [6], Denardo [7], Kreps [24][25], Porteus [29][30], Mitten [27], Iwamoto [14][15] and others). The study is concerned with the sequential optimization of *additive* function as objective function, which includes the discouted case. Especially, in the field of economics, discounted dynamic programming has been extensively applied (Sargeant [33] and Stokey and Lucas [37]).

In this paper, we study stochastic optimization of *associative* function, called *associative problem*. Especially, three typical associative functions – additive, multiplicative and minimum functions –, generate *additive problem*, *multiplicative problem* and *minimum problem*, respectively.

As was mentioned above, the additive problem has been extensively studied. It is tacitly known that there exists an optimal policy which is Markov - Markov policy is enough - for the additive problems ([3, pp.152,*l*.19-22],[5, pp.6,*l*.20-23], and others). In fact, some papers have at the outset restricted to the set of all Markov policies. And then they have tried to find an optimal policy for the problems under consideration. However, first of all, it should be clarified that the plausibility for this restriction is reasonable. Sometimes, for some reason or other, the clarification is omitted.

The multiplicative problem has also been studied under the restriction that each stage-return function takes nonnegative values. Similar results to additive problem are obtained.

The minimum problem has been originally proposed by Bellman and Zadeh in their seminal paper [3], which has encouraged the study of decision-making in a fuzzy environment ([9], [21], [22], [23]). Recently pointing out a mathematical inconsistency in [3], Iwamoto

and Fujita [18] have derived a valid recursive formula through an invariant imbedding method ([2], [26], [36]).

Throughout the study of these three problems, it has been focussed to derive a recursive equation for a given class of (perhaps not general but Markov) policies. In this paper, we rather raise the question whether there exists an optimal policy for the associative problem with the class or not. If it exists, we further focus our attention on the question whether the optimal policy is Markov or not.

Sections 2 discusses additive problem. We derive recursive equations both for the general class and for the Markov class. Identifying both optimal value functions, we show that Markov policy is enough.

Section 3 discusses multiplicative problem *with negative returns*. Since *regular* dynamic programming does not appply to this multiplicative problem, we propose another two methods – bynamic programming method and invariant imbedding method –. It is shown that neither the general class nor the Markov class does admit the recursive equation. In general, both optimal value functions do not coincide. Nevertherless, it is shown through an invariant imbedding method that there exist an optimal policy in the general class and not necessarily in the Markov class.

Section 4 discusses minimum problem through the invariant imbedding method. Both formulation of and results for minimum problem are same for multiplicative one with negative returns. It is also shown that neither the general class nor the Markov class admits the recursive equation. In general, both optimal value functions do not coincide. Nevertherless, it is shown that there exist an optimal policy in the general class.

As a summary, in the last section, we discusses associative problem. It is emphasized that the invariant imbedding method does in general apply and is essential for associative problem. It is also pointed that same formulation and results as the forementioned multiplicative and minimum problems are obtained. The main results of the paper are as follows. Though Markov policy is enough for both additive problem and multiplicative problem with *nonnegative* returns, it is not always optimal for *associative* problem. Though associative problem does not necessarily admit the recursive equation, a general policy, which is constructed through invariant imbedding, is optimal in associative problem.

Throughout the paper the following data is given :

$$
\begin{aligned}
&N \geq 2 \text{ is an integer; the } \textit{total number of stages} \\
&X = \{s_1, s_2, \ldots, s_p\} \text{ is a finite } \textit{state space} \\
&U = \{a_1, a_2, \ldots, a_k\} \text{ is a finite } \textit{action space} \\
&r_n : X \times U \to R^1 \text{ is an } \textit{n-th reward function} \quad (0 \leq n \leq N-1) \\
&k : X \to R^1 \text{ is a } \textit{terminal reward function} \qquad (1) \\
&f : X \times U \to X \text{ is a } \textit{deterministic transition law} \\
&\quad ; \ f(x,u) \text{ represents the successor state of } x \text{ for action } u \\
&p \text{ is a } \textit{Markov transition law} \\
&\quad : \ p(y|x,u) \geq 0 \ \ \forall (x,u,y) \in X \times U \times X, \quad \sum_{y \in X} p(y|x,u) = 1 \ \ \forall (x,u) \in X \times U
\end{aligned}
$$

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

2 Additive Processes

We begin to discuss additve problem. The following formulation and analysis play an important role for characterizing associative problem. Let us consider the stochastic maximization problem with additive function as follows :

$$\begin{array}{ll} \text{Maximize} & E[\,r_0(x_0,u_0)+r_1(x_1,u_1)+\cdots+r_{N-1}(x_{N-1},u_{N-1})+k(x_N)\,] \\ \text{subject to} & \text{(i)}\quad x_{n+1}\sim p(\cdot\,|x_n,u_n) \\ & \text{(ii)}\quad u_n\in U \qquad n=0,1,\ldots,N-1. \end{array} \tag{2}$$

2.1 General policies

In this subsection we consider the original problem (2) with the set of all general policies. We call this problem *general problem.* With any general policy $\sigma=\{\sigma_n,\ \ldots,\ \sigma_{N-1}\}$ over the $(N-n)$-stage process starting on n-th stage and terminating at the last stage, we associate the expected value :

$$\begin{aligned} I^n(x_n;\sigma) \;=\; \sum_{(x_{n+1},\ldots,x_N)\in X\times\cdots\times X}\!\!\cdots\sum\{&[r_n(x_n,u_n)+\cdots+r_{N-1}(x_{N-1},u_{N-1})+k(x_N)] \\ &\times p(x_{n+1}|x_n,u_n)\cdots p(x_N|x_{N-1},u_{N-1})\} \end{aligned} \tag{3}$$

where $\{u_n,x_{n+1},\ldots,x_{N-1},u_{N-1},x_N\}$ is stochastically generated through the general policy σ and the starting state x_n as follows :

$$\begin{aligned} &\sigma_n(x_n)=u_n \;\rightarrow\; p(\cdot\,|x_n,u_n)\sim x_{n+1} \;\rightarrow \\ &\sigma_{n+1}(x_n,x_{n+1})=u_{n+1} \;\rightarrow\; p(\cdot\,|x_{n+1},u_{n+1})\sim x_{n+2} \;\rightarrow \\ &\sigma_{n+2}(x_n,x_{n+1},x_{n+2})=u_{n+2} \;\rightarrow\; p(\cdot\,|x_{n+2},u_{n+2})\sim x_{n+3} \;\rightarrow \\ &\cdots\;\rightarrow \\ &\sigma_{N-1}(x_n,x_{n+1},\ldots,x_{N-1})=u_{N-1} \;\rightarrow\; p(\cdot\,|x_{N-1},u_{N-1})\sim x_N. \end{aligned} \tag{4}$$

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

$$\begin{aligned} V^N(x_N) &= k(x_N) \qquad x_N\in X \\ V^n(x_n) &= \operatorname*{Max}_{\sigma} I^n(x_n;\sigma) \qquad x_n\in X,\ \ 0\le n\le N-1. \end{aligned} \tag{5}$$

Note that the general problem (2) is identical to (5) with $n=0$. Then we have the recursive formula for the general subproblems :

Theorem 2. 1 *([19])*

$$\begin{aligned} V^N(x) &= k(x) \qquad x\in X \\ V^n(x) &= \operatorname*{Max}_{u\in U}[r_n(x,u)+\sum_{y\in X} V^{n+1}(y)p(y|x,u)] \qquad x\in X,\ \ 0\le n\le N-1. \end{aligned} \tag{6}$$

2.2 Markov policies

In this subsection we restrict the problem (2) to the set of all Markov policies. We call this problem *Markov problem.* Here a policy

$$\pi = \{\pi_0, \pi_1, \ldots, \pi_{N-1}\} \tag{7}$$

is called *Markov* if

$$\pi_0 : X \to U, \quad \pi_1 : X \to U, \quad \ldots \quad , \quad \pi_{N-1} : X \to U. \tag{8}$$

We remark that the size in (1) yields k^p *n-th* decision functions π_n $(n = 0, 1, \ldots, N-1)$ and k^{Np} Markov policies π.

Note that any Markov policy $\pi = \{\pi_n, \ldots, \pi_{N-1}\}$ over the $(N-n)$-stage process is associated with its expected value $I^n(x_n; \pi)$ defined by (3), where the alternate sequence $\{u_n, x_{n+1}, \ldots, x_{N-1}, u_{N-1}, x_N\}$ is similarly generated through the Markov policy π and the starting state x_n as in (4). Here we rematk that

$$u_n = \pi_n(x_n), \quad u_{n+1} = \pi_{n+1}(x_{n+1}), \quad \cdots, u_{N-1} = \pi_{N-1}(x_{N-1}). \tag{9}$$

We define the corresponding *Markov subproblems* as follows :

$$\begin{aligned} v^N(x_N) &= k(x_N) \qquad x_N \in X \\ v^n(x_n) &= \max_{\pi} I^n(x_n; \pi) \qquad x_n \in X, \ \ 0 \le n \le N-1. \end{aligned} \tag{10}$$

Then (10) with $n = 0$ reduces to the Markov problem (2). We have the recursive formula for the Markov subproblems :

Theorem 2. 2 *([19])*

$$\begin{aligned} v^N(x) &= k(x) \qquad x \in X \\ v^n(x) &= \max_{u \in U}[r_n(x,u) + \sum_{y \in X} v^{n+1}(y) p(y|x,u)] \qquad x \in X, \ \ 0 \le n \le N-1. \end{aligned} \tag{11}$$

Theorem 2. 3 *([19])* (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) :*

$$I^0(x_0; \pi^*) = V^0(x_0) \quad \textit{for all} \ \ x_0 \in X. \tag{12}$$

In fact, letting $\pi_n^*(x)$ be a maximizer of (11) (or (6)) 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 (10) are equal to the optimal value functions for the general problems (5) :*

$$v^n(x) = V^n(x) \qquad x \in X, \ \ 0 \le n \le N. \tag{13}$$

3 Multiplicative Processes

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

$$\begin{array}{ll} \text{Maximize} & E[\, r_0(x_0,u_0)r_1(x_1,u_1)\cdots r_{N-1}(x_{N-1},u_{N-1})k(x_N)\,] \\ \text{subject to} & \text{(i)}\ \ x_{n+1}\sim p(\cdot\,|x_n,u_n) \\ & \text{(ii)}\ \ u_n\in U \quad n=0,1,\ldots,N-1 \end{array} \tag{14}$$

We treats two cases for multiplicative process. One is with nonnegative returns. The other is with negative returns.

3.1 Nonnegative Returns

We assume the nonnegativity of return functions :

$$r_n(x,u)\geq 0 \qquad (x,u)\in X\times U,\ 1\leq n\leq N-1. \tag{15}$$

3.1.1 General policies

In this subsection we consider the original problem (14) 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 :

$$\begin{aligned} I^n(x_n;\sigma) &= \sum\sum\cdots\sum_{(x_{n+1},\ldots,x_N)\in X\times\cdots\times X}\{[r_n(x_n,u_n)\cdots r_{N-1}(x_{N-1},u_{N-1})k(x_N)] \\ &\qquad \times p(x_{n+1}|x_n,u_n)\cdots p(x_N|x_{N-1},u_{N-1})\}. \end{aligned} \tag{16}$$

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

$$\begin{aligned} V^N(x_N) &= k(x_N) \qquad x_N\in X \\ V^n(x_n) &= \operatorname*{Max}_{\sigma} I^n(x_n;\sigma) \qquad x_n\in X,\ \ 0\leq n\leq N-1. \end{aligned} \tag{17}$$

Then we have the recursive formula for the general subproblems :

Theorem 3. 1

$$\begin{aligned} V^N(x) &= k(x) \qquad x\in X \\ V^n(x) &= \operatorname*{Max}_{u\in U}[r_n(x,u)\sum_{y\in X} V^{n+1}(y)p(y|x,u)] \qquad x\in X,\ \ 0\leq n\leq N-1. \end{aligned} \tag{18}$$

3.1.2 Markov policies

In this subsection we restrict the problem (14) 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 $I^n(x_n;\pi)$ defined by (16). For the corresponding *Markov subproblems* :

$$\begin{aligned} v^N(x_N) &= k(x_N) \qquad x_N\in X \\ v^n(x_n) &= \operatorname*{Max}_{\pi} I^n(x_n;\pi) \qquad x_n\in X,\ \ 0\leq n\leq N-1. \end{aligned} \tag{19}$$

We have the recursive formula :

Theorem 3. 2

$$\begin{aligned} v^N(x) &= k(x) \qquad x \in X \\ v^n(x) &= \underset{u \in U}{\mathrm{Max}}[r_n(x,u) \sum_{y \in X} v^{n+1}(y) p(y|x,u)] \qquad x \in X, \ \ 0 \le n \le N-1. \end{aligned} \tag{20}$$

Theorem 3. 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 (14) :*

$$I^0(x_0; \pi^*) = V^0(x_0) \quad \textit{for all} \ \ x_0 \in X. \tag{21}$$

In fact, letting $\pi_n^*(x)$ be a maximizer of (18) (or (20)) 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 (19) are equal to the optimal value functions for the general problems (17) :*

$$v^n(x) = V^n(x) \qquad x \in X, \ \ 0 \le n \le N. \tag{22}$$

3.2 Negative Returns

In this subsection we take away the nonnegativity assumption (15) for return functions. We rather assume that it takes at least a negative value:

$$r_n(x,u) < 0 \ \text{ for some } \ (x,u) \in X \times U, \ 0 \le n \le N-1. \tag{23}$$

Then, in general, neither recursive formula (18) nor (20) holds.

Nevertheless, we have the following positive result :

Theorem 3. 4 *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 (14) :*

$$J^0(x_0; \sigma^*) = V^0(x_0) \quad \textit{for all} \ \ x_0 \in X. \tag{24}$$

Theorem 3. 5 *([10]) 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*

$$V^0(x_0) > J^0(x_0; \pi) \quad \textit{for some} \ \ x_0 \in X. \tag{25}$$

In the following we show two alternatives for the *negative* case, i.e., under assumption (23). One is a bi-decision approach. The other is an invariant imbedding approach.

3.2.1 Bi-decision processes

In this subsection we consider the problem (14) 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 :

$$\begin{aligned} I^n(x_n; \sigma) = \sum\sum \cdots \sum_{(x_{n+1}, \ldots, x_N) \in X \times \cdots \times X} \{[r_n(x_n, u_n) \cdots r_{N-1}(x_{N-1}, u_{N-1}) k(x_N)] \\ \times p(x_{n+1}|x_n, u_n) \cdots p(x_N|x_{N-1}, u_{N-1})\}. \end{aligned} \tag{26}$$

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

$$\begin{aligned} V^N(x_N) &= k(x_N) \qquad x_N \in X \\ V^n(x_n) &= \operatorname*{Max}_{\sigma} I^n(x_n;\sigma) \qquad x_n \in X,\ \ 0 \le n \le N-1 \end{aligned} \tag{27}$$

$$\begin{aligned} W^N(x_N) &= k(x_N) \qquad x_N \in X \\ W^n(x_n) &= \min_{\sigma} I^n(x_n;\sigma) \qquad x_n \in X,\ \ 0 \le n \le N-1. \end{aligned} \tag{28}$$

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

$$U(n,x,-) = \{u \in U | r_n(x,u) < 0\},\ \ U(n,x,+) = \{u \in U | r_n(x,u) \ge 0\}. \tag{29}$$

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

Theorem 3. 6 *(See also Bicursive Formula [15, pp.685,l.13-22])*

$$\begin{aligned} V^N(x) &= W^N(x) = k(x) \qquad x \in X \\ V^n(x) &= \operatorname*{Max}_{u \in U(n,x,-)} [r_n(x,u) \sum_{y \in X} W^{n+1}(y) p(y|x,u)] \\ &\qquad \vee \operatorname*{Max}_{u \in U(n,x,+)} [r_n(x,u) \sum_{y \in X} V^{n+1}(y) p(y|x,u)], \end{aligned} \tag{30}$$

$$\begin{aligned} W^n(x) &= \min_{u \in U(n,x,-)} [r_n(x,u) \sum_{y \in X} V^{n+1}(y) p(y|x,u)] \\ &\qquad \wedge \min_{u \in U(n,x,+)} [r_n(x,u) \sum_{y \in X} W^{n+1}(y) p(y|x,u)] \\ &\qquad\qquad x \in X,\ \ 0 \le n \le N-1. \end{aligned} \tag{31}$$

Let $\pi = \{\pi_0, \ldots, \pi_{N-1}\}$ be a general policy for maximum problem and $\sigma = \{\sigma_0, \ldots, \sigma_{N-1}\}$ be a general policy for minimum problem, respectively. Then the pair (π, σ) is called a *strategy* for *both maximum and minimum problem* (14).

Given any strategy (π, σ), we regenerate two policies, upper policy and lower policy, togetger 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 $X = \{s_1, s_2, \ldots, s_p\}$ ([15, pp.683]) , is defined as follows:

$$\mu_0(x_0) := \pi_0(x_0) \tag{32}$$

$$\mu_1(x_0,x_1) := \begin{cases} \sigma_1(x_0,x_1) \\ \pi_1(x_0,x_1) \end{cases} \text{ for } r_0(x_0,u_0) \begin{cases} \le 0 \\ > 0 \end{cases} \tag{33}$$

where

$$u_0 = \pi_0(x_0).$$

$$\mu_2(x_0,x_1,x_2) := \begin{cases} \pi_2(x_0,x_1,x_2) \\ \sigma_2(x_0,x_1,x_2) \\ \sigma_2(x_0,x_1,x_2) \\ \pi_2(x_0,x_1,x_2) \end{cases} \text{ for } r_1(x_1,u_1) \begin{cases} \leq 0 \\ \leq 0 \\ > 0 \\ > 0 \end{cases} u_1 = \begin{cases} \sigma_1(x_0,x_1) \\ \pi_1(x_0,x_1) \\ \sigma_1(x_0,x_1) \\ \pi_1(x_0,x_1) \end{cases} \tag{34}$$

$$\vdots$$

$$\mu_n(x_0,\ldots,x_n) := \begin{cases} \pi_n(x_0,\ldots,x_n) \\ \sigma_n(x_0,\ldots,x_n) \\ \sigma_n(x_0,\ldots,x_n) \\ \pi_n(x_0,\ldots,x_n) \end{cases} \text{ for } r_{n-1}(x_{n-1},u_{n-1}) \begin{cases} \leq 0 \\ \leq 0 \\ > 0 \\ > 0 \end{cases}$$

$$u_{n-1} = \begin{cases} \sigma_{n-1}(x_0,\ldots,x_{n-1}) \\ \pi_{n-1}(x_0,\ldots,x_{n-1}) \\ \sigma_{n-1}(x_0,\ldots,x_{n-1}) \\ \pi_{n-1}(x_0,\ldots,x_{n-1}). \end{cases} \tag{35}$$

and so on.

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 ([15, pp.684]).

Now let us return to the problem of selecting an optimal policy for *maximum problem* (14) *with the set of all general policies.* We have obtained the bicursive formula (30),(31) 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 (30) and a minimizer for (31), 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 (14). 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 (32)-(35) together with bicursive formula (30),(31), we see that *upper policy* $\mu^* = \{\mu_0^*, \ldots, \mu_{N-1}^*\}$ *is optimal policy for maximum problem* (14). Thus, the general policy μ^* yields the optimal value function $V^0(\cdot)$ in (27) for the general maximum problem.

Similarly, the lower policy $\hat{\nu} = \{\hat{\nu}_0, \ldots, \hat{\nu}_{N-1}\}$ is optimal for minimum problem (14). The general policy $\hat{\nu}$ yields the optimal value function $W^0(\cdot)$ in (27) for the general minimum problem.

Further, restricting the problem (14) to the *set of all Markov policies*, we have the same bicursive formula (30),(31) for the *Markov problem.* It is shown that the corresponding optimal value functions for Markov subproblems $\{v^n(\cdot), w^n(\cdot)\}$ are identical to the optimal value functions $\{V^n(\cdot), W^n(\cdot)\}$ in (27),(28), respectively:

$$V^n(x) = v^n(x) \quad W^n(x) = w^n(x) \quad x \in X \;\; 0 \leq n \leq N. \tag{36}$$

Letting $\pi_n^*(x)$ and $\hat{\sigma}_n(x)$ be a maximizer for the resulting recursive formula for $\{v^n(\cdot), w^n(\cdot)\}$, we have a pair of Markov policies π^* and $\hat{\sigma}$. Then, the regenerated upper policy μ^* is not Markov but optimal for maximum problem (14). Thus, the general policy μ^* yields the optimal value function $V^0(\cdot)$ in (27) for the general maximum problem. However, Markov policy does not always yield the optimal value function $V^0(\cdot)$ in (27) for

the general maximum problem. Because even if the strategy $(\pi^*, \hat{\sigma})$ obtained by selecting both maximizer and minimizer for bicursive formula is Markov, the resulting upper and lower policies μ^* and $\hat{\nu}$ are not necessarily Markov. In general, both the policies constructed through (32)-(35) and its dual from $(\pi^*, \hat{\sigma})$ are general for Markov problems.

Similarly, the lower policy $\hat{\nu}$ is optimal for minimum problem (14). The general policy $\hat{\nu}$ yields the optimal value function $W^0(\cdot)$ in (27) for the general minimum problem. However, Markov policy does not always yield the optimal value function $W^0(\cdot)$ in (28) for the general minimum problem.

3.2.2 Imbedded processes

In this subsection we imbed the problem (14) into a family of *terminal processes on one-dimensionally augumented state space.* We note that the return, which may take negative values, is multiplicatively accumulating.

Let us return to the original stochastic maximization problem (14) with multiplicative function. Without loss of generality, we may assume that

$$\begin{aligned} &-1 \le r_n(x,u) \le 1 \quad (x,u) \in X \times U, \ 0 \le n \le N-1 \\ &-1 \le k(x) \le 1 \qquad x \in X. \end{aligned} \tag{37}$$

Under the condition (37), we imbed the problem (14) into the family of parametrized problems as follows :

$$\begin{aligned} \text{Maximize} \quad & E[\,\lambda_0 r_0(x_0,u_0) r_1(x_1,u_1)\cdots r_{N-1}(x_{N-1},u_{N-1}) k(x_N)\,] \\ \text{subject to} \quad & \text{(i)} \ \ x_{n+1} \sim p(\cdot\,|x_n,u_n) \\ & \text{(ii)} \ \ u_n \in U \quad n = 0,1,\ldots,N-1 \end{aligned} \tag{38}$$

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

First we consider the imbedded problem (38) with the set of all general policies, called *general problem.* Here we note that any general policy :

$$\sigma = \{\sigma_0, \sigma_1, \ldots, \sigma_{N-1}\} \tag{39}$$

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 :

$$\begin{aligned} K^n(x_n,\lambda_n\,;\sigma) \ = \sum_{(x_{n+1},\ldots,x_N)\in X\times\cdots\times X} \sum \cdots \sum \{ & [\lambda_n r_n(x_n,u_n)\cdots r_{N-1}(x_{N-1},u_{N-1})k(x_N)] \\ & \times p(x_{n+1}|x_n,u_n)\cdots p(x_N|x_{N-1},u_{N-1})\} \end{aligned} \tag{40}$$

where the alternating sequence of action and augumented 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 :

$$\begin{aligned}
&\sigma_n(x_n, \lambda_n) = u_n \ \rightarrow \ \begin{cases} p(\cdot\,|x_n, u_n) \sim x_{n+1} \\ \lambda_n r_n(x_n, u_n) = \lambda_{n+1} \end{cases} \\
&\rightarrow\ \sigma_{n+1}(x_n, \lambda_n, x_{n+1}, \lambda_{n+1}) = u_{n+1} \ \rightarrow \ \begin{cases} p(\cdot\,|x_{n+1}, u_{n+1}) \sim x_{n+2} \\ \lambda_{n+1} r_{n+1}(x_{n+1}, u_{n+1}) = \lambda_{n+2} \end{cases} \\
&\rightarrow\ \sigma_{n+2}(x_n, \lambda_n, x_{n+1}, \lambda_{n+1}, x_{n+2}, \lambda_{n+2}) = u_{n+2} \qquad (41) \\
&\rightarrow \begin{cases} p(\cdot\,|x_{n+2}, u_{n+2}) \sim x_{n+3} \\ \lambda_{n+2} r_{n+2}(x_{n+2}, u_{n+2}) = \lambda_{n+3} \end{cases} \rightarrow \quad \cdots \\
&\rightarrow\ \sigma_{N-1}(x_n, \lambda_n, x_{n+1}, \lambda_{n+1}, \ldots, x_{N-1}, \lambda_{N-1}) = u_{N-1} \\
&\rightarrow \begin{cases} p(\cdot\,|x_{N-1}, u_{N-1}) \sim x_N \\ \lambda_{N-1} r_{N-1}(x_{N-1}, u_{N-1}) = \lambda_N. \end{cases}
\end{aligned}$$

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

$$\begin{aligned}
V^N(x_N, \lambda_N) &= \lambda_N k(x_N) \qquad x_N \in X,\ \ -1 \le \lambda_N \le 1 \\
V^n(x_n, \lambda_n) &= \operatorname*{Max}_{\sigma} K^n(x_n, \lambda_n\,; \sigma) \qquad x_n \in X,\ \ -1 \le \lambda_n \le 1,\ \ 0 \le n \le N-1 \qquad (42)
\end{aligned}$$

Then the general problem (38) is identical to (42) with $n = 0$. We have the recursive formula for the general subproblems :

Theorem 3. 7

$$\begin{aligned}
V^N(x, \lambda) &= \lambda k(x) \qquad x \in X,\ \ \lambda \in [-1, 1] \\
V^n(x, \lambda) &= \operatorname*{Max}_{u \in U} \sum_{y \in X} V^{n+1}(y, \lambda r_n(x, u)) p(y|x, u) \qquad (43) \\
&\qquad\qquad x \in X,\ \ \lambda \in [-1, 1],\ \ \ 0 \le n \le N-1.
\end{aligned}$$

Second we consider the *Markov problem.* That is, we restrict the imbedded problem (38) to the set of all Markov policies. Here Markov policy

$$\pi = \{\pi_0,\ \pi_1,\ \ldots,\ \pi_{N-1}\} \qquad (44)$$

consists in turn of two-variable decision functions :

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

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 (40). The alternating sequence of action and augumented 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 (41), where

$$\begin{aligned} &\pi_n(x_n, \lambda_n) = u_n \\ &\pi_{n+1}(x_{n+1}, \lambda_{n+1}) = u_{n+1} \\ &\cdots \\ &\pi_{N-1}(x_{N-1}, \lambda_{N-1}) = u_{N-1}. \end{aligned} \tag{45}$$

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

$$\begin{aligned} v^N(x_N, \lambda_N) &= \lambda_N k(x_N) \qquad x_N \in X, \ -1 \le \lambda_N \le 1 \\ v^n(x_n, \lambda_n) &= \operatorname*{Max}_{\pi} K^n(x_n, \lambda_n ; \pi) \qquad x_n \in X, \ -1 \le \lambda_n \le 1, \ 0 \le n \le N-1. \end{aligned} \tag{46}$$

Note that the Markov problem (38) is also (46) with $n = 0$. Then we have the recursive formula for the Markov subproblems :

Theorem 3. 8

$$\begin{aligned} v^N(x, \lambda) &= \lambda k(x) \qquad x \in X, \ \lambda \in [-1, 1] \\ v^n(x, \lambda) &= \operatorname*{Max}_{u \in U} \sum_{y \in X} v^{n+1}(y, \lambda r_n(x, u)) p(y|x, u) \\ &\qquad\qquad x \in X, \ \lambda \in [-1, 1], \quad 0 \le n \le N-1. \end{aligned} \tag{47}$$

Theorem 3. 9 (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 (38):*

$$V^0(x_0, \lambda_0) = K^0(x_0, \lambda_0 ; \pi^*) \quad \textit{for all } (x_0, \lambda_0) \in X \times [-1, 1]. \tag{48}$$

In fact, letting $\pi_n^*(x, \lambda)$ be a maximizer of (47) (or (43)) 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 (46) are equal to the optimal value functions for the general problems (42) :*

$$v^n(x, \lambda) = V^n(x, \lambda) \qquad (x, \lambda) \in X \times [-1, 1], \ 0 \le n \le N. \tag{49}$$

4 Minimum Processes

In this section we consider two types of minimum problems. One is deterministic optimization of minimun function. The other is stochastic. We summarize only results. More detailed analysis and a related example are given in [20].

4.1 Deterministic Dynamics

Let us consider the deterministic maximization problem for minimum function :

$$\begin{aligned} \text{Maximize} \quad & r_0(x_0, u_0) \wedge r_1(x_1, u_1) \wedge \cdots \wedge r_{N-1}(x_{N-1}, u_{N-1}) \wedge k(x_N) \\ \text{subject to} \quad & \text{(i)} \ \ f(x_n, u_n) = x_{n+1} \\ & \text{(ii)} \ \ u_n \in U \quad n = 0, 1, \ldots, N-1. \end{aligned} \tag{50}$$

4.1.1 General policies

In this subsection we consider the *general problem* (50), which is accompanied with the set of all general policies. We associate any general policy $\sigma = \{\sigma_n, \ldots, \sigma_{N-1}\}$ for the $(N-n)$-stage process with its expected value :

$$J^n(x_n;\sigma) = r_n(x_n,u_n) \wedge \cdots \wedge r_{N-1}(x_{N-1},u_{N-1}) \wedge k(x_N) \tag{51}$$

where $\{u_n, x_{n+1}, \ldots, x_{N-1}, u_{N-1}, x_N\}$ is uniquely determined by the deteministic transition law f together with general policy σ and x_n.

We consider the following family of *general subproblems* :

$$\begin{aligned} V^N(x_N) &= k(x_N) \qquad x_N \in X \\ V^n(x_n) &= \operatorname*{Max}_{\sigma} J^n(x_n;\sigma) \qquad x_n \in X,\ 0 \le n \le N-1. \end{aligned} \tag{52}$$

Note that the general problem (50) is identical to (52) with $n = 0$. Further we should remark that the maximization for the subproblems above is taken for all general policies, namely, in problem (52)

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

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

Theorem 4. 1

$$\begin{aligned} V^N(x) &= k(x) \qquad x \in X \\ V^n(x) &= \operatorname*{Max}_{u \in U}[r_n(x,u) \wedge V^{n+1}(f(x,u))] \qquad x \in X,\ 0 \le n \le N-1. \end{aligned} \tag{53}$$

4.1.2 Markov policies

We consider the problem (50) with the set of all Markov policies, as Bellman and Zadeh [3, §4] have done. We call this problem *Markov problem.* Note that any Markov policy $\pi = \{\pi_n, \ldots, \pi_{N-1}\}$ for the $(N-n)$-stage process is associated with its value $J^n(x_n;\pi)$ through (51).

We consider the following family of *Markov subproblems* :

$$\begin{aligned} v^N(x_N) &= k(x_N) \qquad x_N \in X \\ v^n(x_n) &= \operatorname*{Max}_{\pi} J^n(x_n;\pi) \qquad x_n \in X,\ 0 \le n \le N-1. \end{aligned} \tag{54}$$

Thus (54) with $n = 0$ reduces to the Markov problem (50). Further we remark that the maximization for the above subproblems is restricted to the set of all Markov policies, namely, in problem (54)

$$\pi_m : X \to U \quad n \le m \le N-1.$$

Then we have the backward recursive formula for the Markov subproblems :

Theorem 4. 2 (Bellman and Zadeh [3, §4])

$$\begin{aligned} v^N(x) &= k(x) \qquad x \in X \\ v^n(x) &= \operatorname*{Max}_{u \in U}[r_n(x,u) \wedge v^{n+1}(f(x,u))] \qquad x \in X,\ 0 \le n \le N-1. \end{aligned} \tag{55}$$

Furthermore we have

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 (50) :*

$$J^0(x_0;\pi^*) = V^0(x_0) \quad \textit{for all} \quad x_0 \in X. \tag{56}$$

In fact, letting $\pi_n^*(x)$ be a maximizer of (55) (or (53)) 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 (54) are equal to the optimal value functions for the general subproblems (52) :*

$$v^n(x) = V^n(x) \qquad x \in X, \quad 0 \le n \le N. \tag{57}$$

4.2 Stochastic Dynamics

Let us consider the stochastic maximization problem with minimum function :

$$\begin{array}{ll} \text{Maximize} & E[\,r_0(x_0,u_0) \wedge r_1(x_1,u_1) \wedge \cdots \wedge r_{N-1}(x_{N-1},u_{N-1}) \wedge k(x_N)\,] \\ \text{subject to} & \text{(i)} \quad x_{n+1} \sim p(\cdot\,|x_n,u_n) \\ & \text{(ii)} \quad u_n \in U \quad n = 0,1,\ldots,N-1. \end{array} \tag{58}$$

4.2.1 General policies

In this subsection we consider the problem (58) with the set of all general policies, called *general problem.* Any general policy $\sigma = \{\sigma_n, \ldots, \sigma_{N-1}\}$ over the $(N-n)$-stage process yields its expected value :

$$\begin{aligned} J^n(x_n;\sigma) \;=\; & \sum_{(x_{n+1},\ldots,x_N)\in X\times\cdots\times X} \cdots \sum \{[r_n(x_n,u_n) \wedge \cdots \wedge r_{N-1}(x_{N-1},u_{N-1}) \wedge k(x_N)] \\ & \times p(x_{n+1}|x_n,u_n)\cdots p(x_N|x_{N-1},u_{N-1})\} \end{aligned} \tag{59}$$

where $\{u_n, x_{n+1}, \ldots, x_{N-1}, u_{N-1}, x_N\}$ is stochastically generated by (4) through σ and x_n.

We define the following family of *general subproblems* :

$$\begin{aligned} V^N(x_N) &= k(x_N) \qquad x_N \in X \\ V^n(x_n) &= \operatorname*{Max}_{\sigma} J^n(x_n;\sigma) \qquad x_n \in X, \quad 0 \le n \le N-1. \end{aligned} \tag{60}$$

Thus the general problem (58) is identical to (60) with $n = 0$. However, in general, the recursive formula for the general subproblems :

$$\begin{aligned} V^N(x) &= k(x) \qquad x \in X \\ V^n(x) &= \operatorname*{Max}_{u\in U}[r_n(x,u) \wedge \sum_{y\in X} V^{n+1}(y)p(y|x,u)] \qquad x \in X, \quad 0 \le n \le N-1 \end{aligned} \tag{61}$$

does not hold.

Nevertheless, we have the following positive result :

Theorem 4. 4 *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 (58) :*

$$J^0(x_0;\sigma^*) = V^0(x_0) \quad \textit{for all} \quad x_0 \in X. \tag{62}$$

In fact, an invariant imbedding approach ([2],[16],[26]) for the general problem (58) yields an optimal general policy $\sigma^* = \{\sigma_0^*, \ldots, \sigma_{N-1}^*\}$.

4.2.2 Markov policies

In this subsection we consider the problem (58) restricted to the set of all Markov policies as Bellman and Zadeh [3, §5] have done. We call this problem *Markov problem.* Any Markov policy $\pi = \{\pi_n, \ldots, \pi_{N-1}\}$ over the $(N-n)$-stage process yields its expected value $J^n(x_n;\pi)$ thorough (59).

We define the corresponding *Markov subproblems* as follows :

$$\begin{aligned} v^N(x_N) &= k(x_N) \qquad x_N \in X \\ v^n(x_n) &= \operatorname*{Max}_{\pi} J^n(x_n;\pi) \qquad x_n \in X, \ \ 0 \le n \le N-1. \end{aligned} \tag{63}$$

Then the Markov problem (58) becomes (63) with $n = 0$. In general, the recursive formula for the Markov subproblems :

$$\begin{aligned} v^N(x) &= k(x) \qquad x \in X \\ v^n(x) &= \operatorname*{Max}_{u\in U}[r_n(x,u) \wedge \sum_{y\in X} v^{n+1}(y)p(y|x,u)] \qquad x \in X, \ \ 0 \le n \le N-1 \end{aligned} \tag{64}$$

does not hold. We remark that Bellman and Zadeh derive the recursive formula for $\{v^0(\cdot), v^1(\cdot), \ldots, v^N(\cdot)\}$ ([3, §5]). (See also ([9],[21],[23],[22]). However, the recursive formula (64) does not hold, as is shown by Iwamoto and Fujita ([18]).

Theorem 4. 5 *([20]) 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 minimum function such that for any Markov policy π*

$$V^0(x_0) > J^0(x_0;\pi) \quad \textit{for some} \quad x_0 \in X. \tag{65}$$

4.3 Imbedded Process

Let us return to the original stochastic maximization problem (58) with minimum function. Note that, without loss of generality, we may assume that

$$\begin{aligned} &0 \le r_n(x,u) \le 1 \quad (x,u) \in X \times U, \ \ 0 \le n \le N-1 \\ &0 \le k(x) \le 1 \qquad x \in X. \end{aligned} \tag{66}$$

In this section we, under the condition (66), imbed the problem (58) into the family of parametrized problems as follows :

$$\begin{aligned} \text{Maximize} \quad & E[\,\lambda_0 \wedge r_0(x_0,u_0) \wedge r_1(x_1,u_1) \wedge \cdots \wedge r_{N-1}(x_{N-1},u_{N-1}) \wedge k(x_N)\,] \\ \text{subject to} \quad & \text{(i)} \ \ x_{n+1} \sim p(\cdot\,|x_n,u_n) \\ & \text{(ii)} \ \ u_n \in U \qquad n = 0,1,\ldots,N-1 \end{aligned} \tag{67}$$

where the papameter ranges over $\lambda_0 \in [0,1]$.

4.3.1 General policies

First we consider the imbedded problem (67) with the set of all general policies, called *general problem.* Here we note that any general policy :

$$\sigma = \{\sigma_0, \sigma_1, \ldots, \sigma_{N-1}\} \tag{68}$$

consists of the following decision functions

$$\begin{aligned} \sigma_0 &: X \times [0,1] \to U \\ \sigma_1 &: (X \times [0,1]) \times (X \times [0,1]) \to U \\ &\cdots \\ \sigma_{N-1} &: (X \times [0,1]) \times (X \times [0,1]) \times \cdots \times (X \times [0,1]) \to U. \end{aligned}$$

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

$$\begin{aligned} K^n(x_n, \lambda_n ; \sigma) = \sum_{(x_{n+1},\ldots,x_N)\in X\times\cdots\times X} \sum \cdots \sum \{ & [\lambda_n \wedge r_n(x_n, u_n) \wedge \cdots \wedge r_{N-1}(x_{N-1}, u_{N-1}) \wedge k(x_N)] \\ & \times p(x_{n+1}|x_n, u_n) \cdots p(x_N|x_{N-1}, u_{N-1}) \} \end{aligned} \tag{69}$$

where the alternating sequence of action and augumented 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 :

$$\begin{aligned} & \sigma_n(x_n, \lambda_n) = u_n \ \to \ \begin{cases} p(\cdot\,|x_n, u_n) \sim x_{n+1} \\ \lambda_n \wedge r_n(x_n, u_n) = \lambda_{n+1} \end{cases} \\ & \to \ \sigma_{n+1}(x_n, \lambda_n, x_{n+1}, \lambda_{n+1}) = u_{n+1} \ \to \ \begin{cases} p(\cdot\,|x_{n+1}, u_{n+1}) \sim x_{n+2} \\ \lambda_{n+1} \wedge r_{n+1}(x_{n+1}, u_{n+1}) = \lambda_{n+2} \end{cases} \\ & \to \ \sigma_{n+2}(x_n, \lambda_n, x_{n+1}, \lambda_{n+1}, x_{n+2}, \lambda_{n+2}) = u_{n+2} \\ & \to \ \begin{cases} p(\cdot\,|x_{n+2}, u_{n+2}) \sim x_{n+3} \\ \lambda_{n+2} \wedge r_{n+2}(x_{n+2}, u_{n+2}) = \lambda_{n+3} \end{cases} \ \to \ \cdots \\ & \to \ \sigma_{N-1}(x_n, \lambda_n, x_{n+1}, \lambda_{n+1}, \ldots, x_{N-1}, \lambda_{N-1}) = u_{N-1} \\ & \to \ \begin{cases} p(\cdot\,|x_{N-1}, u_{N-1}) \sim x_N \\ \lambda_{N-1} \wedge r_{N-1}(x_{N-1}, u_{N-1}) = \lambda_N. \end{cases} \end{aligned} \tag{70}$$

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

$$\begin{aligned} V^N(x_N, \lambda_N) &= \lambda_N \wedge k(x_N) \qquad x_N \in X, \ \ 0 \le \lambda_N \le 1 \\ V^n(x_n, \lambda_n) &= \operatorname*{Max}_{\sigma} K^n(x_n, \lambda_n ; \sigma) \qquad x_n \in X, \ \ 0 \le \lambda_n \le 1, \ \ 0 \le n \le N-1. \end{aligned} \tag{71}$$

Then the general problem (67) is identical to (71) with $n = 0$. We have the recursive formula for the general subproblems :

Theorem 4. 6

$$\begin{aligned} V^N(x, \lambda) &= \lambda \wedge k(x) \qquad x \in X, \ \ \lambda \in [0,1] \\ V^n(x, \lambda) &= \operatorname*{Max}_{u\in U} \sum_{y\in X} V^{n+1}(y, \lambda \wedge r_n(x,u)) p(y|x,u) \\ & \qquad\qquad x \in X, \ \ \lambda \in [0,1], \ \ 0 \le n \le N-1. \end{aligned} \tag{72}$$

4.3.2 Markov policies

Second we consider the *Markov problem.* That is, we restrict the imbedded problem (67) to the set of all Markov policies. Here Markov policy

$$\pi = \{\pi_0, \pi_1, \ldots, \pi_{N-1}\} \tag{73}$$

consists in turn of two-variable decision functions :

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

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 (69). The alternating sequence of action and augumented 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 (70), where

$$\begin{aligned} &\pi_n(x_n, \lambda_n) = u_n \\ &\pi_{n+1}(x_{n+1}, \lambda_{n+1}) = u_{n+1} \\ &\cdots \\ &\pi_{N-1}(x_{N-1}, \lambda_{N-1}) = u_{N-1}. \end{aligned} \tag{74}$$

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

$$\begin{aligned} v^N(x_N, \lambda_N) &= \lambda_N \wedge k(x_N) \qquad x_N \in X, \ \ 0 \le \lambda_N \le 1 \\ v^n(x_n, \lambda_n) &= \operatorname*{Max}_{\pi} K^n(x_n, \lambda_n ; \pi) \qquad x_n \in X, \ \ 0 \le \lambda_n \le 1, \ \ 0 \le n \le N-1. \end{aligned} \tag{75}$$

Note that the Markov problem (67) is also (75) with $n = 0$. Then we have the recursive formula for the Markov subproblems :

Theorem 4. 7

$$\begin{aligned} v^N(x, \lambda) &= \lambda \wedge k(x) \qquad x \in X, \ \ \lambda \in [0,1] \\ v^n(x, \lambda) &= \operatorname*{Max}_{u \in U} \sum_{y \in X} v^{n+1}(y, \lambda \wedge r_n(x,u)) p(y|x,u) \\ &\qquad\qquad x \in X, \ \ \lambda \in [0,1], \ \ 0 \le n \le N-1. \end{aligned} \tag{76}$$

Theorem 4. 8 (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 (67):*

$$V^0(x_0, \lambda_0) = K^0(x_0, \lambda_0 ; \pi^*) \ \ \textit{for all} \ (x_0, \lambda_0) \in X \times [0,1]. \tag{77}$$

In fact, letting $\pi_n^*(x, \lambda)$ be a maximizer of (76) (or (72)) for each $(x, \lambda) \in X \times [0,1]$, $0 \le n \le N-1$, we have the optimal Markov policy $\pi^* = \{\pi_0^*, \ldots, \pi_{N-1}^*\}$.
(ii) *The optimal value functions for the Markov subproblems (75) are equal to the optimal value functions for the general problems (71) :*

$$v^n(x, \lambda) = V^n(x, \lambda) \qquad (x, \lambda) \in X \times [0,1], \ \ 0 \le n \le N. \tag{78}$$

5 Associative Processes

In this section, as a summary, we discuss associative problem. Without loss of generality, we may assume that

$$\begin{aligned} &a \le r_n(x,u) \le b \quad (x,u) \in X \times U,\ 0 \le n \le N-1 \\ &a \le k(x) \le b \qquad x \in X \end{aligned} \tag{79}$$

where

$$-\infty < a < b < \infty.$$

Let $\circ : [a, b] \times [a, b] \to [a, b]$ be an *associative* binary relation with a *left-identity element* ι :

$$\lambda \circ (\mu \circ \nu) = (\lambda \circ \mu) \circ \nu \quad \forall \lambda, \mu, \nu \in [a, b] \tag{80}$$

$$\iota \circ \lambda = \lambda \quad \forall \lambda \in [a, b]. \tag{81}$$

The common value (80) is denoted by $\lambda \circ \mu \circ \nu$. We also use the notation $r_1 \circ r_2 \circ \cdots \circ r_n$. Then the equality

$$r_1 \circ r_2 \circ \cdots \circ r_n = \iota \circ r_1 \circ r_2 \circ \cdots \circ r_n \tag{82}$$

plays an essential role in imbedding. The binary relation is said to be *monotone* (resp.*strictly monotone*) if

$$\mu < \nu \implies \lambda \circ \mu \le \lambda \circ \nu \ (resp.\ \lambda \circ \mu < \lambda \circ \nu). \tag{83}$$

Thus we see that Sections 2,3 and 4 have the following triplets $([a, b], \circ, \iota)$:

(i) (addition) $[a, b] = [-M, M]$ for some $M > 0$, $\circ = +$, $\iota = 0$ (84)

(ii) (multiplication) $[a, b] = [-1, 1]$, $\circ = \times$, $\iota = 1$ (85)

(iii) (minimum) $[a, b] = [0, 1]$, $\circ = \wedge$, $\iota = 1$ (86)

, respectively. Further, the addition $+$ is strictly monotone, the multiplication $\times$ is not necessarily monotone (is rather *bitone* in the sense of [15]), and the minimum $\wedge$ is monotone.

In addition, we have five more triplets as follows ([16]) :

(iv) (multiplication-addition) $[a, b] = [-M, M]$ for some $M > 1$,
$a \circ b = ab + a + b$, $\iota = -1$ (87)

(v) (maximum) $[a, b] = [0, 1]$, $a \circ b = a \vee b$, $\iota = 0$ (88)

(vi) (additive fraction) $[a, b] = [0, 1]$, $a \circ b = \dfrac{a+b}{1+ab}$, $\iota = 0$ (89)

(vii) (multiplicative fraction) $[a, b] = [0, 1]$, $a \circ b = \dfrac{ab}{1+\overline{a}\overline{b}}$,
where $\overline{x} = 1 - x$, $\iota = 1$ (90)

(viii) (terminal) $[a, b] = [0, 1]$, $a \circ b = b$, $\iota =$ any element $\in [0, 1]$. (91)

Further, the multiplication-addition is not necesarrily monotone. It is rather bitone. The maximum is monotone. The additive fraction is strictly monotone except for at $a = 1$, so is the multiplicative fraction except for at $a = 0$. Finally, the terminal is strictly monotone.

5.1 Deterministic Dynamics

Let us consider the deterministic maximization of associative function :

$$\begin{array}{ll} \text{Maximize} & r_0(x_0,u_0)\circ r_1(x_1,u_1)\circ\cdots\circ r_{N-1}(x_{N-1},u_{N-1})\circ k(x_N) \\ \text{subject to} & \text{(i)}\ \ f(x_n,u_n)=x_{n+1} \\ & \text{(ii)}\ \ u_n\in U \quad n=0,1,\ldots,N-1. \end{array} \tag{92}$$

Then we have

Theorem 5. 1 (i) *Under the monotonicity*

(i-1) *both the recursive formulae for general problem and for Markov problem hold,*

(i-2) *both the optimal value functions are coincident,*

and

(i-3) *there exist an optimal policy in Markov class.*

(ii) *However, in general,*

(ii-1) *neither the recursive formula for general problem nor for Markov problem holds, and*

(ii-2) *there exists an optimal policy in general class.*

Remark 1. The general optimal policy for (92) is constructed through an invariant imbedding with additional one-dimensional parameter just as was shown both for multiplicative problem with negative returns and for minimum problem. Needless to say, regular dyanamic programming approach applies for associative problem with monotonicity (83). Without introducing an additional one-dimensional parameter, we can derive the recursive equation both for general problem and for Markov problem.

5.2 Stochastic Dynamics

Let us consider the stochastic maximization of associative function:

$$\begin{array}{ll} \text{Maximize} & E[\,r_0(x_0,u_0)\circ r_1(x_1,u_1)\circ\cdots\circ r_{N-1}(x_{N-1},u_{N-1})\circ k(x_N)\,] \\ \text{subject to} & \text{(i)}\ \ x_{n+1}\sim p(\cdot\,|x_n,u_n) \\ & \text{(ii)}\ \ u_n\in U \quad n=0,1,\ldots,N-1. \end{array} \tag{93}$$

Then we have

Theorem 5. 2 (i) *In general, neither the recursive formula for general problem nor for Markov problem holds.*

(ii) *Nevertheless, there always exists an optimal policy in general class.*

Remark 2. The general optimal policy is also constructed through the invariant imbedding approach. Even if associative problem (93) satisfies the monotonicity, regular dynamic programming does not apply. It does not always yield recursive formula for general and Markov problems. Thus, as far as stochastic optimization, the invariant imbedding approach is a fundamental tool for deriving a valid recursive formula for an one-dimensionally extended problem. An optimal Markov policy for the extended problem generates in tern an optimal general policy for the *original* general problem (93). The method is called *stochastic final state approach* [16]. (For the details on *deterministic* final state approach, see [34], [35] and [36, pp.300]).

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, Vol. **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, "Dynamic Programming and Stochastic Control," Academic Press, New York, 1976.

[5] D.P. Bertsekas and S.E. Shreve, "Stochastic Optimal Control," Academic Press, New York, 1978.

[6] D. Blackwell, Discounted dynamic programming, *Ann. Math. Stat.* **36**(1965), 226-235.

[7] E.V. Denardo, Contraction mappings in the theory underlying dynamic programming, *SIAM Review* **9**(1968), 165-177.

[8] E.V. Denardo, "Dynamic Programming : Models and Applications," Prentice-Hall, N.J., 1982.

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

[10] T. Fujita and Y. Tsurusaki, Stochastic optimization of multiplicative functions with negative value, under consideration.

[11] R. Hartley, L.C. Thomas and D.J. White (ed.), "Recent Development in Markov Decision Processes"; Proceedings of an International Conference on Markov Decision Processes, Academic Press, New York, 1980.

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

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

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

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

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

[17] S. Iwamoto, Multi-stage decision processes: minimum criterion, In M. Kodama and S. Iwamoto (ed.), "Multi-media Environment and Economics; Chap. 9" (*in Japanese*), Kyushu Univ. Press, Fukuoka, 1996.

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

[19] S. Iwamoto and Y. Tsurusaki, Multi-stage decision processes: additive criterion, In M. Kodama and S. Iwamoto (ed.), "Multi-media Environment and Economics; Chap. 8" (*in Japanese*), Kyushu Univ. Press, Fukuoka, 1996.

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

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

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

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

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

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

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

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

[28] G.L. Nemhauser, "Introduction to Dynamic Programming," Wiley, New York, 1966.

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

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

[31] M. L. Puterman (ed.), "Dynamic Programming and Its Applications"; Proceedings of the International Conference on Dynamic Programming and Its Appications, Academic Press, New York, 1978.

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

[33] T.J. Sargent, "Dynamic Macroeconomic Theory," Harvard Univ. Press, Cambridge, MA,1987.

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

[35] M. Sniedovich, Analysis of a class of fractional programming problems, Math. Prog. **43**(1989), 329-347.

[36] M. Sniedovich, "Dynamic Programming," Marcel Dekker, Inc. NY, 1992.

[37] N.L. Stokey and R.E. Lucas Jr., "Recursive Methods in Economic Dynamics," Harvard Univ. Press, Cambridge, MA,1989.

[38] D.J. White, "Dynamic Programming," Holden-Day, San Francisco, Calif., 1969.

[39] D.J. White, "Finite Dynamic Programming," Wiley & Sons, New York, 1978.

[40] P. Whittle, "Optimization Over Time, Vol. I,II : Dynamic Programming and Stochastic Control," Wiley & Sons, New York, 1982,1983.