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 discounted 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.152,19-22],[5, pp.6,20-23], 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 apply to this multiplicative problem, we propose another two methods — dynamic 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. Nevertheless, 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. Nevertheless, 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 aforementioned 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:

\[ N \geq 2 \text{ is an integer; the total number of stages} \]

\[ X = \{s_1, s_2, \ldots, s_p\} \text{ is a finite state space} \]

\[ U = \{a_1, a_2, \ldots, a_k\} \text{ is a finite action space} \]

\[ r_n : X \times U \rightarrow R^1 \text{ is an } n\text{-th reward function } (0 \leq n \leq N - 1) \]

\[ k : X \rightarrow R^1 \text{ is a terminal reward function } \]

\[ f : X \times U \rightarrow X \text{ is a deterministic transition law} \]

\[ p(y|x, u) \text{ represents the successor state of } x \text{ for action } u \]

\[ p \text{ is a Markov transition law} \]

\[ p(y|x, u) \geq 0 \quad \forall (x, u, y) \in X \times U \times X, \quad \sum_{y \in X} p(y|x, u) = 1 \quad \forall (x, u) \in X \times U \]
$y \sim p(\cdot|x, u)$ denotes that next state $y$ conditioned on state $x$ and action $u$ appears with probability $p(y|x, u)$.

2 Additive Processes

We begin to discuss additive 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:

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

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:

$$I^n(x_n; \sigma) = \sum_{(x_{n+1}, \ldots, x_N) \in X \times \cdots \times X} \sum_{(y_n, \ldots, y_N) \in X \times \cdots \times X} \sum_{(u_n, \ldots, u_N) \in U^N} \left[ r_n(x_n, u_n) + \cdots + r_{N-1}(x_{N-1}, u_{N-1}) + k(x_N) \right] p(x_{n+1}|x_n, u_n) \cdots p(x_N|x_{N-1}, u_{N-1})$$

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

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

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

$$V^N(x_N) = k(x_N) \quad x_N \in X$$
$$V^n(x_n) = \max_{\sigma} I^n(x_n; \sigma) \quad x_n \in X, \quad 0 \leq n \leq N-1.$$ (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])**

$$V^N(x) = k(x) \quad x \in X$$
$$V^n(x) = \max_{u \in U} \left[ r_n(x, u) + \sum_{y \in X} V^{n+1}(y)p(y|x, u) \right] \quad x \in X, \quad 0 \leq n \leq N-1.$$ (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

$$\pi = \{\pi_0, \pi_1, \ldots, \pi_{N-1}\}$$  \hspace{0.5cm} (7)

is called Markov if

$$\pi_0 : X \rightarrow U, \quad \pi_1 : X \rightarrow U, \quad \ldots, \quad \pi_{N-1} : X \rightarrow U.$$  \hspace{0.5cm} (8)

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

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 $\pi$ and the starting state $x_n$ as in (4). Here we remark that

$$u_n = \pi_n(x_n), \quad u_{n+1} = \pi_{n+1}(x_{n+1}), \quad \ldots, u_{N-1} = \pi_{N-1}(x_{N-1}).$$  \hspace{0.5cm} (9)

We define the corresponding Markov subproblems as follows:

$$v^N(x_N) = k(x_N), \quad x_N \in X$$
$$v^n(x_n) = \max_{\pi} I^n(x_n; \pi), \quad x_n \in X, \quad 0 \leq n \leq N - 1.$$  \hspace{0.5cm} (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])

$$v^N(x) = k(x), \quad x \in X$$
$$v^n(x) = \max_{u \in U} [r_n(x, u) + \sum_{y \in X} v^{n+1}(y) p(y|x, u)]$$  \hspace{0.5cm} (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 $\pi^*$ for the general problem (2):

$$I^0(x_0; \pi^*) = V^0(x_0) \quad \text{for all} \quad x_0 \in X.$$  \hspace{0.5cm} (12)

In fact, letting $\pi^*_n(x)$ be a maximizer of (11) (or (6)) for each $x \in X$, $0 \leq n \leq N - 1$, 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):

$$v^n(x) = V^n(x), \quad x \in X, \quad 0 \leq n \leq N.$$  \hspace{0.5cm} (13)
3 Multiplicative Processes

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

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

We treat 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:

$r_n(x, u) \geq 0 \quad (x, u) \in X \times U, \ 1 \leq n \leq N - 1$. (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 $\sigma = \{\sigma_1, \ldots, \sigma_{N-1}\}$, we associate the corresponding expected value:

$I^n(x_n; \sigma) = \sum_{x_{n+1}, \ldots, x_N} \sum_{x_n} \sum_{x_{n+1}} \cdots \sum_{x_N} [r_0(x_0, u_0)\cdots r_{N-1}(x_{N-1}, u_{N-1})k(x_N)]$

$\times p(x_{n+1}|x_n, u_n)\cdots p(x_N|x_{N-1}, u_{N-1})$. (16)

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

$v^N(x_N) = k(x_N) \quad x_N \in X$
$v^n(x_n) = \max_{\sigma} I^n(x_n; \sigma) \quad x_n \in X, \ 0 \leq n \leq N - 1$. (17)

Then we have the recursive formula for the general subproblems:

Theorem 3.1

$v^N(x) = k(x) \quad x \in X$
$v^n(x) = \max_{u \in U} [r_n(x, u) \sum_{y \in X} V^{n+1}(y)p(y|x, u)] \quad x \in X, \ 0 \leq n \leq N - 1$. (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 $\pi = \{\pi_1, \ldots, \pi_{N-1}\}$ 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:

$v^N(x_N) = k(x_N) \quad x_N \in X$
$v^n(x_n) = \max_{\pi} I^n(x_n; \pi) \quad x_n \in X, \ 0 \leq n \leq N - 1$. (19)

We have the recursive formula:


Theorem 3.2

\begin{align*}
v^N(x) & = k(x) \quad x \in X \\
v^n(x) & = \max_{u \in U} \left[ r_n(x, u) \sum_{y \in X} v^{n+1}(y) p(y|x, u) \right] \quad x \in X, \ 0 \leq n \leq N - 1. \tag{20}
\end{align*}

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 \( \pi^* \) for the general problem (14):

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

In fact, letting \( \pi^*_n(x) \) be a maximizer of (18) (or (20)) for each \( x \in X, \ 0 \leq n \leq N - 1 \), we have the optimal Markov policy \( \pi^* = \{ \pi^*_0, \ldots, \pi^*_N \} \).

(ii) The optimal value functions for the Markov subproblems (19) are equal to the optimal value functions for the general problems (17):

\[ v^n(x) = V^n(x) \quad x \in X, \ 0 \leq n \leq N. \tag{22} \]

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:

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

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 \( \sigma^* \) for the general problem (14):

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

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 \( \pi \)

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

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 \emph{general problem}. With any general policy \( \sigma = \{ \sigma_n, \ldots, \sigma_{N-1} \} \), we associate the corresponding expected value:

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

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

\[
\begin{align*}
V^N(x_N) &= k(x_N) \quad x_N \in X \\
V^n(x_n) &= \max_\sigma I^n(x_n; \sigma) \quad x_n \in X, \quad 0 \leq n \leq N - 1
\end{align*}
\]

\[
\begin{align*}
W^N(x_N) &= k(x_N) \quad x_N \in X \\
W^n(x_n) &= \min_\sigma I^n(x_n; \sigma) \quad x_n \in X, \quad 0 \leq n \leq N - 1.
\end{align*}
\]

(27)

(28)

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

\[
U(n, x, -) = \{u \in U | r_n(x, u) < 0\}, \quad U(n, x, +) = \{u \in U | r_n(x, u) \geq 0\}.
\]

(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])

\[
\begin{align*}
V^N(x) &= W^N(x) = k(x) \quad x \in X \\
V^n(x) &= \max_{u \in U(n, x, -)} [r_n(x, u) \sum_{y \in X} W^{n+1}(y)p(y|x, u)] \\
&\quad \vee \min_{u \in U(n, x, +)} [r_n(x, u) \sum_{y \in X} V^{n+1}(y)p(y|x, u)],
\end{align*}
\]

\[
\begin{align*}
W^n(x) &= \min_{u \in U(n, x, -)} [r_n(x, u) \sum_{y \in X} V^{n+1}(y)p(y|x, u)] \\
&\quad \wedge \min_{u \in U(n, x, +)} [r_n(x, u) \sum_{y \in X} W^{n+1}(y)p(y|x, u)]
\end{align*}
\]

(30)

(31)

\[x \in X, \quad 0 \leq n \leq N - 1.\]

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 \((\pi, \sigma)\) is called a strategy for both maximum and minimum problem (14).

Given any strategy \((\pi, \sigma)\), 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 \(X = \{s_1, s_2, \ldots, s_p\}\) ([15, pp.683]), is defined as follows:

\[
\mu_0(x_0) := \pi_0(x_0)
\]

\[
\mu_1(x_0, x_1) := \begin{cases} 
\sigma_1(x_0, x_1) & \text{for } r_0(x_0, u_0) \leq 0 \\
\pi_1(x_0, x_1) & > 0
\end{cases}
\]

(32)

(33)

where

\[u_0 = \pi_0(x_0)\]
The minimum strategy for regeneration formula for general corresponding both the problem. Then, the pair \( \{v^{*}(\cdot), w^{*}(\cdot)\} \) is optimal policy for maximum problem (14). Thus, the general policy \( \mu^{*} \) 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:

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

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 \( \pi^{*} \) and \( \hat{\sigma} \). Then, the regenerated upper policy \( \mu^{*} \) is not Markov but optimal for maximum problem (14). Thus, the general policy \( \mu^{*} \) 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 \(\mu^*\) 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 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 (14) with multiplicative function. Without loss of generality, we may assume that

\[
-1 \leq r_n(x, u) \leq 1 \quad (x, u) \in X \times U, \quad 0 \leq n \leq N - 1
\]

\[
-1 \leq k(x) \leq 1 \quad x \in X.
\] (37)

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

\[
\text{Maximize } \quad E[\lambda_0 r_0(x_0, u_0) r_1(x_1, u_1) \cdots r_{N-1}(x_{N-1}, u_{N-1}) k(x_N)]
\]

subject to

(i) \(x_{n+1} \sim p(\cdot|x_n, u_n)\)

(ii) \(u_n \in U \quad n = 0, 1, \ldots, N - 1\) (38)

where the parameter 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:

\[
\sigma = \{\sigma_0, \sigma_1, \ldots, \sigma_{N-1}\}
\] (39)

consists of the following decision functions

\[
\sigma_0 : X \times [-1,1] \rightarrow U
\]

\[
\sigma_1 : (X \times [-1,1]) \times (X \times [-1,1]) \rightarrow U
\]

\[
\ldots
\]

\[
\sigma_{N-1} : (X \times [-1,1]) \times (X \times [-1,1]) \times \cdots \times (X \times [-1,1]) \rightarrow U.
\]

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

\[
K^n(x_n, \lambda_n; \sigma) = \sum_{(x_{n+1}, \ldots, x_N) \in X \times \cdots \times X} \sum_{\lambda_n, \ldots, \lambda_{N-1}} \sum_{\lambda_n r_n(x_n, u_n) \cdots r_{N-1}(x_{N-1}, u_{N-1}) k(x_N)} \times p(x_{n+1}|x_n, u_n) \cdots p(x_N|x_{N-1}, u_{N-1})
\] (40)
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 \(\sigma\) and the starting state \((x_n, \lambda_n)\) as follows:

\[
\sigma_n(x_n, \lambda_n) = u_n \rightarrow \begin{cases} p(\cdot|x_n, u_n) \sim x_{n+1} \\ \lambda_n r_n(x_n, u_n) = \lambda_{n+1} \end{cases}
\]

\[
\rightarrow \sigma_{n+1}(x_n, \lambda_n, x_{n+1}, \lambda_{n+1}) = u_{n+1} \rightarrow \begin{cases} p(\cdot|x_{n+1}, u_{n+1}) \sim x_{n+2} \\ \lambda_{n+1} r_{n+1}(x_{n+1}, u_{n+1}) = \lambda_{n+2} \end{cases}
\]

\[\Rightarrow \sigma_{n+2}(x_n, \lambda_n, x_{n+1}, \lambda_{n+1}, x_{n+2}, \lambda_{n+2}) = u_{n+2} \ldots (41)\]

\[
\Rightarrow \sigma_{N-1}(x_n, \lambda_n, x_{n+1}, \lambda_{n+1}, \ldots, x_{N-1}, \lambda_{N-1}) = u_{N-1} \rightarrow \begin{cases} p(\cdot|x_{N-1}, u_{N-1}) \sim x_N \\ \lambda_{N-1} r_{N-1}(x_{N-1}, u_{N-1}) = \lambda_N \end{cases}
\]

We define the family of the corresponding general subproblems:

\[
V^N(x_N, \lambda_N) = \lambda_N k(x_N) \quad x_N \in X, \ -1 \leq \lambda_N \leq 1
\]

\[
V^n(x_n, \lambda_n) = \max_{\sigma} K^n(x_n, \lambda_n; \sigma) \quad x_n \in X, \ -1 \leq \lambda_n \leq 1, \ 0 \leq n \leq N - 1 (42)
\]

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

**Theorem 3.7**

\[
V^N(x, \lambda) = \lambda k(x) \quad x \in X, \ \lambda \in [-1, 1]
\]

\[
V^n(x, \lambda) = \max_{u \in U} \sum_{y \in X} V^{n+1}(y, \lambda r_n(x, u)) p(y|x, u) \quad x \in X, \ \lambda \in [-1, 1], \ 0 \leq n \leq N - 1 (43)
\]

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

\[
\pi = \{\pi_0, \pi_1, \ldots, \pi_{N-1}\} \quad (44)
\]

consists in turn of two-variable decision functions:

\[
\pi_n : X \times [-1, 1] \rightarrow U \quad 0 \leq n \leq N - 1.
\]

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 (40). 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 \( \pi \) and the state \((x_n, \lambda_n)\) as in (41), where

\[
\begin{align*}
\pi_n(x_n, \lambda_n) &= u_n \\
\pi_{n+1}(x_{n+1}, \lambda_{n+1}) &= u_{n+1} \\
&\vdots \\
\pi_{N-1}(x_{N-1}, \lambda_{N-1}) &= u_{N-1}.
\end{align*}
\]  

(45)

We define the family of the corresponding Markov subproblems:

\[
\begin{align*}
v^N(x_N, \lambda_N) &= \lambda_N k(x_N) \quad x_N \in X, \quad -1 \leq \lambda_N \leq 1 \\
v^n(x_n, \lambda_n) &= \max_{\pi} K^n(x_n, \lambda_n; \pi) \quad x_n \in X, \quad -1 \leq \lambda_n \leq 1, \quad 0 \leq n \leq N - 1.
\end{align*}
\]  

(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**

\[
\begin{align*}
v^N(x, \lambda) &= \lambda k(x) \quad x \in X, \quad \lambda \in [-1, 1] \\
v^n(x, \lambda) &= \max_{u \in U} \sum_{y \in X} v^{n+1}(y, \lambda r_n(x, u)) p(y|x, u) \quad x \in X, \quad \lambda \in [-1, 1], \quad 0 \leq n \leq N - 1.
\end{align*}
\]  

(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 \( \pi^* \) for the general problem (38):

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

(48)

In fact, letting \( \pi^*_n(x, \lambda) \) be a maximizer of (47) (or (43)) for each \((x, \lambda) \in X \times [-1, 1], \quad 0 \leq n \leq N - 1\), we have the optimal Markov policy \( \pi^* = \{\pi^*_0, \ldots, \pi^*_N\} \).

(ii) The optimal value functions for the Markov subproblems (46) are equal to the optimal value functions for the general problems (48):

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

(49)

4 Minimum Processes

In this section we consider two types of minimum problems. One is deterministic optimization of minimum 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:

Maximize \( r_0(x_0, u_0) \land r_1(x_1, u_1) \land \cdots \land r_{N-1}(x_{N-1}, u_{N-1}) \land k(x_N) \)

subject to \( \begin{align*}
\text{(i)} \quad f(x_n, u_n) &= x_{n+1} \\
\text{(ii)} \quad u_n &\in U \quad n = 0, 1, \ldots, N - 1.
\end{align*} \)  

(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:

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

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

We consider the following family of general subproblems:

\[
V^N(x_N) = k(x_N) \quad x_N \in X
\]
\[
V^n(x_n) = \max_{\sigma} J^n(x_n; \sigma) \quad x_n \in X, \ 0 \leq n \leq N-1.
\] (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)

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

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

**Theorem 4.1**

\[
V^N(x) = k(x) \quad x \in X
\]
\[
V^n(x) = \max_{u \in U} [r_n(x, u) \land V^{n+1}(f(x, u))] \quad x \in X, \ 0 \leq n \leq N-1.
\] (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:

\[
v^N(x_N) = k(x_N) \quad x_N \in X
\]
\[
v^n(x_n) = \max_{\pi} J^n(x_n; \pi) \quad x_n \in X, \ 0 \leq n \leq N-1.
\] (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)

\[
\pi_m : X \to U \quad n \leq m \leq N-1.
\]

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

**Theorem 4.2** (Bellman and Zadeh [3, §4])

\[
v^N(x) = k(x) \quad x \in X
\]
\[
v^n(x) = \max_{u \in U} [r_n(x, u) \land v^{n+1}(f(x, u))] \quad x \in X, \ 0 \leq n \leq N-1.
\] (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 $\pi^*$ for the general problem (50):

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

In fact, let $\pi^*_n(x)$ be a maximizer of (55) (or (53)) for each $x \in X$, $0 \leq n \leq N-1$, we have the optimal Markov policy $\pi^* = \{\pi^*_0, \ldots, \pi^*_N\}$.

(ii) The optimal value functions for the Markov subproblems (54) are equal to the optimal value functions for the general subproblems (52):

$$v^n(x) = V^n(x) \quad x \in X, \ 0 \leq n \leq N. \quad (57)$$

### 4.2 Stochastic Dynamics

Let us consider the stochastic maximization problem with minimum function:

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

subject to

(i) $x_{n+1} \sim p(\cdot|x_n, u_n)$

(ii) $u_n \in U \quad n = 0, 1, \ldots, N-1$. \quad (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 $\sigma = \{\sigma_0, \ldots, \sigma_{N-1}\}$ over the $(N-n)$-stage process yields its expected value:

$$J^n(x_n; \sigma) = \sum_{(x_{n+1}, \ldots, x_N) \in X \times \cdots \times X} \sum_{u_n \in U} \{r_n(x_n, u_n) \wedge \cdots \wedge r_{N-1}(x_{N-1}, u_{N-1}) \wedge k(x_N)\}$$

$$\times p(x_{n+1}|x_n, u_n) \cdots p(x_N|x_{N-1}, u_{N-1}) \} \quad (59)$$

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

We define the following family of *general subproblems*:

$$V^N(x_N) = k(x_N) \quad x_N \in X$$

$$V^n(x_n) = \max_{\sigma} J^n(x_n; \sigma) \quad x_n \in X, \ 0 \leq n \leq N-1. \quad (60)$$

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

$$V^N(x) = k(x) \quad x \in X$$

$$V^n(x) = \max_{u \in U} \{r_n(x, u) \wedge \sum_{y \in X} V^{n+1}(y)p(y|x, u)\} \quad x \in X, \ 0 \leq n \leq N-1 \quad (61)$$

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 $\sigma^*$ for the general problem (58):

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

(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 $J^n(x_n; \pi)$ through (59).

We define the corresponding Markov subproblems as follows:

$$v^N(x_N) = k(x_N) \quad x_N \in X$$

$$v^n(x_n) = \max_{\pi} J^n(x_n; \pi) \quad x_n \in X, \ 0 \leq n \leq N - 1. \quad (63)$$

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

$$v^N(x) = k(x) \quad x \in X$$

$$v^n(x) = \max_{u \in U} \left[ r_n(x, u) \land \sum_{y \in X} v^{n+1}(y)p(y|x, u) \right] \quad x \in X, \ 0 \leq n \leq N - 1 \quad (64)$$

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 $\pi$

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

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

$$0 \leq r_n(x, u) \leq 1 \quad (x, u) \in X \times U, \ 0 \leq n \leq N - 1$$

$$0 \leq k(x) \leq 1 \quad x \in X. \quad (66)$$

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

Maximize $E[\lambda_0 \land r_0(x_0, u_0) \land r_1(x_1, u_1) \land \cdots \land r_{N-1}(x_{N-1}, u_{N-1}) \land k(x_N)]$

subject to

(i) $x_{n+1} \sim p(\cdot|x_n, u_n)$

(ii) $u_n \in U \quad n = 0, 1, \ldots, N - 1 \quad (67)$

where the parameter 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:

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

consists of the following decision functions

\[ \sigma_0 : X \times [0,1] \to U \]
\[ \sigma_1 : (X \times [0,1]) \times (X \times [0,1]) \to U \]
\[ \vdots \]
\[ \sigma_{N-1} : (X \times [0,1]) \times (X \times [0,1]) \times \cdots \times (X \times [0,1]) \to U. \]

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

\[ K^n(x_n, \lambda_n; \sigma) = \sum_{x_{n+1}} \sum_{x_{n+2}} \cdots \sum_{x_N} \{ [\lambda_n \land r_n(x_n, u_n) \land \cdots \land r_{N-1}(x_{N-1}, u_{N-1}) \land k(x_N)] \times p(x_{n+1}|x_n, u_n) \cdots p(x_N|x_{N-1}, u_{N-1}) \} \]  

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 \( \sigma \) and the starting state \((x_n, \lambda_n)\) as follows:

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

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

\[ V^N(x_N, \lambda_N) = \lambda_N \land k(x_N) \quad x_N \in X, \quad 0 \leq \lambda_N \leq 1 \]
\[ V^n(x_n, \lambda_n) = \max_{\sigma} K^n(x_n, \lambda_n; \sigma) \quad x_n \in X, \quad 0 \leq \lambda_n \leq 1, \quad 0 \leq n \leq N-1. \]  

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

**Theorem 4.6**

\[ V^N(x, \lambda) = \lambda \land k(x) \quad x \in X, \quad \lambda \in [0,1] \]
\[ V^n(x, \lambda) = \max_{u \in U} \sum_{y \in X} V^{n+1}(y, \lambda \land r_n(x, u)) p(y|x, u) \]  

\[ x \in X, \quad \lambda \in [0,1], \quad 0 \leq n \leq N - 1. \]
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

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

(73)

consists in turn of two-variable decision functions:

\[ \pi_n : X \times [0,1] \rightarrow U \quad 0 \leq n \leq N - 1. \]

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 (69). 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 \( \pi \) and the state \((x_n, \lambda_n)\) as in (70), where

\[ \begin{align*}
\pi_n(x_n, \lambda_n) &= u_n \\
\pi_{n+1}(x_{n+1}, \lambda_{n+1}) &= u_{n+1} \\
&\quad \vdots \\
\pi_{N-1}(x_{N-1}, \lambda_{N-1}) &= u_{N-1}.
\end{align*} \]

(74)

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

\[ \begin{align*}
v^N(x_N, \lambda_N) &= \lambda_N \land k(x_N) \quad x_N \in X, \ 0 \leq \lambda_N \leq 1 \\
v^n(x_n, \lambda_n) &= \max_{\pi} K^n(x_n, \lambda_n; \pi) \quad x_n \in X, \ 0 \leq \lambda_n \leq 1, \ 0 \leq n \leq N - 1.
\end{align*} \]

(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**

\[ \begin{align*}
v^N(x, \lambda) &= \lambda \land k(x) \quad x \in X, \ \lambda \in [0,1] \\
v^n(x, \lambda) &= \max_{u \in U} \sum_{y \in X} v^{n+1}(y, \lambda \land \tau_n(x, u)) p(y|x, u) \\
&\quad x \in X, \ \lambda \in [0,1], \ 0 \leq n \leq N - 1.
\end{align*} \]

(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 \( \pi^* \) for the general problem (67):*

\[ V^0(x_0, \lambda_0) = K^0(x_0, \lambda_0; \pi^*) \]  

for all \((x_0, \lambda_0) \in X \times [0,1] \). \hspace{1cm} (77)

In fact, letting \( \pi^*_n(x, \lambda) \) be a maximizer of (76) (or (72)) for each \((x, \lambda) \in X \times [0,1] \), \( 0 \leq n \leq N - 1 \), 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) :*

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

(78)
5 Associative Processes

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

\[ a \leq r_n(x, u) \leq b \quad (x, u) \in X \times U, \quad 0 \leq n \leq N - 1 \]

\[ a \leq k(x) \leq b \quad x \in X \]

(79)

where

\[-\infty < a < b < \infty.\]

Let \( \circ : [a, b] \times [a, b] \rightarrow [a, b] \) be an associative binary relation with a left-identity element \( \iota \):

\[ \lambda \circ (\mu \circ \nu) = (\lambda \circ \mu) \circ \nu \quad \forall \lambda, \mu, \nu \in [a, b] \]  

(80)

\[ \iota \circ \lambda = \lambda \quad \forall \lambda \in [a, b]. \]  

(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

\[ r_1 \circ r_2 \circ \cdots \circ r_n = \iota \circ r_1 \circ r_2 \circ \cdots \circ r_n \]

(82)

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

\[ \mu < \nu \quad \Rightarrow \quad \lambda \circ \mu \leq \lambda \circ \nu \quad (\text{resp. } \lambda \circ \mu < \lambda \circ \nu). \]  

(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 \( M > 0 \), \( \circ = + \), \( \iota = 0 \) \hspace{1cm} (84)

(ii) (multiplication) \( [a, b] = [-1, 1] \), \( \circ = \times \), \( \iota = 1 \) \hspace{1cm} (85)

(iii) (minimum) \( [a, b] = [0, 1] \), \( \circ = \wedge \), \( \iota = 1 \) \hspace{1cm} (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 \( M > 1 \),

\[ a \circ b = ab + a + b, \quad \iota = -1 \]  

(87)

(v) (maximum) \( [a, b] = [0, 1] \), \( a \circ b = a \vee b, \quad \iota = 0 \) \hspace{1cm} (88)

(vi) (additive fraction) \( [a, b] = [0, 1] \), \( a \circ b = \frac{a + b}{1 + ab}, \quad \iota = 0 \) \hspace{1cm} (89)

(vii) (multiplicative fraction) \( [a, b] = [0, 1] \), \( a \circ b = \frac{ab}{1 + ab} \),

where \( x = 1 - x \), \( \iota = 1 \) \hspace{1cm} (90)

(viii) (terminal) \( [a, b] = [0, 1] \), \( a \circ b = b, \quad \iota = \text{any element } \in [0, 1] \). \hspace{1cm} (91)

Further, the multiplication-addition is not necessarily 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:
\[
\text{Maximize } \quad r_0(x_0, u_0) \circ r_1(x_1, u_1) \circ \cdots \circ r_{N-1}(x_{N-1}, u_{N-1}) \circ k(x_N)
\]
subject to
\[
\begin{align*}
(i) & \quad f(x_n, u_n) = x_{n+1} \\ (ii) & \quad u_n \in U \quad n = 0, 1, \ldots, N - 1.
\end{align*}
\]
(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 dynamic 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:
\[
\text{Maximize } \quad E[r_0(x_0, u_0) \circ r_1(x_1, u_1) \circ \cdots \circ r_{N-1}(x_{N-1}, u_{N-1}) \circ k(x_N)]
\]
subject to
\[
\begin{align*}
(i) & \quad x_{n+1} \sim p(\cdot | x_n, u_n) \\ (ii) & \quad u_n \in U \quad n = 0, 1, \ldots, N - 1.
\end{align*}
\]
(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 turn 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


