J. Operations Research Soc. of Japan
Vol. 18, No. 3 & No. 4, September 1975

LINEAR PROGRAMMING ON RECURSIVE ADDITIVE DYNAMIC PROGRAMMING

SEIICHI IWAMOTO

Kyushu University

(Received October 26, 1973; Revised March 19, 1974)

Abstract

We study, by using linear programming (LP), an infinite-horizon stochastic dynamic programming (DP) problem with the recursive additive reward system. Since this DP problem has discount factors which may depend on the transition, it includes the "discounted" Markovian decision problem. It is shown that this problem can also be formulated as one of LP problems and that the optimal stationary policy can be obtained by the usual LP method. Some interesting examples of DP models and their mumerical solutions by LP algorithm are illustrated. Furthermore, it is verified that these solutions coincides with ones obtained by Howard's policy iteration algorithm.

1. Introduction

We are concerned a certain class of the discrete, stochastic and infinite-horizon DP's. In general DP problems, the word "reward" or "return" is to be understood in a very broad sense ; it is not limited to any particular economic connotation (see [1 ; pp.74]). In some cases, for example, in the fields of engineering we shall be concerned the maximizing some sort of summation of reward [2 ; pp.58, 59, 102]. From this view point, Nemhauser [9 ; Chap.II-IV] introduced the deterministic DP's with recursive (not necessarily additive) return. In this paper we use the "reward system" (RS) in stead of the "return". He also treated the stochastic DP's. But their RS is restricted to only additive or multiplicative one [9 ; pp. 152-158]. Furukawa and Iwamoto [5] have extended the continuous stochastic DP's into ones with recursive (including additive and multiplicative) RS.

In 1960, Howard [6] established the policy iteration algorithm (PIA) for the discrete stochastic DP with the discounted additive RS. Recently, the author [7] proved that Howard's PIA remains valid for the discrete stochastic DP with the recursive additive (including the discounted additive but being included by recursive) RS. This DP is a discrete, stochastic and infinite-horizon version of examples in [2 ; pp.58, 59, 102].

On the other hand, Manne [8] originated an approach to Markovian decision problems by LP method. Since then, LP approach has been used in order to find optimal policies for discounted Markovian, average Markovian or semi-Markovian decision problems by D'Epenoux [4], De Ghellinck and Eppen [3] and Osaki and Mine [10, 11].

In this paper we shall discuss DP with recursive additive RS (hereafter abbreviated as "recursive additive DP") by LP method. In section 2, we describe this DP and give some preliminary notations and definitions used throughout this paper. In section 3, we give a formulation of this DP problem into a LP problem and show a correspondence between solutions of two problems. Section 4 is devoted to illustrate numerical examples by LP. It is shown that the optimal solution by LP algorithm is the same as one by the algorithm in [7]. Further comments are given in the last section. The method used in our proofs of results is mainly due to that of [3].

2. Description of recursive additive DP

A recursive additive DP is defined by six-tuple $\{S, A, p, r, \beta, t\}$. $S = \{1, 2, \ldots, N\}$ is a set of states, $A = (A_1, A_2, \ldots, A_N)$ is an N-tuple, each $A_i = \{1, 2, \ldots, K_i\}$ is a set of actions available at state $i \in S$, $p = (p_{ij}^k)$ is a

transition law, that is,

$$\sum_{j=1}^{N} p_{ij}^{k} = 1, \quad p_{ij}^{k} \geqslant 0, \qquad i \in S, \ j \in S, \ k \in A_i,$$

$r = (r_{ij}^{k})$, $i, j \in S$, $k \in A_i$ is a set of stage-wise reward, $\beta = (\beta_{ij}^{k})$, $i, j \in S$, $k \in A_i$ is a generalized accumulator whose element β_{ij}^{k} is a discount factor depending on transition (i, k, j), and t is a translator from R^1 to R^1.

Throughout this paper we call the recursive additive DP defined by $\{S, A, p, r, \beta, t\}$ simply "recursive additive DP". We sometimes use the convenient notations $\beta(i, k, j)$, $r(i, k, j)$ and $p(i, k, j)$ in stead of β_{ij}^{k}, r_{ij}^{k} and p_{ij}^{k} respectively.

When the system starts from an initial state $s_1 \in S$ at the 1-st stage and the decision maker takes an action $a_1 \in A_{s_1}$ on this state s_1, the system moves to the next state $s_2 \in S$ with probability $p(s_1, a_1, s_2)$ at the 2-nd stage and it yields a stage-wise reward $r(s_1, a_1, s_2)$ and a discount factor $\beta(s_1, a_1, s_2)$. However, at the end of the 1-st stage the decision maker obtains the translated reward $t(r(s_1, a_1, s_2))$. The system is then repeated from the new state $s_2 \in S$ at the 2-nd stage. If he chooses an action $a_2 \in A_{s_2}$ on state s_2, it moves to state s_3 with probability $p(s_2, a_2, s_3)$ at the 3-rd stage. Then the system also yields a stage-wise reward $r(s_2, a_2, s_3)$ and a discount factor $\beta(s_2, a_2, s_3)$ at the end

of the 2-nd stage and he really receives the discounted reward $\beta(s_1, a_1, s_2) \cdot t(r(s_2, a_2, s_3))$. Similarly at the end of the 3-rd stage he gets a reward $\beta(s_1, a_1, s_2)\, \beta(s_2, a_2, s_3)\, t(r(s_3, a_3, s_4))$. In general when he undergoes the history $(s_1, a_1, s_2, a_2, \cdots, s_n, a_n, s_{n+1})$ of the system up to the n-th stage, he is to receive a reward $\beta(s_1, a_1, s_2)\, \beta(s_2, a_2, s_3) \cdots \beta(s_{n-1}, a_{n-1}, s_n)\, t(r(s_n, a_n, s_{n+1}))$ at the end of the n-th stage.

Furthermore, the process goes on the (n+1)-st stage, the (n+2)-nd stage and so on.

Since we are considering a sequential nonterminating decision process, the decision maker continues to take actions infinitely. Consequently if he undergoes the history $h = (s_1, a_1, s_2, a_2, \ldots)$, he is to receive the recursive additive reward

$$V(h) = t(r(s_1, a_1, s_2)) + \beta(s_1, a_1, s_2)t(r(s_2, a_2, s_3))$$

$$+ \beta(s_1, a_1, s_2)\, \beta(s_2, a_2, s_3)t(r(s_3, a_3, s_4))$$

$$+ \cdots + \beta(s_1, a_1, s_2)\, \beta(s_2, a_2, s_3) \cdots \beta(s_{n-1},$$

$$a_{n-1}, s_n)t(r(s_n, a_n, s_{n+1})) + \cdots .$$

We call $V = V(h)$ recursive additive RS([7]).

The decision maker wishes to maximize his expected reward

over the infinite future.

We are assumed that he has a complete information on his history consisted of states and actions up to date and that he knows not only the stage-wise reward $r = (r_{ij}^k)$, its translator $t = t(\cdot)$ and the generalized accumulator $\beta = (\beta_{ij}^k)$ but also the recursive additivity of RS.

Let for integer $m \geq 1$ $\Delta_m = \{(p_1, p_2, \cdots, p_m) ; \sum_{i=1}^{m} p_i = 1, p_1 \geq 0, p_2 \geq 0, \cdots, p_m \geq 0\}$. We say a sequence $\pi = \{f_1, f_2, \cdots\}$ randomized policy if $f_n(i) \in \Delta_{K_i}$ for all $i \in S$, $n \geq 1$. Then we write $f_n(i)$ as a stochastic vector $f_n(i) = (f_n^1(i), f_n^2(i), \cdots, f_n^{K_i}(i))$ for $i \in S$, $n \geq 1$.
Using randomized policy $\pi = \{f_1, f_2, \cdots\}$ means that the decision maker chooses action $k \in A_i$ with probability $f_n^k(i)$ in state $i \in S$ at n-th stage.
A stationary randomized policy (S-randomized policy) is the randomized policy $\pi = \{f_1, f_2, \cdots\}$ such that $f_1 = f_2 = \cdots = f$. Such a S-randomized policy is denoted by $\pi = f^{(\infty)}$. The randomized policy $\pi = \{f_1, f_2, \cdots\}$ is called nonrandomized if for each $n \geq 1$ and $i \in S$ $f_n(i)$ is degenerate at some $k \in A_i$, that is, $f_n(i) = (0, 0, \ldots, \underbrace{1}_{k\text{-th}}, 0, \ldots, 0)$.

We associate with each f such that $f(i) = (f^1(i), f^2(i), \ldots, f^{K_i}(i)) \in \Delta_{K_i}$ for $i \in S$ (i) the $N \times 1$ column vector $\bar{r}(f)$ whose i-th element $\bar{r}(f)(i)$ is

$$\bar{r}(f)(i) = \sum_{k \in A_i} \sum_{j \in S} p_{ij}^k t(r_{ij}^k) f^k(i), \qquad i \in S,$$

and (ii) the $N\times N$ matrix $\bar{P}(f)$ whose (i,j) element $\bar{P}(i,j)$ is

$$\bar{P}(f)(i,j) = \sum_{k\in A_i} p_{ij}^k \beta_{ij}^k f^k(i), \qquad i,j \in S.$$

If the decision maker uses a randomized policy $\pi = \{f_1, f_2, \cdots\}$ and the system starts in $i \in S$ at 1-st stage, his recursive additive expected reward from π is the column vector

$$V(\pi) = \sum_{n=0}^{\infty} \bar{P}_n(\pi)\bar{r}(f_{n+1}),$$

where $\bar{P}_0(\pi) = I$, the $N\times N$ identity matrix, and for $n \geq 1$

$$\bar{P}_n(\pi) = \bar{P}(f_1)\bar{P}(f_2)\cdots\bar{P}(f_n).$$

That is, i-th element of $V(\pi)$ is

$$V(\pi)(i) = \bar{r}(f_1)(i) + \sum_{k\in A_i, j\in S} p_{ij}^k \beta_{ij}^k f_1^k(i)\bar{r}(f_2)(j)$$

$$+ \sum_{k\in A_i,\ j\in S,\ m\in A_j,\ l\in S} p_{ij}^k p_{jl}^m \beta_{ij}^k \beta_{jl}^m f_1^k(i) f_2^m(j)\bar{r}(f_3)(l) +$$

$$\cdots + \sum_{k\in A_i, j\in S, m\in A_j, l\in S, \ldots, t\in A_r, s\in S} p_{ij}^k p_{jl}^m \cdots p_{rs}^t \beta_{ij}^k \beta_{jl}^m$$

$$\cdots \beta_{rs}^t f_1^k(i) f_2^m(j) \cdots f_n^t(r)\bar{r}(f_{n+1})(s) + \cdots$$

3. Formulation and algorithm by LP

Let $\{S, A, p, r, \beta, t\}$ be a fixed recursive additive

DP defined at section 2, and $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_N)$ a fixed initial (at 1-st stage) distribution of state, that is,

$$\sum_{i=1}^{N} \alpha_i = 1, \quad \alpha_i \geq 0, \qquad i=1,2,\ldots, N.$$

Let $\{\mu_i^k(n) \; ; \; n \geq 1, \; k \in A_i, \; i \in S\}$ be any set of nonnegative numbers satisfying the recursive relation ;

$$(1) \qquad \sum_{\ell \in A_j} \mu_j^{\ell}(n) = \begin{cases} \alpha_j, & n=1, \; j \in S, \\ \sum_{i \in S} \sum_{k \in A_i} p_{ij}^k p_{ij}^k \mu_i^k(n-1), & n \geq 2, \; j \in S. \end{cases}$$

In the remainder of this paper we shall assume the following assumption :

ASSUMPTION (I). $0 \leq p_{ij}^k < 1$ for any $i, j \in S$, $k \in A_i$.

LEMMA 3.1. Under the Assumption (I), any nonnegative $\{\mu_i^k(n) \; ; \; n \geq 1, \; k \in A_i, \; i \in S\}$ satisfying (1) has the following properties :

$$(i) \qquad \sum_{i \in S} \sum_{k \in A_i} \mu_i^k(1) = 1, \; \sum_{k \in A_i} \mu_i^k \geq 0, \qquad i \in S,$$

$$(ii) \qquad \beta_*^{n-1} \leq \sum_{i \in S} \sum_{k \in A_i} \mu_i^k(n) \leq \beta^{*n-1}, \quad n \geq 2.$$

Therefore, we have

$$\frac{r_*}{1-\beta_*} \leq \sum_{n\geq 1} \sum_{i\in S} \sum_{k\in A_i} \sum_{j\in S} p_{ij}^k t(r_{ij}^k)\mu_i^k(n) \leq \frac{r^*}{1-\beta^*} ,$$

where $r_* = \min_{i,j\in S,\ k\in A_i} t(r_{ij}^k)$, $r^* = \max_{i,j\in S,\ k\in A_i} t(r_{ij}^k)$,

$\beta_* = \min_{i,j\in S, k\in A_i} \beta_{ij}^k$ and $\beta^* = \max_{i,j\in S, k\in A_i} \beta_{ij}^k$.

PROOF. Property (i) is a trivial consequence. Property (ii) is to be proved by induction on n.

LEMMA 3.2. Under the Assumptipn (I), (i) any randomized policy $\pi = \{f_1, f_2, \cdots\}$ gives a nonnegative solution $\{\mu_i^k(n)\}$ of (1) and vice versa, and, furthermore,

(ii) $$\sum_{i\in S} \alpha_i V(\pi)(i) = \sum_{n=1}^{\infty} \sum_{i\in S} \sum_{k\in A_i} \bar{r}_i^k \mu_i^k(n),$$

where $\bar{r}_i^k = \sum_{j\in S} p_{ij}^k t(r_{ij}^k)$.

PROOF. Let $\pi = \{f_1, f_2, \cdots\}$ be any randomized policy. Then we can give a nonnegative $\mu_j^l(n)$ for $n\geq 1$, $l\in A_j$, $j\in S$ as follows ;

(1)'
$$\begin{cases} \mu_j^l(1) = \alpha_j f_1^l(j), & l\in A_j,\ j\in S, \\ \mu_j^l(n+1) = \sum_{i\in S} \sum_{k\in A_i} p_{ij}^k \beta_{ij}^k \mu_i^k(n) f_{n+1}^l(j), & n\geq 1, \\ & l\in A_j,\ j\in S. \end{cases}$$

Obviously, these $\{\mu_j^l(n)\ ;\ n\geq 1,\ l\in A_j,\ j\in S\}$ satisfy (1).

Conversely, let nonnegative $\{\mu_i^k(n)\}$ satisfy (1). Then, we can define f_n as follows ;

$$\begin{cases} f_1^k(i) = \dfrac{\mu_i^k(1)}{\alpha_i}, & n=1,\ k\in A_i,\ i\in S, \\ f_n^l(j) = \dfrac{\mu_j^l(n)}{\sum_{i\in S}\sum_{k\in A_i} \beta_{ij}^k p_{ij}^k \mu_i^k(n-1)}, & n\geq 2,\ l\in A_j,\ j\in S, \end{cases}$$

where $\frac{0}{0} = 0$. Then the policy $\pi = \{f_1, f_2, \cdots\}$ is a randomized policy. Moreover, we have, by using (1)' and exchanging the summation,

$$\sum_{i\in S}\sum_{k\in A_i} \bar{r}_i^k \mu_i^k(n+1)$$

$$= \sum_{i\in S} \alpha_i \sum_{k\in A_i,\ j\in S,\ m\in A_j,\ l\in S,\ \cdots,\ t\in A_r,\ s\in S} p_{ij}^k p_{jl}^m \cdots$$

$$p_{rs}^t \beta_{ij}^k \beta_{jl}^m \cdots \beta_{rs}^t f_1^k(i) f_2^k(j) \cdots f_n^t(r) \bar{r}(f_{n+1})(s), \quad n \geq 0.$$

Hence (ii) holds. This completes the proof.

We note that $\sum_{n=1}^{\infty} \sum_{i\in S} \sum_{k\in K_i} \bar{r}_i^k \mu_i^k(n)$

is the total expected recursive additive reward obtained from the randomized policy $\pi = \{f_1, f_2, \cdots\}$ corresponding $\{\mu_i^k(n)\}$, started in the initial distribution α.

Consequently, above lemmas and note enable us to give a maximization problem (P_0) :

Problem (P_0) : Maximize $\sum_{n=1}^{\infty} \sum_{i\in S} \sum_{k\in A_i} \bar{r}_i^k \mu_i^k(n)$

under

$$(1) \quad \sum_{l\in A_j} \mu_j^l(n) = \begin{cases} \alpha_j, & n=1, \\ \sum_{i\in S}\sum_{k\in A_i} \beta_{ij}^k p_{ij}^k \mu_i^k(n-1), & n\geq 2,\ j\in S, \end{cases}$$

(2) $\mu_i^k(n) \geqslant 0, \qquad n \geqslant 1,\ k \in A_i,\ i \in S.$

By Lemma 3.1, we can define a set of the new variables $\{y_i^k\}$ as follows :

$$y_i^k = \sum_{n=1}^{\infty} \mu_i^k(n), \qquad k \in A_i,\ i \in S.$$

Hence, we have a modified maximization problem (P_T) :

Problem (P_T) : Maximize

(3) $$\sum_{i \in S} \sum_{k \in A_i} \bar{r}_i^k y_i^k$$

under

(4) $$\sum_{\ell \in A_j} y_j^{\ell} - \sum_{i \in S} \sum_{k \in A_i} \beta_{ij}^k p_{ij}^k y_i^k = \alpha_j, \quad j \in S,$$

(5) $y_i^k \geqslant 0, \quad k \in A_i,\ i \in S.$

Next lemma states the relationship between Problem (P_0) and Problem (P_T).

LEMMA 3.3. If $\{\mu_i^k(n)\}$ is a nonnegative solution of (1), then $\{y_i^k\}$ is a solution of Problem (P_T), and $\sum_{i \in S} \sum_{k \in A_i} \bar{r}_i^k y_i^k$ is the expected recursive additive reward which corresponds to $\{\mu_i^k(n)\}$.

PROOF. It is easy to show that $\{y_i^k\}$ satisfies (4) and (5).

We can define a S-nonrandomized policy $\pi = f^{(\infty)}$ by a

function f such that for each $i \in S$ selects exactly one variable y_i^k $k \in A_i$. This fact is easy to check.

THEOREM 3.1. Let Assumption (I) be satisfied. If the equation (4) is restricted to the variables y_i^k selected by any S-nonrandomized policy, then : (i) the corresponding subsystem has a unique solution,

(ii) if $\alpha_i \geqslant 0$ $i \in S$, then $y_i^k \geqslant 0$ $i \in S$,

(iii) if $\alpha_i > 0$ $i \in S$, then $y_i^k > 0$ $i \in S$.

PROOF. This theorem corresponds to Proposition 2.3 in [3] which treated the case of $\beta_{ij}^k \equiv \beta$. The proof is similar to that of Proposition 2.3.

LEMMA 3.4. Let Assumption (I) be satisfied and $\alpha_i > 0$ for $i \in S$. Then there exists an one to one correspondence between S-nonrandomized policies and basic feasible solutions of (4), (5). Moreover, any basic feasible solution is nondegenerate.

PROOF. The proof follows in the same way as in Proposition 2.4 of [3], and the details are omitted.

Lemma 3.4 yields the following definition of optimality.

A S-nonrandomized policy $\pi = f^{(\infty)}$ is optimal if its corresponding basic feasible solution is optimal.

THEOREM 3.2. Let Assumption (I) be satisfied. Whenever $\alpha_i > 0$ for $i \in S$, the Problem (P_T) has an optimal basic solution and its dual problem has a unique optimal solution. Any optimal S-policy associated with it remains optimal for any $(\alpha_1, \alpha_2, \cdots, \alpha_N)$ such that $\alpha_i \geqslant 0$ for $i \in S$.

PROOF. The proof is similar to that of Proposition 3.5 of [3], and the details are omitted.

COROLLARY For $\alpha_i > 0$ for $i \in S$ (say $\alpha_i = \frac{1}{N}$, $i \in S$) there exists an optimal basic solution such that for each $i \in S$ there is exactly one k such that $y_i^k > 0$ and $y_i^k = 0$ for k otherwise.

PROOF. This is a straightforward from Lemma 3.4 and Theorem 3.2.

4. Numerical examples

We now illustrate correspondence between the optimal solution by PIA and the optimal solution by LP algorithm. As for the definition, reward system and optimal solution by PIA of the following DP's, see the corresponding example in [5].

EXAMPLE 1 (General Additive DP)

In the general additive DP $\{ S, A, p, r, \beta \}$, the objective function is the expected value of the general additive RS

$$V(h) = r_1 + \beta_1 r_2 + \beta_1\beta_2 r_3 + \cdots + \beta_1\beta_2 \cdots \beta_{n-1} r_n + \cdots,$$

since this is the case where $t(r) = r$ in the recursive additive DP$\{S, A, p, r, \beta, t\}$.

Following data is a slightly modified one from Howard [4]. Of course Assumption (I) is satisfied.

TABLE 4.1.
Data for general additive DP

state	action	transition probability			stage-wise reward			generalized accumulator		
i	k	p_{i1}^k	p_{i2}^k	p_{i3}^k	r_{i1}^k	r_{i2}^k	r_{i3}^k	β_{i1}^k	β_{i2}^k	β_{i3}^k
1	1	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$	10	4	8	.95	.98	.98
	2	$\frac{1}{16}$	$\frac{3}{4}$	$\frac{3}{16}$	8	2	4	.90	.90	.93
	3	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{5}{8}$	4	6	4	.98	.96	.98
2	1	$\frac{1}{2}$	0	$\frac{1}{2}$	14	0	18	.85	.90	.95
	2	$\frac{1}{16}$	$\frac{7}{8}$	$\frac{1}{16}$	6	16	8	.80	.80	.95
	3	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	-5	-5	-5	.95	.95	.95
3	1	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$	10	2	8	.75	.90	.95
	2	$\frac{1}{8}$	$\frac{3}{4}$	$\frac{1}{8}$	6	4	2	.95	.70	.80
	3	$\frac{3}{4}$	$\frac{1}{16}$	$\frac{3}{16}$	4	0	8	.95	.95	.95

Then PIA yields an optimal S-policy $f^{(\infty)}$, where $f = \begin{Bmatrix} 1 \\ 1 \\ 3 \end{Bmatrix}$ and an optimal

return $V(f^{(\infty)}) = \begin{pmatrix} 169. & 490 \\ 166. & 129 \\ 164. & 411 \end{pmatrix}$.

On the other hand, for an initial vector $\alpha = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ the LP Problem (P_T) becomes :

$$\text{Maximize } 8y_1^1 + \frac{11}{4}y_1^2 + \frac{17}{4}y_1^3 + 16y_2^1 + \frac{31}{2}y_2^2 + (15)y_2^3 + 7y_3^1 + 4y_3^2 + \frac{9}{2}y_3^3$$

subject to

$$\frac{105}{200}y_1^1 + \frac{1510}{1600}y_1^2 + \frac{302}{400}y_1^3 - \frac{85}{200}y_2^1 - \frac{80}{1600}y_2^2 - \frac{95}{300}y_2^3 - \frac{75}{400}y_3^1 - \frac{95}{800}y_3^2 - \frac{285}{400}y_3^3 = \frac{1}{3},$$

$$-\frac{98}{400}y_1^1 - \frac{270}{400}y_1^2 - \frac{96}{800}y_1^3 + y_2^1 + \frac{30}{100}y_2^2 + \frac{205}{300}y_2^3 - \frac{90}{400}y_3^1 - \frac{210}{400}y_3^2 - \frac{95}{1600}y_3^3 = \frac{1}{3},$$

$$-\frac{98}{400}y_1^1 - \frac{279}{1600}y_1^2 - \frac{490}{800}y_1^3 - \frac{95}{200}y_2^1 - \frac{95}{1600}y_2^2 - \frac{95}{300}y_2^3 + \frac{105}{200}y_3^1 + \frac{90}{100}y_3^2 + \frac{1315}{1600}y_3^3 = \frac{1}{3},$$

$y_1^1, y_1^2, y_1^3, y_2^1, y_2^2, y_2^3, y_3^1, y_3^2, y_3^3 \geqq 0.$

The optimal solution of this LP problem is

$$(y_1^1, y_1^2, y_1^3, y_2^1, y_2^2, y_2^3, y_3^1, y_3^2, y_3^3)$$

$$= (10.9688, 0.0, 0.0, 3.3540, 0.0, 0.0, 0.0, 0.0.5.6138)$$

and its (optimal) value of the objective function is 166.6768. Note that this value is nearly equal to

$$\alpha V(f^{(\infty)}) = (\tfrac{1}{3}, \tfrac{1}{3}, \tfrac{1}{3}) \begin{pmatrix} 169.490 \\ 166.129 \\ 164.411 \end{pmatrix} = 166.6767.$$

Furthermore this optimal solution shows that $f = \begin{bmatrix} 1 \\ 1 \\ 3 \end{bmatrix}$ is optimal.

EXAMPLE 2(Multiplicative Additive DP)

The multiplicative additive DP $\{S, A, p, r\}$ is the case where $\beta_{ij}^k \equiv r_{ij}^k$, $t(r) = r$ in the recursive additive DP. Then, the objective function of this DP $\{S, A, p, r\}$ is the expected value of the multiplicative additive RS

$$V(h) = r_1 = r_1 r_2 + r_1 r_2 r_3 + \cdots + r_1 r_2 \cdots r_n + \cdots.$$

The following data satisfies Assumption (I).

TABLE 4.2.

Data for multiplicative additive DP

state	action	transition probability			state-wise reward		
i	k	p_{i1}^k	p_{i2}^k	p_{i3}^k	r_{i1}^k	r_{i2}^k	r_{i3}^k
1	1	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{5}$	$\frac{2}{5}$
	2	$\frac{1}{16}$	$\frac{3}{4}$	$\frac{3}{16}$	$\frac{2}{5}$	$\frac{1}{10}$	$\frac{1}{5}$
2	1	$\frac{1}{2}$	0	$\frac{1}{2}$	$\frac{7}{10}$	$\frac{1}{20}$	$\frac{9}{10}$
	2	$\frac{1}{16}$	$\frac{7}{8}$	$\frac{1}{16}$	$\frac{2}{5}$	$\frac{4}{5}$	$\frac{2}{5}$
	3	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{20}$	$\frac{1}{20}$	$\frac{1}{20}$
3	1	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{10}$	$\frac{2}{5}$
	2	$\frac{1}{8}$	$\frac{3}{4}$	$\frac{1}{8}$	$\frac{3}{10}$	$\frac{1}{5}$	$\frac{1}{10}$
	3	$\frac{3}{4}$	$\frac{1}{16}$	$\frac{3}{16}$	$\frac{1}{5}$	$\frac{1}{20}$	$\frac{2}{5}$

Then, by PIA, we have an optimal S-policy $f^{(\infty)}$, where $f = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$, and the optimal return

$$V(f^{(\infty)}) = \begin{pmatrix} 0.7938 \\ 2.6198 \\ 0.6434 \end{pmatrix}.$$

The LP problem (P_T) for $\alpha = (\frac{1}{4}, \frac{1}{4}, \frac{1}{2})$ has an optimal solution $(y_1^1, y_1^2, y_1^3, y_2^1, y_2^2, y_2^3, y_3^1, y_3^2, y_3^3)$ =(0.4851, 0.0,0.0,0.0,0.9739,0.0,0.7161,0.0,0.0) and an optimal value 1.1751. Note that this optimal value is equal

$$\alpha V(f^{(\infty)}) = (\tfrac{1}{4}, \tfrac{1}{4}, \tfrac{1}{2}) \begin{pmatrix} 0.7938 \\ 2.6198 \\ 0.6434 \end{pmatrix}.$$

$$= 1.1751.$$

EXAMPLE 3 (Divided Additive DP)

The divided additive DP$\{S, A, p, r\}$ has the divided additive RS

$$V(h) = r_1 + \frac{r_2}{r_1} + \frac{r_3}{r_1 r_2} + \cdots + \frac{r_n}{r_1 r_2 \cdots r_{n-1}} + \cdots,$$

since this is the case where $\beta_{ij}^k \equiv 1/r_{ij}^k$, $t(r) = r$ in the recursive additive DP. We can illustrate a DP with $\beta_{ij}^k \equiv 1/r_{ij}^k$, $r_{ij}^k \equiv k$, $t(r)=r^b$ $(b>0)$ in [2;pp.58]. This DP has continuous state-action spaces, deterministic transition law and finite horizon. In the divided additive DP Assumption (I) means $r_{ij}^k > 1$ for $i \in S$, $k \in A_i$, $j \in S$, which is satisfied by the following data.

TABLE 4.3.
Data for divided additive DP

state	action	transition probability			stage-wise reward		
i	k	p_{i1}^k	p_{i2}^k	p_{i3}^k	r_{i1}^k	r_{i2}^k	r_{i3}^k
1	1	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{3}{2}$	$\frac{6}{5}$	$\frac{7}{5}$
	2	$\frac{1}{16}$	$\frac{3}{4}$	$\frac{3}{16}$	$\frac{7}{5}$	$\frac{11}{10}$	$\frac{6}{5}$
	3	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{5}{8}$	$\frac{6}{5}$	$\frac{13}{10}$	$\frac{6}{5}$
2	1	$\frac{1}{2}$	0	$\frac{1}{2}$	$\frac{17}{10}$	$\frac{21}{20}$	$\frac{19}{10}$
	2	$\frac{1}{16}$	$\frac{7}{8}$	$\frac{1}{16}$	$\frac{7}{5}$	$\frac{9}{5}$	$\frac{26}{25}$
	3	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{21}{20}$	$\frac{21}{20}$	$\frac{21}{20}$
3	1	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{3}{2}$	$\frac{11}{10}$	$\frac{7}{5}$
	2	$\frac{1}{8}$	$\frac{3}{4}$	$\frac{1}{8}$	$\frac{13}{10}$	$\frac{6}{5}$	$\frac{11}{10}$
	3	$\frac{3}{4}$	$\frac{1}{16}$	$\frac{3}{16}$	$\frac{6}{5}$	$\frac{21}{20}$	$\frac{7}{5}$

Then optimal S-policy is specified by $f = \begin{bmatrix} 2 \\ 3 \\ 2 \end{bmatrix}$,

and optimal return is $V(f^{(\infty)}) = \begin{pmatrix} 11.8020 \\ 12.2804 \\ 11.2934 \end{pmatrix}$.

If process strates at initial distribution $\alpha = (\frac{1}{5}, \frac{2}{5}, \frac{2}{5})$, the LP algorithm yields an optimal solution $(y_1^1, y_1^2, y_1^3, y_2^1, y_2^2, y_2^3, y_3^1, y_3^2, y_3^3) = (0.0, 2.3176, 0.0, 0.0, 0.0, 5.4885, 0.0, 2.8256, 0.0)$ and an optimal value 11.7899. Note that

$$\alpha V(f^{(\infty)}) = (\frac{1}{5}, \frac{2}{5}, \frac{2}{5}) \begin{pmatrix} 11.8020 \\ 12.2804 \\ 11.2934 \end{pmatrix}$$

$$= 11.7899.$$

EXAMPLE 4 (Exponential Additive DP)

The exponential additive DP $\{S, A, p, r\}$ has the exponential additive RS

$$V(h) = r_1 + e^{r_1} \cdot r_2 + e^{r_1 + r_2} \cdot r_3 + \cdots + e^{r_1 + r_2 + \cdots + r_{n-1}} \cdot r_n + \cdots,$$

since this is the case where $\beta_{ij}^k \equiv e^{r_{ij}^k}$, $t(r) = r$ in the recursive additive DP. We have a DP with $\beta_{ij}^k \equiv e^{r_{ij}^k}$, $t(r) = (1-r)e^r$ [2; pp.102]. But this DP has continuous action space, deterministic transition law and finite horizon. If $r_{ij}^k < 0$ for $i \in S$, $k \in A_i$, $j \in S$ then the exponential additive DP satisfies Assumption (I). The following data satisfies Assumption (I).

TABLE 4.4.

Data for exponential additive DP

state	action	transition probability			stage-wise reward		
i	k	p_{i1}^k	p_{i2}^k	p_{i3}^k	r_{i1}^k	r_{i2}^k	r_{i3}^k
1	1	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$	$-\frac{1}{2}$	$-\frac{1}{5}$	$-\frac{2}{5}$
	2	$\frac{1}{16}$	$\frac{3}{4}$	$\frac{3}{16}$	$-\frac{2}{5}$	$-\frac{1}{10}$	$-\frac{1}{5}$
	3	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{5}{8}$	$-\frac{1}{5}$	$-\frac{3}{10}$	$-\frac{1}{5}$
2	1	$\frac{1}{2}$	0	$\frac{1}{2}$	$-\frac{7}{10}$	$-\frac{1}{20}$	$-\frac{9}{10}$
	2	$\frac{1}{16}$	$\frac{7}{8}$	$\frac{1}{16}$	$-\frac{2}{5}$	$-\frac{4}{5}$	$-\frac{2}{5}$
	3	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	$-\frac{1}{20}$	$-\frac{1}{20}$	$-\frac{1}{20}$
3	1	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$	$-\frac{1}{2}$	$-\frac{1}{10}$	$-\frac{2}{5}$
	2	$\frac{1}{8}$	$\frac{3}{4}$	$\frac{1}{8}$	$-\frac{3}{10}$	$-\frac{1}{5}$	$-\frac{1}{10}$
	3	$\frac{3}{4}$	$\frac{1}{16}$	$\frac{3}{16}$	$-\frac{1}{5}$	$-\frac{1}{20}$	$-\frac{2}{5}$

We have optimal stationary policy $f^{(\infty)}$, where $f = \begin{bmatrix} 2 \\ 3 \\ 2 \end{bmatrix}$

and optimal return $V(f^{(\infty)}) = \begin{pmatrix} -1.0831 \\ -1.0807 \\ -1.0867 \end{pmatrix}$.

If $\alpha = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$, then the LP problem (P_T) yields an optimal solution $(y_1^1, y_1^2, y_1^3, y_2^1, y_2^2, y_2^3, y_3^1, y_3^2, y_3^3)$ = (0.0, 2.2768,0.0,0.0,0.0,5.0739,0.0,2.5839,0.0) and an optimal value -1.0835. We can verify that

$$\alpha V(f^{(\infty)}) = (\tfrac{1}{3}, \tfrac{1}{3}, \tfrac{1}{3}) \begin{pmatrix} -1.0831 \\ -1.0807 \\ -1,0867 \end{pmatrix}$$

$$= -1.0835$$

coincides with optimal value.

EXAMPLE 5 (Logarithmic Additive DP)

This is the case where $\beta_{ij}^k \equiv \log r_{ij}^k$, $t(r) = r$ in the recursive additive DP $\{ S, A, p, r, \beta, t \}$. Then, the logarithmic additive RS is given as follows :

$$V(h) = r_1 + (\log r_1) \cdot r_2 + (\log r_1 \cdot \log r_2) r_3 + \cdots +$$

$$(\log r_1 \cdot \log r_2 \cdots \log r_{n-1}) r_n + \cdots .$$

In this DP Assumption (I) means that $1 \leq r_{ij}^k < e$ for $i \in S$, $k \in A_i$, $j \in S$. The following data satisfies Assumption (I).

TABLE 4.5.
Data for logarithmic additive DP

state	action	transition probability			stage-wise reward		
i	k	p_{i1}^k	p_{i2}^k	p_{i3}^k	r_{i1}^k	r_{i2}^k	r_{i3}^k
1	1	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$	2.3	2.7	2.4
	2	$\frac{1}{16}$	$\frac{3}{4}$	$\frac{3}{16}$	2.7	2.3	2.6
	3	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{5}{8}$	2.5	2.4	2.6
2	1	$\frac{1}{2}$	0	$\frac{1}{2}$	2.7	2.3	2.4
	2	$\frac{1}{16}$	$\frac{7}{8}$	$\frac{1}{16}$	2.6	2.4	2.7
	3	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	2.4	2.1	2.5
3	1	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$	2.6	2.5	2.7
	2	$\frac{1}{8}$	$\frac{3}{4}$	$\frac{1}{8}$	2.7	2.6	2.4
	3	$\frac{3}{4}$	$\frac{1}{16}$	$\frac{3}{16}$	2.6	2.7	2.5

Then optimal S-policy is $f^{(\infty)}$ and optimal return is

$$V(f^{(\infty)}) = \begin{pmatrix} 52.3188 \\ 52.0526 \\ 53.7307 \end{pmatrix}, \text{ where } \quad f = \begin{bmatrix} 3 \\ 1 \\ 1 \end{bmatrix} .$$

The LP problem (P_T) with an initial distribution $\alpha = (\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$ gives an optimal solution $(y_1^1, y_1^2, y_1^3, y_2^1, y_2^2, y_2^3, y_3^1, y_3^2, y_3^3) = (0.0, 0.0, 6.1654, 3.3892, 0.0, 0.0, 10.7585, 0.0, 0.0)$ and an optimal value 52.6052.

Note that this value is

$$\alpha V(f^{(\infty)}) = (\tfrac{1}{2}, \tfrac{1}{4}, \tfrac{1}{4}) \begin{pmatrix} 52.3188 \\ 52.0526 \\ 53.7307 \end{pmatrix} .$$

We remark that above five examples are the case $t(r)=r$ in the recursive additive DP $\{ S, A, p, r, \beta, t \}$. But we can treat, for example, the case where $t(r)=\frac{1}{r}$, $t(r)=e^r$, $t(r)=(1-r)e^r$, $t(r)=\log r$, etc., ([7]).

5. Further remarks

In this section we shall give some remarks on the recursive additive DP.

Let $\{ S, A, p, r, \beta, t \}$ be the recursive additive DP satisfying Assumption (I). We define DP $\{ \bar{S}, \bar{A}, \bar{p}, \bar{r} \}$ in which

$$\bar{S} = S \cup \{0\} , \quad 0 (\notin S) \text{ is a fictitious state,}$$

$$\bar{A} = (A_0, A_1, \cdots, A_N),\ A_0 = \{1\},$$

$$\bar{p}_{ij}^{k} = \begin{cases} 1, & i=0,\ k=1,\ j=0, \\ 1 - \sum_{j\in S} \beta_{ij}^{k} p_{ij}^{k}, & i\in S,\ k\in A_i,\ j=0, \\ \beta_{ij}^{k} p_{ij}^{k}, & i\in S,\ k\in A_i,\ j\in S, \end{cases}$$

and

$$\bar{r}_{ij}^{k} = \begin{cases} 0, & i=0, k=1, j=0, \\ t(r_{ij}^{k}), & i\in S, k\in A_i,\ j\in S. \end{cases}$$

Note that $\bar{P}(X_{n+1}=j \mid X_n = i,\ Y_n=k)=\bar{p}_{ij}^{k}$ for $i\in\bar{S}$, $k\in A_i$, $j\in\bar{S}$,

where $\bar{P}$ is a probability law associated with DP$\{\bar{S}, \bar{A}, \bar{p}, \bar{r}\}$, and X_n, $Y_n (n\geq 1)$ denote observed state and action at n-th stage. In other words, nonnegative $\mu_i^k(n)$ satisfying (1) is the joint probability of being in state $i\in\bar{S}$ and making decision $k\in A_i$ at the n-th stage regarding to above probability law $\bar{P}$.

Furthermore above $\{\bar{S}, \bar{A}, \bar{p}, \bar{r}\}$ gives DP with an absorved state $\{0\}$. We can also apply the LP method for DP$\{\bar{S}, \bar{A}, \bar{p}, \bar{r}\}$ as well as DP $\{S, A, p, r, \beta, t\}$ with Assumption (I). But it is rather difficut to get five examples in section 4 from the reduced DP $\{\bar{S}, \bar{A}, \bar{p}, \bar{r}\}$.

Acknowledgement

The author wishes to express his hearty thanks to Prof. N. Furukawa for his advices. He also thanks the referee for his various comments and suggestions for improving this paper.

References

[1] Aris, R., Discrete Dynamic Programming, Blaisdell, Publishing Company, New York Tront London, (1964).

[2] Bellman, R., Dynamic Programming, Princeton Univ. Press, Princeton, New Jersey, (1957).

[3] DeGhellinck, G.T. and Eppen, G.D., "Linear programming solutions for separable Markovian decision problems", Mangt. Sci., 13, 371-394, (1967).

[4] D'Epenoux, F., "A probabilistic production and inventory problem", Mangt. Sci., 10, 98-108 (1963).

[5] Furukawa, N. and Iwamoto, S., "Markovian decision processes with recursive reward functions", Bull. Math. Statist., 15, 3-4, 79-91, (1973).

[6] Howard, R.A., Dynamic Programming and Markov Processes, M.I.T. Press, Cambridge, Massachusetts, (1960).

[7] Iwamoto, S., "Discrete dynamic programming with recursive additive system", Bull. Math. Statist., 16, 1-2, 49-66, (1974).

[8] Manne, A.S., "Linear programming and sequential decisions", Mangt. Sci., 6, 259-267, (1960).

[9] Nemhauser, G.L., Introduction to Dynamic Programming, John Wiley and Sons, NewwYork London Sydney, (1966).

[10] Osaki, S. and Mine, H., "Linear programming algorithm for semi-Markovian decision processes", J. Math. Anal. Appl., 22, 356-381, (1968).

[11] Osaki, S. and Mine, H., "Some remarks on a Markovian decision problem with an absorbing state", J. Math. Anal. Appl., 23, 327-333, (1968).