Journal of the Operations Research
Society of Japan
Vol. 23, No. 2, June 1980

PIECEWISE LINEAR DYNAMIC PROGRAMS WITH APPLICATIONS

Katsushige Sawaki
Nanzan University

(Received May 22, 1978; Revised May 11, 1979)

Abstract This paper considers a special class of dynamic programs which satisfies the monotone and contraction assumptions. This class of dynamic programs is characterized by the piecewise linearity that the cost function is piecewise linear whenever the terminal cost function is piecewise linear. Partially observable Markov decision processes have this property.

An algorithm based on policy improvement is developed to construct ϵ-optimal policies and ϵ-optimal cost functions. This algorithm has the advantage of involving only linear functions. A numerical example is also presented.

1. Introduction

Blackwell [1], Denardo [4], Strauch [14] et al consider a general class of monotone contractive dynamic programs. In this paper we consider a special class of Denardo's dynamic programs which satisfies the monotone and contraction assumptions. This class of dynamic programs, called piecewise linear dynamic programs, has the property that only piecewise linear cost functions and piecewise constant policies are involved. A partially observable Markov decision process ([6],[11],[12],[13]) has this property.

Larson [9] develops a theory, as well as an algorithm, for a state increment dynamic programming which is applied to the continuous time model where the state dynamics is described by differential equations. The concept of "state increment" is similar to the one of simple partition in this paper in the sense that a convex polyhedral cell of a simple partition corresponds to a rectangular block of a state increment dynamic programming. Fox [7] studies

essentially finite-state dynamic programs to approximate denumerable state dynamic programs. Blackwell [1] and Strauch [14] also introduce the concept of essentially finiteness of the action space to approximate uncountable (Polish) state dynamic programs, but they neither mention how to constract an optimal policy and optimal cost, nor provide an algorithm. In this paper we show how to generate and construct ε-optimal costs and ε-optimal policies over simple partitions of which cells are convex polyhedrons. Furthermore, an algorithm is provided. It includes policy improvement and successive approximation as special cases. Its advantage is that we only have to involve linear equalities and linear inequalities.

In Section 2 piecewise linear dynamic programs with an abstract state space and finite action set over an infinite horizon will be discussed. Section 3 will introduce several examples having piecewise linearity. Model 1 is actually a basic mathematical model which includes Model 2, a partially observable Markov decision process as a special case. Model 3 is a machine maintenance model under uncertainty. Model 4 is rather a dynamic economic linear model. Section 4 explicitly develops the algorithm based on a modified policy improvement and a proof for the convergence. Section 5 provides a numerical example and concluding remarks.

2. Piecewise Linear Dynamic Programs

First, we shall formulate a general dynamic programming problem under the setting of Denardo [4]. Secondly, a piecewise linear dynamic program will be defined. It is a special class of general dynamic programs which satisfies the monotonicity and contraction assumptions.

The state space Ω is an arbitrary set of a real linear space. For each $x \in \Omega$ there is a set A_x of actions. Let Δ be the Cartesian product $\underset{x\in\Omega}{X} A_x$. An element $\delta \in \Delta$ is a policy. There is always an optimal stationary policy among

a general class of policies in a contractive monotone dynamic program by Denardo [4] or Blackwell [1]. It suffices to consider only the class of stationary policies. Let V be the set of all bounded real valued functions on Ω. An element of V is a cost function. V is a Banach space with the norm $||v|| = \sup_{x\in\Omega}|v(x)|$. For u, v $\in$ V we write $u \le v$ if $u(x) \le v(x)$ for all $x \in \Omega$. The loss function h is defined to be a mapping from $\bigcup_{x\in\Omega} \{x\} \times A_x \times V$ to a real number. Our objective function to be minimized is somehow ambiguous, unless that the loss function h is specified. In a Markov decision process, however, $h(x,a,v)$ can be written as $h(x,a,v) = c(x,a) + \beta\int_\Omega v(y)q(dy|x,a)$ where $c(x,a)$ is the immediate cost, β the discount factor and $q(\cdot|x,a)$ the transition probability on Ω given x and a. Therefore, note that the system dynamics as well as the objective function is concealed behind our formulation. Assume that the loss function satisfies the monotonicity and contraction assumptions, that is, for each $x \in \Omega$ and $a \in A_x$ $h(x,a,u) \le h(x,a,v)$ whenever $u \le v$ in V, and for some $\beta\in[0,1)$, $|h(x,a,u) - h(x,a,v)| \le \beta||u-v||$ for each $u,v \in V, x \in \Omega$ and $a \in A_x$. For $\delta \in \Delta$ define $U_\delta : V \to V$ by $(U_\delta v)(x) = h(x,\delta(x),v)$ for $v \in V$ and $x \in \Omega$. Assume that there is some $\bar{\delta} \in \Delta$ such that $U_{\bar{\delta}} v = \inf_{\delta\in\Delta} U_\delta v$. Also, define $U_*: V \to V$ by $U_* v = \inf_{\delta\in\Delta} U_\delta v$. If $\delta(x) = a$ for each $x \in \Omega$, then we write $U_a = U_\delta$. Denardo [4] verifies that U_* and U_δ are monotone contraction operators. By Banch's fixed point theorem, for each $\delta \in \Delta$ there is a unique $v^\delta \in V$ such that $U_\delta v^\delta = v^\delta$. Similarly there is $v^* \in V$ such that $U_* v^* = v^*$. Such v^δ and v^* are called the cost of the policy δ and the optimal cost, respectively. Denardo [4] shows that $v^* = \inf_{\delta\in\Delta} v^\delta$. If $||v^\delta - v^*|| \le \varepsilon$, then δ is an ε-optimal policy, and if $||v-v^*|| \le \varepsilon$, then v is an ε-optimal cost function. Our purpose is to find such ε-optimal policy and ε-optimal cost function.

Any set of the form $\{x\in\Omega : \ell_{ij}(x) < (\text{or } \le)\ d_j, j=1,2,\ldots,n_i\}, i=1,2,\ldots,m$, where ℓ_{ij} is a linear functional and d_j a real number is called a convex polyhedron. A collection $P = \{E_1,E_2,\ldots,E_m\}$ of subsets of Ω is a partition if $E_i \cap E_j = \phi$ for $i \ne j$ and $\bigcup_{i=1}^m E_i = \Omega$. Each member of a partition is a cell.

m is the number of cells in the partition. If each cell of a partition is a convex polyhedron, then the partition is called simple. The product of two partitions P^1 and P^2 is $P^1 \cdot P^2 = \{E \cap D : E \in P^1, D \in P^2\}$. The product of $P^1 \cdot P^2 \cdot \cdots \cdot P^m$ is defined inductively by $\prod_{i=1}^{m} P^i = P^m \cdot \prod_{i=1}^{m-1} P^i$. Plainly, the finite product of simple partitions is again simple. A vector valued function v on Ω is piecewise linear (abbreviated, hereafter, by p.w.) if there exists a simple partition $\{E_1, E_2, \ldots, E_m\}$ of Ω and m linear functions $v_1, v_2, \ldots, v_m$ such that $v(x) = v_i(x)$ for all $x \in E_i$, $i=1,2,\ldots,m$. A policy δ is piecewise (p.w.) constant if there is a simple partition $\{E_1, E_2, \ldots, E_m\}$ of Ω and m actions $a_1, a_2, \ldots, a_m$, such that $\delta(x) = a_i$ for all $x \in E_i$, $i=1,2,\ldots,m$. A p.w. constant policy is simple and easily represented in a computer. For example, a bang bang control is such p.w. constant policy. The paper Denardo and Rothblum [5] discusses affine (but not piecewise) dynamic programs.

Although v^* is not necessarily p.w. linear and δ^* is not necessarily p.w. constant, we will show for a class of dynamic programs having the structure described in the following assumption that there are ε-optimal p.w. linear cost functions and p.w. constant policies.

Assumption I (A.I.). For each a, $(U_a v)(x)$ is p.w. linear on Ω, provided that v is p.w. linear on Ω.

The following theorem shows that the structure in Assumption I implies how U_* and U_δ preserve the p.w. linearity of loss functions and the p.w. constant of policies. Assume from now on that $A_x = A = \{a_1, a_2, \ldots, a_p\}$ for all $x \in \Omega$ is finite.

Theorem 1. Suppose that (A.I.) holds and that v is p.w. linear. Then

(i) $U_\delta v$ is p.w. linear whenever δ is p.w. constant;

(ii) $U_* v$ is p.w. linear; and

(iii) there exists a p.w. constant policy δ such that $U_\delta v = U_* v$.

Proof.

(i) Suppose that δ is p.w. constant with respect to a simple partition $\{E_i\}$. Let E_i be an arbitrary but fixed cell from the partition and suppose that $\delta(x) = a$ for $x \in E_i$. Then

$$(U_\delta v)(x) = (U_a v)(x) \quad \text{for} \quad x \in E_i.$$

From (A.I.), $U_a v$ is p.w. linear for each a. Hence $U_\delta v$ is p.w. linear on each cell E_i, and is consequently p.w. linear on Ω.

(ii)& (iii) The functions $U_a v$ are each p.w. linear by (A.I.). Suppose that $U_a v$ is p.w. linear with respect to the simple partition P^a. Let $P = \prod_{a \in A} P^a$. Then P is finer than each P^a, and so each $U_a v$ is p.w. linear with respect to P. For each $F \in P$ and $a \in A$, there is some linear functional α_F^a such that

$$(U_a v)(x) = \alpha_F^a(x) \quad \text{for} \quad x \in F.$$

For each $F \in P$, define the sets G_F^b, $b \in A = \{1,2,\ldots,p\}$, by
$G_F^b = \{x : \alpha_F^b x < \alpha_F^a x,\ a=1,2,\ldots,b-1 \text{ and } \alpha_F^b x \leq \alpha_F^a x,\ a=b+1,\ldots,p\}$.
Then $\{G_F^a : a \in A\} = P^F$ is a partition of F and $\hat{P} = \prod_{F \in P} P^F$ is a partition of Ω with the property that

$$(U_* v)(x) = \alpha_F^a(x) \quad \text{if} \quad x \in G_F^a \in \hat{P}.$$

The policy δ defined by $\delta(x) = a$ for $x \in G_F^a \in \hat{P}$ satisfies $U_\delta v = U_* v$.

Corollary. Suppose that (A.I.) holds and that $v^0 \in V$ is p.w. linear.

(i) Define $v^n(x) = (U_\delta v^{n-1})(x)$, $n=1,2,\ldots$, for p.w. constant δ.

(ii) Define $v^n(x) = (U_* v^{n-1})(x)$, $n=1,2,\ldots$.

Then v^n is p.w. linear and there exists a p.w. constant stationary policy, δ_n, satisfying $U_{\delta_n} v^{n-1} = U_* v^{n-1}$.

We next consider the effects of iterating monotone contraction mappings such as U_* and U_δ, citing some results of Denardo [4].

Lemma 1. Suppose that U is a contraction mapping on V with contraction coefficient $\beta < 1$. Let $v^0 \in V$ be given and define the functions v^n, $n=1,2,\ldots$ by

$$v^n(x) = (Uv^{n-1})(x).$$

Then

(i) $\{v^n\}$ converges in norm to the fixed point $\hat{v}$ of U; i.e., $U\hat{v} = \hat{v}$.

Now assume that U is also monotone.

(ii) If $v^1 \leq v^0$, then $\{v^n\}$ is monotonically decreasing to $\hat{v}$.

(iii) If $v^1 \geq v^0$, then $\{v^n\}$ is monotonically increasing to $\hat{v}$.

Remark. The fixed point $\hat{v}$ need not to be p.w. linear since the cells in the limiting partition are not necessarily finite in number nor polyhedral.

3. Examples

Model 1. A Markov decision process (Blackwell [1])

Let Ω be a bounded convex polyhedron in R^N and the loss function $h(x,a,v) = c(x,a) + \beta \int_\Omega v(x')q(dx'|x,a)$ as mentioned in the preceding section. Assume that $c(x,a) = c^a \cdot x$, which may be interpreted to be the expectation of c^a if x is a probability vector. Also assume that for each convex polyhedron $B \subset \Omega$

$$q^a(B,x) = \int_B x'q(dx'|x,a)$$

is p.w. linear in x with respect to a simple partition $P^a(B) = \{E_j(a,B),\ j=1,2,\ldots,m_{a,B}\}$ for each a where the integral of the vector x' is defined componentwise. These two assumptions imply (A.I.).

We explicitly check that (A.I.) is satisfied. Let $a \in A$ be arbitrary but fixed and suppose that v is p.w. linear with respect to a simple partition $\{E_i,\ i=1,2,\ldots,m\}$. Let $P^a = \prod_{i=1}^{m} P^a(E_i) = \{\tilde{E}^a_j;\ j=1,2,\ldots,r\}$, the product partition, which is again simple.

$$\begin{aligned}
(U_a v)(x) &= c^a \cdot x + \beta \int_\Omega v(x')q(dx'|x,a) \\
&= c^a \cdot x + \beta \sum_{i=1}^{m} \int_{E_i} (v_i x')q(dx'|x,a) \\
&= c^a x + \beta \sum_{i=1}^{m} v_i \cdot \left(\int_{E_i} x'q(dx'|x,a)\right) \\
&= c^a \cdot x + \beta \sum_{i=1}^{m} v_i q^a(E_i,x) \\
&= \left[c^a + \beta \sum_{i=1}^{m} v_i \lambda^a_{ij}\right]x \quad \text{for} \quad x \in E_j(a,E_i)
\end{aligned}$$

where $\lambda^a_{ij} \cdot x = q^a(E_i,x)$ for $x \in E_j(a,E_i)$ and the index j depends on i for each $a \in A$. The third equality is obtained from the fact that the integral of the inner product is equal to the inner product of the integral if v_i does not depend on x, a and each componentwise integral is well defined. $U_a v$ is linear on each $\tilde{E}^a_j$. Hence $U_a v$ is p.w. linear with respect to the simple partition $P^a=\{\tilde{E}^a_j, j=1,2,\ldots,r\}$, which satisfies (A.I.).

Model 2. A partially observable Markov Decision Process (Sawaki and Ichikawa [11], Dynkin [6])

We will show that a partially observable Markov decision process, model 2, is a special case of model 1. Consider a Markov decision process with state space $\{1,2,\ldots,N\}$, with finite action set A, with probability transition matrices p^a and with immediate cost vectors c^a. Let Z_n be the state at the n-th transition. Assume that the process $\{Z_n,\ n=0,1,2,\ldots.\}$ cannot be observed, but at each transition a signal θ is transmitted to the decision maker. The set of possible signals θ is assumed to be finite. For each n, given that $Z_n = j$ and that action a is to be implemented, the signal θ_n is independent of the history of the signals and actions $\{\theta_0,a_0,\theta_1,a_1,\ldots,\theta_{n-1},a_{n-1}\}$ prior to the n-th transition and has conditional probability denoted by $\gamma^a_{j\theta} = P[\theta_n=\theta \,|\, Z_n=j, a_{n-1}=a]$.

Let $\Omega = \{x=(x_1,x_2,\ldots,x_N): \sum_{i=1}^{N} x_i=1, x_i\geq 0, \forall_i\} \subset R^N$. Define the i-th component of X_n, the random variable of x, to be

$$P[Z_n = i \,|\, \theta_0,a_0,\theta_1,\ldots,\theta_{n-1},a_{n-1},\theta_n],\ i=1,2,\ldots,N.$$

It can be shown (see Dynkin [6]) that

$$P[Z_{n+1} = j \,|\, \theta_0,a_0,\theta_1,\ldots,\theta_n,a_n,\theta_{n+1}] = P[Z_{n+1} = j \,|\, \theta_{n+1},a_n,X_n].$$

Thus X_n represents a sufficient statistics for the complete past history $\{\theta_0,a_0,\ldots,a_{n-1},\theta_n\}$. It follows that $\{X_n : n=0,1,2,\ldots\}$ is a Markov process (see Dynkin [6]), called the observed process. Its immediate cost is $c(x,a) = c^a \cdot x$. Its probability transition function is determined by the following calculation. For each measurable subset $B \subseteq \Omega$, $x \in \Omega$, and $a \in A$,

$$q(B|x,a) = P[X_{n+1} \in B | X_n = x,\ a_n = a]$$

$$= \sum_{\theta} P[X_{n+1} \in B | \theta_{n+1} = \theta,\ X_n = x,\ a_n = a] \cdot \sum_{j} P[\theta_{n+1} = \theta | Z_{n+1} = j,\ X_n = x, a] \cdot P[Z_{n+1} = j | X_n = x,\ a_n = a]$$

$$= \sum_{\theta} P[X_{n+1} \in B | \theta_{n+1} = \theta,\ X_n = x,\ a_n = a] \cdot \sum_{j} \gamma^{a}_{j\theta} \sum_{i} P[Z_{n+1} = j | Z_n = i,\ X_n = x,\ a_n = a] P[Z_n = i | X_n = x,\ a_n = a]$$

$$= \sum_{\theta} P[X_{n+1} \in B | \theta_{n+1} = \theta,\ X_n = x,\ a_n = a] \sum_{j} \gamma^{a}_{j\theta} \sum_{i} P^{a}_{ij} x_i$$

$$= \sum_{\theta} P[X_{n+1} \in B | \theta_{n+1} = \theta,\ X_n = x,\ a_n = a]\ \underline{1} P^{a}(\theta) x$$

where $\underline{1}$ $(1,1,\ldots,1)$ and $P^{a}(\theta) = [P^{a}_{ij(\theta)}] = [P^{a}_{ji}\gamma^{a}_{i\theta}]$. Define the vector $T(x|\theta,a)$ by

$$T(x|\theta,a) = \frac{P^{a}(\theta)x}{\underline{1}P^{a}(\theta)x}$$

Note that $T(X_n|\theta,a) = X_{n+1}$, and that

$$P[X_{n+1} \in B | \theta_{n+1} = \theta, X_n = x,\ a_n = a] = \begin{cases} 1 & \text{if } T(x|\theta,a) \in B \\ 0 & \text{if } T(x|\theta,a) \notin B. \end{cases}$$

So,

$$q(B|x,a) = \sum_{\theta \in \Phi^{a}(B,x)} \underline{1}\ P^{a}(\theta)x$$

where $\Phi^a(b,x) = \{\theta: T(x|\theta,a) \in B\}$.

Finally, we show that the observed process $\{X_n\}$ is a special case of Model 1; i.e., $q^a(B,x) = \int_B x'q(dx'|x,a)$ is p.w. linear in x for each convex polyhedral set $B \subset \Omega$ and action $a \in A$. Using the previously computed $q(B|x,a)$ we have

$$q^a(B,x) = \int_B x'q(dx'|x,a)$$

$$= \sum_{\theta\in\Phi^a(B,x)} T[x|\theta,a] \underline{1} P^a(\theta)x$$

$$= \sum_{\theta\in\Phi^a(B,x)} \frac{P^a(\theta)x}{\underline{1}P^a(\theta)x} \underline{1} P^a(\theta)x$$

$$= \sum_{\theta\in\Phi^a(B,x)} P^a(\theta)x$$

which can be shown to be p.w. linear (see Brumelle and Sawaki [2]).

Model 3. A Machine Replacement Model with Partially Observable States

This model is an application of partially observable models into the quality control model modified from Sawaki and Ichikawa [11]. A machine consists of n internal components. The state of the machine is the number of working components. The machine produces constant finished products (say M units) and the machine cannot be inspected, that is, the state of the machine is unobservable. Let θ be the number of defective products out of M finished products. Assume that the conditional probabilities of finding θ given the machine in state i are given by

$$P\{\theta|X_t = i\} = \gamma_{i\theta} \quad i=0,1,\ldots,n, \quad \theta=0,1,\ldots,M,$$

when the machine is not replaced. Assume that $\gamma_{00} = 1$ and $\gamma_{0\theta} = 0$ for $\theta > 1$. Thus, the only available information about the state is the posterior probability x_i and θ. If the probability at time t is $x = (x_0,\ldots,x_n)$ and θ has been observed, then it will be $T(x|\theta)$ given in Model 2. Let c_i be the operating cost if the machine is in state i and let $c = (c_0,\ldots,c_n)$. The replacement cost is $R > 0$ and the daily expected operating cost is

$$cx = \sum_{i=0}^{n} c_i x_i$$

which is linear in x. It is shown from Blackwell [1] that the minimal expected total discounted cost $V^*(x)$ is the unique solution to the optimal equation

$$V^*(x) = \min \{cx + \beta \Sigma \underline{1} P_{\gamma_\theta} V^*(T(x|\theta)) \ ; \ R + \beta V^*(eP)\}.$$

Model 4. A classical linear economic model

Let x be a price vector of N commodities (or N securities) in the market and assume that a new price vector x' can be written as

$$x' = Q_\theta^a x$$

where Q_θ^a is an $N \times N$ matrix depending on the present economic situation θ and on an economic alternative a. Let $P[\theta|x,a]$ be the conditional probability of θ forecasted, given x and a. Assume that there exists a simple partition $\{E_i\}$ of the set of price vectors x such that

$$P[\theta|x,a] = p_{\theta i}^a \quad \text{for} \quad x \in E_i,$$

which is p.w. constant with respect to $\{E_i\}$. Therefore, the model belongs to

the class of model 1, provided the immediate cost is well defined.

4. Algorithm and Its Convergence

If the state space Ω is uncountable, or even countably infinite, then the dynamic program is difficult to implement on a computer. However, if the dynamic program has the structure of (A.I.) and v is p.w. linear, then $U_\delta^n v$ is p.w. linear and each δ^n constructed as in Theorem 1(iii) is p.w. constant. In this case, the cost functions and policies can be specified by a finite number of items - the inequalities describing each cell of a simple partition and the corresponding action or linear function.

In this section, we discuss the algorithm in general terms, choosing the parameters $\{k_n\}$ which specify the degree of approximation of v^δ in the n-th iteration, terminating the algorithm, and a proof that the algorithm converges. The algorithm includes policy improvement and successive approximation as special cases.

Algorithm

Start with a simple policy δ^0 and a p.w. linear function $y^0 \in V$ satisfying $y^0 \geq U_{\delta^0} y^0$.

An iteration of the algorithm is described as follows: n=0,1,2,... . At the start of the n-th iteration, we have a simple policy δ^n and a p.w. linear function $y^n \in V$ satisfying $y^n \geq U_{\delta^n} y^n$. β is a contraction operator's coefficient.

(i) Compute $U_{\delta^n}^{k_n} y^n$ where the integer k_n is the number of iterations of U_{δ^n} which are to be performed.

(ii) Set $y^{n+1} = U_{\delta^n}^{k_n} y^n$ and find a policy δ^{n+1} such that $U_{\delta^{n+1}} y^{n+1} = U_* y^{n+1}$.

(iii) If $||y^n - y^{n+1}|| \leq (1-\beta)\varepsilon$, then stop wity y^n ε-optimal and δ_n ε-optimal.

Moreover, $v^* \leq v^{\delta n} \leq y^{n+1}$.

(iv) If $||y^n - y^{n+1}|| > (1-\beta)\varepsilon$, then increment n by 1 and perform another iteration.

To start, the algorithm needs a simple policy δ and a p.w. linear function y satisfying $y \geq U_\delta y$. There is no difficulty in finding a simple policy; for example, $\delta(x) = a$ for all $x \in \Omega$ is satisfactory and one can choose $y^0(x) = 0$ for each $x \in \Omega$ which satisfies $y^0 \geq U_\delta y^0$ in a Markov decision process with $c(x,\delta) \leq 0$. Note that if each $k_n = 1$ in step (i), the algorithm is successive approximation and that if each $k_n = \infty$, the algorithm is reduced into policy improvement.

Theorem 2. For each iteration, n=0,1,2,..., in the algorithm,

$$y^n \geq U_{\delta^n} y^n \geq U^2_{\delta^n} y^n \geq \ldots \geq U^{k_n}_{\delta^n} y^n = y^{n+1} .$$

In other wards, $\{y^n\}$ is a decreasing sequence.

Proof. First, it is true for n = 0. Since $y^0 \geq U_{\delta^0} y^0$ and since U_{δ^0} is monotone, it follows that $y^0 \geq U_{\delta^0} y^0 \geq U^2_{\delta^0} y^0 \geq \ldots \geq U^{k_0}_{\delta^0} y^0 = y^1 \geq U_{\delta^0} y^1$. By definition δ_1 satisfies $U_{\delta_1} y^1 = U_* y^1$. However, $U_* y^1 \leq U_{\delta^0} y^1 \leq y^1$, and so not only is the Theorem established for n = 0, but we have also shown that $U_{\delta^1} y^1 \leq y^1$.

Now suppose $U_{\delta^n} y^n \leq y^n$. The same argument as in the first paragraph establishes the Theorem for n and also that $U_{\delta^{n+1}} y^{n+1} \leq y^{n+1}$. Hence the proof is completed by induction.

Corollary. $y^n \geq v^*$ for $n = 1,2,\ldots$.

Proof. For an arbitrary n, $y^n \geq U_{\delta^n} y^n \geq U_* y^n$. Since U_* is monotone, $y^n \geq U_*^j y^n$ for each j. By Lemma 1, $U_*^j y^n$ decreases monotonically and converges to v^* as $j \to \infty$. Consequently, $y^n \geq v^*$ and the proof is complete.

We next show that if the algorithm terminates then it will provide an ε-optimal cost function and an ε-optimal policy.

Theorem 3. If $||y^n - y^{n+1}|| \leq (1 - \beta)\varepsilon$, then $||y^n - v^*|| \leq \varepsilon$, i.e., y^n is ε-optimal. Moreover, δ_n is also ε-optimal and $v^* \leq v^{\delta_n} \leq y^n$.

Proof. Note that $U_{\delta^n} y^n = u_* y^n$ and that by the previous corollary $y^n \geq v^*$.

$$||y^n - v^*|| \leq ||y^n - U_* y^n|| + ||U_* y^n - U_* v^*||$$

$$\leq ||y^n - U_{\delta^n} y^n|| + \beta ||y^n - v^*||$$

$$\leq ||y^n - U_{\delta^n}^m y^n|| + \beta ||y^n - v^*|| \quad \text{for } m=1,2,\ldots ,$$

because $y^n \geq U_{\delta^n} y^n \geq U_{\delta^n}^m y^n$ for $m = 1,2,\ldots$. (Theorem 2.)

Thus $(1-\beta)||y^n - v^*|| \leq ||y^n - U_{\delta^n}^m y^n|| = ||y^n - y^{n+1}|| \leq (1-\beta)\varepsilon$,

and so $||y^n - v^*|| \leq \varepsilon$.

The last statement in the Theorem follows by Theorem 2 and Corollary.

Theorem 4. Suppose that $\{y^n\}$ is a sequence of costs generated by the algorithm.

(i) y^n converges pointwise to $y \in V$.

(ii) $y = U_* y$, i.e., y is optimal.

In other words, the algorithm converges.

Proof.

(i) First of all we shall show that $\{y^n\}$ is bounded below. By Theorem 2 we have $y^n \geq U^m_{\delta^n} y^n$ for each $m = 1,2,\ldots$. It is well known (see [1] and [4]) that $U^m_{\delta^n} y^n \to v^{\delta^n}$ as $m \to \infty$. Therefore $y^n \geq v^{\delta^n}$. From $v^{\delta_n} \geq v^* \in V$, there exists an M such that $||v^{\delta_n}|| \leq M$. Hence $y^n(x) \geq -M$ for all x. From Theorem 2 y^n is a decreasing sequence. Hence y^n converges pointwise.

(ii) By a choice of y^0 and Theorem 2 we know that

$$y^n \geq U_{\delta^n} y^n \geq U_* y^n \quad . \tag{1}$$

To show the other way we have

$$\begin{aligned} y^n &= U^m_{\delta^{n-1}} y^{n-1} && \text{(By definition of } y^n\text{)} \\ &\leq U_{\delta^{n-1}} y^{n-1} && (U^m_\delta y \leq Uy,\ \forall y \in V) \\ &= U_* y^{n-1} && \text{(By definition of } \delta^{n-1}\text{)}. \end{aligned} \tag{2}$$

Then, from (1) and (2), we obtain

$$U_* y^n \leq y^n \leq U_* y^{n-1} \quad .$$

From the statement (i) $y^n \to y$. Since a contraction mapping U_* is continuous, $U_* y^n \to U_* y$. Therefore, we must have

$$U_* y = y$$

which completes the proof.

5. A Numerical Example and Conclusion

This section presents a numerical example for Model 2, partially observable Markov decision processes, especially in the case of two demensions, $\Omega = \{(x_1,x_2) \mid x_1 + x_2 = 1,\ x_1,\ x_2 \geq 0\}$, $A = \{1,2\}$ and $\theta = \{1,2\}$. The necessary data are shown in Table I. To specify the stopping rule we choose $\beta = 0.8$ and $\varepsilon = 0.01$. Therefore, if $||y^n - y^{n-1}|| \geq 0.002$, then the algorithm stops and y^n is ε-optimal.

Set $x_1 = x$. To start the algorithm an initial p.w. constant policy δ^0 and an initial p.w. linear cost function y^0 satisfying $y_0 \geq U_{\delta^0} y^0$ must be found. Choose a policy δ^0 minimizing $c^a \cdot x$; thus $\delta^0(x) = 1$ if $x \leq 2/3$, $\delta^0(x) = 2$ if $x > 2/3$. Set an initial cost function $y^0(x) = (0,0)[\frac{x}{1-x}]$ with the partition $\{[0,1]\}$, which is p.w. linear and satisfies $y^0 \geq H_{\delta^0} y^0$. Also set $k_n = 1$ for all n. The computational results programmed in FORTRAN are shown in Table II. We may observe from Table II that the algorithm converges at period n=35, and an ε-optimal cost is $-15.166-3.826x$ if $x \leq 0.571$ and $-16.732-1.086x$ if $x > 0.571$. An ε-optimal policy $\delta^{30}(x) = 1$ if $x \leq 0.571$ and $\delta^{30}(x) = 2$ if $x > 0.571$. Table II also shows that an ε-optimal policy converges (at n=10) much faster than an ε-optimal cost does.

The goal of this paper is to generate and construct ε-optimal costs and ε-optimal policies in a sequential fashion for a general class of dynamic programs. Toward this end we have taken advantage of the properties of p.w. linear cost functions and p.w. constant policies. These properties guarantee that the algorithm involves only p.w. linear and constant functions which belong to the class of linear programs. Finally we should also emphasis the importance of the algorithm capable for solving continuous state dynamic programs. Many sequential decision problems under uncertainty often turn out to have a probability vector as their state space, which is no longer finite nor countably infinite, but continuous. Therefore, the algorithm developed

in this paper will become more important in the field of sequential decision problems under uncertainty.

TABLE I

actions	c^a	P^a	γ^a
a = 1	(-5,-1)	0.7 0.3	0.75 0.25
		0.9 0.1	0.60 0.40
a = 2	(-4,-3)	0.5 0.5	0.3 0.70
		0.4 0.6	0.40 0.60

TABLE II

Periods n	cost functions y^n	Policies and Partitions	$\|\|y^n-y^{n-1}\|\|$
1	-1-4x -3-x	1 [0.00,0.666] 2 (0.666,1.00]	5.00
2	-4.12-3.84x -5.719-1.08x	1 [0.00,0.579] 2 (0.579,1.00]	2.96
3	-6.35-3.827x -7.92-1.086x	1 [0.00,0.572] 2 (0.572,1.00]	2.22
5	-9.53-3.826x -11.095-1.086x	1 [0.00,0.571] 2 (0.571,1.00]	1.411
10	-13.324-3.826x -14.889-1.086x	1 [0.00,0.571] 2 (0.571,1.00]	0.462
20	-14.975-3.826x -16.540-1.086x	1 [0.00,0.571] 2 (0.571,1.00]	0.05
30	-15.152-3.826x -16.717-1.086x	1 [0.00,0.571] 2 (0.571,1.00]	0.005
35	-15.166-3.826x -16.732-1.086x	1 [0.00,0.571] 2 (0.571,1.00]	0.001

Acknowledgement

The author would like to express his thanks to the referees for their helpful comments and suggestions. Thanks also to Dr. Seiichi Iwamoto for informing him the references [7] and [9]. The author is also grateful to Professor Y. Iihara for his suggestions and encouragement.

This research was partially supported by the Nanzan University Research Grant A (1979).

References

[1] Blackwell, D.: Discounted Dynamic Programming, *Annals of Mathematical Statistics*, 36(1965), 226-235.

[2] Brumelle, S.L. and Sawaki, K.: Generalized Policy Improvement for Simple Dynamic Programs, Working Paper 546, Faculty of Commerce, University of British Columbia, Vancouver (1978).

[3] Brumelle, S.L. and Puterman, M.L.: On the Convergence of Newton's Method for Operators with Supports, University of California, Berkeley (1976).

[4] Denardo, E.V.: Contraction Mappings in the Theory Underlying Dynamic Programming, *SIAM Review* 9(1967), 165-177.

[5] Denardo, E.V. and Rothblum, U.G.: Affine Dynamic Programming, Proceedings of the International Conference on Dynamic Programming and Its Applications, University of British Columbia (1977): *Dynamic Programming and Its Applications* (ed. M.L. Puterman), Academic Press (1978), 255-267.

[6] Dynkin, E.B.: Controlled Random Sequences, *Theory of Probability and Its Applications* X(1965), 1-14.

[7] Fox, B.L.: Finite-State Approximations to Denumerable-State Dynamic Programs, *Journal of Mathematical Analysis and Applications*, 34(1971), 665-670.

[8] Howard, R.A.: *Dynamic Programming and Markov Processes*, Wiley, New York (1960).

[9] Larson, R.E.: *State Increment Dynamic Programming*, Elsevier, New York (1968).

[10] Sawaki, K.: Piecewise Linear Markov Decision Processes with An Application into Partially Observable Models, presented at the 1978 International Conference on Markov Decision Processes, University of Manchester (1978).

[11] Sawaki, K. and Ichikawa, A.: Optimal Control for Partially Observable Markov Decision Processes Over an Infinite Horizon, *Journal of Operations Research Society of Japan*, 21(1978), 1-16.

[12] Smallwood, R.D. and Sondik, E.J.: The Optimal Control of Partially Observable Markov Processes over a Finite Horizon, *Operations Research*, 21(1973),1071-1088.

[13] Sondik, E.J.: The Optimal Control of Partially Observable Markov Processes over the Infinite Horizon: Discounted Costs, *Operations Research*, 26(1978), 282-304.

[14] Strauch, R.E.: Negative Dynamic Programming, *Annals of Mathematical Statistics*, 37(1966), 871-890.

Katsushige SAWAKI: Faculty of
Business Administration,
Nanzan University, 18,
Yamazato-cho, Showa-ku
Nagoya, 466 Japan