JOURNAL OF MATHEMATICAL ANALYSIS AND APPLICATIONS **89**, 639–647 (1982)

Multi-Objective Infinite-Horizon Discounted Markov Decision Processes

D. J. WHITE

Department of Decision Theory, University of Manchester, Manchester M139PL. England

Submitted by Richard Bellman

1. INTRODUCTION

In a recent paper, Daellenbach and de Kluyver [1] discussed a multi-objective routing problem, using an extension of Bellman's [6] successive-approximations equation for vector-minimisation problems. This approach is also studied in White [2, p. 190], Henig [3], and Hartley [4]. The extension of the usual method of successive-approximations method works, in effect, because only a finite number of iterations are required, and only a finite number of actions are allowable at each iteration. For other problems the situation is not so straightforward.

In this paper we shall examine the extended method in the context of a class of infinite-horizon discounted Markov decision processes. The alternative weighting-factor method is fully discussed in Hartley [7] for this class of problems, and in White [10] for the multi-objective routing problem referred to earlier. We shall make the following assumptions, some of which may be relaxed (e.g., finite-state requirements), given appropriate measurability conditions.

There is a finite set I, of states, $i = 1, 2, \ldots, N$; for each state $i \in I$, there is a compact-action set K_i; for each $i \in I$, $k \in K_i$, there is a single-stage loss vector $f_i^k \in \mathbb{R}^m$, continuous on K_i, with components f_{il}^k, $l = 1, 2, \ldots, m$, and a transition probability p_{ij}^k, continuous on K_i, from state i to state j; there is a discount factor ρ, $0 \leq \rho < 1$.

A decision rule δ tells us which action in K_i to take for any state $i \in I$ which we may realise. The set of all such decision rules Δ is compact. A policy γ is a sequence of such decision rules, $(\delta_1, \delta_2, \ldots, \delta_n, \ldots)$, which tells us, at the beginning of the appropriate period, which action to take when we know the state $i \in I$ at the beginning of that period. If the sequence is infinite, then the set of all such sequences is Γ. If the sequence contains n members, we let Γ^n be the set of all such sequences.

For each $\delta \in \Delta$, there will be a transition probability matrix P^δ with the

0022-247X/82/100639–09$02.00/0

(i,j)th element written as $[P^\delta]_{ij}$, a general convention we shall use. For each $\gamma \in \Gamma$, or $\gamma \in \Gamma^n$, we shall have expected discounted vector-objective function values as given below.

$$\gamma \in \Gamma, \qquad v_i^\gamma = \sum_{s=0}^{\infty} \rho^s \sum_{j=1}^{N} \left[\prod_{t=0}^{s} P^{\delta_t} \right]_{ij} f_j^{\delta_{s+1}(j)}, \tag{1}$$

$$\gamma \in \Gamma^n, \quad n \geqq 1, \qquad v_i^\gamma = \sum_{s=0}^{n} \rho^s \sum_{j=1}^{N} \left[\prod_{t=0}^{s} P^{\delta_t} \right]_{ij} f_j^{\delta_{s+1}(j)}, \tag{2}$$

$$\gamma \in \Gamma^n, \quad n = 0, \qquad v_i^\gamma = 0. \tag{3}$$

We shall use v_i^γ, whether or not $\gamma \in \Gamma$ or $\gamma \in \Gamma^n$, and the usage will be seen from the context. We shall also not differentiate, in typeface, between vectors in $\mathbb{R}^p$, for any p, and scalars. Again the context will make the usage clear.

The sets of all such vectors given by (1) or (2), (3), are V_i and V_i^n, respectively.

Let us now consider the efficiency aspects. For any set, $X \subseteq \mathbb{R}^p$, we let the efficient set be defined as follows:

$$\mathscr{E}(X) = \{x \in X : y \in X, y \leqq x \rightarrow y = x\}.$$

For future use, we note that we adopt the standard convention for inequalities, viz. if x, $y \in \mathbb{R}^p$, then $y \leqq x \rightleftarrows y_i \leqq x_i$, $\forall i$; $y \leqslant x \rightleftarrows y \leqq x$, $y \neq x$; $y < x \rightleftarrows y_i < x_i$, $\forall_i$; $x = y \rightleftarrows x \leqq y$, $y \leqq x$.

In this paper we shall be concerned with, in particular, $\{\mathscr{E}(V_i)\}$ and $\{\mathscr{E}(V_i^n)\}$, and the vector value method of successive approximations given below (see Theorem 2 for detailed analysis and definitions).

$$n \geqq 1, \qquad W_i^n = \mathscr{E}\left(\bigcup_{k \in K_i} \left(f_i^k \oplus \rho \sum_{j=1}^{N} p_{ij}^k W_j^{n-1} \right) \right), \qquad \forall i \in I$$

$$n = 0, \qquad W_i^0 = \{0\}, \qquad \forall i \in I.$$

It will be shown, in particular, that $\{\mathscr{E}(V_i^n)\}$ is the unique solution to these equations. Now, for scalar-valued Markov decision processes, the limiting solution, as $n \to \infty$, to these equations would be the unique solution to the infinite-horizon problem, giving the minimal-expected discounted value, over Γ. Unfortunately this limiting result does not carry over to vector-valued Markov decision processes. Thus, consider the following example: $I = \{1, 2, 3\}$; $K_1 = \{1, 2\}$, $K_2 = \{1\}$, $K_3 = \{1\}$; $p_{12}^1 = 1$, $p_{13}^2 = 1$, $p_{22}^1 = 1$, $p_{33}^1 = 1$, $p_{ij}^k = 0$, otherwise; $f_1^1 = (1, 1 + \rho^{-1})$, $f_1^2 = (2, \rho^{-1})$, $f_2^1 = (1, 1)$, $f_3^1 = (2, \rho^{-1})$. We then have

$$
\begin{aligned}
n \geqq 1, \quad & \mathscr{E}(V_1^n) = \{u^n, v^n\},\\
& u^n = ((1-\rho^n)/(1-\rho), (1-\rho^{n+1})/\rho(1-\rho)),\\
& v^n = (2(1-\rho^n)/(1-\rho), (1-\rho^n)/\rho(1-\rho)),\\
& \mathscr{E}(V_1) = \{u\}, \qquad u = (1/(1-\rho), 1/\rho(1-\rho)).
\end{aligned}
$$

Hence, the limit of the sequence $\{v^n\}$ is not efficient in V_1, but each of its members is efficient in V_1^n, for each n.

Theorem 1 specifically deals with this problem and, in effect, says that $\mathscr{E}(V_i)$ is identical with the efficient set of the limit points, which gives the usual scalar-valued result when $m = 1$.

In our theoretical analysis we merely assume Δ to be compact. In Hartley [7], K_i is the set of all probability distributions over a finite set of actions, and this makes Δ compact. In Hartley's case, he shows that, when finding $\mathscr{E}(V_i)$, we may restrict ourselves to stationary members of Γ, i.e., those $\gamma \in \Gamma$ which can be expressed as an infinite repetition of a single $\delta \in \Delta$. This is useful if one is using a weighting-factor approach to find $\mathscr{E}(V_i)$. In our case, we are not studying the weighting-factor approach, and since our computations will involve finite iterations in arriving at some suitable termination point, we may have a nonstationary sequence $\gamma \in \Gamma^n$ at the termination point. We shall not study the stationary problem for the general compact Δ, but will comment on this when $\#\Delta < \infty$.

Let us consider the stationary/nonstationary issue via an example, viz. $I = \{1\}$; $K_1 = \{1, 2, 3\}$; $\Delta = \{\delta_1, \delta_2, \delta_3\}$, where $\delta_q(1) = q$, $\forall q$; $f_1^1 = (3, 0)$, $f_1^2 = (0, 3)$, $f_1^3 = (2, 2)$. Then, if $\gamma_q = (\delta_q, \delta_q, ..., \delta_q, ...)$ we have the following:

$$
v_1^{\gamma_1} = (3/(1-\rho), 0), \qquad v_1^{\gamma_2} = (0, 3/(1-\rho)), \qquad v_1^{\gamma_3} = (2/(1-\rho), 2/(1-\rho)).
$$

If $\gamma = (\delta_1, \delta_2, \delta_1, \delta_2, ...)$, i.e., δ_1, δ_2 alternating, we have the following:

$$
v_1^{\gamma} = (3/(1-\rho^2), 3\rho/(1-\rho^2)).
$$

We then see that $v_1^{\gamma} < v_1^{\gamma_3}$ if $\rho > \frac{1}{2}$. At the same time, had we been interested only in stationary policies, none of $\{v_1^{\gamma_q}\}$ dominates any other member of $\{v_1^{\gamma_q}\}$, but $v_1^{\gamma_3}$ is dominated by v_1^{γ}. Clearly, it would be inadvisable to restrict onself to stationary policies in this case.

For the same example, if ρ is small enough, it is possible to show that $\mathscr{E}(V_1) = V_1$, and again it would be inadvisable to restrict oneself to nonstationary policies in this case, if all members of $\mathscr{E}(V_1)$ are required.

Finally, if it is admissible in principle to extend K_i to all probability mixtures over K_i, and correspondingly, extend Δ, we could get wrong results by confining ourselves to the original Δ for efficiency purposes. Thus, if k_4 is an equiprobability mixture of $k_1(=1)$, and k_2 $(=2)$, and $\gamma = (\delta, \delta, ..., \delta, ...)$, where δ uses action k_4 always, we shall have $v_1^{\gamma} < v_1^{\gamma_3}$, whereas γ_3 would be undominated in the original problem.

The net effect is that if Δ is obtained by taking probability mixtures over a finite number of actions, for each state, we must ensure that our calculations cover these probability mixtures, and, at the same time, we must cater for nonstationary policies, for the general compact Δ, although, in Hartley's case, this is not necessary.

Let us now consider our theoretical results.

2. Theoretical Results

Let us first define the terms we shall use, in addition to those already defined.

M: $\max\limits_{i\in I} \max\limits_{1\leqslant l\leqslant m} \sup\limits_{k\in K_i} [\|f_{il}^k\|]$.

L_i^V: the set of limit points of all sequences $\{v_i^n\}$, with $v_i^n \in V_i^n$, for all n in the sequence.

$L_i^{\mathscr{E}}$: the set of limit points of all sequences $\{v_i^n\}$, with $v_i^n \in \mathscr{E}(V_i^n)$ for all n in the sequence.

Convergence in all cases is in the usual sup–norm sense. We may now establish the following results.

Lemma 1.

$$L_i^V = V_i, \qquad \forall i.$$

Proof. Let $v \in V_i$. Then there is a $\gamma \in \Gamma$ such that $v_i = v_i^{\gamma}$, and v_i^{γ} is the limit of a sequence, $\{v_i^n\}$, with $v_i^n \in V_i^n$, $\forall i$, given by

$$n \geqq 1, \qquad v_i^n = f_i^{\delta_n(i)} + \rho \sum_{j=1}^{N} p_{ij}^{\delta_n(i)} v_j^{n-1} \tag{4}$$

$$v_i^0 = 0 \tag{5}$$

for some sequence $\{\delta_n\} \subseteq \Delta$.

Hence $V_i \subseteq L_i^V$, $\forall i$.

Now let $v \in L_i^V$. Then v is a point of convergence of a subsequence, $\{v^n\}$, $n \in \mathscr{N}_1$, and $v^n \in V_i^n, \forall n \in \eta_1$. There also exists, for each such $n \geqq 1$, $\delta_n \in \Delta$, and a set $\{v_j^{n-1}\}, j \in I$, with $v_j^{n-1} \in V_j^{n-1}$, $\forall j$, such that

$$v^n = f_i^{\delta_n(i)} + \rho \sum_{j=1}^{N} p_{ij}^{\delta_n(i)} v_j^{n-1}. \tag{6}$$

Since Δ is compact, there exists a subsequence $\mathcal{N}_2 \subseteq \mathcal{N}_1$, and a $\delta^1 \in \Delta$ such that $\{\delta_n\}$ converges to δ^1 as $n \to \infty$ in $\mathcal{N}_2$. Let us now consider $n \in \mathcal{N}_2$, and the set $\{v_j^{n-1}\}$. We see that, for each $j \in I$, v_j^{n-1} may likewise be expressed in a form similar to (6), in terms of a set $\{v_{js}^{n-2}\}$, $s \in I$, and some $\delta_{n,n-1} \in \Delta$. Again we may choose a subsequence $\mathcal{N}_3 \subseteq \mathcal{N}_2$, and a $\delta^2 \in \Delta$, such that $\delta_{n,n-1}$ converges to δ^2 as $n \to \infty$ in $\mathcal{N}_3$. We may repeat the process r times to produce a sequence of subsequences $\mathcal{N}_r$ with $\mathcal{N}_{r+1} \subseteq \mathcal{N}_r$, with the property that, for any given $\varepsilon > 0$, and any $r \geqq 2$, for n sufficiently large in $\mathcal{N}_r$, we have

$$\left| v_l^n - \sum_{s=0}^{r-2} \rho^s \sum_{j=1}^{N} \left[\prod_{t=0}^{s} P^{\delta^t} \right]_{lj} f_{jl}^{\delta^{s+1}(j)} - \rho^{r-1} \sum_{j=1}^{N} \left[\prod_{t=0}^{r-1} P^{\delta^t} \right]_{lj} z_{jl} \right| \leqq \varepsilon, \qquad \forall l, \quad (7)$$

where $\{\delta^t\} \subseteq \Delta$, P^{δ^0} is equal to the unit matrix, and $z_j \in V_j^{n-r+1}$, $\forall j$. Since $|z_{jl}| \leqq M/(1-\rho)$, $\forall j, l$, the last term inside the modulus signs in (4) may be made arbitrarily small if r is made large enough. Since ε is arbitrary, we obtain $v_i^\sigma = \lim_{n \to \infty} [v^n] = v$, where $\sigma = (\delta^1, \delta^2, ..., \delta^t, ...)$.

Hence $L_i^V \subseteq V_i$. ∎

LEMMA 2.

(i) If $v \in V_i$, there exists a $u \in \mathscr{E}(V_i)$ with $u \leqq v$.

(ii) If $v \in V_i^n$, there exists a $u \in \mathscr{E}(V_i^n)$ with $u \leqq v$.

(iii) If $v \in L_i^V$, there exists a $u \in L_i^{\mathscr{E}}$ with $u \leqq v$.

Proof. (i) This follows from White [5, Theorem 4], if we can show that V_i is compact, and the set $S_i(v) = \{u \in V_i : u \leqq v\}$ is closed, $\forall v \in V_i$. Clearly, V_i is bounded. Now let $\{u^\alpha\}$ be a sequence of points in V_i, converging to a point $u \in \mathbb{R}^m$. Since $v \in V_i$, and $\{u^\alpha\} \subseteq V_i$, we have $v = v_i^\gamma$, and $u^\alpha = v_i^{\gamma^\alpha}$, $\forall \alpha$, and $\gamma \in \Gamma$, $\{\gamma^\alpha\} \subseteq \Gamma$. Let δ^α be the first decision rule in γ^α. We may choose a subsequence $\mathscr{A}_1$ such that $\{\delta^\alpha\}$ converges to some $\delta^1 \in \Delta$ as $\alpha \to \infty$ in $\mathscr{A}_1$. We then repeat the same sort of procedure as in Lemma 1, and produce a sequence of subsequences $\mathscr{A}_r$ such that we have a similar result to that in (8), and since each limit point of $\{u^\alpha\}$, $\alpha \in \mathscr{A}_r$, must be the same as u, we deduce that $u = v_i^\sigma$. This proves that V_i is closed. Now, if $u^\alpha \leqq v$, $\forall \alpha$, it follows that $u \leqq v$, and, together with the closure of V_i, it follows that $S_i(v)$ is closed.

(ii) In line with the proof in (i), all we need to show is that V_i^n is closed, since the other requirements follow as in (i). Let $\{u^\alpha\}$ be a sequence of points in V_i^n, converging to a point $u \in \mathbb{R}^m$. For each α, there exists a $\gamma^\alpha \in \Gamma^n$ such that $u^\alpha = v_i^{\gamma^\alpha}$. We may now choose a subsequence $\mathscr{A}_1$ such that if δ^α is the first decision rule in γ^α, $\{\delta^\alpha\}$ converges to some $\delta' \in \Delta$. We may repeat the same sort of analysis as in (i), but in this case we only have n subsequences, and can set $z_j = 0$, $\forall j \in I$, in the corresponding expression to (7) when $r = n$, and the required result follows.

(iii) Let $v \in L_i^V$. Then there exists a subsequence $\{v^n\}$, with $v^n \in V_i^n$, $\forall n \in \mathcal{N}_1$, and $\{v^n\}$ converges to v. From part (ii), there exists a subsequence $\{u^n\}$ $n \in \mathcal{N}_1$, with $u^n \leqq v^n$, and $u^n \in \mathscr{E}(V_i^n)$, $\forall n \in \mathcal{N}_1$. Since $\bigcup_n V_i^n$ is bounded, there exists a subsequence $\mathcal{N}_2 \subseteq \mathcal{N}_1$ such that $\{u^n\}$ converges to some point u in $\mathbb{R}^m$ for $n \in \mathcal{N}_2$, and, by definition, $u \in L_i^{\mathscr{E}}$. Clearly, $u \leqq v$ and the requisite result follows. ∎

We may now prove our main theorems.

THEOREM 1.

$$\mathscr{E}(L_i^\varepsilon) = \mathscr{E}(V_i), \forall i.$$

Proof. Let $v \in \mathscr{E}(L_i^\varepsilon)/\mathscr{E}(V_i)$. Then $v \in L_i^V$, and, from Lemma 1, there is a $\sigma \in \Gamma$, such that $v = v_i^\sigma$. Since $v \notin \mathscr{E}(V_i)$, there is a $\gamma \in \Gamma$, such that $v_i^\gamma \leqslant v$. Now v_i^γ is the limit of a sequence $\{v_i^n\}$, given by (4), (5), with $v_i^n \in V_i^n$, $\forall n$. From Lemma 2, for each n there exists a $u^n \in \mathscr{E}(V_i^n)$ such that $u^n \leqq v^n$ (where we replace v_i^n by v^n, for a specific i, for notational convenience). Since $\bigcup_n V_i^n$ is bounded, there exists a subsequence $\mathcal{N}_1$ of n, for which $\{u^n\}$ converges to some point $u \in \mathbb{R}^m$, for $n \in \mathcal{N}_1$. From Lemma 1, since $u \in L_i^V$, there exists a $\tau \in \Gamma$ such that $u = v_i^\tau$. Then, since clearly, $u \leqq v_i^\gamma \leqslant v = v_i^\sigma$, we have $v_i^\tau \leqslant v_i^\sigma$. Since $v_i^\tau \in L_i^\varepsilon$, by construction, and $v_i^\sigma \in \mathscr{E}(L_i^\varepsilon)$, by assumption, we have a contradiction.

Hence $\mathscr{E}(L_i^\varepsilon) \subseteq \mathscr{E}(V_i)$, $\forall i$.

Now let $v \in \mathscr{E}(V_i)/\mathscr{E}(L_i^\varepsilon)$. Then there is a $\sigma \in \Gamma$ such that $v = v_i^\sigma$. From Lemma 1, $v \in V_i = L_i^V$, and, from Lemma 2, there is a $u \in L_i^\varepsilon$, with $u \leqq v$. Then $u \in L_i^\varepsilon \subseteq L_i^V = V_i$, however, and hence we must have $u = v$, i.e., $v \in L_i^\varepsilon$. Since $v \notin \mathscr{E}(L_i^\varepsilon)$, there is a $w \in L_i^\varepsilon$ with $w \leqslant v$. Then $w \in L_i^\varepsilon \subseteq L_i^V = V_i$, and this contradicts $v \in \mathscr{E}(V_i)$.

Hence $\mathscr{E}(V_i) \subseteq \mathscr{E}(L_i^\varepsilon)$. ∎

THEOREM 2. *For all $n \geqq 1$, $\{\mathscr{E}(V_i^n)\}$ is the unique solution $\{W_i^n\}$ to the following equation, which is in $\oplus$ sum-set form.*

$$n \geqq 1, \qquad W_i^n = \mathscr{E}\left(\bigcup_{k \in K_i}\left(f_i^k \oplus \sum_{j=1}^{N} p_{ij}^k W_j^{n-1}\right)\right), \qquad \forall i = I \tag{8}$$

$$n = 0, \qquad W_i^0 = \{0\}, \qquad \forall i \in I. \tag{9}$$

Proof. The theorem is clearly true for $n = 0$. Let us assume that it is true for $n - 1$, for some $n \geqq 2$.

Let $v \in \mathscr{E}(V_i^n)$. Then there is $\gamma \in \Gamma^n$ such that $v = v_i^\gamma$. If δ_n is the first decision rule in γ, we shall obtain, as in (4), the following equation for some set $\{v_j^{n-1}\}$, $j \in I$, with $v_j^{n-1} \in V_j^{n-1}$, $\forall j \in J$.

$$v = f_i^{\delta_n(i)} + \rho \sum_{j=1}^{N} p_{ij}^{\delta_n(i)} v_j^{n-1}. \tag{10}$$

From Lemma 2, for each $j \in I$, there is a $u_j^{n-1} \in \mathscr{E}(V_j^{n-1})$ such that $u_j^{n-1} \leqq v_j^{n-1}$. Let

$$w = f_i^{\delta_n(i)} + \rho \sum_{j=1}^{N} p_{ij}^{\delta_n(i)} u_j^{n-1}. \tag{11}$$

Then $w \in V_i^n$, and $w \leqq v$. Hence $w = v$. Since by assumption, $\{\mathscr{E}(V_j^{n-1})\}$ is the unique solution W_j^{n-1} to (8), (9), for $n-1$,

$$\mathscr{E}(V_i^n) \subseteq W_i^n, \qquad \forall i \in I.$$

Now let $w \in W_i^n$. Then there is a $\gamma \in \Gamma^n$, such that $v = v_i^\gamma$, and if δ_n is the first decision rule in γ, we have, again, expression (10), with $v_j^{n-1} \in W_j^{n-1}$, $\forall j \in J$. By assumption, $v_j^{n-1} \in \mathscr{E}(V_j^{n-1})$, $\forall j \in I$. Now suppose $w \notin \mathscr{E}(V_i^n)$. Then there is a $\tau \in \Gamma^n$, such that $v_i^\tau \leqslant w$, and by Lemma 2, we may assume that $v_i^\tau \in \mathscr{E}(V_i^n)$. Hence, from the first part, we have $v_i^\tau \in W_i^n$. This is not possible since W_i^n is an efficient set, and we cannot have $v_i^\tau \leqslant w$.

Hence $W_i^n \subseteq \mathscr{E}(V_i^n)$, $\forall i \in I$.

The uniqueness of $\{V_i^n\}$ as a solution to (8), (9) is obvious. ∎

LEMMA 3. *Let $f_i^k \leqq 0$, $\forall i \in I$, $k \in K_i$. Then, for each $i \in I$, $n \geqq 1$, and each $v \in \mathscr{E}(V_i^{n-1})$ there is a $u \in \mathscr{E}(V_i^n)$ with $u \leqq v$.*

Proof. The lemma is clearly true for $n = 1$. Let us assume it is true for $n-1$, for some $n \geqq 2$, and let $v \in \mathscr{E}(V_i^n)$. Then, for some $k \in K_i$ and some set $\{w_j\}, j \in I$, with $w_j \in V_j^{n-2}$, for all such j, we have

$$v = f_i^k + \rho \sum_{j=1}^{N} p_{ij}^k w_j. \tag{12}$$

Let

$$t = f_i^k + \rho \sum_{j=1}^{N} p_{ij}^k z_j, \tag{13}$$

where by assumption, we may choose $z_j \in \mathscr{E}(V_j^{n-1})$ with $z_j \leqq w_j$, for all $j \in I$ and $t \in V_i^n$.

Then, clearly, $t \leqq v$. From Lemma 2, there is a $u \in \mathscr{E}(V_i^n)$ with $u \leqq t$, and the requisite result follows. ∎

We may now prove Theorem 3, where $\tilde{L}_i^\varepsilon$ is now the set of all limit points of monotonic-decreasing sequences $\{v^n\}$, i.e., such that $v^n \leqq v^{n-1}$, for $n \geqq 1$, and $v^n \in (V_i^n)$, for $n \geqq 1$.

THEOREM 3.

$$\mathscr{E}(\tilde{L}_i^\varepsilon) \subseteq \mathscr{E}(V_i).$$

Proof. This follows in a similar manner to that of Theorem 1, replacing L_i^ε by $\tilde{L}_i^\varepsilon$, and noting that, by virtue of Lemma 3, the sequence $\{u^n\}$ may be chosen so that $u^n \leqq u^{n-1}$, for all $n \geqq 1$, and hence, that the limit u is in $\tilde{L}_i^\varepsilon$. ▮

CONCLUDING REMARKS

The basic purpose of the paper is to show that, in principle, the vector generalisation of the usual scalar method of successive approximations may be used to tackle the problem of finding efficient solutions for infinite-horizon discounted Markov decision processes. The introductory section indicates some of the difficulties one meets, which do not arise in the scalar-valued case.

The analysis provides a framework for an extension of bound, and elimination of action, analyses as described, in general, in White [8], and clearly considerable development is needed if this approach is to be used in a manner analogous to the established approach for scalar-valued problems. Although there is the problem of *explosion*, as we increase the value of n, the scalar analyses have indicated that n need not be too large before acceptable approximations are reached.

There is still the problem of interpretting the use of the $\mathscr{E}(V_i^n)$ analysis to obtain *approximations* to $\mathscr{E}(V_i)$.

First of all, if $v^n = v_i^\gamma$, $u^n = v_i^\tau$, for some $\gamma, \tau \in \Gamma^n$, we know that neither of v^n or u^n dominates each other. We are seeking, however, for the infinite-horizon problem, policies in Γ. If $\gamma' = (\gamma, \sigma)$, $\tau' = (\tau, r)$, where $\sigma, r \in \Gamma$, it is easily seen that $\|v^n - v_i^{\gamma'}\| \leqq \rho^n M/(1-\rho)$, and $\|u^n - v_i^{\gamma'}\| \leqq \rho^n M/(1-\rho)$, and hence, although either $v_i^{\gamma'}$ or $v_i^{\tau'}$ may dominate each other, they can be made arbitrarily close to v^n and u^n, respectively, which do not dominate each other, if n is large enough.

Secondly, if the sequence $\{v^n\}$, $v^n \in \mathscr{E}(V_i^n)$, has a limit point v, then either $v \in \mathscr{E}(V_i)$, in which case v^n is close to a member of $\mathscr{E}(V_i)$ if n is large enough, or $v \notin \mathscr{E}(V_i)$, in which case, if $u \in \mathscr{E}(V_i)$ and $u \leqslant v$, there is a $u^n \in \mathscr{E}(V_i^n)$ close to u, which does not dominate v^n. That is, there are points in $\mathscr{E}(V_i^n)$ close enough to v, which are not dominated by some points in $\mathscr{E}(V_i^n)$ close enough to u.

Hence our *approximation* must be defined in terms of the above concepts combined, which show that if n is large enough, and we use the policies γ' or τ' as determined above, even though the one limit point may dominate another, and either $v_i^{\gamma'}$ or $v_i^{\tau'}$ may dominate each other, there are points in $\mathbb{R}^m$ which are close enough to the limit points or to $v_i^{\gamma'}$ and $v_i^{\tau'}$, which do not dominate each other.

In using Theorem 2 there is the problem of picking out convergent subsequences. For monotone decreasing sequences we have no problem, since they will converge in any case. Hence, Theorem 3 may be useful. It will, of course, only produce, in general, a subset of $\mathscr{E}(V_i)$. All Markov decision processes can be transformed to ones in which $f_i^k \leqq 0$, $\forall i \in I$, $k \in K_i$ (see White [2]).

Finally, we have only been concerned with finding the individual $\mathscr{E}(V_i)$; and associated policies, or approximating policies. We might be interested in policies $\gamma \in \Gamma$, such that $v_i^\gamma \in \mathscr{E}(V_i)$, $\forall i \in I$. In this case, if $\lambda \in \mathbb{R}^m$, $\lambda > 0$, it is well known that, if $\# \Delta < \infty$, if we minimise $[\lambda v_i^\gamma]$, where γ is a repeated application some δ, over Δ (see White [9]), then $v_i^\gamma \in \mathscr{E}(V_i)$ for any such optimisers δ. Now, if P^δ has no transient states, for any $\delta \in \Delta$, then if γ minimises $[\lambda v_i^\gamma]$ over Δ, for some i, γ will also minimise $[\lambda v_i^\gamma]$ over Δ for all i. Hence, in this case, the weighting-factor approach will produce uniformly efficient policies, i.e., for all $i \in I$. Hartley [7] deals with the weighting-factor approach when Δ is the convex hull of a finite set of policies as described in Section 1.

References

1. H. G. Daellenbach and C. A. de Kluyver, Note on multiple objective dynamic programming, *J. Oper. Res. Soc.* **31** (1980), 591–594.
2. D. J. White, "Finite Dynamic Programming," Wiley, New York, 1978.
3. M. I. Henig, "Dynamic Programming with Returns in Partially Ordered Sets," Faculty of Commerce and Business Administration, University of British Columbia, Vancouver, 1980.
4. R. Hartley, "Dynamic Programming in Vector Networks," Notes in Decision Theory, No. 86, Department of Decision Theory, Manchester University, England, 1979.
5. D. J. White, Kernels of preference structures, *Econometrica* **45** (1977), 91–100.
6. R. Bellman "Dynamic Programming," Princeton Univ. Press, Princeton, N.J., 1957.
7. R. Hartley, "Finite Discounted, Vector Markov Decision Processes," Notes in Decision Theory, No. 85, Department of Decision Theory, Manchester University, England, 1979.
8. D. J. White, Elimination of non-optimal actions in Markov decision processes, *in* "Dynamic Programming and Its Applications" (M. Puterman, Ed.), Academic Press, New York, 1978.
9. D. J. White, "Fundamentals of Decision Theory," Elsevier, Amsterdam/New York, 1976.
10. D. J. White, "The Set of Efficient Solutions for Multiple Objective Shortest Path Problems," *Comp. & Ops. Res.* **9** (1982), 101–107.