ACCELERATING PROCEDURES OF THE VALUE ITERATION ALGORITHM FOR DISCOUNTED MARKOV DECISION PROCESSES, BASED ON A ONE-STEP LOOKAHEAD ANALYSIS

MEIR HERZBERG

Telecom Australia Research Laboratories, Clayton, Victoria, Australia

URI YECHIALI

Tel Aviv University, Tel Aviv, Israel

(Received April 1991; revisions received December 1991, November 1992; accepted February 1993)

Accelerating procedures for solving *discounted* Markov decision processes problems are developed based on a one-step lookahead analysis of the value iteration algorithm. We apply the criteria of minimum difference and minimum variance to obtain good adaptive relaxation factors that speed up the convergence of the algorithm. Several problems (including Howard's automobile replacement) are tested and a preliminary numerical evaluation reveals considerable reductions in computation time when compared to existing value iteration schemes.

The purpose of this paper is to derive and analyze accelerating procedures for convergence of the value iteration algorithm (VIA) used when solving *discounted* Markov decision processes (MDP). (For a survey of MDP applications see White 1985.)

Various value iteration schemes have been developed in the literature, aimed at reducing the computational effort required for solving such problems. The main schemes are (see Thomas, Harley and Lavecombe 1983):

Pre-Jacobi (PJ) (see Blackwell 1965):

$$V_n(i) = \min_{a \in A_i} \left\{ C_i(a) + \beta \sum_{j \in I} P_{ij}(a) V_{n-1}(j) \right\} \quad i \in I. \tag{1}$$

Jacobi (J) (see Porteus and Totten 1978):

$$V_n(i) = \min_{a \in A_i} \left\{ \left[C_i(a) + \beta \sum_{j \neq i} P_{ij}(a) V_{n-1}(j) \right] \Big/ \left[1 - \beta P_{ii}(a) \right] \right\} \quad i \in I. \tag{2}$$

Pre-Gauss-Seidel (PGS) (see Porteus 1975):

$$V_n(i) = \min_{a \in A_i} \left\{ C_i(a) + \beta \sum_{j=1}^{i-1} P_{ij}(a) V_n(j) + \beta \sum_{j=i}^{|I|} P_{ij}(a) V_{n-1}(j) \right\} \quad i \in I. \tag{3}$$

Gauss-Seidel (GS) (see Kushner and Kleinman 1971):

$$V_n(i) = \min_{a \in A_i} \left\{ \left[C_i(a) + \beta \sum_{j=1}^{i-1} P_{ij}(a) V_n(j) + \beta \sum_{j=i+1}^{|I|} P_{ij}(a) V_{n-1}(j) \right] \Big/ \left[1 - \beta P_{ii}(a) \right] \right\} \quad i \in I, \tag{4}$$

where,

- $V_0(i), i \in I$ is an arbitrary chosen cost function;
- $V_n(i)$ *is the minimal* total expected *discounted* cost when starting at state i, moving n periods and paying a terminal cost $V_0(j)$ if the process ends up at state j;
- A_i denotes the set of possible actions admissible in state i;
- $C_i(a)$ is the immediate (one-step) expected payment when selecting action $a \in A_i$ while in state i;
- $P_{ij}(a)$ is the one-step transition probability from state i to state j when selecting action $a \in A_i$;
- I is a finite set of states with cardinality $|I|$; and
- β is the discount factor $\in (0, 1)$.

Subject classifications: Dynamic programming/optimal control: discounted Markov decision processes, value iteration algorithm.
Area of review: STOCHASTIC PROCESSES AND THEIR APPLICATIONS.

0030-364X/94/4204-0940 $01.25

All four schemes are directed at calculating the (optimal) values $\{V(i)\}$, $i \in I$, that satisfy the optimality conditions:

$$V(i) = \min_{a \in A_i} \left\{ C_i(a) + \beta \sum_{j \in I} P_{ij}(a) V(j) \right\} \quad i \in I. \quad (5)$$

The actions satisfying (5) comprise the *optimal stationary policy*.

Each value iteration procedure prescribed by (1)–(4) stops at the first iteration n when estimators $\hat{V}(j)$ are achieved which are within a predetermined tolerance error ϵ of $V(j)$.

Letting $\delta_n(j) \equiv V_n(j) - V_{n-1}(j)$, and defining

$$m_n = \min_{j \in I} \{\delta_n(j)\}, \qquad M_n = \max_{j \in I} \{\delta_n(j)\},$$

one can calculate bounds on the required $V(j)$'s.

As the implied transition matrix in the form of (5) might not have equal row sums for all the schemes, one should use the general form suggested by Porteus (1975) to find the $V(j)$ bounds as:

$$V_n(j) + \frac{\beta'(n)}{1 - \beta'(n)} m_n \leq V(j) \leq V_n(j) + \frac{\beta''(n)}{1 - \beta''(n)} M_n \quad j \in I, \quad (6)$$

where

$$\beta'(n) = \begin{cases} \rho'(n) & \text{if } m_n \geq 0 \\ \rho''(n) & \text{otherwise} \end{cases}$$

and

$$\beta''(n) = \begin{cases} \rho''(n) & \text{if } M_n \geq 0 \\ \rho'(n) & \text{otherwise.} \end{cases}$$

The values $\rho'(n)$ and $\rho''(n)$ represent the minimum and maximum row sums, respectively, derived from the transition matrix associated with the policy obtained at each iteration n. For the **PJ** scheme, one can use $\rho'(n) \equiv \rho''(n) \equiv \beta = \beta'(n) = \beta''(n)$ for all $n \geq 1$.

To ensure $|\hat{V}(j) - V(j)| \leq \epsilon$ for all $j \in I$ we use the stopping criterion:

$$\frac{\beta''(n)}{1 - \beta''(n)} M_n - \frac{\beta'(n)}{1 - \beta'(n)} m_n \leq 2\epsilon \quad (7)$$

so that, when the algorithm stops at iteration n, the values $\hat{V}(j)$, $j \in I$, are calculated by

$$\hat{V}(j) = V_n(j) + \frac{1}{2} \left[\frac{\beta''(n)}{1 - \beta''(n)} M_n + \frac{\beta'(n)}{1 - \beta'(n)} m_n \right]. \quad (8)$$

Procedures for solving *un*discounted Markov or semi-Markov decision processes use the method of an adaptive relaxation factor (ARF) to speed up the convergence of the VIA.

The idea is to replace the values $V_n(j)$, obtained at iteration n, by the values $\bar{V}_n(j)$, formed as a linear combination of $V_n(j)$ and $V_{n-1}(j)$: $\bar{V}_n(j) = wV_n(j) + (1 - w)V_{n-1}(j)$. The parameter w, whose value is determined anew at each iteration, is known as the ARF. Popyack, Brown and White (1979) suggested a "dynamic relaxation factor" depending on M_n and m_n. In Herzberg and Yechiali (1991) we introduced two new criteria for selecting the ARF, w, when solving *un*discounted MDPs. The criteria are termed *minimum ratio* and *minimum variance*. By using these criteria a good ARF is calculated in each iteration, so that the total number of iterations required for convergence becomes smaller, and the total computational effort is reduced, even though each iteration requires the extra work of determining the ARF. For the discounted MDP, Kushner and Kleinman and Porteus and Totten tested the effect of using a *constant* overrelaxation factor for various discounted VIAs. Porteus and Totten also pointed out that the order of calculation of the $V_n(j)$ values might affect the convergence of the VIA when using the **PGS** or **GS** scheme.

Another means of reducing computational efforts is to use the concept of *action elimination*. The idea is to exclude from the computations those actions that cannot be part of the optimal policy. (See MacQueen 1967, Porteus 1975, Hastings and Van-Nunen 1977, and Puterman and Shin 1982). Usually, action elimination does not affect the *number* of iterations performed until reaching the stopping criterion. Additional ideas regarding improved iterative computation are presented in Porteus (1980).

This paper extends our ideas from Herzberg and Yechiali, developed originally for the *un*discounted MDPs, and apply them to the *discounted* processes. We use the method presented in Herzberg and Yechiali and introduce *additional* improvement so that the total time required for convergence is reduced (in the problems tested) by up to 76%. The main idea, based on a one-step lookahead analysis, is to replace $V_n(j)$ by a modified value $W_n(j) = V_n(j) + w\beta g(j)$, where $g(j)$ is a function of the differences $\delta_n(j)$'s and the one-step transition probabilities.

Following a one-step lookahead analysis (presented in Section 1), we modify the discounted value iteration schemes defined by (1)–(4). It is interesting that the *same* type of analysis applies to *all* four

procedures **PJ**, **J**, **PGS**, and **GS**. In Section 2 we use the criterion of minimum difference to develop a method for calculating a good ARF. In Section 3 we apply the minimum variance method to obtain a good relaxation factor. In Section 4 we present numerical results for several problems tested, and discuss various computational aspects.

1. MODIFIED VALUE ITERATION SCHEMES

Suppose that after calculating the values $V_n(i)$, $i \in I$, at the nth iteration of the VIA, we apply the concept of relaxation and consider the values $\bar{V}_n(i)$, $i \in I$, where

$$\begin{aligned}\bar{V}_n(i) &= wV_n(i) + (1-w)V_{n-1}(i) \\ &= V_{n-1}(i) + w\delta_n(i) \quad i \in I. \end{aligned} \tag{9}$$

Here w is the ARF such that, for $w = 1$, $\bar{V}_n(i) = V_n(i)$.

We now look one step ahead and examine an *estimator* of the *future* value of $V_{n+1}(i)$. This estimator, denoted $W_n(i)$, will replace $V_n(i)$ in the $(n+1)$st iteration. Such an estimator has the prospect of being close to the next calculated value, $V_{n+1}(i)$, thus causing the VIA to converge faster.

Denote by R_i the selected action for state i determined by the VIA at iteration n. Then, for the **PJ** scheme,

$$\begin{aligned}W_n^{PJ}(i) &= C_i(R_i) + \beta \sum_{j\in I} P_{ij}(R_i)\bar{V}_n(j) \\ &= C_i(R_i) + \beta \sum_{j\in I} P_{ij}(R_i)V_{n-1}(j) \\ &\quad + \beta w \sum_{j\in I} P_{ij}(R_i)\delta_n(j).\end{aligned}$$

That is,

$$W_n^{PJ}(i) = V_n(i) + \beta w g_n^{PJ}(i), \tag{10}$$

where,

$$g_n^{PJ}(i) = \sum_{j\in I} P_{ij}(R_i)\delta_n(j). \tag{11}$$

Now, an estimator of $\delta_{n+1}(i)$ would be

$$\hat{\delta}_{n+1}^{PJ}(i) = W_n^{PJ}(i) - \bar{V}_n(i).$$

Substituting expressions (9), (10), and (11) we obtain

$$\hat{\delta}_{n+1}^{PJ}(i) = \delta_n(i) + w[\beta g_n^{PJ}(i) - \delta_n(i)]. \tag{12}$$

Clearly, for $w = 1$, $\hat{\delta}_{n+1}^{PJ}(i) = \beta g_n^{PJ}(i)$.

For the **Jacobi** VIA scheme,

$$W_n^J(i) = \left[C_i(R_i) + \beta \sum_{j\neq i} P_{ij}(R_i)\bar{V}_n(j)\right] \Big/ [1 - \beta P_{ii}(R_i)].$$

Using relation (9) we derive,

$$W_n^J(i) = V_n(i) + \beta w g_n^J(i), \tag{13}$$

where,

$$g_n^J(i) = \sum_{j\neq i} P_{ij}(R_i)\delta_n(j)/[1 - \beta P_{ii}(R_i)]. \tag{14}$$

Also, it readily follows that $\hat{\delta}_{n+1}^J(i)$ is given by (12), where $g_n^{PJ}(i)$ is replaced by $g_n^J(i)$.

Considering next the **PGS** procedure, one can show (by induction) that if $V_n(i)$ are replaced by $\bar{V}_n(i)$, then

$$W_n^{PGS}(i) = V_n(i) + \beta w g_n^{PGS}(i), \tag{15}$$

where,

$$g_n^{PGS}(i) = \beta \sum_{j=1}^{i-1} P_{ij}(R_i)g_n^{PGS}(j) + \sum_{j=i}^{|I|} P_{ij}(R_i)\delta_n(j). \tag{16}$$

Again, $\hat{\delta}_n^{PGS}(i)$ is derived from (12) by using $g_n^{PGS}(i)$.

Finally, performing similar operations on the **GS** scheme, it follows that (10) (as well as (13) and (15)) holds for $W_n^{GS}(i)$, with $g_n^{GS}(i)$ replacing $g_n^{PJ}(i)$, where

$$g_n^{GS}(i) = \left[\beta \sum_{j=1}^{i-1} P_{ij}(R_i)g_n^{GS}(j) + \sum_{j=i+1}^{|I|} P_{ij}(R_i)\delta_n(j)\right] \Big/ [1-\beta P_{ii}(R_i)], \tag{17}$$

and $\hat{\delta}_n^{GS}(i)$ is once again given by (12) by substituting $g_n^{GS}(i)$ instead of $g_n^{PJ}(i)$.

To summarize: For *all four schemes* we have:

$$\begin{aligned}&W_n(i) = V_n(i) + \beta w g_n(i) \\ &\text{and} \\ &\hat{\delta}_{n+1}(i) = \delta_n(i) + w[\beta g_n(i) - \delta_n(i)],\end{aligned} \tag{18}$$

where $g_n(i)$ is calculated by (11), (14), (16), or (17) for the **PJ**, **J**, **PGS**, or **GS** procedure, respectively.

As stated in the Introduction, the main new idea in our modified VIA is to replace $V_n(i)$ by the estimator $W_n(i) = V_n(i) + \beta w g_n(i)$. In addition, we extend our methods of selecting a good ARF, w, so that the overall modified VIA will result in a considerably improved procedure.

2. MINIMUM DIFFERENCE CRITERION

Equation 7 would serve as our stopping condition for the modified VIA. If this condition has not been satisfied by iteration n, it seems plausible for the $(n+1)$st iteration to find an ARF that will *minimize* the difference

$$D(w) = \pi_1(w) - \pi_2(w), \tag{19}$$

where $\pi_1(w) = \max_i\{\hat{\delta}_{n+1}(i)\}$, and $\pi_2(w) = \min_i\{\hat{\delta}_{n+1}(i)\}$. This is so, because $\pi_1(w)$ and $\pi_2(w)$ are the estimators of M_{n+1} and m_{m+1}, respectively.

Now, using (18), we have

$$\pi_1(w) = \max_i\{\delta_n(i) + w\alpha_n(i)\}$$
$$\pi_2(w) = \min_i\{\delta_n(i) + w\alpha_n(i)\}, \tag{20}$$

where,

$$\alpha_n(i) = \beta g_n(i) - \delta_n(i). \tag{21}$$

Clearly, $\pi_1(0) = M_n$, $\pi_2(0) = m_n$, and $D(0) = M_n - m_n$.

Observe also that $\pi_1(w)$ ($\pi_2(w)$) is a piecewise linear convex (concave) function, being the max (min) of a set of linear functions. Therefore, $D(w)$ is also piecewise linear, and hence, it is sufficient to examine only the endpoints of its segments when searching for w^* that minimizes $D(w)$. This implies that w^* is found on one of the breakpoints, either of $\pi_1(w)$ or of $\pi_2(w)$. Figure 1 depicts the case where w^* is attained at a breakpoint of $\pi_2(w)$.

For practical considerations we use the fact that the search over the segments on $\pi_2(w)$ is a "mirror reflection" of the search over the segments on $\pi_1(w)$ (see Herzberg and Yechiali), so by multiplying both values of $\delta_n(i)$ and $\alpha_n(i)$, $i \in I$, by (-1) the search procedure over $\pi_2(w)$ is identical to that over $\pi_1(w)$. A systematic procedure for finding w^* is as follows:

STEP 0. Set $w^* = 0$, $\delta = M_n$. Let h be the state for which $\delta_n(h) = M_n$. If h is not unique select the state with the highest value of $\alpha_n(\cdot)$. Set $\alpha = \alpha_n(h)$.

STEP 1. Find $w_1 = \min_{j:\alpha_n(j)>\alpha}\{(\delta - \delta_n(j))/(\alpha_n(j) - \alpha)\} = (\delta - \delta_n(k))/(\alpha_n(k) - \alpha) > 0$.

STEP 2. Find $\gamma = \alpha_n(r)$ where $\min_j\{\delta_n(j) + w_1\alpha_n(j)\} = \delta_n(r) + w_1\alpha_n(r)$.

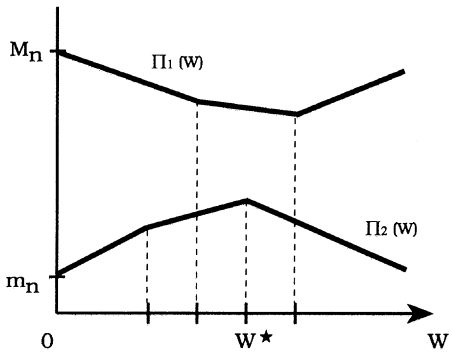

Figure 1. w^* is attained at a breakpoint of $\pi_2(w)$.

STEP 3. If $\alpha \leq \gamma$ and $\alpha_n(k) \geq \gamma$, set w^* equal to $w^* + w_1$ and stop. If $\alpha_n(k) < \gamma$ go to Step 4 (continue the search). If $\alpha > \gamma$ go to Step 5 (move to $\pi_2(w)$).

STEP 4. Update $\delta_n(j) = \delta_n(j) + w_1\alpha_n(j)$, $j \in I$ and $w^* = w^* + w_1$. Set $\delta = \delta_n(k)$, $\alpha = \alpha_n(k)$ and go to Step 1.

STEP 5. Update $\delta_n(j) = -\delta_n(j)$, $\alpha_n(j) = -\alpha_n(j)$, $j \in I$. Find $\delta = \text{Max}_{j\in I}\{\delta_n(j)\} = \delta_n(u)$; set $\alpha = \alpha_n(u)$, and go to Step 1.

3. MINIMUM VARIANCE CRITERION

In Herzberg and Yechiali we introduced the minimum variance criterion for selecting the ARF. By this criterion we select the value w^* that minimizes the *variance* of the terms $\hat{\delta}_{n+1}(i)$, $i \in I$. This criterion takes into consideration the *entire* set of $\hat{\delta}_{n+1}(i)$'s and tries to keep them close together so that $M_{n+1} - m_{n+1}$ will be small.

Consider the vectors $\delta_n = \{\delta_n(i), i \in I\}$ and $\alpha_n = \{\alpha_n(i), i \in I\}$. Then the vector $\hat{\Delta}_{n+1}(w) = \delta_n + w\alpha_n$ has components $\{\delta_n(i) + w\alpha_n(i)\}$. Clearly,

$$Var[\hat{\Delta}_{n+1}(w)] = Var[\delta_n] + w^2 Var[\alpha_n] + 2wCov[\delta_n, \alpha_n]. \tag{22}$$

Setting the derivative of $Var[\hat{\Delta}_{n+1}(w)]$ to zero, one gets

$$w^* = \frac{-Cov[\delta_n, \alpha_n]}{Var[\alpha_n]}. \tag{23}$$

Since $d^2/dw^2\{Var[\hat{\Delta}_{n+1}(w)]\} = 2Var[\alpha_n] > 0$, the optimal value w^* minimizes the variance of $\hat{\Delta}_{n+1}$. In fact, w^* represents the value of the regression coefficient in the linear regression of δ_n on $(-\alpha_n)$, and is easy to calculate using (26):

$$w^* = \frac{-\left\{\sum_i \delta_n(i)\alpha_n(i) - \frac{\left[\sum_i \delta_n(i)\right]\left[\sum_i \alpha_n(i)\right]}{|I|}\right\}}{\sum_i [\alpha_n(i)]^2 - \frac{\left[\sum_i \alpha_n(i)\right]^2}{|I|}}. \tag{24}$$

Usually $Cov[\delta_n, \alpha_n] \leq 0$, and, consequently, $w^* \geq 0$. This happens as $\alpha_n(i) = \beta g_n(i) - \delta_n(i)$, and $\delta_n(i)$ increasing (decreasing) usually results in $\alpha_n(i)$ decreasing (increasing).

4. COMPUTATIONAL CONSIDERATIONS AND NUMERICAL RESULTS

The effort per iteration of the various value iteration schemes, for a fully-dense case of state-to-state transition probability matrix, is of the order $A|I|^2$ (where A is the average number of admissible actions per state). Real-dimensional MDP problems are usually sparse with an average of $N \ll |I|$ possible one-step transitions, so that the effort per iteration is of the order $NA|I|$. The proposed procedures, which are aimed at reducing the number of iterations, try to achieve this goal at the expense of increasing the effort per iteration, resulting from calculating the terms $g_{\hat{n}}(i)$, $i \in I$ and the ARF w^*.

The computational effort for calculating the value $g_{\hat{n}}(i)$ is of the order $N|I|$. The computational effort for calculating an ARF depends on the criteria selected. When using the minimum variance criterion the order is of $4|I|$ (see (24)), while it is usually between $4|I|$ and $12|I|$ when selecting the minimum difference criterion. The additional computation per iteration of the proposed procedures is therefore in the range $(N + 4)|I| - (N + 12)|I|$. This is paid-off because reducing the number of iterations by 1 saves an effort of the order $NA|I|$. Thus, the method is particularly attractive for cases where A is large and $NA|I| > (N + 4)|I|$.

It is worth noting that values of the ARF, calculated anew for each iteration by the proposed criteria, can be either less than or greater than 1, and at certain iterations may even reach the range 2–3. As a result, the relationship between bounds of consecutive iterations cannot be defined fully. However, the concept of *action elimination* can still be applied, e.g., using McQueen's test after each value iteration phase.

Several problems dealing with optimal resource allocation in telecommunication networks and Howard's well-known automobile replacement problem (HARP), numbered as problem 5, were tested. The results are summarized in Table I.

Each problem was solved three times for *every* procedure (**PJ**, **J**, **PGS** or **GS**): First by using the classical VIA; then by using the minimum difference criterion for the corresponding VIA; and finally, by applying the minimum variance method. The same set of calculations were performed for two values of the discount factor: $\beta = 0.8$ and $\beta = 0.9$. For the stopping criterion we use a tolerance error $\epsilon = 10^{-3}$ (see (7)).

For each problem five entries have been defined:

i. The number of iterations (NOI) when using the classical scheme (denoted ST NOI).
ii. NOI when using the minimum difference (MD) criterion (denoted: MD NOI).
iii. Percentage of CPU time-savings when using MD (denoted: MD %TS).
iv. NOI when applying the minimum variance (MV) procedure (denoted: MV NOI).
v. Percentage of CPU time-savings when using MV (denoted: MV %TS).

From the table we see that improved results are achieved for problems where A is large (see problem 6). The performance of the proposed procedures for the cases where $\beta = 0.9$ is usually better than for the cases where $\beta = 0.8$. This is so, because the coefficients $\alpha_n(i)$ are restrained for small values of β (see (20) and (21)), while for high values of β the functions $\pi_1(w)$ and $\pi_2(w)$ are more sensitive to changes in the values of w. Therefore, the effectiveness of our procedures, when selecting the ARF w^*, is increased for high values of β.

5. CONCLUSION

We have introduced new methods for selecting the ARF in value iteration algorithms used for solving *discounted* MDP problems. By applying a one-step lookahead analysis, we further modified the VIA schemes by replacing $V_n(i)$ with an estimator $W_n(i) = V_n(i) + \beta w g_n(i)$. These methods result in computational time-savings of up to 76% (for the problems tested). In the majority of cases the minimum difference criterion appears to be slightly better than the minimum variance method. The methods are attractive and the use of lookahead analysis seems to be promising. In particular, this approach may be useful for cases where the number of decisions considered per state is large and for cases where the discount factor is close to 1, for which convergence of the VIA is usually slow (see Scherer and White 1988). It seems that this approach and the new ARF criteria developed may enhance convergence of successive approximation procedures in general and therefore have the potential to be incorporated in the modified policy iteration algorithm developed by Puterman and Shin (1978) for discounted MDPs, for which a successive substitution technique has been used instead of solving sets of linear equations.

Table I

Number of Iterations (NOI) When Using the Standard (ST) VIA Procedure, the Minimum Difference (MD) Criterion and the Minimum Variance (MV) Rule (Denoted ST NOI, MD NOI and MV NOI, Respectively), and Percentage of CPU Time Savings With Respect to the Standard Procedure (Denoted MD %TS and MV %TS) When Applying the MD and MV Criteria

		PJ		J		PGS		GS	
Case	Scheme	$\beta = 0.8$	$\beta = 0.9$	$\beta = 0.8$	$\beta = 0.9$	$\beta = 0.8$	$\beta = 0.9$	$\beta = 0.8$	$\beta = 0.9$
Prob. No. 1	ST NOI	27	44	37	81	27	53	23	44
A = 2	MD NOI	10	18	10	18	9	16	8	16
|I| = 8	MD %TS	24.9	19.8	42.6	56.2	28.7	38.3	28.0	26.5
N = 2	MV NOI	13	18	12	34	9	17	10	15
	MV %TS	3.7	15.1	35.3	18.4	29.3	34.0	17.4	32.1
Prob. No. 2	ST NOI	28	43	38	80	28	57	24	45
A = 4	MD NOI	11	16	13	14	10	16	10	16
|I| = 10	MD %TS	34.7	39.4	48.4	71.3	44.4	54.6	36.5	46.3
N = 2	MV NOI	14	20	15	23	10	16	10	14
	MV %TS	25.0	29.6	30.3	62.9	44.2	57.5	39.6	52.2
Prob. No. 3	ST NOI	28	45	38	81	29	59	26	48
A = 8	MD NOI	11	18	12	19	10	16	10	15
|I| = 12	MD %TS	44.7	47.0	57.0	68.1	55.1	62.4	46.8	57.3
N = 3	MV NOI	15	22	18	27	10	17	10	17
	MV %TS	31.6	38.3	41.8	59.3	56.4	63.5	50.5	54.5
Prob. No. 4	ST NOI	31	45	39	79	32	65	28	48
A = 16	MD NOI	14	19	15	25	11	20	11	17
|I| = 15	MD %TS	47.4	50.3	57.1	64.2	59.2	64.4	53.0	59.0
N = 3	MV NOI	16	19	20	37	12	21	12	19
	MV %TS	40.3	53.8	40.7	45.8	56.0	62.1	50.6	54.3
Prob. No. 5	ST NOI	36	69	37	68	50	107	49	105
(HARP)	MD NOI	20	35	19	36	23	47	22	47
A = 41	MD %TS	38.5	41.7	43.7	39.8	48.5	50.9	50.0	50.0
|I| = 40	MV NOI	19	36	21	37	23	48	22	48
N = 2	MV %TS	42.9	44.4	46.3	42.8	50.8	54.8	52.1	51.0
Prob. No. 6	ST NOI	33	59	37	83	29	57	29	54
A = 90	MD NOI	15	22	12	19	11	15	12	16
|I| = 250	MD %TS	52.1	60.8	69.1	76.1	60.3	72.4	57.0	68.9
N = 8	MV NOI	17	30	19	40	11	17	11	19
	MV %TS	47.2	48.0	47.6	50.9	61.3	69.6	61.1	64.0
Prob. No. 7	ST NOI	32	67	31	67	29	58	29	58
A = 10	MD NOI	11	20	10	20	9	14	9	14
|I| = 950	MD %TS	54.6	60.5	58.1	61.6	59.8	68.2	59.8	69.3
N = 8	MV NOI	16	33	16	33	10	18	10	18
	MV %TS	41.2	41.4	35.2	41.3	59.6	63.3	58.5	62.9

ACKNOWLEDGMENT

The authors thank the referees for valuable suggestions and remarks. The permission of the Director of Research, Telecom Australia, to publish this paper is hereby acknowledged.

REFERENCES

D. BLACKWELL. 1965. Discounted Dynamic Programming. *Ann. Math. Statist.* **36,** 226–235.

N. A. J. HASTINGS, AND J. A. E. VAN-NUNEN. 1977. The Action Elimination Algorithm for Markov Decision Processes. In *Markov Decision Processes*, H. C. Tijms and J. Wessels (eds.). Mathematical Centre Tract 93, Amsterdam, 161–170.

M. HERZBERG, AND U. YECHIALI. 1991. Criteria for Selecting the Relaxation Factor of the Value Iteration Algorithm for Undiscounted Markov and Semi-Markov Decision Processes. *O.R. Letts.* **10,** 193–202.

H. KUSHNER, AND A. J. KLEINMAN. 1971. Accelerated Procedures for the Solution of Discrete Markov Control Problems. *IEEE Trans. Autom. Control* **16,** 147–152.

J. MACQUEEN. 1967. A Test of Suboptimal Actions in Markovian Decision Problems. *Opns. Res.* **15,** 559–561.

J. L. POPYACK, R. L. BROWN AND C. C. WHITE. 1979. Discrete Versions of an Algorithm Due to Varaiya. *IEEE Trans. Autom. Control* **24,** 503–504.

E. L. PORTEUS. 1975. Bounds and Transformation of Finite Markov Decision Chains. *Opns. Res.* **23,** 761–784.

E. L. PORTEUS, AND J. TOTTEN. 1978. Accelerated Computation of the Expected Discounted Return in a Markov Chain. *Opns. Res.* **26,** 350–358.

E. L. PORTEUS. 1980. Improved Iterative Computation of the Expected Discounted Return in Markov and Semi-Markov Chains. *Zeitschrift f. Opns. Res.* **24,** 155–170.

M. L. PUTERMAN, AND M. C. SHIN. 1978. Modified Policy Iteration Algorithms for Discounted Markov Decision Problems. *Mgmt. Sci.* **24,** 1127–1137.

M. L. PUTERMAN, AND M. C. SHIN. 1982. Action Elimination Procedures for Modified Policy Iteration Algorithms. *Opns. Res.* **30,** 301–307.

L. C. THOMAS, R. HARLEY AND A. C. LAVERCOMBE. 1983. Computational Comparison of Value Iteration Algorithms for Discounted Markov Decision Processes. *O. R. Letts.* **2,** 72–76.

W. T. SCHERER, AND D. J. WHITE. 1988. The Convergence of Value Iteration in Discounted Markov Decision Processes. ORSA/TIMS Conference, Washington D.C.

D. J. WHITE. 1985. Real Applications of Markov Decision Processes. *Interfaces* **15**(6).