JOURNAL OF MATHEMATICAL ANALYSIS AND APPLICATIONS 71, 297–302 (1979)

# A Priori Bounds for Approximations of Markov Programs

WARD WHITT

*Bell Laboratories, Holmdel, New Jersey 07733*

*Submitted by M. Aoki*

This note determines a priori bounds for B. L. Fox's [*J. Math. Anal. Appl.* 34 (1971), 665–670] scheme of approximating discounted Markov programs, thus refining bounds recently obtained by D. J. White (Notes in Descision Theory No. 43, University of Manchester, 1977). The approximation scheme focuses careful attention on only a subset of the state space and uses a fixed function to characterize future returns outside the designated subset. The a priori bounds are useful to design the specific approximation, that is, to select the appropriate subset on which the approximation is based.

## 1. INTRODUCTION AND SUMMARY

The purpose of this note is to extend recent work by D. J. White [5] determining a priori bounds for B. L. Fox's [2] method of approximating discounted Markov programs. The basic idea in Fox's scheme is to focus careful attention on only a subset of the state space and use a fixed function to characterize future returns outside the designated subset. Hopefully good decisions can be determined for many states inside the designated subset without examining the behavior outside the subset in detail. As noted in Whitt [6], two-sided a posteriori bounds can be obtained by considering more than one fixed function characterizing future returns outside the designated subset. It is also significant that the approximation scheme is not limited to a finite subset of a countably infinite state space. For example, we could work with a finite subset of a large finite state space or a compact subset of a noncompact uncountably infinite state space.

The object here, as in White [5], is to determine bounds on the difference between the optimal return function in the original model and the return associated with a policy generated from the approximate model, depending on the designated subset on which the approximation is based. The bounds can be used in turn to select the designated subset.

As in White [5] the bounds on the error in the optimal return function here are based on uniform bounds (over all possible actions) of the Markov transition function. By means of such a "bounding transition function," we can bound the

0022-247X/79/090297–06$02.00/0

probabilities of leaving the designated subset, and thus bound the error in the optimal return function. We extend White [5] by providing sharper and more flexible bounds.

For a broad survey of approximation methods in dynamic programming, see Morin [3].

## 2. The Basic Model

We consider the standard discounted Markov program with finite or countably infinite state and action spaces. (It is easy to see that the cardinality assumption can be relaxed.) Thus, let the state space $I$ and the action spaces $K_i$ for $i \in I$ be nonempty subsets of the positive integers. Let the space $\Delta$ of stationary policies be the Cartesian product of the action spaces. For each $\delta \in \Delta$, $\delta(i)$ is the action in $K_i$ used each time the process is in state $i$. Let $r(i, k)$ be the one-step reward associated with using action $k$ in state $i$, assumed to satisfy $|r(i, k)| \leqslant M$ for all $k \in K_i$ and $i \in I$, and let $g(j \mid i, k)$ be the probability of a one-step transition to state $j$ from state $i$ using action $k$. Let $c$ be the discount factor, $0 \leqslant c < 1$; let $v_\delta$ be the return function associated with policy $\delta \in \Delta$ and let $f = \sup\{v_\delta : \delta \in \Delta\}$ be the optimal return function. A basic result (see Denardo [1]) is that the optimal return function $f$ is the unique solution in the space of bounded functions (in the space $l_\infty$) to the functional equation

$$f(i) = \sup_{k \in K_i} \left\{ r(i, k) + c \sum_j g(j \mid i, k) f(j) \right\}, \qquad i \in I. \tag{1}$$

## 3. The Approximation

In the language of Whitt [6], we work with the lower approximation here. This is obtained by assuming, without loss of generality, that all one-step rewards are nonnegative and that the fixed function characterizing future returns outside the designated subset always assigns the value zero. We still let $M$ represent the bound on the one-step rewards.

Let the designated subset of the state space $I$ be $S$. Then the return function $v_\delta{}^S$ associated with the policy $\delta$ in the approximate model is the unique bounded solution of the functional equation

$$\begin{aligned} v_\delta{}^S(i) &= 0, && i \in I - S, \\ v_\delta{}^S(i) &= r(i, \delta(i)) + c \sum_{j \in I} g(j \mid i, \delta(i))\, v_\delta{}^S(i), && i \in S. \end{aligned} \tag{2}$$

Let the approximate optimal return function be $\hat{f} = \sup\{v_\delta^S, \delta \in \Delta\}$. Then $\hat{f}$ is the unique bounded solution of the functional equation

$$\begin{aligned} \hat{f}(i) &= 0, && i \in I - S, \\ \hat{f}(i) &= \sup_{k \in K_i} \left\{ r(i, k) + c \sum_{j \in I} g(j \mid i, k) \hat{f}(j) \right\}, && i \in S. \end{aligned} \tag{3}$$

## 4. The Bounds

Since $r(i, k) \geqslant 0$ for all $k \in K_i$ and $i \in I$, $v_\delta(i) \geqslant v_\delta^S(i)$ for all $i$, $\delta$ and $S$. Hence, if $\delta$ restricted to $S$ is $\epsilon$-optimal in the approximation, then

$$f(i) \geqslant v_\delta(i) \geqslant v_\delta^S(i) \geqslant \hat{f}(i) - \epsilon, \qquad i \in S. \tag{4}$$

Hence, to study $| v_\delta(i) - f(i)|$ for a policy $\delta$ which is $\epsilon$-optimal in the approximation, it suffices to focus on $| f(i) - \hat{f}(i)|$.

As an immediate consequence of (1) and (3), we obtain

$$| f(i) - \hat{f}(i)| \leqslant \sup_{k \in K_i} \left\{ c \sum_{j \in I} g(j \mid i, k) \, | f(j) - \hat{f}(j)| \right\} \tag{5}$$

cf. (15) of White [5]. We shall generate bounds using (5) in conjunction with a finite collection of nested subsets of the designated subset $S$. In particular, let $S_1, \ldots, S_m$ be subsets of $S$ such that

$$\phi = S_0 \subseteq S_1 \subseteq S_2 \subseteq \cdots \subseteq S_m = S \subseteq I = S_{m+1}.$$

Let $S_j^c$ be the complement of $S_j$ in $I$. We will bound the errors and transtions over these subsets by means of

$$\begin{aligned} x_j &= \sup\{| f(i) - \hat{f}(i)| : i \in S_j\}, \\ \pi(S_{l_2}^c \mid S_{l_1}) &= \sup \left\{ \sum_{j \in S_{l_2}^c} g(j \mid i, k) : k \in K_i \, , i \in S_{l_1} \right\}, && 1 \leqslant l_i \leqslant m, \\ P(m + 1 \mid i) &= \pi(S_m{}^c \mid S_i), \\ P(j \mid i) &= \pi(S_{j-1}^c \mid S_i) - \pi(S_j{}^c \mid S_i), && 1 \leqslant i, \ \ j \leqslant m. \end{aligned} \tag{6}$$

The function $P$ is the bounding one-step transition function for transitions between the subsets $S_1, \ldots, S_m$: $P(j \mid i)$ is the probability of a transition to $S_j - S_{j-1}$ from $S_i$. Of course, we need some structure in order for $P$ to be useful. In the worst case, $P(m + 1 \mid i) = 1$ for all $i$, which makes the bound below trivial.

Let $\mathbf{x}$ be the $m$-vector with components $x_j$; let $\mathbf{P}$ be the $m \times m$ substochastic matrix with $(i, j)$th entry $P(j \mid i)$; let $\mathbf{b}$ be the $m$-vector with components $b_j = (1 - c)^{-1} MP(m + 1 \mid j)$. Since $0 \leqslant r(i, k) \leqslant M$ for all $k \in K_i$ and $i \in I$, $|f(i) - \hat{f}(i)| \leqslant (1 - c)^{-1} M$. The component $b_j$ obviously is a bound on the one-step probability of leaving $S$ multiplied by a bound on the total error. Let $\mathbf{I}$ be the $m \times m$ identity matrix.

THEOREM. $\mathbf{x} \leqslant (\mathbf{I} - c\mathbf{P})^{-1} c\mathbf{b}$.

*Proof.* From (5),

$$x_l = \sup_{i \in S_l} \sup_{k \in K_i} \left\{ c \sum_{j \in I} g(j \mid i, k) \, |f(j) - \hat{f}(j)| \right\}$$

$$\leqslant \sup_{i \in S_l} \sup_{k \in K_i} \left\{ c \sum_{j \in I - S} g(j \mid i, k) \, |f(j)| + c \sum_{n=1}^{n=m} \sum_{j \in S_n - S_{n-1}} g(j \mid i, k) \, |f(j) - \hat{f}(j)| \right\}$$

$$\leqslant \sup_{i \in S_l} \sup_{k \in K_i} \left\{ \sum_{j \in I - S} g(j \mid i, k) \, c(1 - c)^{-1} M + \sum_{n=1}^{m} \sum_{j \in S_n - S_{n-1}} g(j \mid i, k) \, cx_n \right\}$$

$$\leqslant P(m + 1 \mid l) \, c(1 - c)^{-1} M + \sum_{n=1}^{m} P(n \mid l) \, cx_n$$

$$\leqslant cb_l + c \sum_{n=1}^{m} P(n \mid l) \, x_n .$$

The semi-penultimate inequality holds because $x_j \leqslant x_{j+1} \leqslant (1 - c)^{-1} M$ and the penultimate inequality holds because of the stochastic dominance associated with (6): for $i \in S_l$ and $k \in K_i$,

$$\sum_{n=1}^{m+1} y_n \sum_{j \in S_n - S_{n-1}} g(j \mid i, k) \leqslant \sum_{n=1}^{m+1} y_n P(n \mid l)$$

if $y_1 \leqslant y_2 \leqslant \cdots \leqslant y_{m+1}$, cf. p. 769 of Veinott [4]. ▮

EXAMPLE 1. Suppose $S_j = \{1, \ldots, j\}$, $1 \leqslant j \leqslant m$. Then $P(\cdot \mid i)$ is the supremum in the sense of stochastic order among the subprobability distributions on $\{1, \ldots, m\}$ in the collection $\{g(\cdot \mid j, k) \colon 1 \leqslant j \leqslant i, k \in K_j\}$. In other words,

$$\sum_{l=1}^{l=n} P(l \mid i) = \inf \left\{ \sum_{l=1}^{n} g(l \mid j, k) \colon 1 \leqslant j \leqslant i, k \in K_j \right\}, \qquad 1 \leqslant n \leqslant m.$$

EXAMPLE 2. Suppose there is some state in $S$, say $s$, such that there is a probability of at least $\epsilon > 0$ of reaching this state from every state in $S$ in one transition under every action, i.e., $g(s \mid i, k) > \epsilon$ for all $i \in S$ and $k \in K_i$. If we construct $S_1$ such that $s \in S_1$, then $P(1, i) \geqslant \epsilon$ for all $i$ and $j$, $1 \leqslant i, j \leqslant m$.

EXAMPLE 3. Following White (1977), let

$$\epsilon(N) = \sup \left\{ \sum_{j>N+i} g(j \mid i, k) : i \in I, k \in K_i \right\}.$$

If, for $i$ given, $S_j = \{1, \ldots, (j-1)N + i\}$, then $P(1 \mid j) + \cdots + P(j+1 \mid j) \geqslant 1 - \epsilon(N)$ for all $j$. Based on this limited information, the best possible bounding function $P$ is $P(m+1 \mid i) = \epsilon(N) = 1 - P(i+1 \mid i)$ for all $i$. The theorem here with this $P$ yields White's [5] bound.

EXAMPLE 4. Suppose we wish to concentrate on a fixed initial state $i$ and an approximation based on $k$ iterations. We can select the designated subset $S$ to achieve a specified bound on the error by using a cruder approximation which is similar to the one suggested by White [5]. First choose probabilities $p_j$, $1 \leqslant j \leqslant k$, so that the bound $B = (1-c)^{-1} M (\sum_{j=1}^{k} c^j p_j + c^{k+1})$ is satisfactory. (For example, we could set $p_j = p$ for all $j$.) Then choose subsets $S_j$, $1 \leqslant j \leqslant k$, such that $S_j = \{1, 2, \ldots, n_j\}$ and

$$n_1 = \min \left\{ n \geqslant 1 : \sup \left\{ \sum_{j=n}^{\infty} g(j \mid i, k) : k \in K_i \right\} \leqslant p_1 \right\}$$

$$n_j = \min \left\{ n \geqslant n_{j-1} : \sup \left\{ \sum_{j=n}^{\infty} g(j \mid l, k) : k \in K_l, l \in S_{j-1} \right\} \leqslant p_j \right\}.$$

This procedure guarantess that the bound $B$ is met, but the Theorem applied to the collection $\{S_1, \ldots, S_k\}$ gives a better bound.

## ACKNOWLEDGMENTS

I would like to thank R. E. Thomas of Bell Laboratories and D. J. White of the University of Manchester for helpful comments.

## REFERENCES

1. E. V. DENARDO, Contraction mappings in the theory underlying dynamic programming, *SIAM Rev.* **9** (1967), 165–177.
2. B. L. FOX, Finite-state approximation to denumerable-state dynamic programs, *J. Math. Anal. Appl.* **34** (1971), 665–670.

3. T. L. Morin, Computational advances and reduction of dimensionality in dynamic programming: A survey, *in* "Proceedings of the International Conference on Dynamic Programming, University of British Columbia, Vancouver" (M. Puterman, Ed.), 1979, in press.
4. A. F. Veinott, Jr., Optimal policy in a dynamic single product, nonstationary inventory model with several demand classes, *Operations Res.* 13 (1965), 761–778.
5. D. J. White, Finite state approximations for denumerable state infinite horizon discounted markov decision processes, Notes in Decision Theory Number 43, Department of Decision Theory, University of Manchester, 1977.
6. W. Whitt, Approximations of dynamic programs, II, *Math. Operations Res.* 4 (1979), 179–185.