*J. Opl Res. Soc.* Vol. 39, No. 2, pp. 217–220, 1988
Printed in Great Britain. All rights reserved

0160-5682/88 $3.00 + 0.00

# *Letters and Viewpoints*

(Contributions to this section are invited but we reserve the right to edit at our discretion—Ed.)

## NOTE ON "A PARTIALLY OBSERVABLE MARKOV DECISION PROCESS WITH LAGGED INFORMATION"

Kim and Jeong[1] state that '. . . using the information of lagged observation improves the total expected reward. . .' of a partially observed Markov decision process. This statement is supported by a numerical example but is not proved. In this note, we provide a proof of this important and interesting observation, indicating also the possibility that lagged information may not have a beneficial effect on the optimal cost-to-go function.

Throughout, we use the notation found in Kim and Jeong[1]. Let $d_{ij\delta\theta} = b_{j\theta} l_{i\delta}$. Then

$$d_{ij\delta\theta} = Pr\{S_c(t+1) = \theta, S_l(t+1) = \delta \mid S(t+1) = j, S(t) = i\}.$$

We remark that $b_{j\theta}$, $l_{i\delta}$, and hence $d_{ij\delta\theta}$ can depend on the action taken at control interval $t$. Let $V_n(\pi)(V'_n(\pi))$ be the maximum expected reward for $n$ remaining control intervals and initial state vector $\pi$, given an information source described by $\{d_{ij\delta\theta}\}$ $(\{b_{j\theta}\})$. A precise interpretation of the above statement due to Kim and Jeong, modified as stated above, is:

$$V_n(\pi) \geqslant V'_n(\pi) \quad \text{for all } n \text{ and } \pi.$$

Proof of this statement is based on Corollary 3.2 in White and Harrington,[2] which we note is easily extended to include observations dependent on both the current and the delayed (by one control interval) state. It is sufficient to show that there exists an array $\{\gamma(\delta, \theta, \delta', \theta')\}$ such that:

$$\gamma(\delta, \theta, \delta', \theta') \geqslant 0 \quad \text{for all } (\delta, \theta, \delta', \theta'), \tag{1a}$$

$$\sum_{\delta', \theta'} \gamma(\delta, \theta, \delta', \theta') = 1 \quad \text{for all } (\delta, \theta), \tag{1b}$$

and

$$\sum_{\delta, \theta} d_{ij\delta\theta}\, \gamma(\delta, \theta, \delta', \theta') = b_{j\theta'}/M \tag{1c}$$

for all $i, j, \delta'$ and $\theta'$. Let

$$\gamma(\delta, \theta, \delta', \theta') = H(\theta, \theta')/M,$$

where $H(\theta, \theta') = 1 (= 0)$ if $\theta = \theta'$ (if $\theta \neq \theta'$). Then it is easily shown that (1) holds.

*University of Virginia*

CHELSEA C. WHITE III

## REPLY

I would like to thank Professor White for his constructive comment.

Following the notation in the paper and the note, it is true that the information source $d_{ij\theta\delta}$ has better measurement quality than that of the information source $b_{j\theta}$. However, the maximum total expected reward does not always increase by adding lagged information. For instance, if the $L$ matrix has a relatively worse measurement quality than the $B$ matrix

$$\left(\text{e.g. } b_{j\theta} = \begin{bmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{bmatrix}, \quad l_{i\delta} = \begin{bmatrix} 0.6 & 0.4 \\ 0.4 & 0.6 \end{bmatrix}\right),$$

the maximum total expected reward does not increase (i.e. $V_n(\pi) = V'_n(\pi)$), and there is no change in the optimal policy. In other words, the value of the lagged information becomes zero in this