Correction to Hidden Markov Model Multi-arm Bandits: A Methodology for Bearm Scheduling in Multi-Target Tracking

Vikram Krishnamurthy

Department of Electrical Engineering

University of Melbourne,

Victoria 3010, Australia

email vikram@ee.mu.oz.au

August 19, 2002

We have discovered an error in the return-to-state formulation of the HMM multi-armed bandit problem in our recently published paper [4]. This note briefly outlines the error in [4] and describes a computationally simpler solution. Complete details including proofs of this simpler solution appear in the already submitted paper [3].

The error in [4] is in the return-to-state argument given in Result 1, equation (12). It should read:

\[ V^{(p)}(x^{(p)}, \bar{x}^{(p)}) = \min \left[ R'(p)x^{(p)} + \beta \sum_{m=1}^{M_p} V^{(p)} \left( \frac{B^{(p)}(m)A^{(p)}x^{(p)}}{\bar{Y}B^{(p)}(m)A^{(p)}x^{(p)}}, \bar{x}^{(p)} \right) 1'B^{(p)}(m)A^{(p)}x^{(p)} \right], \]

\[ R'(p)\bar{x}^{(p)} + \beta \sum_{m=1}^{M_p} V^{(p)} \left( \frac{B^{(p)}(m)A^{(p)}\bar{x}^{(p)}}{1'YB^{(p)}(m)A^{(p)}\bar{x}^{(p)}}, \tilde{x}^{(p)} \right) 1'B^{(p)}(m)A^{(p)}\bar{x}^{(p)} \]  

(1)

Unfortunately, there appears no obvious way of obtaining a finite dimensional characterization of the above value function in the return-to-state argument.

In this note, it is shown that by introducing the retirement formulation [2] of the multi-armed bandit problem option, a finite dimensional value iteration algorithm can be obtained for computing

*This work was supported by an ARC large grant and the Centre of Expertise in Networked Decision Systems
the Gittins index of a POMDP bandit. To avoid repetition, we use exactly the same notation as [4].

For each project $p$, let $M^{(p)}$ denote a positive real number such that

$$0 \leq M^{(p)} \leq \bar{M}^{(p)} \triangleq \max_{i \in \mathcal{N}_p} R(p, i).$$

(2)

To simplify subsequent notation, we omit the superscript $p$ in $M^{(p)}$ and $\bar{M}^{(p)}$.

In terms of a parameterized retirement reward $M$, the Gittins index [1], [2] of project $p$ with information state $x^{(p)}$ can be defined as

$$\gamma^{(p)}(x^{(p)}) \triangleq \min \{ M : V^{(p)}(x, M) = M \}$$

(3)

where $V^{(p)}(x, M)$ satisfies the functional Bellman’s recursion

$$V^{(p)}(x, M) = \max \left\{ R^{(p)}x^{(p)} + \beta \sum_{m=1}^{M_p} V^{(p)} \left( \frac{B^{(p)}(m) A^{(p)} x^{(p)}}{1 - \mathbf{1}_{\mathcal{N}_p} B^{(p)}(m) A^{(p)} x^{(p)}}, M \right) \mathbf{1}_{\mathcal{N}_p} B^{(p)}(m) A^{(p)} x^{(p)}, M \right\}$$

(4)

The $N$th order approximation of $V^{(p)}(x^{(p)}, M)$ is obtained as the following value iteration algorithm $k = 1, \ldots, N$:

$$V^{(p)}_{k+1}(x^{(p)}, M) = \min \left[ R^{(p)}x^{(p)} + \beta \sum_{m=1}^{M_p} V^{(p)}_{k} \left( \frac{B^{(p)}(m) A^{(p)} x^{(p)}}{1 - \mathbf{1}_{\mathcal{N}_p} B^{(p)}(m) A^{(p)} x^{(p)}}, M \right) \mathbf{1}_{\mathcal{N}_p} B^{(p)}(m) A^{(p)} x^{(p)}, M \right]$$

(5)

Note that Corollary 1 of [4] holds for Gittins index. Similar to Sec.2.4 in [4] we need to compute a finite dimensional characterization of the value iteration algorithm (5). This is done as follows: Define the $(N_p + 1)$ dimensional augmented information state $\bar{x} \in \{ [x', 0]', [0', 1] \}$ where $x \in \mathcal{X}^{(p)}$. Define an augmented observation process $\bar{y}_k \in \{1, \ldots, M_p + 1\}$ and the corresponding $(N_p + 1) \times (N_p + 1)$
transition and observation probability matrices as

\[
A_1^{(p)} = \begin{bmatrix} A^{(p)} & 0_N \end{bmatrix}, \quad B_1^{(p)}(m) = \begin{bmatrix} B^{(p)}(m) & 0_N \end{bmatrix},
\]

\[
A_2^{(p)} = \begin{bmatrix} 0_{N_p \times N_p} & 1_N \end{bmatrix}, \quad B_2^{(p)}(m) = I_{(N_p+1) \times (N_p+1)}
\]  

\[
\begin{equation}
B_1^{(p)}(m) = \text{diag}(\text{column } m \text{ of } B_1^{(p)}); \quad B_2^{(p)}(m) = \text{diag}(\text{column } m \text{ of } B_2^{(p)}), \quad m \in \{1, \ldots, M_p+1\}
\end{equation}
\]

\[
\begin{align*}
\pi & = z \otimes \bar{x} \\
\bar{A}_1 = I_{2 \times 2} \otimes A_1 = \begin{bmatrix} A_1 & 0 \\ 0 & A_1 \end{bmatrix}, & \quad \bar{A}_2 = I_{2 \times 2} \otimes A_2 = \begin{bmatrix} A_2 & 0 \\ 0 & A_2 \end{bmatrix}, \\
\bar{R}_1(p) = \begin{bmatrix} R^c(p) & 0 \\ R^c(p) & 0 \end{bmatrix}, & \quad \bar{R}_2(p) = \begin{bmatrix} M \lambda_{i,N}^{c'} & 0 \\ 0 & 0 \end{bmatrix}
\end{align*}
\]

\[
\begin{equation}
\gamma_N^{(p)}(x^{(p)}) = \max_{\lambda_{i,N} \in \Lambda_N} \frac{M \lambda_{i,N}^{c'}(3) x^{(p)}}{(\lambda_{i,N}(3) - \lambda_{i,N}(1)) x^{(p)} + M}
\end{equation}
\]

where each vector \( \lambda_{i,N} \in \Lambda_N^{(p)} \) is of the form

\[
\lambda_{i,k} = \begin{bmatrix} \lambda_{i,k}(1) & 0 \\ \lambda_{i,k}(3) & 0 \end{bmatrix}, \quad \text{where } \lambda_{i,k}(1), \lambda_{i,k}(3) \in \mathbb{R}^{N_p}
\]
Statement 3 above gives an explicit formula for the Gittins index of the HMM multi-armed bandit problem. Recall $x_k^{(p)}$ is the information state computed by the $p$th HMM filter at time $k$. Given that we can compute set of vectors $A_N^{(p)}$, (9) gives an explicit expression for the Gittins index $\gamma_N^{(p)}(x_k^{(p)})$ at any time $k$ for project $p$. Note if all elements of $R(p)$ are identical, then $\gamma^{(p)}(x) = \bar{M}$ for all $x$.

Sections 2.5, 3 and 4 of the paper [4], including the beam scheduling algorithm for a hybrid sensor of Sec. 3.2, still hold.

It is worthwhile noting that the above solution is computationally simpler as the information state $\pi$ is a $2(N_p + 1)$ dimensional vector, whereas the information state $\pi$ considered in [4] is a $N_p^2$ dimensional vector.

References


