Online Supplement to "Computing a Classic Index for Finite-Horizon Bandits"

José Niño-Mora
Department of Statistics, Carlos III University of Madrid, C/ Madrid 126, 28903 Getafe (Madrid), Spain, {jnimora@alum.mit.edu}

This online supplement illustrates the workings of the RAG index algorithm presented in the companion paper by means of a small example. It can be of use to researchers and to programmers seeking to check the algorithm's workings step by step.

Consider the project instance with $n = 3$ states, horizon $T = 3$, discount factor $\beta = 1$, rewards $\mathbf{R} = (0.5, 0.9, 0.3)$, and transition probability matrix

$$\mathbf{P} = \begin{bmatrix} 0.25 & 0.40 & 0.35 \\ 0.05 & 0.45 & 0.50 \\ 0.30 & 0.60 & 0.10 \end{bmatrix}.$$

We start with horizon $d = 1$, constructing the initial table

	$k=1$	2	3
(s_1^k, j_1^k)	$(1,2)$	$(1,1)$	$(1,3)$
$\lambda^*(s_1^k, j_1^k)$	0.9	0.5	0.3
$w_1^{k-1}(1)$	1	$\boxed{1}$	1
$w_1^{k-1}(2)$	$\boxed{1}$	1	1
$w_1^{k-1}(3)$	1	1	$\boxed{1}$
$r_1^{k-1}(1)$	0.5	$\boxed{0.5}$	0.5
$r_1^{k-1}(2)$	$\boxed{0.9}$	0.9	0.9
$r_1^{k-1}(3)$	0.3	0.3	$\boxed{0.3}$

For horizon $d = 2$, we obtain the updated table

	$k=1$	2	3	4
(s_2^k, j_2^k)	$(2,2)$	$(1,2)$	$(2,1)$	$(2,3)$
$\lambda^*(s_2^k, j_2^k)$	0.9	0.9	0.6143	0.525
$w_2^{k-1}(1)$	1	1	$\boxed{1.400}$	1.400
$w_2^{k-1}(2)$	$\boxed{1}$	1	1.450	1.450
$w_2^{k-1}(3)$	1	1	1.600	$\boxed{1.600}$
$r_2^{k-1}(1)$	0.5	0.5	$\boxed{0.860}$	0.860
$r_2^{k-1}(2)$	$\boxed{0.9}$	0.9	1.305	1.305
$r_2^{k-1}(3)$	0.3	0.3	0.840	$\boxed{0.840}$

As for the quantities $\widehat{w}_2^{k-1}(i)$ and $\widehat{r}_2^{k-1}(i)$ in the algorithm's block implementation, they are obtained by the updates

$$\begin{bmatrix} \widehat{w}_2^0(1) & \widehat{w}_2^1(1) & \widehat{w}_2^2(1) & \widehat{w}_2^3(1) \\ \widehat{w}_2^0(2) & \widehat{w}_2^1(2) & \widehat{w}_2^2(2) & \widehat{w}_2^3(2) \\ \widehat{w}_2^0(3) & \widehat{w}_2^1(3) & \widehat{w}_2^2(3) & \widehat{w}_2^3(3) \end{bmatrix} = \begin{bmatrix} 0.25 & 0.40 & 0.35 \\ 0.05 & 0.45 & 0.50 \\ 0.30 & 0.60 & 0.10 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 & 1.400 \\ 0 & 1 & 1.450 & 1.450 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$
$$= \begin{bmatrix} 0 & 0.40 & 0.5800 & 0.9300 \\ 0 & 0.45 & 0.6525 & 0.7225 \\ 0 & 0.60 & 0.8700 & 1.2900 \end{bmatrix}$$

and

$$\begin{bmatrix} \widehat{r}_2^0(1) & \widehat{r}_2^1(1) & \widehat{r}_2^2(1) & \widehat{r}_2^3(1) \\ \widehat{r}_2^0(2) & \widehat{r}_2^1(2) & \widehat{r}_2^2(2) & \widehat{r}_2^3(2) \\ \widehat{r}_2^0(3) & \widehat{r}_2^1(3) & \widehat{r}_2^2(3) & \widehat{r}_2^3(3) \end{bmatrix} = \begin{bmatrix} 0.25 & 0.40 & 0.35 \\ 0.05 & 0.45 & 0.50 \\ 0.30 & 0.60 & 0.10 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 & 0.860 \\ 0 & 0.9 & 1.305 & 1.305 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$
$$= \begin{bmatrix} 0 & 0.3600 & 0.5220 & 0.7370 \\ 0 & 0.4050 & 0.5872 & 0.6302 \\ 0 & 0.5400 & 0.7830 & 1.0410 \end{bmatrix}.$$

For horizon $d = 3$, we obtain the following final table, which gives the index values for all states. Thus, e.g., $\lambda^*(2, 1) = 0.6143$.

	$k = 1$	2	3	4	5	6
(s_3^k, j_3^k)	$(3, 2)$	$(2, 2)$	$(1, 2)$	$(3, 1)$	$(2, 1)$	$(3, 3)$
$\lambda^*(s_3^k, j_3^k)$	0.9	0.9	0.9	0.6468	0.6143	0.5856
$w_3^{k-1}(1)$	1	1	1.400	$\boxed{1.58000}$	1.58000	1.93000
$w_3^{k-1}(2)$	$\boxed{1}$	1	1.450	1.65250	1.65250	1.72250
$w_3^{k-1}(3)$	1	1	1.600	1.87000	1.87000	$\boxed{2.29000}$
$r_3^{k-1}(1)$	0.5	0.5	0.860	$\boxed{1.02200}$	1.02200	1.23700
$r_3^{k-1}(2)$	$\boxed{0.9}$	0.9	1.305	1.48725	1.48725	1.53025
$r_3^{k-1}(3)$	0.3	0.3	0.840	1.08300	1.08300	$\boxed{1.34100}$