Understanding kernel computations

Gianvito Losapio

October 26, 2022

Abstract

Kernels are explained

Contents

1 Data

Problem setting

- States: $|S| = 2 \longrightarrow S = \{0, 1\}$
- Actions: $|A| = 2 \longrightarrow A = \{0, 1\}$
- Observations: $|\Omega| = 2 \longrightarrow \Omega = \{0, 1\}$

MDP

- 1D transition matrix: $T(s, a) = s'$

$$T = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}_{|A||S|}$$

- 1D inverse transition matrix: $T_{\text{inv}}(s', a) = s$

$$T_{\text{inv}} = \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix}_{|A||S|}$$

- Observation matrix: $O(a, s', o) = \Pr(o \,|\, a_t = a, s_{t+1} = s')$

$$O = \begin{bmatrix} \begin{bmatrix} 0.1 & 0.2 \\ 0.1 & 0.2 \end{bmatrix}, & \begin{bmatrix} 0.1 & 0.2 \\ 0.1 & 0.2 \end{bmatrix} \end{bmatrix}_{|A||S||\Omega|}$$

- Reward: $R(s, a)$

$$R = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}_{|S||A|}$$

Alpha vectors + belief set

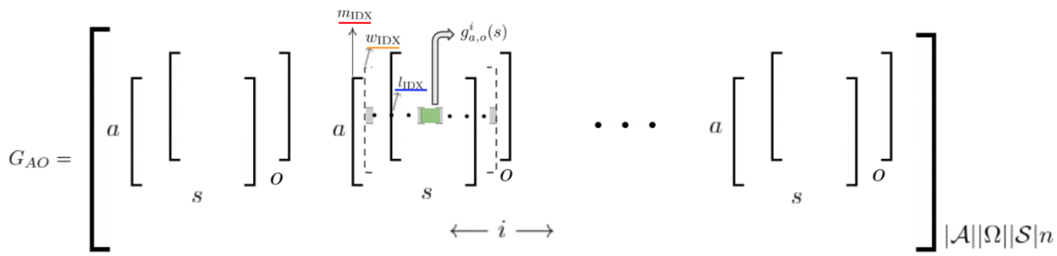
- Alpha vectors: $n = 2$

$$\Gamma = \begin{bmatrix} \begin{bmatrix} 1 & 2 \end{bmatrix} \\ \begin{bmatrix} 0 & 3 \end{bmatrix} \end{bmatrix}_{n|S|}$$

- Belief set: $|B| = 2$

$$B = \begin{bmatrix} \begin{bmatrix} 0.7 & 0.3 \end{bmatrix} \\ \begin{bmatrix} 0.4 & 0.6 \end{bmatrix} \end{bmatrix}_{|B||S|}$$

2 Compute G_{AO}



$$
\begin{aligned}
g^i_{a,o}(s) &= \sum_{s'} O(a,s',o)T(s,a,s')\alpha_i(s') \\
&= O(a,s',o)\underbrace{T(s,a,s')}_{=1}\alpha_i(s') \quad \text{with } s' = T(s,a)
\end{aligned}
$$

Access:

$$g^i_{a,o}(s) = G_{AO}\underbrace{[\,i\,]}_{\text{4D}}\underbrace{[\,o\,]}_{\text{3D}}\underbrace{[\,a\,][\,s\,]}_{\text{2D}}$$

$$O(a,s',o) = O\underbrace{[\,o\,]}_{\text{3D}}\underbrace{[\,a\,][\,s'\,]}_{\text{2D}}$$

$$\alpha_i(s') = \Gamma[\,i\,][\,s'\,]$$

Computation:

$i = 0, o = 0$

$$G_{AO}[i=0][o=0][a=0][s=0] \leftarrow O[o=0][a=0]\underbrace{[s'=0]}_{T[a=0][s=0]}\cdot\Gamma[i=0][s'=0] = 0.1\cdot 1 = 0.1$$

$$G_{AO}[i=0][o=0][a=0][s=1] \leftarrow O[o=0][a=0]\underbrace{[s'=1]}_{T[a=0][s=1]}\cdot\Gamma[i=0][s'=1] = 0.2\cdot 2 = 0.4$$

$$G_{AO}[i=0][o=0][a=1][s=0] \leftarrow O[o=0][a=1]\underbrace{[s'=1]}_{T[a=1][s=0]}\cdot\Gamma[i=0][s'=1] = 0.2\cdot 2 = 0.4$$

$$G_{AO}[i=0][o=0][a=1][s=1] \leftarrow O[o=0][a=1]\underbrace{[s'=1]}_{T[a=1][s=1]}\cdot\Gamma[i=0][s'=1] = 0.2\cdot 2 = 0.4$$

$i = 0, o = 1$

$$G_{AO}[i=0][o=1][a=0][s=0] \leftarrow O[o=1][a=0]\underbrace{[s'=0]}_{T[a=0][s=0]}\cdot\Gamma[i=0][s'=0] = 0.1\cdot 1 = 0.1$$

$$G_{AO}[i=0][o=1][a=0][s=1] \leftarrow O[o=1][a=0]\underbrace{[s'=1]}_{T[a=0][s=1]}\cdot\Gamma[i=0][s'=1] = 0.2\cdot 2 = 0.4$$

$$G_{AO}[i=0][o=1][a=1][s=0] \leftarrow O[o=1][a=1]\underbrace{[s'=1]}_{T[a=1][s=0]}\cdot\Gamma[i=0][s'=1] = 0.2\cdot 2 = 0.4$$

$$G_{AO}[i=0][o=1][a=1][s=1] \leftarrow O[o=1][a=1]\underbrace{[s'=1]}_{T[a=1][s=1]}\cdot\Gamma[i=0][s'=1] = 0.2\cdot 2 = 0.4$$

$i = 1, o = 0$

$G_{AO}[i=1][o=0][a=0][s=0] \leftarrow O[o=0][a=0] \underbrace{[s'=0]}_{T[a=0][s=0]} \cdot\Gamma[i=1][s'=0] = 0.1 \cdot 0 = 0$

$G_{AO}[i=1][o=0][a=0][s=1] \leftarrow O[o=0][a=0] \underbrace{[s'=1]}_{T[a=0][s=1]} \cdot\Gamma[i=1][s'=1] = 0.2 \cdot 3 = 0.6$

$G_{AO}[i=1][o=0][a=1][s=0] \leftarrow O[o=0][a=1] \underbrace{[s'=1]}_{T[a=1][s=0]} \cdot\Gamma[i=1][s'=1] = 0.2 \cdot 3 = 0.6$

$G_{AO}[i=1][o=0][a=1][s=1] \leftarrow O[o=0][a=1] \underbrace{[s'=1]}_{T[a=1][s=1]} \cdot\Gamma[i=1][s'=1] = 0.2 \cdot 3 = 0.6$

$i = 1, o = 1$

$G_{AO}[i=1][o=1][a=0][s=0] \leftarrow O[o=1][a=0] \underbrace{[s'=0]}_{T[a=0][s=0]} \cdot\Gamma[i=1][s'=0] = 0.1 \cdot 0 = 0$

$G_{AO}[i=1][o=1][a=0][s=1] \leftarrow O[o=1][a=0] \underbrace{[s'=1]}_{T[a=0][s=1]} \cdot\Gamma[i=1][s'=1] = 0.2 \cdot 3 = 0.6$

$G_{AO}[i=1][o=1][a=1][s=0] \leftarrow O[o=1][a=1] \underbrace{[s'=1]}_{T[a=1][s=0]} \cdot\Gamma[i=1][s'=1] = 0.2 \cdot 3 = 0.6$

$G_{AO}[i=1][o=1][a=1][s=1] \leftarrow O[o=1][a=1] \underbrace{[s'=1]}_{T[a=1][s=1]} \cdot\Gamma[i=1][s'=1] = 0.2 \cdot 3 = 0.6$

The final matrix is

$$G_{AO} = \begin{bmatrix} \begin{bmatrix} \begin{array}{cc} & o=0 \\ a & \begin{bmatrix} 0.1 & 0.4 \\ 0.4 & 0.4 \end{bmatrix} \\ & s \end{array}, & \begin{array}{c} o=1 \\ \begin{bmatrix} 0.1 & 0.4 \\ 0.4 & 0.4 \end{bmatrix} \\ s \end{array} \end{bmatrix} & i=0 \\ \begin{bmatrix} \begin{array}{cc} & o=0 \\ a & \begin{bmatrix} 0 & 0.6 \\ 0.6 & 0.6 \end{bmatrix} \\ & s \end{array}, & \begin{array}{c} o=1 \\ \begin{bmatrix} 0 & 0.6 \\ 0.6 & 0.6 \end{bmatrix} \\ s \end{array} \end{bmatrix} & i=1 \end{bmatrix}$$

Kernel computation example: element $G_{AO}[i=1][o=1][a=0][s=1] = 0.6$

$l_{\text{IDX}} = 13$

$i = l_{\text{IDX}}/(|A||\Omega||S|) = 13/8 = 1$

$m_{\text{IDX}} = i|A||\Omega|S| = 1 * 8 = 8$

$o = (l_{\text{IDX}} - m_{\text{IDX}})/|A|S| = (13-8)/4 = 5/4 = 1$

$a = (l_{\text{IDX}} - m_{\text{IDX}} - o|A||S|)/|S| = (13 - 8 - 1 * 4)/2 = 1/2 = 0$

$s = (l_{\text{IDX}} - m_{\text{IDX}} - o|A||S|) - a|S| = 1 - 0 = 1$

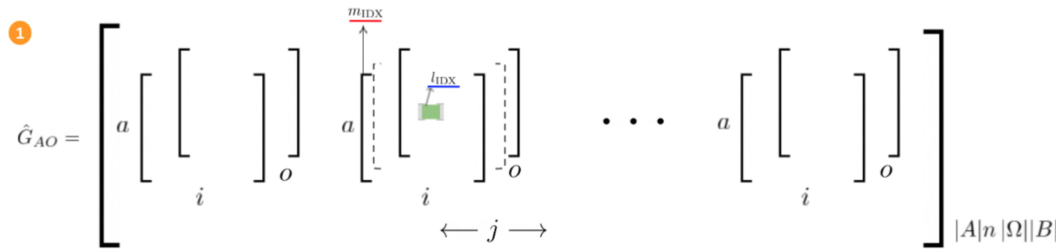
$T[a|S| + s] = T[0+1] = T[1] = s' = 1$

$O[o|A||S| + a|S| + s'] = O[1 * 4 + 0 + 1] = O[5] = 0.2$

$\Gamma[i|S| + s] = \Gamma[1 * 2 + 1] = \Gamma[3] = 3$

$G_{AO}[l_{\text{IDX}}] = G_{AO}[13] = O[5] * T[3] = 0.2 * 3 = 0.6$

3 Compute $\hat{G}_{AO}$



$$g_a^b = r_a + \gamma \sum_o \underset{\{g_{a,o}^i\}_i}{\arg\max} \boxed{b \cdot g_{a,o}^i}$$

Access:

$$\hat{G}_{AO} \underbrace{[\,j\,]}_{\text{4D}} \underbrace{[\,o\,]}_{\text{3D}} \underbrace{[\,a\,][\,i\,]}_{\text{2D}} = b_j \cdot g_{a,o}^i$$

$g_{a,o}^i = G_{AO}[\,i\,][\,o\,][\,a\,][\,:\,] \longrightarrow$ find linear range

$b_j = B[\,j\,][\,:\,] \longrightarrow$ find linear range

Computation:

$j = 0, o = 0$

$\hat{G}_{AO}[j = 0][o = 0][a = 0][i = 0] \leftarrow B[j = 0][\,:\,] \cdot G_{AO}[i = 0][o = 0][a = 0][\,:\,] = [0.7\ 0.3] \cdot [0.1\ 0.4]^T = 0.07 + 0.12 = 0.19$

$\hat{G}_{AO}[j = 0][o = 0][a = 0][i = 1] \leftarrow B[j = 0][\,:\,] \cdot G_{AO}[i = 1][o = 0][a = 0][\,:\,] = [0.7\ 0.3] \cdot [0\ 0.6]^T = 0 + 0.18 = 0.18$

$\hat{G}_{AO}[j = 0][o = 0][a = 1][i = 0] \leftarrow B[j = 0][\,:\,] \cdot G_{AO}[i = 0][o = 0][a = 1][\,:\,] = [0.7\ 0.3] \cdot [0.4\ 0.4]^T = 0.28 + 0.12 = 0.4$

$\hat{G}_{AO}[j = 0][o = 0][a = 1][i = 1] \leftarrow B[j = 0][\,:\,] \cdot G_{AO}[i = 1][o = 0][a = 1][\,:\,] = [0.7\ 0.3] \cdot [0.6\ 0.6]^T = 0.42 + 0.18 = 0.6$

$j = 0, o = 1$

$\hat{G}_{AO}[j = 0][o = 1][a = 0][i = 0] \leftarrow B[j = 0][\,:\,] \cdot G_{AO}[i = 0][o = 1][a = 0][\,:\,] = [0.7\ 0.3] \cdot [0.1\ 0.4]^T = 0.07 + 0.12 = 0.19$

$\hat{G}_{AO}[j = 0][o = 1][a = 0][i = 1] \leftarrow B[j = 0][\,:\,] \cdot G_{AO}[i = 1][o = 1][a = 0][\,:\,] = [0.7\ 0.3] \cdot [0\ 0.6]^T = 0 + 0.18 = 0.18$

$\hat{G}_{AO}[j = 0][o = 1][a = 1][i = 0] \leftarrow B[j = 0][\,:\,] \cdot G_{AO}[i = 0][o = 1][a = 1][\,:\,] = [0.7\ 0.3] \cdot [0.4\ 0.4]^T = 0.28 + 0.12 = 0.4$

$\hat{G}_{AO}[j = 0][o = 1][a = 1][i = 1] \leftarrow B[j = 0][\,:\,] \cdot G_{AO}[i = 1][o = 1][a = 1][\,:\,] = [0.7\ 0.3] \cdot [0.6\ 0.6]^T = 0.42 + 0.18 = 0.6$

$j = 1, o = 0$

$\hat{G}_{AO}[j = 1][o = 0][a = 0][i = 0] \leftarrow B[j = 1][\,:\,] \cdot G_{AO}[i = 0][o = 0][a = 0][\,:\,] = [0.4\ 0.6] \cdot [0.1\ 0.4]^T = 0.04 + 0.24 = 0.28$

$\hat{G}_{AO}[j=1][o=0][a=0][i=1] \leftarrow B[j=1][\,:\,] \cdot G_{AO}[i=1][o=0][a=0][\,:\,] = [0.4\ 0.6] \cdot [0\ 0.6]^T = 0 + 0.36 = 0.36$

$\hat{G}_{AO}[j=1][o=0][a=1][i=0] \leftarrow B[j=1][\,:\,] \cdot G_{AO}[i=0][o=0][a=1][\,:\,] = [0.4\ 0.6] \cdot [0.4\ 0.4]^T = 0.16 + 0.24 = 0.4$

$\hat{G}_{AO}[j=1][o=0][a=1][i=1] \leftarrow B[j=1][\,:\,] \cdot G_{AO}[i=1][o=0][a=1][\,:\,] = [0.4\ 0.6] \cdot [0.6\ 0.6]^T = 0.24 + 0.36 = 0.6$

$j = 1, o = 1$

$\hat{G}_{AO}[j=1][o=1][a=0][i=0] \leftarrow B[j=1][\,:\,] \cdot G_{AO}[i=0][o=1][a=0][\,:\,] = [0.4\ 0.6] \cdot [0.1\ 0.4]^T = 0.04 + 0.24 = 0.28$

$\hat{G}_{AO}[j=1][o=1][a=0][i=1] \leftarrow B[j=1][\,:\,] \cdot G_{AO}[i=1][o=1][a=0][\,:\,] = [0.4\ 0.6] \cdot [0\ 0.6]^T = 0 + 0.36 = 0.36$

$\hat{G}_{AO}[j=1][o=1][a=1][i=0] \leftarrow B[j=1][\,:\,] \cdot G_{AO}[i=0][o=1][a=1][\,:\,] = [0.4\ 0.6] \cdot [0.4\ 0.4]^T = 0.16 + 0.24 = 0.4$

$\hat{G}_{AO}[j=1][o=1][a=1][i=1] \leftarrow B[j=1][\,:\,] \cdot G_{AO}[i=1][o=1][a=1][\,:\,] = [0.4\ 0.6] \cdot [0.6\ 0.6]^T = 0.24 + 0.36 = 0.6$

The final matrix is

$$\hat{G}_{AO} = \begin{bmatrix} \begin{bmatrix} \begin{array}{cc} & o=0 \\ a & \begin{bmatrix} 0.19 & 0.18 \\ 0.4 & 0.6 \end{bmatrix} \\ & i \end{array}, & \begin{array}{c} o=1 \\ \begin{bmatrix} 0.19 & 0.18 \\ 0.4 & 0.6 \end{bmatrix} \\ i \end{array} \end{bmatrix} & j=0 \\ \begin{bmatrix} \begin{array}{cc} & o=0 \\ a & \begin{bmatrix} 0.28 & 0.36 \\ 0.4 & 0.6 \end{bmatrix} \\ & i \end{array}, & \begin{array}{c} o=1 \\ \begin{bmatrix} 0.28 & 0.36 \\ 0.4 & 0.6 \end{bmatrix} \\ i \end{array} \end{bmatrix} & j=1 \end{bmatrix}$$

Kernel computation example: element $\hat{G}_{AO}[j=1][o=1][a=1][i=0] = 0.4$

$l_{\text{IDX}} = 14$

$j = l_{\text{IDX}}/(|A|n|\Omega|) = 14/8 = 1$

$m_{\text{IDX}} = j|A|n|\Omega| = 1 * 8 = 8$

$o = (l_{\text{IDX}} - m_{\text{IDX}})/|A|n = (14-8)/4 = 6/4 = 1$

$a = (l_{\text{IDX}} - m_{\text{IDX}} - o|A|n)/n = (14 - 8 - 1*4)/2 = 2/2 = 1$

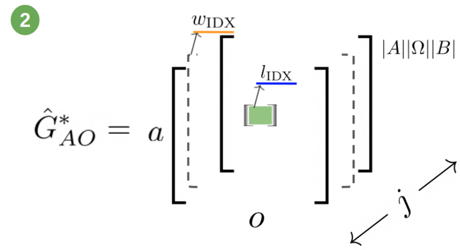
$i = (l_{\text{IDX}} - m_{\text{IDX}} - o|A|n) - an = 2 - 1*2 = 0$

$B[\,j|S|\ :\ ... + |S| - 1] = B[1*2\ :\ ... + 1] = B[2:3] = [0.4\ 0.6]$

$G_{AO}[\,i|A||\Omega||S| + o|A||S| + a|S|\ :\ ... + |S| - 1] = G_{AO}[0 + 1*4 + 1*2\ :\ +1] = G_{AO}[6:7] = [0.4\ 0.4]$

$\hat{G}_{AO}[l_{\text{IDX}}] = G_{AO}[14] = B[2:3] \cdot G_{AO}[6:7] = [0.4\ 0.6] \cdot [0.4\ 0.4]^T = 0.16 + 0.24 = 0.4$

4 Compute $\hat{G}^*_{AO}$



$$g_a^b = r_a + \gamma \sum_o \boxed{\arg\max_{\{g^i_{a,o}\}_i} b \cdot g^i_{a,o}}$$

Access:

$$\hat{G}^*_{AO} \underbrace{[\,j\,]}_{\text{3D}} \underbrace{[\,a\,][\,o\,]}_{\text{2D}} = i^*_{a,o,j} = \arg\max_i \; b_j \cdot g^i_{a,o}$$

$$b_j \cdot g^i_{a,o} = \hat{G}_{AO}[\,j\,][\,o\,][\,a\,][\,i\,]$$

Computation:

$$\hat{G}^*_{AO}[j=0][a=0][o=0] = \arg\max_i \left\{ \hat{G}_{AO}[j=0][o=0][a=0][\,i\,] \,\middle|\, i=0,1 \right\} = 0$$

$$\hat{G}^*_{AO}[j=0][a=0][o=1] = \arg\max_i \left\{ \hat{G}_{AO}[j=0][o=1][a=0][\,i\,] \,\middle|\, i=0,1 \right\} = 0$$

$$\hat{G}^*_{AO}[j=0][a=1][o=0] = \arg\max_i \left\{ \hat{G}_{AO}[j=0][o=0][a=1][\,i\,] \,\middle|\, i=0,1 \right\} = 1$$

$$\hat{G}^*_{AO}[j=0][a=1][o=1] = \arg\max_i \left\{ \hat{G}_{AO}[j=0][o=1][a=1][\,i\,] \,\middle|\, i=0,1 \right\} = 1$$

$$\hat{G}^*_{AO}[j=1][a=0][o=0] = \arg\max_i \left\{ \hat{G}_{AO}[j=1][o=0][a=0][\,i\,] \,\middle|\, i=0,1 \right\} = 1$$

$$\hat{G}^*_{AO}[j=1][a=0][o=1] = \arg\max_i \left\{ \hat{G}_{AO}[j=1][o=1][a=0][\,i\,] \,\middle|\, i=0,1 \right\} = 1$$

$$\hat{G}^*_{AO}[j=1][a=1][o=0] = \arg\max_i \left\{ \hat{G}_{AO}[j=1][o=0][a=1][\,i\,] \,\middle|\, i=0,1 \right\} = 1$$

$$\hat{G}^*_{AO}[j=1][a=1][o=1] = \arg\max_i \left\{ \hat{G}_{AO}[j=1][o=1][a=1][\,i\,] \,\middle|\, i=0,1 \right\} = 1$$

The final matrix is

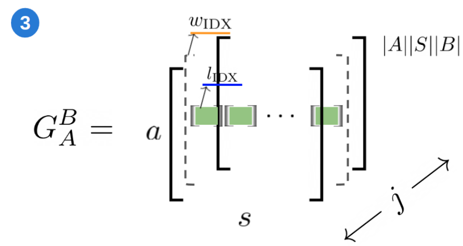
$$\hat{G}^*_{AO} = \begin{bmatrix} \begin{array}{ccc} a & \begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix} & j=0 \\ & o & \end{array} \\ \\ \begin{array}{ccc} a & \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} & j=1 \\ & o & \end{array} \end{bmatrix}$$

Kernel computation example: element $\hat{G}^*_{AO}[j=1][a=0][o=0] = 1$

$l_{\text{IDX}} = 4$

$$j = l_{\text{IDX}}/|A||\Omega| = 4/4 = 1$$
$$w_{\text{IDX}} = j|A||\Omega| = 1 * 4 = 4$$
$$a = (l_{\text{IDX}} - w_{\text{IDX}})/|\Omega| = (4-4)/2 = 0$$
$$o = (l_{\text{IDX}} - w_{\text{IDX}}) - a|\Omega| = 0$$
$$i_k = \{j|A|n|\Omega| + o|A|n + an + k \mid k = 0,1\} = \{1*8+k \mid k=0,1\} = \{8,9\}$$
$$\hat{G}^*_{AO}[l_{\text{IDX}}] = \arg\max_k\ \hat{G}_{AO}[\,i_k\,] = \arg\max_k\{G_{AO}[i_{k=0}=8], G_{AO}[i_{k=1}=9]\} = \arg\max_k\{\underbrace{0.28}_{k=0}, \underbrace{0.36}_{k=1}\} = 1$$

5 Compute G_A^B



$$g_a^b = \boxed{r_a + \gamma \sum_o \arg\max_{\{g^i_{a,o}\}_i} b \cdot g^i_{a,o}}$$

Access:

$$G_A^B[\,j\,][\,a\,][\,:\,] = \underbrace{r_a}_{\text{access } R_{SA}} + \gamma \underbrace{\sum_o \arg\max_{\{g^i_{a,o}\}_i} g^i_{a,o}}_{\text{access } \hat{G}^*_{AO},\, G_{AO}}$$

$r_a = R_{SA}[\,a\,][\,:\,] \longrightarrow$ find linear range

$\arg\max_{\{g^i_{a,o}\}_i} g^i_{a,o} = G_{AO}[i^*_{a,o,j}][\,o\,][\,a\,][\,:\,]$ with $i^*_{a,o,j} = \hat{G}^*_{AO}[\,j\,][\,a\,][\,o\,]$

$\sum_o \arg\max_{\{g^i_{a,o}\}_i} g^i_{a,o} = G_{AO}[i^*_{a,o,j}][\,:\,][\,a\,][\,:\,] \longrightarrow$ find double linear range, one for the index s, one for the index o

Computation:

$$G_A^B[j=0][a=0][s=0] = R_{SA}[a=0][s=0] +$$
$$+\gamma\Big(G_{AO}\underbrace{[i=0]}_{\hat{G}^*_{AO}[j=0][a=0][o=0]}[o=0][a=0][s=0] + G_{AO}\underbrace{[i=0]}_{\hat{G}^*_{AO}[j=0][a=0][o=1]}[o=1][a=0][s=0]\Big) =$$
$$= 0 + (0.1 + 0.1) = 0.2$$

$$G_A^B[j=0][a=0][s=1] = R_{SA}[a=0][s=1] +$$
$$+\gamma\Big(G_{AO}\underbrace{[i=0]}_{\hat{G}^*_{AO}[j=0][a=0][o=0]}[o=0][a=0][s=1] + G_{AO}\underbrace{[i=0]}_{\hat{G}^*_{AO}[j=0][a=0][o=1]}[o=1][a=0][s=1]\Big) =$$
$$= 1 + (0.4 + 0.4) = 1.8$$

$$G_A^B[j=0][a=1][s=0] = R_{SA}[a=1][s=0] +$$
$$+\gamma\Big(G_{AO}\underbrace{[i=1]}_{\hat{G}^*_{AO}[j=0][a=1][o=0]}[o=0][a=1][s=0] + G_{AO}\underbrace{[i=1]}_{\hat{G}^*_{AO}[j=0][a=1][o=1]}[o=1][a=1][s=0]\Big) =$$
$$= 1 + (0.6 + 0.6) = 2.2$$

$$G_A^B[j=0][a=1][s=1] = R_{SA}[a=1][s=1] +$$
$$+\gamma\Big(G_{AO}\underbrace{[i=1]}_{\hat{G}^*_{AO}[j=0][a=1][o=0]}[o=0][a=1][s=1] + G_{AO}\underbrace{[i=1]}_{\hat{G}^*_{AO}[j=0][a=1][o=1]}[o=1][a=1][s=1]\Big) =$$
$$= 0 + (0.6 + 0.6) = 1.2$$

$$G_A^B[j=1][a=0][s=0] = R_{SA}[a=0][s=0] +$$
$$+\gamma\Big(G_{AO}\underbrace{[i=1]}_{\hat{G}^*_{AO}[j=1][a=0][o=0]}[o=0][a=0][s=0] + G_{AO}\underbrace{[i=1]}_{\hat{G}^*_{AO}[j=1][a=0][o=1]}[o=1][a=0][s=0]\Big) =$$
$$= 0 + (0 + 0) = 0$$

$$G_A^B[j=1][a=0][s=1] = R_{SA}[a=0][s=1] +$$
$$+\gamma\Big(G_{AO}\underbrace{[i=1]}_{\hat{G}^*_{AO}[j=1][a=0][o=0]}[o=0][a=0][s=1] + G_{AO}\underbrace{[i=1]}_{\hat{G}^*_{AO}[j=1][a=0][o=1]}[o=1][a=0][s=1]\Big) =$$
$$= 1 + (0.6 + 0.6) = 2.2$$

$$G_A^B[j=1][a=1][s=0] = R_{SA}[a=1][s=0] +$$
$$+\gamma\Big(G_{AO}\underbrace{[i=1]}_{\hat{G}^*_{AO}[j=1][a=1][o=0]}[o=0][a=1][s=0] + G_{AO}\underbrace{[i=1]}_{\hat{G}^*_{AO}[j=1][a=1][o=1]}[o=1][a=1][s=0]\Big) =$$
$$= 1 + (0.6 + 0.6) = 2.2$$

$$G_A^B[j=1][a=1][s=1] = R_{SA}[a=1][s=1] +$$
$$+\gamma\Big(G_{AO}\underbrace{[i=1]}_{\hat{G}^*_{AO}[j=1][a=1][o=0]}[o=0][a=1][s=1] + G_{AO}\underbrace{[i=1]}_{\hat{G}^*_{AO}[j=1][a=1][o=1]}[o=1][a=1][s=1]\Big) =$$
$$= 0 + (0.6 + 0.6) = 1.2$$

The final matrix is

$$G_A^B = \begin{bmatrix} \begin{array}{ccc} a & \begin{bmatrix} 0.2 & 1.8 \\ 2.2 & \color{orange}{1.2} \end{bmatrix} & j=0 \\ & s & \end{array} \\ \\ \begin{array}{ccc} a & \begin{bmatrix} 0 & 2.2 \\ 2.2 & 1.2 \end{bmatrix} & j=1 \\ & s & \end{array} \end{bmatrix}$$

Kernel computation example: element $G_A^B[j=0][a=1][s=1] = \color{orange}{1.2}$

$l_{\text{IDX}} = 3$

$j = l_{\text{IDX}}/|A||S| = 3/4 = 0$

$w_{\text{IDX}} = j|A||S| = 0$

$$
\begin{aligned}
a &= (l_{\text{IDX}} - w_{\text{IDX}})/|S| = (3-0)/2 = 1 \\
s &= (l_{\text{IDX}} - w_{\text{IDX}}) - a|S| = (3-0) - 1*2 = 1 \\
\{t_k\} &= \Big\{ j|A||\Omega| + a|\Omega| + k \,\Big|\, k = 0,\dots,n-1 \Big\} = \Big\{ 0 + 1*2 + k \,\Big|\, k = 0,1 \Big\} = \Big\{ 2,3 \Big\} \\
\{p_k\} &= \Big\{ i^{*(k)}_{a,o,j}|A||S||\Omega| + k|A||S| + a|S| + s \,\Big|\, k = 0,1 \Big\} \text{ with } i^{*(k)}_{a,o,j} = \hat{G}^*_{AO}[\,t_k\,] \\
&= \Big\{ \hat{G}^*_{AO}[\,2\,]*8 + 0*4 + 1*2 + 1 \quad , \quad \hat{G}^*_{AO}[\,3\,]*8 + 1*4 + 1*2 + 1 \Big\} \\
&= \Big\{ 1*8+3 \quad , \quad 1*8+7 \Big\} = \Big\{ 11, 15 \Big\} \\
G^B_A[l_{\text{IDX}}] &= R[a|S|+s] + \gamma \textstyle\sum_k G_{AO}[\,p_k\,] = R[1*2+1] + G_{AO}[11] + G_{AO}[15] = 0 + 0.6 + 0.6 = 1.2
\end{aligned}
$$

6 Compute $\hat{G}^B_A$

$$
\hat{G}^B_A = \begin{array}{c c} & l_{\text{IDX}} \\ j & \begin{bmatrix} & \boxed{} & \\ & & \\ & & \end{bmatrix}_{|B||A|} \\ & a \end{array}
$$

$$
\text{backup}(b) = \underset{\{g^b_a\}_{a \in \mathcal{A}}}{\arg\max} \boxed{b \cdot g^b_a}
$$

Access:

$$\hat{G}^B_A[\,j\,][\,a\,] = b_j \cdot g^{b_j}_a$$

$$b_j = B[\,j\,][\,:\,] \longrightarrow \text{find linear range}$$

$$g^{b_j}_a = G^B_A[\,j\,][\,a\,][\,:\,] \longrightarrow \text{find linear range}$$

Computation:

$$\hat{G}^B_A[j=0][a=0] = B[j=0][\,:\,] \cdot G^B_A[j=0][a=0][\,:\,] = [0.7\ 0.3] \cdot [0.2\ 1.8]^T = 0.14 + 0.54 = 0.68$$

$$\hat{G}^B_A[j=0][a=1] = B[j=0][\,:\,] \cdot G^B_A[j=0][a=1][\,:\,] = [0.7\ 0.3] \cdot [2.2\ 1.2]^T = 1.54 + 0.36 = 1.9$$

$$\hat{G}^B_A[j=1][a=0] = B[j=1][\,:\,] \cdot G^B_A[j=1][a=0][\,:\,] = [0.4\ 0.6] \cdot [0\ 2.2]^T = 0 + 1.32 = 1.32$$

$$\hat{G}^B_A[j=1][a=1] = B[j=1][\,:\,] \cdot G^B_A[j=1][a=1][\,:\,] = [0.4\ 0.6] \cdot [2.2\ 1.2]^T = 0.88 + 0.72 = 1.6$$

The final matrix is

$$
\hat{G}^B_A = \begin{array}{c c} j & \begin{bmatrix} 0.68 & 1.9 \\ 1.32 & 1.6 \end{bmatrix} \\ & a \end{array}
$$

Kernel computation example: element $\hat{G}^B_A[j=1][a=1] = 1.6$

$$l_{\text{IDX}} = 3$$

$j = l_{\text{IDX}}/|A| = 3/2 = 1$

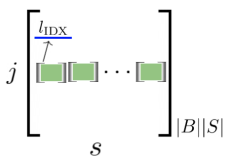
$a = l_{\text{IDX}} - j|A| = 3 - 1*2 = 1$

$g_a^{b_j}[\,s\,] = G_A^B[j|A||S| + a|S| + s] = G_A^B[1*4 + 1*2 + s] = G_A^B[6+s]$, with $s = 0, \ldots, |S| - 1$

$b_j = B[j|S| + s] = B[1*2+s] = B[2+s]$, with $s = 0, \ldots, |S| - 1$

$\hat{G}_A^B[l_{\text{IDX}}] = B[2+s] \cdot G_A^B[6+s]$ with $s = 0, 1 = [0.4\ 0.6] \cdot [2.2\ 1.2]^T = 0.88 + 0.72 = 1.6$

7 Compute B_{KP}

$B_{KP} =$ 

$$\text{backup}(b) = \boxed{\underset{\{g_a^b\}_{a\in\mathcal{A}}}{\arg\max}\ b \cdot g_a^b}$$

Access:

$B_{KP}[\,j\,][\,:\,] = \arg\max_{\{g_a^b\}_{a\in\mathcal{A}}}\ b_j \cdot g_a^b = G_A^B[\,j\,][\,a_{b_j}^*\,][\,:\,] \longrightarrow$ find linear range

$a_{b_j}^* = \hat{G}_A^{B*}[\,j\,]$

Computation:

$B_{KP}[j=0][\,:\,] = G_A^B[j=0]\ \underbrace{[a=1]}_{\hat{G}_A^{B*}[j=0]}\,[\,:\,] = [2.2\ 1.2]$

$B_{KP}[j=1][\,:\,] = G_A^B[j=1]\ \underbrace{[a=1]}_{\hat{G}_A^{B*}[j=1]}\,[\,:\,] = [2.2\ 1.2]$

The final matrix is

$$B_{KP} = \begin{array}{cc} j & \begin{bmatrix} 2.2 & 1.2 \\ 2.2 & 1.2 \end{bmatrix} \\ & s \end{array}$$

Kernel computation example: element $B_{KP}[j=0][s=1] = 1.2$

$l_{\text{IDX}} = 1$

$j = l_{\text{IDX}}/|S| = 1/2 = 0$

$a_{b_j}^* = \hat{G}_A^{B*}[j] = \hat{G}_A^{B*}[0] = 1$

$s = l_{\text{IDX}} - j|S| = 1 - 0 = 1$

$B_{KP}[l_{\text{IDX}}] = G_A^B[j|A||S| + a_{b_j}^*|S| + s] = G_A^B[0 + 1*2 + 1] = G_A^B[3] = 1.2$

$$b^{a,o} = \Big[\overset{l_{\text{IDX}}}{\square} \qquad \underset{s'}{} \qquad \Big]_{|S|}$$

8 Compute $b^{a,o}$

$$\begin{aligned} b_{a,o}(s') &= \beta_{\text{norm}}\ O(a,s',o)\sum_{s\in\mathcal{S}} T(s,a,s')b(s) \\ &= \beta_{\text{norm}}\ O(a,s',o)\underbrace{T(s,a,s')}_{=1} b(s) \\ &= \beta_{\text{norm}}\ \hat{b}_{a,o}(s') \end{aligned}$$

Access:

$T_{\text{inv}}[\,a\,][\,s'\,] = s$

$b = [0.1\ 0.9]$

Computation:

$\hat{b}_{a,o}[s'=0]$ with $a=1, o=0$ given $\longrightarrow O[o=0][a=1][s'=0]\ *\ b\ \underbrace{[s=0]}_{T_{\text{inv}}[a=1][s'=0]} = 0.1*0.1 = 0.01$

$\hat{b}_{a,o}[s'=1]$ with $a=1, o=0$ given $\longrightarrow O[o=0][a=1][s'=1]\ *\ b\ \underbrace{[s=1]}_{T_{\text{inv}}[a=1][s'=1]} = 0.2*0.9 = 0.18$

$$\beta_{\text{norm}} = \frac{1}{\sum_{s'}\hat{b}_{a,o}(s')} = \frac{1}{0.01+0.18} = 5.26$$

The final array is:

On CPU:

$$b_{a,o} = \beta_{\text{norm}}\ [0.01\ 0.18] = [0.05\ 0.95]$$

Kernel computation example: element $b_{a,o}[s'=0] = 0.05$

$l_{\text{IDX}} = 0 = s'$

$a = 1, o = 0$ given

$s = T_{\text{inv}}[a|S| + s'] = T_{\text{inv}}[1*2+0] = 0$

$b_{a,o}[l_{\text{IDX}}] = O[o|A||S| + a|S| + s'] * b[l_{\text{IDX}}] = O[0+1*2+0]*b[0] = 0.1*0.1 = 0.01$

On CPU: $b_{a,o}[s'=0] = \beta_{\text{norm}} * \hat{b}_{a,o}[s'=0] = 5.26*0.01 = 0.05$

9 Compute $t_{a,o}$

$$t_{a,o}(s) = \Big[\quad \overset{l_{\text{IDX}}}{\boxed{}} \quad {}_{s} \qquad \Big]_{|S|}$$

$$\begin{aligned} t_{a,o}(s) &= \sum_{s' \in \mathcal{S}} T(s,a,s')O(a,s',o) \\ &= \underbrace{T({\color{red}s},a,s')}_{=1} O(a,s',o) \end{aligned}$$

Access:

$T_{\text{inv}}[\,a\,][\,{\color{red}s}\,] = s'$

$b = [0.1 \;\, 0.9]$

Computation:

$t_{a,o}[s=0]$ with $a=1, o=0$ given $\longrightarrow O[o=0][a=1] \;\; \underbrace{[s'=1]}_{T[a=1][s=0]} \;\; = 0.2$

$t_{a,o}[s=1]$ with $a=1, o=0$ given $\longrightarrow O[o=0][a=1] \;\; \underbrace{[s'=1]}_{T[a=1][s=1]} \;\; = 0.2$

The final array is:

$$t_{a,o} = [0.2 \;\, {\color{orange}0.2}]$$

Kernel computation example: element $t_{a,o}[s=1] = {\color{orange}0.2}$

$l_{\text{IDX}} = 1 = s$

$a = 1, o = 0$ given

$s' = T[a|S| + s] = T[1 * 2 + 1] = T[3] = 1$

$t_{a,o}[l_{\text{IDX}}] = O[o|A||S| + a|S| + s'] = O[0 + 1 * 2 + 1] = O[3] = {\color{orange}0.2}$

On CPU: $p(o \,|\, b, a) = b \cdot t_{a,o} = [0.1 \;\, 0.9] \cdot [0.2 \;\, 0.2]^T = 0.02 + 0.18 = 0.2$

10 Pseudocode

Algorithm 1: Full backup on CUDA

Input

- B : belief set (size $|B| \times |S|$)
- γ : discount factor
- R : reward matrix (size $|A| \times |S|$)
- T : transition matrix (size $|A| \times |S|$)
- O : observation matrix (size $|\Omega| \times |A| \times |S|$)
- Γ : set of alpha vector (size $n \times |S|$)

Output

- B_{KP} : the set of new alpha vectors for each belief in the belief set B (size $|B| \times |S|$)

Pseudo-code

Each step corresponds to a CUDA kernel

1. Compute the matrix G_{AO} (size $|A||\Omega||S|n$) element-wise

$$g^i_{a,o}(s) = \sum_{s'} O(a,s',o)T(s,a,s')\alpha_i(s')$$

2. Compute the matrix $\hat{G}_{AO}$ (size $|A|n|\Omega||B|$) element-wise

$$\hat{G}_{AO} \underbrace{[\,j\,]}_{\text{4D}} \underbrace{[\,o\,]}_{\text{3D}} \underbrace{[\,a\,][\,i\,]}_{\text{2D}} = b_j \cdot g^i_{a,o}$$

3. Compute the matrix $\hat{G}^*_{AO}$ (size $|A||\Omega||B|$) element-wise

$$\hat{G}^*_{AO} \underbrace{[\,j\,]}_{\text{3D}} \underbrace{[\,a\,][\,o\,]}_{\text{2D}} = i^*_{a,o,j} = \arg\max_i \; b_j \cdot g^i_{a,o}$$

4. Compute the matrix G_{AB} (size $|A||S||B|$) element-wise

$$G^B_A \underbrace{[\,j\,]}_{\text{3D}} \underbrace{[\,a\,][\,:\,]}_{\text{2D}} = r_a + \gamma \sum_o \arg\max_{\{g^i_{a,o}\}_i} b \cdot g^i_{a,o}$$

5. Compute the matrix $\hat{G}_{AB}$ (size $|B||A|$) element-wise

$$\hat{G}^B_A[\,j\,][\,a\,] = b_j \cdot g^{b_j}_a$$

6. Compute the vector $\hat{G}^*_{AB}$ (size $|B|$) on CPU element-wise:

$$\hat{G}^*_{AB}[\,j\,] = a^*_{b_j} = \arg\max_a b_j \cdot g^{b_j}_a$$

7. Compute the matrix B_{KP} (size $|B||S|$) element-wise

$$B_{KP}[\,j\,][\,s\,] = \arg\max_{\{g^b_a\}_{a \in \mathcal{A}}} \; b_j \cdot g^b_a = G^B_A[\,j\,][\,a^*_{b_j}\,][\,s\,]$$

Algorithm 2: Policy evaluation on CUDA

Input

- b : current belief (size $|S|$)
- γ : discount factor
- R : reward matrix (size $|A| \times |S|$)
- T : transition matrix (size $|A| \times |S|$)
- T_{inv} : inverse transition matrix (size $|A| \times |S|$)
- O : observation matrix (size $|\Omega| \times |A| \times |S|$)
- Γ : set of alpha vector (size $n \times |S|$)

Output

- a^* : best action given the current belief

Pseudo-code

Split sum:

$$\pi_*(b) = \arg\max_{a \in \mathcal{A}} \underbrace{\sum_{s \in \mathcal{S}} b(s) R(s,a)}_{c_1} + \gamma \underbrace{\sum_{o \in \Omega} p(o|b,a) v_*(b^{a,o})}_{c_2}$$

foreach *action* $a \in \mathcal{A}$ **do**

1. Compute $\sum_{s \in \mathcal{S}} b(s) R(s,a)$ as a dot product on GPU $\longrightarrow c_1 = b \cdot R[\,a\,][\,:\,]$
2. Initialize $c_2 = 0$
3. **foreach** *observation* $o \in \Omega$ **do**
    4. Compute $b_{a,o}$ on GPU $(b, O, T_{\text{inv}}, a, o)$
    5. Compute $v_*(b_{a,o}) = \max_{\{\alpha_i\}_i} b_{a,o} \cdot \alpha_i \longrightarrow$ each dot product $b_{a,o} \cdot \alpha_i$ on GPU
    6. Compute $p(o \,|\, b, a) \longrightarrow$ dot product $b \cdot t_{a,o}$ on GPU
    7. $c_2 {+}{=}\; p(o \,|\, b, a)\, v_*(b_{a,o})$

   end

   $\text{sum}_a = c_1 + \gamma\, c_2$

end

return $a^* = \arg\max_a \text{sum}_a$