5

(Markov Decision Process)

State i = A, B, C

Decision k = 1, 2, 3

(In state B, decision 3 can not be chosen)

Transition probability matrices are given

$P(k) = \{P_{ij}(k)\}$, e.g. $P_{CA}(3) = 3/4$

Region

prob. end up in

$P(1)$, $P(2)$, $P(3)$

Expected cost C_{ik} given.

$\beta = \alpha = 0{,}5$

Initial policy: $R = [d_A, d_B, d_C] = [1, 2, 2]$

Policy improvement algorithm (with discounting)

First iteration

1 Compute V_i, i.e. V_A, V_B and V_C:

$$\begin{cases} V_A = C_{A d_A^1} + \alpha\left[P_{AA}(d_A^1)V_A + P_{AB}(d_A^1)V_B + P_{AC}(d_A^1)V_C\right] \\ \quad = 8 + \frac{1}{2}\left[\frac{1}{2}V_A + \frac{1}{4}V_B + \frac{1}{4}V_C\right] \\ V_B = C_{B d_B^2} + \alpha\left[P_{BA}(d_B^2)V_A + P_{BB}(d_B^2)V_B + P_{BC}(d_B^2)V_C\right] \\ \quad = 15 + \frac{1}{2}\left[\frac{1}{16}V_A + \frac{7}{8}V_B + \frac{1}{16}V_C\right] \\ V_C = C_{C d_C^2} + \alpha\left[P_{CA}(d_C^2)V_A + P_{CB}(d_C^2)V_B + P_{CC}(d_C^2)V_C\right] \\ \quad = 4 + \frac{1}{2}\left[\frac{1}{8}V_A + \frac{3}{4}V_B + \frac{1}{8}V_C\right] \end{cases}$$

$$\begin{cases} \left(1-\frac{1}{4}\right)V_A - \frac{1}{8}V_B - \frac{1}{8}V_C = 8 & (1) \\ -\frac{1}{32}V_A + \left(1-\frac{7}{8}\right)V_B - \frac{1}{32}V_C = 15 & (2) \\ -\frac{1}{16}V_A - \frac{3}{8}V_B + \left(1-\frac{1}{16}\right)V_C = 4 & (3) \end{cases}$$

$(1)\cdot 8$, $(2)\cdot 32$, $(3)\cdot 16$ gives

<u>Gauss elimination</u>

$$\left[\begin{array}{ccc|c} 6 & -1 & -1 & 64 \\ -1 & 4 & -1 & 15\cdot 32 \\ -1 & -6 & 15 & 4\cdot 16 \end{array}\right] \quad 1/6 \quad 1/6$$

$$\left[\begin{array}{ccc|c} 6 & -1 & -1 & 64 \\ 0 & 23/6 & -7/6 & 1472/3 \\ 0 & -37/6 & 89/6 & 224/3 \end{array}\right] \quad \begin{array}{l} \cdot 1/6 \\ \cdot 6/23 \\ \ \end{array}$$

$$\left[\begin{array}{ccc|c} 1 & -1/6 & -1/6 & 64/6 \\ 0 & 1 & -7/23 & 128 \\ 0 & 0 & 298/23 & 864 \end{array}\right] \quad 37/6 \quad 23/298$$

$$\left[\begin{array}{ccc|c} 1 & -1/6 & -1/6 & 64/6 \\ 0 & 1 & -7/23 & 128 \\ 0 & 0 & 1 & 9936/149 \end{array}\right] \quad 7/23 \quad 1/6$$

<u>Gauss-Jordan</u>

$$\left[\begin{array}{ccc|c} 1 & -1/6 & 0 & 9736/447 \\ 0 & 1 & 0 & 22096/149 \\ 0 & 0 & 1 & 9936/149 \end{array}\right] \quad 1/6$$

$$\left[\begin{array}{ccc|c} 1 & 0 & 0 & 6928/149 \\ 0 & 1 & 0 & 22096/149 \\ 0 & 0 & 1 & 9936/149 \end{array}\right] \quad \begin{array}{l} \approx 46,4966 \\ \approx 148,2953 \\ \approx 66,6846 \end{array}$$

$$\begin{cases} (1-1/4)\,V_A - \frac{1}{8} V_B - \frac{1}{8} V_C = 8 \\ -\frac{1}{32} V_A + \left(1-\frac{7}{8}\right) V_B - \frac{1}{32} V_C = 15 \\ -\frac{1}{16} V_A - \frac{3}{8} V_B + \left(1-\frac{1}{16}\right) V_C = 4 \end{cases}$$

solve gives

$$\begin{cases} V_A \approx 46{,}4966 \\ V_B \approx 148{,}2953 \\ V_C \approx 66{,}6846 \end{cases}$$

(2) Policy improvement (note $\tilde{V}_{ik}$ is a reward)

state(i)	Decision(k)	
A	1	$\tilde{V}_{A1} = C_{A1} + d\left(P_{AA}(1)V_A + P_{AB}(1)V_B + P_{AC}(1)V_C\right)$ $= 8 + 0{,}5\left(\frac{1}{2}V_A + \frac{1}{4}V_B + \frac{1}{4}V_C\right) \approx 46{,}4966$
	2	$\tilde{V}_{A2} = 2{,}75 + 0{,}5\left(\frac{1}{16}V_A + \frac{3}{4}V_B + \frac{3}{16}V_C\right) \approx 66{,}0654$
	3	$\tilde{V}_{A3} = 4{,}25 + 0{,}5\left(\frac{1}{4}V_A + \frac{1}{8}V_B + \frac{5}{8}V_C\right) \approx 40{,}1695$
B	1	$\tilde{V}_{B1} = 16 + 0{,}5\left(\frac{1}{2}V_A + 0 + \frac{1}{2}V_C\right) \approx 44{,}2953$
	2	$\tilde{V}_{B2} = 15 + 0{,}5\left(\frac{1}{16}V_A + \frac{7}{8}V_B + \frac{1}{16}V_C\right) \approx 83{,}4161$
C	1	$\tilde{V}_{C1} = 7 + 0{,}5\left(\frac{1}{4}V_A + \frac{1}{4}V_B + \frac{1}{2}V_C\right) \approx 48{,}0201$
	2	$\tilde{V}_{C2} = 4 + 0{,}5\left(\frac{1}{8}V_A + \frac{3}{4}V_B + \frac{1}{8}V_C\right) \approx 66{,}6846$
	3	$\tilde{V}_{C3} = 4{,}5 + 0{,}5\left(\frac{3}{4}V_A + \frac{1}{16}V_B + \frac{3}{16}V_C\right) \approx 32{,}8221$

Let d_i = maximizing k $\left(\max_k \{\tilde{g}_{ik}\} \quad \forall i\right)$

gives $d_A = 2$, $d_B = 2$, $d_C = 2$

so $R_1 = [2, 2, 2] \neq R_0 \Rightarrow$ new iteration.

But we know that R_1 is better than R_0,
hence the solution is improved (and we can stop).