Our example: Let R_1 be the policy s.t. $d(R_1) = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

Value determination: $g(R_1) = C_{ik} + \sum_{j=1}^{2} p_{ij}(k)\, v_j(R_1) - v_i(R_1) \quad i=1,2.$

$$C = [C_{11} \;\; C_{12} \;\; C_{21} \;\; C_{22}]^T = [-350 \;\; -400 \;\; -50 \;\; -250]$$

$$P(1) = \begin{bmatrix} 0.9 & 0.1 \\ 0.7 & 0.3 \end{bmatrix} \qquad P(2) = \begin{bmatrix} 0.6 & 0.4 \\ 0.2 & 0.8 \end{bmatrix}$$

VDE:
$$\begin{cases} g(R_1) = C_{11} + p_{11}(1)v_1(R_1) + p_{12}(1)v_2(R_1) - v_1(R_1) \Leftarrow \boxed{k = d_1(R_1) = 1} \\ g(R_2) = C_{21} + p_{21}(1)v_1(R_1) + p_{22}(1)v_2(R_1) - v_2(R_1) \Leftarrow \boxed{k = d_2(R_1) = 1} \end{cases}$$

Let $v_2(R_1) = 0$

$$\begin{cases} g(R_1) = -350 + 0.9\, v_1(R_1) + 0.1 \cdot 0 - v_1(R_1) \\ g(R_1) = -50 + 0.7\, v_1(R_1) + 0.3 \cdot 0 - 0 \end{cases}$$

$$\begin{cases} g(R_1) + 0.1\, v_1(R_1) = -350 \\ g(R_1) - 0.7\, v_1(R_1) = -50 \end{cases} \Rightarrow \begin{matrix} g(R_1) = -312.5 \\ v_1(R_1) = -375 \end{matrix}$$

Note: This is the same as we got last time.

Policy improvement: $\underset{k=1,2}{\text{Minimize}} \left\{ C_{ik} + \sum_{j=1}^{2} p_{ij}(k)\, v_j(R_1) - v_i(R_1) \right\} \quad i=1,2$

$$\underset{k=1,2}{\text{Minimize}} \left\{ C_{1k} + p_{11}(k)\, v_1(R_1) + p_{12}(k)\, v_2(R_1) - v_1(R_1) \right\}$$

$$= \text{Minimize} \left\{ \overset{k=1}{\underbrace{-350 + 0.9(-375) + 0.1 \cdot 0 - (-375)}_{= -256.25}}, \; \overset{k=2}{\underbrace{-400 + 0.6 \cdot (-375) + 0.4 \cdot 0 - (-375)}_{= -212.5}} \right\}$$

Let $d_1(R_2) = 1$

$$\underset{k=1,2}{\text{Minimize}} \left\{ C_{2k} + p_{21}(k)\, v_1(R_1) + p_{22}(k)\, v_2(R_1) - v_2(R_1) \right\}$$

$$= \text{Minimize} \left\{ \overset{k=1}{\underbrace{-50 + 0.7(-375) + 0.3 \cdot 0 - 0}_{= -268.75}}, \; \overset{k=2}{\underbrace{-250 + 0.2 \cdot (-375) + 0.8 \cdot 0 - 0}_{= -312.5}} \right\}$$

Let $d_2(R_2) = 2$

$R_1 \neq R_2$ we have not converged...

Value determination:

$$g(R_2) = C_{11} + p_{11}(1)v_1(R_2) + p_{12}(1)v_2(R_2) - v_1(R_2) \Leftarrow \boxed{k = d_1(R_2) = 1}$$

$$g(R_2) = C_{22} + p_{21}(2)v_1(R_2) + p_{22}(2)v_2(R_2) - v_2(R_2) \Leftarrow \boxed{k = d_2(R_2) = 2}$$

Let $v_2(R_2) = 0$

$$\begin{cases} g(R_2) = -350 + 0{,}9\, v_1(R_2) + 0{,}1 \cdot 0 - v_1(R_2) \\ g(R_2) = -250 + 0{,}2\, v_1(R_2) + 0{,}8 \cdot 0 - 0 \end{cases}$$

$$\begin{cases} g(R_2) + 0{,}1\, v_1(R_2) = -350 \\ g(R_2) - 0{,}2\, v_1(R_2) = -250 \end{cases} \Rightarrow \begin{matrix} g(R_2) = -316{,}7 \\ v_1(R_2) = -333{,}3 \end{matrix}$$

Same as the optimal from last time.

Note that $g(R_2) < g(R_1)$ so we have a better policy.

Policy improvement:

$$\underset{k=1,2}{\text{Minimize}} \left\{ C_{1k} + p_{11}(k)v_1(R_2) + p_{12}(k)v_2(R_2) - v_1(R_1) \right\}$$

$$= \text{Minimize} \left\{ \overset{k=1}{\underbrace{-350 + 0{,}9 \cdot (-333{,}3) + 0{,}1 \cdot 0 - (-333{,}3)}_{-301{,}7}}, \overset{k=2}{\underbrace{-400 + 0{,}6(-333{,}3) + 0{,}4 \cdot 0 - (-333{,}3)}_{-256{,}7}} \right\}$$

Let $d_1(R_3) = 1$

$$\underset{k=1,2}{\text{Minimize}} \left\{ C_{2k} + p_{21}(k)v_1(R_2) + p_{22}(k)v_2(R_2) - v_2(R_2) \right\}$$

$$= \text{Minimize} \left\{ \overset{k=1}{\underbrace{-50 + 0{,}7(-333{,}3) + 0{,}3 \cdot 0 - 0}_{-271{,}7}}, \overset{k=2}{\underbrace{-250 + 0{,}2 \cdot (-333{,}3) + 0{,}8 \cdot 0 - 0}_{-313{,}3}} \right\}$$

Let $d_2(R_3) = 2$

$\Rightarrow R_3 = R_2$ and the algorithm has converged to the optimal policy, same as before, and optimal cost $= -316{,}7$