Our example again: Now introduce discount factor $\alpha = 0.9$

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

Value determination: $V_i(R) = C_{ik} + \alpha \sum_{j=1}^{2} p_{ij}(k) V_j(R)$ $i = 1, 2.$

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

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

VDE: $$\begin{cases} V_1(R_1) = c_{11} + 0.9\left(p_{11}(1) V_1(R_1) + p_{12}(1) V_2(R_1)\right) \Leftarrow \boxed{k = d_1(R_1) = 1} \\ V_2(R_1) = c_{21} + 0.9\left(p_{21}(1) V_1(R_1) + p_{22}(1) V_2(R_1)\right) \Leftarrow \boxed{k = d_2(R_1) = 1} \end{cases}$$

$$\begin{cases} V_1(R_1) = -350 + 0.9\left(0.9 V_1(R_1) + 0.1 \cdot V_2(R_1)\right) \\ V_2(R_1) = -50 + 0.9\left(0.7 \cdot V_1(R_1) + 0.3 \, V_2(R_1)\right) \end{cases}$$

$$\begin{cases} -0.19 V_1(R_1) + 0.09 V_2(R_1) = 350 \\ 0.63 V_1(R_1) - 0.73 V_2(R_1) = 50 \end{cases} \Rightarrow \begin{matrix} V_1(R_1) \approx -3171 \\ V_2(R_1) \approx -2805 \end{matrix}$$

Policy improvement: $\underset{k=1,2}{\text{Minimize}} \left\{ C_{ik} + \alpha \sum_{j=1}^{2} p_{ij}(k) V_j(R_n) \right\}.$

$i=1$

$$\underset{k=1,2}{\text{Minimize}} \left\{ C_{1k} + \alpha\left(p_{11}(k) V_1(R_1) + p_{12}(k) V_2(R_1)\right)\right\}$$

$$= \text{Minimize} \left\{ \underbrace{-350 + 0.9\left(0.9 \cdot (-3171) + 0.1(-2805)\right)}_{= -3171}^{k=1}, \underbrace{-400 + 0.9\left(0.6(-3171) + 0.4(-2805)\right)}_{= -3122}^{k=2} \right\}$$

Let $d_1(R_2) = 1$

$i=2$

$$\underset{k=1,2}{\text{Minimize}} \left\{ C_{2k} + \alpha\left(p_{21}(k) V_1(R_1) + p_{22}(k) V_2(R_1)\right)\right\}$$

$$= \text{Minimize} \left\{ \underbrace{-50 + 0.9\left(0.7 \cdot (-3171) + 0.3 \cdot (-2805)\right)}_{= -2805}^{k=1}, \underbrace{-250 + 0.9\left(0.2(-3171) + 0.8(-2805)\right)}_{= -2840}^{k=2} \right\}$$

Let $d_2(R_2) = 2.$

Value determination

$$\begin{cases} V_1(R_2) = C_{11} + 0{,}9\left(p_{11}(1)\,V_1(R_2) + p_{12}(1)\,V_2(R_2)\right) \Leftarrow \boxed{k = d_1(R_2) = 1} \\ V_2(R_2) = C_{22} + 0{,}9\left(p_{21}(2)\,V_1(R_2) + p_{22}(2)\,V_2(R_2)\right) \Leftarrow \boxed{k = d_2(R_2) = 2} \end{cases}$$

$$V_1(R_2) = -350 + 0{,}9\left(0{,}9\,V_1(R_2) + 0{,}1\,V_2(R_2)\right)$$
$$V_2(R_2) = -250 + 0{,}9\left(0{,}2\,V_1(R_2) + 0{,}8\,V_2(R_2)\right)$$

$$\begin{cases} -0{,}19\,V_1(R_2) + 0{,}09\,V_2(R_2) = 350 \\ 0{,}18\,V_1(R_2) - 0{,}28\,V_2(R_2) = 250 \end{cases} \Rightarrow \begin{array}{l} V_1(R_2) \approx -3257 \\ V_2(R_2) \approx -2986 \end{array}$$

Note that $V_1(R_2) < V_1(R_1)$ and $V_2(R_2) < V_2(R_1)$ so the policy has been improved.

Policy improvement:

$i=1$ $$\underset{k=1,2}{\text{Minimize}} \left\{ C_{1k} + \alpha\left(p_{11}(k)\,V_1(R_2) + p_{12}(k)\,V_2(R_2)\right)\right\}$$

$$= \text{Minimize}\left\{ \underbrace{-350 + 0{,}9\left(0{,}9(-3257) + 0{,}1(-2986)\right)}_{= -3257}^{k=1},\ \underbrace{-400 + 0{,}9\left(0{,}6(-3257) + 0{,}4(-2986)\right)}_{= -3234}^{k=2}\right\}$$

Let $d_1(R_3) = 1$

$i=2$ $$\underset{k=1,2}{\text{Minimize}} \left\{ C_{2k} + \alpha\left(p_{21}(k)\,V_1(R_2) + p_{22}(k)\,V_2(R_2)\right)\right\}$$

$$= \text{Minimize}\left\{ \underbrace{-50 + 0{,}9\left(0{,}7(-3257) + 0{,}3(-2986)\right)}_{= -2908}^{k=1},\ \underbrace{-250 + 0{,}9\left(0{,}2(-3257) + 0{,}8(-2986)\right)}_{= -2986}^{k=2}\right\}$$

Let $d_2(R_3) = 2$

$\Rightarrow R_2 = R_3$ and the algorithm converged to the optimal policy.