1 Policy Iteration

Iter 1, Step 1

$$g(1, a_1) = 2, \; g(1, a_2) = 0.5, \; g(2, a_1) = 1, \; g(2, a_2) = 3 \tag{1}$$

$$\mu^0(1) = a_1, \mu^0(2) = a_2, \gamma = 0.9 \tag{2}$$

$$P(u^1) = \begin{pmatrix} p_{11}(u^1) & p_{12}(u^1) \\ p_{21}(u^1) & p_{22}(u^1) \end{pmatrix} = \begin{pmatrix} 3/4 & 1/4 \\ 3/4 & 1/4 \end{pmatrix} \tag{3}$$

$$P(u^2) = \begin{pmatrix} p_{11}(u^2) & p_{12}(u^2) \\ p_{21}(u^2) & p_{22}(u^2) \end{pmatrix} = \begin{pmatrix} 1/4 & 3/4 \\ 1/4 & 3/4 \end{pmatrix} \tag{4}$$

Iter 1, Step 2

In this case $p_{ii} + p_{ij} = 1$, thus e.g.

$$\begin{aligned} v_1 &= p_{11}(g(1,a_1) + \gamma v_1) + p_{12}(g(1,a_1) + \gamma v_2) \\ &= g(1,a_1) + \gamma(p_{11}(a_1)v_1 + p_{12}(a_1)v_2) \end{aligned} \tag{5}$$

$s = 1, \mu_0 = a_1,\ s = 2, \mu_0 = a_2,$

$$v_1 = g(1,a_1) + \gamma(p_{11}(a_1)v_1 + p_{12}(a_1)v_2) \tag{6}$$
$$v_2 = g(2,a_2) + \gamma(p_{21}(a_2)v_1 + p_{22}(a_2)v_2) \tag{7}$$

$$v_1 = 2 + 0.9(3/4v_1 + 1/4v_2) \tag{8}$$
$$v_2 = 3 + 0.9(1/4v_1 + 3/4v_2) \tag{9}$$

Solve by hand or e.g. with Wolfram Alpha

$$v_1 = 265/11 = 24.0909\ldots \tag{10}$$
$$v_2 = 285/11 = 25.9091\ldots \tag{11}$$

Iter 1, Step 3

$$\mu^1(s) = \arg\min_{a \in A} \left[g(s,a) + \gamma \sum_{s' \in S} P(s,a,s') V_{\mu^0}(s') \right], \ \forall s \in S \tag{12}$$

$s = 1, \mu_1 = a_2$

$$\begin{aligned} v_1 &= g(1, a_2) + \gamma(p_{11}(a_2)v_1 + p_{12}(a_2)v_2) \\ &= 0.5 + 0.9(1/4 \cdot 265/11 + 3/4 \cdot 285/11) \\ &= 512/22 = 23.4090\ldots \end{aligned} \tag{13}$$

$s = 2, \mu_1 = a_1$

$$\begin{aligned} v_2 &= g(2, a_1) + \gamma(p_{21}(a_1)v_1 + p_{22}(a_1)v_2) \\ &= 1 + 0.9(3/4 \cdot 265/11 + 1/4 \cdot 285/11) \\ &= 254/11 = 23.0909\ldots \end{aligned} \tag{14}$$

$$\mu^1(1) = \arg\min_{a \in A} [24.10, 23.41] = 23.41 \tag{15}$$

$$\mu^1(2) = \arg\min_{a \in A} [23.10, 25.91] = 23.10 \tag{16}$$

Thus $\mu^1(1) = a_2$ and $\mu^1(2) = a_1$. Now $\mu^1(1) = a_2 \neq a_1 = \mu^0(1)$ so we need to continue.

Iter 2, Step 2

$s = 1, \mu_1 = a_2,\ s = 2, \mu_1 = a_1,$

$$v_1 = g(1, a_2) + \gamma(p_{11}(a_2)v_1 + p_{12}(a_2)v_2) \tag{17}$$
$$v_2 = g(2, a_1) + \gamma(p_{21}(a_1)v_1 + p_{22}(a_1)v_2) \tag{18}$$

$$v_1 = 0.5 + 0.9(1/4v_1 + 3/4v_2) \tag{19}$$
$$v_2 = 1 + 0.9(3/4v_1 + 1/4v_2) \tag{20}$$

Solve by hand or e.g. with Wolfram Alpha

$$v_1 = 425/58 = 7.3275\ldots \tag{21}$$
$$v_2 = 445/58 = 7.6724\ldots \tag{22}$$

Iter 3, Step 3

$s = 1, \mu_1 = a_1$

$$\begin{aligned} v_1 &= g(1, a_1) + \gamma(p_{11}(a_1)v_1 + p_{12}(a_1)v_2) \\ &= 2 + 0.9(3/4 \cdot 425/58 + 1/4 \cdot 445/58) \\ &= 503/58 = 8.6724\ldots \end{aligned} \tag{23}$$

$s = 2, \mu_1 = a_2$

$$\begin{aligned} v_2 &= g(2, a_2) + \gamma(p_{21}(a_2)v_1 + p_{22}(a_2)v_2) \\ &= 3 + 0.9(1/4 \cdot 425/58 + 3/4 \cdot 445/58) \\ &= 285/29 = 9.8275\ldots \end{aligned} \tag{24}$$

$$\mu^2(1) = \arg\min_{a \in A} [8.67, 7.33] = 7.33 \tag{25}$$

$$\mu^2(2) = \arg\min_{a \in A} [7.67, 9.83] = 7.67 \tag{26}$$

Thus $\mu^2(1) = a_2$ and $\mu^2(2) = a_1$. Now $\mu^2(1) = a_2 = \mu^1(1)$ and $\mu^2(2) = a_1 = \mu^1(2)$, so we can stop.

Therefore $\mu^*(1) = a_2$, $\mu^*(2) = a_1$, $V^*(1) \approx 7.33$ and $V^*(2) \approx 7.67$.