when $f(s,a)=6 \Rightarrow s=5,14,7$

1) $V^*(s)=\max_a E[r(s,a)+\alpha V^*(f(s,a))] = \max_a\left(E[r(s,a)]+\alpha V^*(f(s,a))\right)$ since $f(s,a)$ is deterministic

If $s=5,14,7$ then the correct action always gives a reward of 10 and so we take it and transition to $s=6$. Otherwise no matter what we do we always have a reward of $E[r(s,a)]=0.5\cdot-1+0.5\cdot-2=-1.5$

$$V^*(s)=\begin{cases}10 & \text{if } s=5,14,7\\ \max_a\left(-1.5+\alpha V^*(f(s,a))\right) & \forall \text{ includes } s=6 \text{ as we must move out and back in}\end{cases}$$

2) $V_0=0_{64}$ $\quad V_1(s)$ where $s\neq5,14,7 = \max_a(-1.5+\alpha V_0(\sim)) = \max_a(-1.5+0) = -1.5$

$V_1(s)$ where $s=5,14,7 = \max\begin{cases}-1.5+\alpha\cdot0\\ 10+\alpha\cdot0\end{cases} = 10$

3) $\mu_0(s)=\text{up}$ for all s $\quad V_{\mu_0}(s)$ where $s\neq14$ satisfy $V_{\mu_0}(s)=-1.5+\alpha V_{\mu_0}(s) \Rightarrow V_{\mu_0}(s)=\frac{3}{2(\alpha-1)}$

$V_{\mu_0}(s)$ where $s=14$ $=10$

13 and 15 now see that $a=1$ and 3 respectively got expected reward $-1.5+\alpha\cdot10$ and swap to that.

5 and 7 see $a=1$ and 3 respectively got reward 10 and swap to that

Now $\mu_1(5 \text{ or } 13)=1$ $\quad \mu_1(7 \text{ or } 13)=3$ and $\mu_1(s)$ for all other $s=0$

2. Since we have a cost instead of a reward we will always seek to minimize the sum of future costs.

$$\tilde{Q}(i,u) \leftarrow \tilde{Q}(i,u) + 0.1\left(c(i,u) + 0.9 \min_{v} \tilde{Q}(j,v) - \tilde{Q}(i,u)\right)$$

i. $(i=1,\ u=1,\ c(i,u)=1) \to j=2$

$\tilde{Q}(1,1)=0$ initially. $\min_{v} \tilde{Q}(2,v) = \min\{0.3, 0\} = 0$

$\tilde{Q}(1,1) \leftarrow 0 + 0.1(1 + 0.9 \cdot 0 - 0) = 0.1 \Rightarrow \tilde{Q} = \begin{pmatrix} 0.1 & 0.5 \\ 0.3 & 0 \\ 0.2 & 0.1 \end{pmatrix}$

ii. $(i=2,\ u=1,\ c(i,u)=0) \to j=3$

$\tilde{Q}(2,1) = 0.3$ initially. $\min_{v} \tilde{Q}(3,v) = \min\{0.2, 0.1\} = 0.1$

$\tilde{Q}(2,1) \leftarrow 0.3 + 0.1(0 + 0.9 \cdot 0.1 - 0.3) = 0.279 \Rightarrow \tilde{Q} = \begin{pmatrix} 0.1 & 0.5 \\ 0.279 & 0 \\ 0.2 & 0.1 \end{pmatrix}$

iii. $(i=3,\ u=2,\ c(i,u)=1) \to j=2$

$\tilde{Q}(3,2) = 0.1$ initially. $\min_{v} \tilde{Q}(2,v) = \min\{0.279, 0\} = 0$

$\tilde{Q}(3,2) \leftarrow 0.1 + 0.1(1 + 0.9 \cdot 0 - 0.1) = 0.19$

$$\tilde{Q} = \begin{pmatrix} 0.1 & 0.5 \\ 0.279 & 0 \\ 0.2 & 0.19 \end{pmatrix}$$