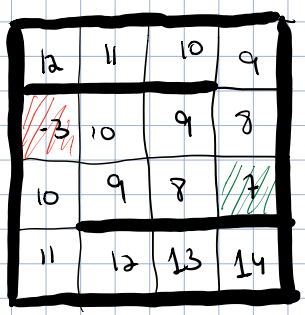
1. a. for all s $r_s = -1$. in this way, to max the profit of the agent, he will choose the path that goes through the min amount of states, i.e the shortest path.

The optimal value of each state would be the number of states to get to g + 5.

b. $-(V^{\pi_g}_{old} - 5) + 7$:



c. $V^{\pi}_{new} = \mathbb{E}\left[\sum_{l=0}^{\infty} \gamma^l r_{t+l} + c \mid S_t = s\right] =$

$= \mathbb{E}\left[\sum_{l=0}^{\infty} \gamma^l r_{t+l} \mid S_t = s\right] + \mathbb{E}\left[\sum_{l=0}^{\infty} \gamma^l c \mid S_t = s\right] =$

$V^{\pi}_{old} + \sum_{l=0}^{\infty} \gamma^l c$

d. Now $r_s = 1$, thus the optimal policy will be the longest path, thus the value will be ∞ and we will get stuck at the edges.

e. We will still wonder around, until γ^t will be small enough and then we'll go to g.

f. $r_s \leq -5$

2. for $t \in [H]$:

$$V^{\pi_1}_t(x_t) - V^{\pi_2}_t(x_t) = Q^{\pi_1}_t(x_t, \pi_1(x_t, t)) - Q^{\pi_2}_t(x_t, \pi_2(x_t, t)) =$$

$$Q^{\pi_1}_t(x_t, \pi_1(x_t, t)) - Q^{\pi_1}_t(x_t, \pi_2(x_t, t)) + \underbrace{Q^{\pi_1}_t(x_t, \pi_2(x_t, t)) - Q^{\pi_2}_t(x_t, \pi_2(x_t, t))}_{\textcircled{*}} =$$

$$\otimes \quad Q_t^{\pi_1}(x_t, \pi_2(x_t,t)) - Q_t^{\pi_2}(x_t, \pi_2(x_t,t)) \overset{P(x_t,\pi_2(x_t,t))=1}{=} \Omega$$

$$r_t(x_t, \pi_2(x_t,t)) + \mathbb{E}_{s'\sim\Omega} V_{t+1}^{\pi_1}(s') - r_t(x_t, \pi_2(x_t,t)) - \mathbb{E}_{s'\sim\Omega} V_{t+1}^{\pi_2}(s')$$

$$= \mathbb{E}_{s'\sim\Omega} V_{t+1}^{\pi_1}(s') - \mathbb{E}_{s'\sim\Omega} V_{t+1}^{\pi_2}(s') = \mathbb{E}_{s'\sim\Omega}\left[V_{t+1}^{\pi_1}(s') - V_{t+1}^{\pi_2}(s')\right]$$

$$\Rightarrow V_t^{\pi_1}(x_t) - V_t^{\pi_2}(x_t) = Q_t^{\pi_1}(x_t, \pi_1(x_t,t)) - Q_t^{\pi_1}(x_t, \pi_2(x_t,t)) + \mathbb{E}_{s'\sim\Omega}\left[V_{t+1}^{\pi_1}(s') - V_{t+1}^{\pi_2}(s')\right]$$

Now, we will take the expectation of $\mathbb{E}_{x_t\sim\pi_2}$ on both sides:

$$\mathbb{E}_{x_t\sim\pi_2}\left[V_t^{\pi_1}(x_t) - V_t^{\pi_2}(x_t)\right] = \mathbb{E}_{x_t\sim\pi_2}\left[Q_t^{\pi_1}(x_t, \pi_1(x_t,t)) - Q_t^{\pi_1}(x_t, \pi_2(x_t,t)) + \mathbb{E}_{s'\sim\Omega}\left[V_{t+1}^{\pi_1}(s') - V_{t+1}^{\pi_2}(s')\right]\right]$$

$$\Rightarrow V_t^{\pi_1}(x_t) - V_t^{\pi_2}(x_t) = \mathbb{E}_{x_t\sim\pi_2}\Big\{Q_t(x_t, \pi_1(x_t,t)) - Q_t^{\pi_1}(x_t, \pi_2(x_t,t))$$

$$+ \mathbb{E}_{s'\sim\Omega}\left[V_{t+1}^{\pi_1}(s') - V_{t+1}^{\pi_2}(s')\right]\Big\} =$$

$$= \mathbb{E}_{x_t\sim\pi_2}\left\{Q_t(x_t, \pi_1(x_t,t)) - Q_t^{\pi_1}(x_t, \pi_2(x_t,t))\right\} +$$

$$\mathbb{E}_{x_{t+1}\sim\pi_2}\left\{V_{t+1}^{\pi_1}(x_{t+1}) - V_{t+1}^{\pi_2}(x_{t+1})\right\} =$$

$$= \sum_{T=t}^{H} \mathbb{E}_{x_T\sim\pi_2}\left\{Q_T(x_T, \pi_1(x_T,T)) - Q_T(x_T, \pi_2(x_T,T))\right\}$$