Ex. 3.1

State-space: $\mathcal{S} = \{S, \bar{S}, O\}$

$S \equiv$ (Location = same, rains)

$\bar{S} \equiv$ (Location = same, doesn't rain)

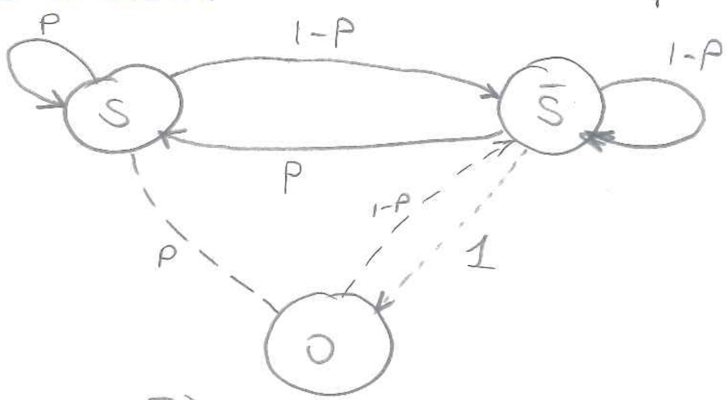
$O \equiv$ Location is different

Actions: $\begin{cases} T = \text{take the umbrella} \text{ ——} \\ \bar{T} = \text{leave the umbrella} \text{ - - - -} \end{cases}$

$A_S = \{T\}$

$A_{\bar{S}} = \{T, \bar{T}\}$

$A_O = \{\bar{T}\}$



Transition probabilities

$P(\bar{S} | S, T) = 1-p$

$P(\bar{S} | \bar{S}, T) = 1-p$

$P(S | S, T) = p$

$P(S | \bar{S}, T) = p$

$P(S | O, \bar{T}) = p$

$P(\bar{S} | O, \bar{T}) = 1-p$

$P(O | \bar{S}, \bar{T}) = 1$

Rewards

$r(S, T) = 0$

$r(\bar{S}, \bar{T}) = 0$

$r(O, \bar{T}) = +pW$

$r(\bar{S}, T) = +U$

Bellman Equation $\forall s \in \mathcal{S}$ $V^*(s) = \min_{a \in A_s} [r(s,a) + \lambda \sum_{j \in S} P(j | s, a) \cdot V^*(j)]$

(1) $V(O) = r(O, \bar{T}) + \lambda \cdot [pV(S) + (1-p)V(\bar{S})] = +pW + \lambda p V(S) + \lambda(1-p)V(\bar{S})$

(2) $V(S) = r(S,T) + \lambda p V(S) + (1-p)V(\bar{S}) = \lambda p V(S) + (1-p)\lambda V(\bar{S})$

(3) $V(\bar{S}) = \min \{ \underbrace{U + \lambda p V(S) + \lambda(1-p)V(\bar{S})}_{T}, \underbrace{0 + \lambda V(O)}_{\bar{T}} \}$

3.1 Cont.

$\Rightarrow$ At $\bar{S}$ she takes an umbrella, if

$$u + \underbrace{\lambda p V(S) + \lambda(1-p)V(\bar{S})}_{V(S) \text{ from } (2)} < \lambda V(0)$$

$$u + V(S) < \lambda \cdot V(0) \quad (*)$$

$\Rightarrow$ Bellman equations will become

$$(1)\ V(0) = pw + V(S)$$

$$(2)\ V(\bar{S}) = V(S) \cdot \frac{1-\lambda p}{(1-p)\lambda}$$

$$(3)\ V(\bar{S}) = u + V(S)$$

$$(2)-(3) \Leftrightarrow V(S)\left[\frac{1-\lambda p}{(1-p)\lambda} - 1\right] = u$$

$$V(S) = \frac{u(1-p)\lambda}{(1-\lambda)} \quad (4)$$

$$(1) \Leftrightarrow V(0) = pw + \frac{u(1-p)\lambda}{(1-\lambda)} \quad (5)$$

Plugging (4) and (5) in $(*)$ we get

$$u + \frac{u(1-p)\lambda}{(1-\lambda)} < \lambda \cdot \left[pw + \frac{u(1-p)\lambda}{(1-\lambda)}\right]$$

$$u(1-\lambda) + u(1-p)\lambda < \lambda(1-\lambda)pw + u\lambda^2(1-p)$$

$$u(1-\lambda) + u\lambda - up\lambda < \lambda(1-\lambda)w \cdot p + u\lambda^2 - u\lambda^2 p$$

$$u - u\lambda^2 < p(u\lambda + \lambda(1-\lambda)w - u\lambda^2)$$

$$p > \frac{u\cancel{(1-\lambda)}(1+\lambda)}{w\lambda\cancel{(1-\lambda)} + u\lambda\cancel{(1-\lambda)}}$$

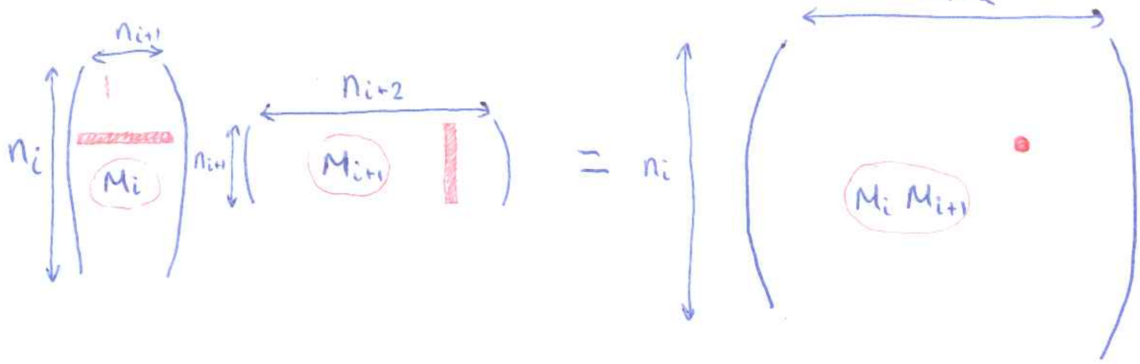
$$\boxed{p > \frac{1+\lambda}{\lambda} \cdot \frac{u}{u+w}}$$

Ex 3.4

Given

- Denote $V(i,j)$ to be the minimum # of multiplications to compute $M_i \cdot M_{i+1} \cdots M_j$,
- We need to find $V(1,N)$ – ?
- Size M_i is $n_i \times n_{i+1}$

Solution How many computations do we need to multiply M_i and M_{i+1}



We need $n_i \cdot n_{i+1} \cdot n_{i+2}$ multiplications

- Example: $M_i \, M_{i+1} \, M_{i+2} \, M_{i+3} \cdots M_j = (M_i \; M_{i+1})(M_{i+2} \; M_{i+3} \cdots M_j)$

$M_i \, M_{i+1} \cdots M_{j-2} \, M_{j-1} \, M_j = (M_i \; M_{i+1} \cdots M_{j-2})(M_{j-1} \; M_j)$

$$V(i,j) = \min_{i \le K < j} \left(V(i,K) + V(K+1,j) + n_i \cdot n_{K+1} \cdot n_j \right) \quad i < j$$

$$V(i,i) = 0 \qquad i = 1, \dots, N$$

- Numerical Example: $N=3$, $n_1 = 5$, $n_2 = 10$, $n_3 = 2$, $n_4 = 1$

Find $V(1,3)$.

$V(1,2) = V(1,1) + V(2,2) + n_1 n_2 n_3 = 100$

$V(2,3) = V(2,2) + V(3,3) + n_2 n_3 n_4 = 20$

$$V(1,3) = \min_{K=\{1,2\}} \left(V(1,K) + V(K+1,3) + n_1 \cdot n_{K+1} \cdot n_4 \right) =$$

$$= \min \Big(\underbrace{0 + 20 + 5 \cdot 10 \cdot 1}_{K=1}, \underbrace{100 + 0 + 5 \cdot 2 \cdot 1}_{K=2} \Big) = 70 \; (K=1)$$

$\Rightarrow$ We will multiply matrices $M_1 M_2 M_3$ in the following order $M_1 \cdot (M_2 M_3)$

Exercise 3.8.

State-space $\mathcal{S} = \{0, 1, 2, \cdots, M\}$

Actions: $A_s = \{-1, +1\}$

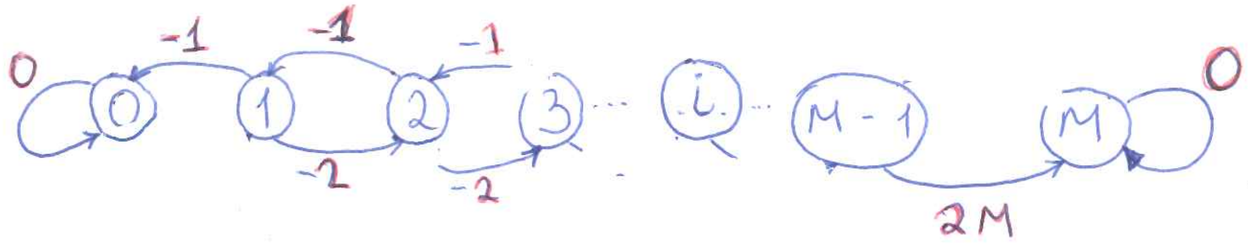
Rewards: $r(s,a) = 0$, when $s = \{0, M\}$

$r(s,-1) = -1$, when $s = \{1, 2, \cdots, M-1\}$

$r(s,+1) = -2$, when $s = \{1, 2, \cdots, M-2\}$

$r(M-1, +1) = 2M$

Trans. probs: $P(s'|s,a) = \mathbb{1}\{s' = s+a\}$, Here we take $\gamma = 1$



Solution: We start with $V_0(s) = 0$ for all s, then we find π_0

- $\pi_0(s) = -1$ for all $s = \{1, \cdots, M-2\}$
- $\pi_0(s) = 1$ if $s = \{M-1\}$

Then the value function at state s is

- $$V_1(s) = \begin{cases} 0, & \text{if } s = 0, M \\ -s & \text{if } s = 1, 2, \cdots, M-2 \\ 2M & \text{if } s = M-1 \end{cases}$$

- For the second step, π_1 differs from π_0 only on state $s = M-2$ where it is optimal to choose $a = +1$.
- For each next step, the policy π_i differs from π_{i-1} only in stat $s = M-i-1$, flipping the optimal action from -1 to $+1$.
- After $M-1$ policy iterations we find the optimal policy $\pi^*(s) = +1$ for all $s = \{1, \cdots, M-1\}$.
- The opimal value function is given by

$$V^*(s) = \begin{cases} 0 & \text{if } s = 0, M \\ 2(s+1) & \text{if } s = 1, 2, 3, \cdots, M-1 \end{cases}$$

3.8 cont.

$$\Rightarrow \cancel{\lambda} \cdot \frac{1-\lambda^{N-1}}{\cancel{1-\lambda}} > \frac{\lambda^{\cancel{2}}}{\cancel{1-\lambda}} \Rightarrow \lambda^{N-1} < 1-\lambda$$

$$(N-1)\log\lambda < \log(1-\lambda)$$

$$N > \frac{\log(1-\lambda)}{\log(\lambda)} + 1$$

Ex. 3.9

Here, instead of considering negative cost, we consider positive rewards and in Bellman equation use minimum over the action space.

State-Space: $S = \{0, 1, 2\}$

Actions: $A = \{1, 2\}$

Rewards: $r(s_1, a_1) = r(0, 1) = 0$

$r(s_2, a_2) = r(0, 2) = \frac{\lambda^2}{1-\lambda}$

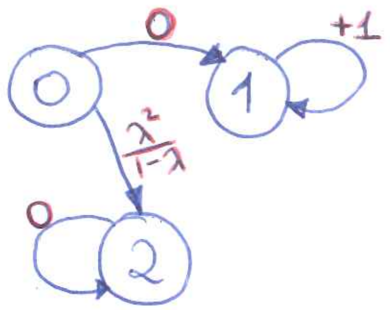
$P(s' | s, a) = P(1 | 0, 1) = 1$

$P(2 | 0, 2) = 1$

States $\{1\}$ and $\{2\}$ are absorbing states

$r(1, \cdot) = 1$

$r(2, \cdot) = 0$



Solution:

$\boxed{\text{Bellman Equation}}$

- $V^*(0) = \min\{\underbrace{r(0,1) + \lambda \cdot V^*(1)}_{a=1}; \underbrace{r(0,2) + \lambda \cdot V^*(2)}_{a=2}\}$

$= \min\{\lambda V^*(1); \frac{\lambda^2}{1-\lambda} + \lambda V^*(2)\}$

- $V^*(1) = 1 + \lambda V^*(1)$
- $V^*(2) = \lambda V^*(2)$

$\boxed{\text{Value Iteration}}$

Choose the next value at each state using the values from previous states and Bellman equations.

- $V_n(0) = \min\{\lambda V_{n-1}(1), \frac{\lambda^2}{1-\lambda} + \lambda V_{n-1}(2)\}$ (*)
- $V_n(1) = 1 + \lambda V_{n-1}(1)$
- $V_n(2) = \lambda V_{n-1}(2)$

- Starting from $V_0(s) = 0$ for $\forall s \in \{0, 1, 2\}$, the value iteration gives:
  - $V_n(1) = 1 + \lambda V_{n-1}(1) = 1 + \lambda[1 + \lambda V_{n-2}(1)] = 1 + \lambda + \lambda^2 \cdot [1 + \lambda V_{n-3}(1)] =$
    $= 1 + \lambda + \lambda^2 + \lambda^3 + \dots + \lambda^{n-1}[1 + V_0(1)] = \sum_{j=0}^{n-1} \lambda^j = \frac{1-\lambda^n}{1-\lambda}$
  - $V_n(2) = \lambda \cdot V_{n-1}(2) = \lambda^2 \cdot V_{n-2}(2) = \dots = \lambda^n \cdot V_0(2) = 0$

a=1 is suboptimal

- From (*) $V_1(0) = \min\{\underbrace{\lambda V_0(1)}_{a=1}; \underbrace{\frac{\lambda^2}{1-\lambda} + \lambda V_0(2)}_{a=2}\} = \min\{\underbrace{0}_{a=1}, \underbrace{\frac{\lambda^2}{1-\lambda}}_{a=2}\} = 0$

Assume that VI converges in N iterations $\Rightarrow$ at iteration N VI chooses Optimal policy, i.e. action 2, at state 0. $\Rightarrow \lambda \cdot V_{N-1}(1) > \frac{\lambda^2}{1-\lambda} + \lambda V_{N-1}(2)$