Question 2 - The $c\mu$ rule

a. we define the following MDP for the server problem:

<u>State space</u>: state S_t will hold all jobs which are still unfinished in the system.

The state space is actually $S = P(\{1 \dots n\})$ - the power set of $\{1 \dots n\}$. Thus $|S| = 2^n$

<u>Action space</u>: action a_t will mark the job chosen by the server. The action space is $A = \{1 \dots n\}$

Thus $|A| = n$

<u>Transition probabilities</u>: when choosing an action $a_t = i$ there are two options - the job is completed and it is taken out of the system or the job is uncompleted and is putted back in the system. Thus

$$P(s \mid S_t = B, a_t = i) = \begin{cases} \mu_i & s = B/\{i\} \\ 1-\mu_i & s = B \\ 0 & \text{else} \end{cases}$$

<u>cost</u>: Cost $c_t(S_t, a_t)$ will be the total cost of all jobs which are still in the system. we notice that the cost doesn't depend on the action or the time.

we have

$$c_t(S_t, a_t) = c_t(S_t) = \sum_{i=1}^{n} I(i \in S_t)\, c_i$$

The total cost we wish to minimize is

$$J^{\pi}(S = \{1 \dots n\}) = E^{\pi, S}\left(\sum_{t=0}^{\infty} c(S_t)\right)$$

we also notice that once we reached ~~out~~ the state $S_t = \{\}$ the process ends and the cost for all $t \geq \tau$ is $c_t(S_t) = 0$

Bellman equation:

$$V(S) = \min_{a = i \in A}\left\{c(S) + \sum_{s'} P(s' \mid S, i) V(s')\right\} = c(S) + \min_{a = i \in A}\left\{\mu_i V(S/\{i\}) + (1-\mu_i) V(S)\right\}$$

b. we wish to prove that the stationary policy $j^* = \arg\max_i \{\mu_i c_i\}$ is an optimal policy. we know that a policy is optimal if its value function satisfies the Bellman equation. we will compute the value function for the above policy and show that it does.

for convenience we mark $S = \{i_1 \dots i_k\}$ s.t. the jobs $i_1 \dots i_k$ are in an ordered way s.t $\mu_{i_1} c_{i_1} \geq \dots \geq \mu_{i_k} c_{i_k}$

In the suggested policy and using the above marking we take job i_1:

$$V(S) = V(\{i_1 \dots i_k\}) = C(S) + \mu_{i_1} V(S/\{i_1\}) + (1-\mu_{i_1}) V(S)$$

$$\Rightarrow V(S) = \frac{C(S)}{\mu_{i_1}} + V(S/\{i_1\}) = \frac{1}{\mu_{i_1}} \sum_{j=1}^{K} C_{i_j} + V(S/\{i_1\})$$

Applying the recursion to $V(S_k)$ we get

$$V(S) = \frac{1}{\mu_{i_1}} \sum_{j=1}^{} C_{i_j} + V(S/\{i_1\}) \overset{\text{rec.}}{=} \frac{1}{\mu_{i_1}} \sum_{j=1}^{} C_{i_j} + \dots + \frac{1}{\mu_{i_k}} \sum_{j=k}^{K} C_{i_j}$$

similarly for $V(S/\{i_a\})$ we take C_{i_a} out from each summand and cancel the $\frac{1}{\mu_{i_a}}$ term

$$V(S/\{i_a\}) = V(S) - \left(C_{i_a} \sum_{j=1}^{a-1} \frac{1}{\mu_{i_j}} + \frac{1}{\mu_{i_a}} \sum_{j=a}^{K} C_{i_j}\right)$$

now we show that for each state S the Bellman equation holds:

$$V(S) = C(S) + \min_{i_a \in A} \{\mu_{i_a} V(S/\{i_a\}) + (1-\mu_{i_a}) V(S)\}$$

$$= C(S) + V(S) + \min_{i_a \in A} \{\mu_{i_a} V(S/\{i_a\}) - \mu_{i_a} V(S)\}$$

$$C(S) = -\min_{i_a \in A} \{\mu_{i_a} (V(S/\{i_a\}) - V(S))\} = \max_{i_a \in A} \{\mu_{i_a} (V(S) - V(S/\{i_a\}))\}$$

$$= \max_{i_a \in A} \left\{ \mu_{i_a} C_{i_a} \sum_{j=1}^{a-1} \frac{1}{\mu_{ij}} + \sum_{j=a}^{K} C_{ij} \right\}$$

The maximum above is achieved for $i_a = i_1$ and thus we get:

$$= \mu_{i_1} C_{i_1} \sum_{j=1}^{1-1} \frac{1}{\mu_{ij}} + \sum_{j=1}^{K} C_{ij} = \sum_{j=1}^{K} C_{ij} = C(S)$$

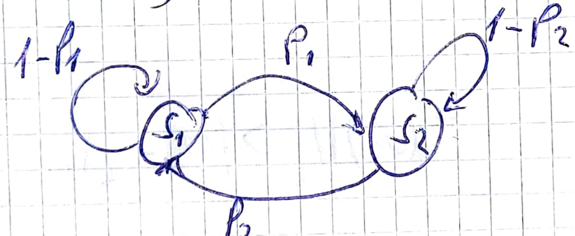
∎

Question 3 - DP operator not contracting in Euclidean norm

The fixed operator T^{π}:

$$(T^{\pi}(J))(s) = r(s, \pi(s)) + \gamma \sum_{s'} P(s' | s, \pi(s)) J(s')$$

We wish to prove that T^{π} is not necessarily a contraction in the Euclidean norm. As hinted, we will use the following MDP:



We also define the rewards as $r(s_1) = r_1$ $r(s_2) = r_2$
we will show that for appropriate P_1, P_2, γ, there exist J_1, J_2 such that:

$$\| T^{\pi}(J_1) - T^{\pi}(J_2) \|_2 \geq \| J_1 - J_2 \|_2$$

First for the above MDP we have:

$$(T^{\pi}(J))(s_i) = r_i + \gamma \left((1-P_i) J(s_i) + P_i J(s_{3-i}) \right)$$

$$\Rightarrow \left(T^{\pi}(J_1) - T^{\pi}(J_2) \right)(s_i) = \gamma \Big((1-P_i)(J_1(s_i) - J_2(s_i)) + P_i (J_1(s_{3-i}) - J_2(s_{3-i})) \Big)$$

$$T^{\pi}(J_1) - T^{\pi}(J_2) = \begin{pmatrix} \gamma\left((1-P_1)(J_1(s_1) - J_2(s_1)) + P_1(J_1(s_2) - J_2(s_2))\right) \\ \gamma\left((1-P_2)(J_1(s_2) - J_2(s_2)) + P_2(J_1(s_1) - J_2(s_1))\right) \end{pmatrix}$$

$$\|T^{\pi}(J_1) - T^{\pi}(J_2)\|_2^2 = \gamma^2\Big(((1-P_1)^2 + P_2^2)(J_1(s_1) - J_2(s_1))^2$$

$$+ 2(P_1(1-P_1) + P_2(1-P_2))(J_1(s_1) - J_2(s_1))(J_1(s_2) - J_2(s_2))$$

$$+ ((1-P_2)^2 + P_1^2)(J_1(s_2) - J_2(s_2))^2\Big)$$

we take the following $J_1 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ $J_2 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$

for these we have:

$$\|J_1 - J_2\|_2^2 = 1$$

$$\|T^{\pi}(J_1) - T^{\pi}(J_2)\|_2^2 = \gamma^2((1-P_2)^2 + P_1^2) = \circledast$$

lets take $P_1 = 1$, $P_2 = 0$, $\gamma = 0.8$:

$$\circledast = 1.28 > 1 = \|J_1 - J_2\|_2^2$$

We have found that for a given J_1, J_2 we get

$$\|T^{\pi}(J_1) - T^{\pi}(J_2)\|_2 > \|J_1 - J_2\|_2$$

Thus T^{π} is not a contraction under the Euclidean norm.