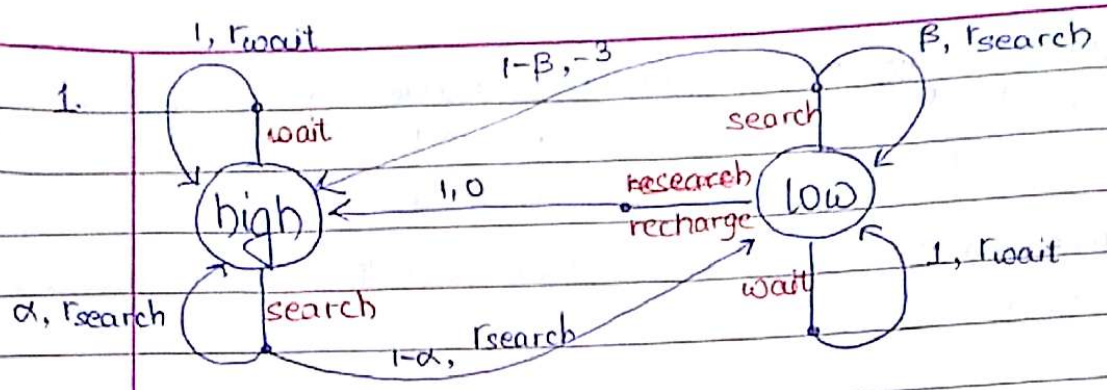
1.



s	a	s'	r	p(s', r | s, a)
high	search	high	rsearch	α
high	search	low	rsearch	1-α
low	search	high	-3	1-β
low	search	low	rsearch	β
high	wait	high	rwait	1
low	wait	low	rwait	1
low	recharge	high	0	1

Explaination,
using the figure above, when is state high, using action search, with the prob of α we go to state high and get reward rsearch.

3.15 We know,

$$G_t = \sum_{K=0}^{\infty} \gamma^K (R_{t+K+1} + c)$$

$$V_\pi(s) = E_\pi [G_t \mid S_t = s]$$

$$= E_\pi \left[\sum_{K=0}^{\infty} \gamma^K (R_{t+K+1} + c) \mid S_t = s \right]$$

$$= E_\pi \left[\sum_{K=0}^{\infty} \gamma^K R_{t+K+1} \mid S_t = s \right] + E_\pi \left[\sum_{K=0}^{\infty} c\gamma^K \mid S_t = s \right]$$

let $V_c = \sum_{K=0}^{\infty} c\gamma^K$

$$\boxed{V_c = \frac{c}{1-\gamma}}$$

thus adding c doesn't relative values of any state

3.16. If we have episodic task, then the value of V_c would change

$$V_c = \sum_{K=0}^{1} c\gamma^K = c\left[\frac{1-\gamma^K}{1-\gamma}\right]$$

hence same value would be added and thus task would be left unchanged.

5. We know,

$$V_*(s) = \max_{\pi} V_\pi(s) \quad \forall s$$

$$q_*(s,a) = \max_{\pi} q_\pi(s,a)$$

} optimal values

using the above equations.

$$V_*(s) = \max_{a \in A(s)} q_*(s,a)$$