Bellman Optimality Equations

1. After reading the question, I got the following information:

   ①. r_s : search reward
   ②. r_w : Wait reward
   ③. α : Probability of staying in high state when searching from high state.
   ④. β : Probability of staying in Low state when searching from low state.
   ⑤. γ : Discount factor

2. Parameter values selected

   $r_s = 6$, $r_w = 1$, $\alpha = 0.8$, $\beta = 0.6$, $\gamma = 0.9$ (This was given)

   The transition probabilities and the expected rewards:

   State : $s = \{high, low\}$ Action : $a = \{search, wait, recharge\}$

   Rerards :

   From high state:
   search → high, $P = \alpha$, $r = r_s$
   search → low, $P = (1-\alpha)$, $r = r_s$
   wait → high, $P = 1$, $r = r_w$
   recharge → low, $P = 1$, $r = 0$

   From low state:
   search → high, $P = (1-\beta)$, $r = -3$
   search → low, $P = \beta$, $r = r_s$
   wait → low, $P = 1$, $r = 0$
   recharge → high, $P = 1$, $r = 0$

   The Bellman Optimality Equations in this problem:

   For high state:

$$V_*(h) = \max \begin{cases} r_s + \gamma[\alpha V_*(h) + (1-\alpha) V_*(l)] & , \text{search} \\ r_w + \gamma V_*(h) & , \text{wait} \end{cases}$$

For **low state:**

$$V_*(l) = \max \begin{cases} -3(1-\beta) + \beta r_s + \gamma(1-\beta)V_*(h) + \beta V_*(l)], & \text{search} \\ r_w + \gamma V_*(l), & \text{wait} \\ 0 + \gamma V_*(h), & \text{recharge} \end{cases}$$

Substituting Parameter Values

For **high state:**

$$V_*(h) = \max \begin{cases} 6 + \gamma[\alpha V_*(h) + (1-\alpha)V_*(l) = 6 + 0.72V_*(h) + 0.18V_*(l) & \text{search} \\ 1 + \gamma V_*(h) = 1 + 0.9V_*(h) & \text{wait} \end{cases}$$

For **low state:**

$$V_*(l) = \max \begin{cases} 0.4 \times 6 - 0.6 \times 3 + \gamma[0.4V_*(h) + 0.6V_*(l)] \\ = 2.4 - 1.8 + 0.36V_*(h) + 0.54V_*(l) \\ = 0.6 + 0.36V_*(h) + 0.54V_*(l) & \text{search} \\ 1 + 0.9V_*(l) & \text{wait} \\ 0.9V_*(h) & \text{recharge} \end{cases}$$

If we set a **hypothesis**: $\pi(h) = \text{search}$

$\pi(l) = \text{recharge}$

From $\pi(l) = \text{recharge}$, we know $V_*(l) = 0.9V_*(h)$

So, take this equation to high state equation:

$$\begin{aligned} V_*(h) &= 6 + 0.72V_*(h) + 0.18V_*(l) \\ &= 6 + 0.72V_*(h) + 0.18(0.9V_*(h)) \\ &= 6 + 0.72V_*(h) + 0.162V_*(h) \\ &= 6 + 0.882V_*(h) \end{aligned}$$

$$V_*(h) - 0.882V_*(h) = 6$$

$$0.118V_*(h) = 6$$

$$V_*(h) \approx 50.85$$

$$\therefore V_*(l) = 0.9 \times 50.85 \approx 45.76$$

Verify optimal actions:

For high state:

search policy: $6 + 0.72(50.85) + 0.18(45.76)$
$= 6 + 36.61 + 8.24 = 50.85$

wait policy: $1 + 0.9(50.85)$
$= 1 + 45.77$
$= 46.77$

$\therefore$ the best action for high state is search

For low state:

search: $0.6 + 0.36(50.85) + 0.54(45.76)$
$= 0.6 + 18.3 + 24.71$
$= 43.62$

wait: $1 + 0.9(45.76) = 1 + 41.18$
$= 42.18$

recharge: $0.9 \times 50.85 = 45.77$

$\therefore$ the best action for low state is recharge

the optimal value function:

$V_*(h) = 50.85$

$V_*(l) = 45.76$

the optimal policy:

$\pi(h) =$ search

$\pi(l) =$ rechavge

Conclusion: So, when in high energy state, the robot should search, taking advantage of low risk and higher rewards.
When in low energy state, the robot should recharge, avoiding the risk of battery depletion and the -3 penalty, and returning a high value state.