1.

For DFS, with noise $w_t = 0$.

State: $x_0 \in \mathcal{X}$, which is a finite space.

goal: $\min_{u_{0:T-1}} q(x_T) + \sum_{t=0}^{T-1} l_t(x_t, u_t)$

s.t. $x_{t+1} = f(x_t, u_t)$

$x_t \in \mathcal{X}$, $u_t \in \mathcal{U}$.

we can transfer it into a DSP problem:

State $\Leftrightarrow$ node. $\mathcal{V} := \left(\bigcup_{t=0}^{T} \{(t, x_t) \mid x_t \in \mathcal{X}\}\right) \cup \{\tau\}$

initial start point $s = (0, x_0)$

τ is an artifical terminal state.

cost $\Leftrightarrow$ edge.

$$\mathcal{C} = \left\{ \left((t, x_t), (t+1, x_{t+1}), c\right) \middle| c = \min_{\substack{u \in \mathcal{U}(x_t) \\ \text{s.t. } x_{t+1} = f(x_t, u_t)}} l_t(x_t, u) \right\} \cup \left\{ (T, x_T), \tau, q(x_T) \right\}$$

a. Thus, for Dijkstra's algorithm, its backward version is

Backward Dijkstra's Algorithm

$OPEN \leftarrow \{\tau\}$, $g_\tau = 0$, $g_i = \infty$ $\forall i \in \mathcal{V} \setminus \{\tau\}$.

while OPEN is not empty do

  Remove $i = \operatorname{argmin}_{j \in OPEN} g_j$ from OPEN

  for $k \in Children(i)$

    if $g_k > g_i + c_{k,i}$

      $g_k \leftarrow g_i + c_{k,i}$

      $Parent(k) \leftarrow i$

      if $k \neq s$ then

        OPEN.append(k)

It actually investigates the same states as Dynamic Programming(DP). For DP.

$$V_t(i) = \min_{j \in \mathcal{V} \setminus \{\tau\}} c_{i,j} + V_{t+1}(j)$$
$$\pi_t(i) = \operatorname{argmin}_{j \in \mathcal{V} \setminus \{\tau\}} c_{i,j} + V_{t+1}(j)$$

will still investigate all children of i to find out the minimum.

(b) Yes, because of the equivalence between DFS and DSP.

Here, a heuristic means a lower bound estimation for cost that takes from state x_t to terminal state x_τ.

Backward A^* Algorithm

OPEN $\leftarrow \{\tau\}$, CLOSED $= \{\}$. $\epsilon \geq 1$

$g_\tau = 0$, $g_i = \infty$ for $i \forall \mathcal{V} \setminus \{\tau\}$

while $s \notin$ CLOSED do

  Remove $i = \operatorname{argmin}_{j \in \text{OPEN}} f_j = \operatorname{argmin}_{j \in \text{OPEN}} g_j + \epsilon h_j$ from OPEN

  CLOSED.append(i)

  for $k \in$ Children(i) and $k \notin$ CLOSED do

    if $g_k > g_i + c_{k,i}$ then

      $g_k \leftarrow g_i + c_{k,i}$

      Parent(k) $\leftarrow i$

      OPEN.append(k)

2.

(a) Because $h^{(1)}$ and $h^{(2)}$ are consistent, then

$$h(x_T) = \max\{h^{(1)}(x_T), h^{(2)}(x_T)\} = 0$$

Also.

$$\begin{aligned} h(x_i) &= \max\{h^{(1)}(x_i), h^{(2)}(x_i)\} \\ &\leq \max\{h^{(1)}(x_j) + cost(x_i, x_j), h^{(2)}(x_j) + cost(x_i, x_j)\} \\ &= cost(x_i, x_j) + \max\{h^{(1)}(x_j), h^{(2)}(x_j)\} \end{aligned}$$

Hence h is also consistent.

(b) If $h^{(1)}$, $h^{(2)}$ are heuristic, then

$$h(x_T) = h^{(1)}(x_T) + h^{(2)}(x_T) = 0.$$

Plus,

$$\begin{aligned} h(x_i) &= h^{(1)}(x_i) + h^{(2)}(x_i) \\ &\leq 2cost(x_i, x_j) + h^{(1)}(x_j) + h^{(2)}(x_j) \\ &= 2cost(x_i, x_j) + h(x_j) \end{aligned}$$

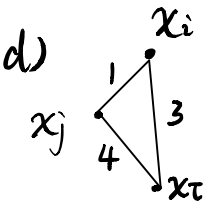
Let $\epsilon = 2$, then h is $2-$consistent.

(c) h is consistent, then

$$\begin{aligned} h(x_i) &\leq cost(x_i, x_T) + \cancel{h(x_T)}^{\,0} \\ &\leq dist(x_i, x_T) \end{aligned}$$

Thus h is also admissible.

(d)



If h is consistent then

$$h(x_i) \leq 1 + h(x_j)$$

$$h(x_j) \leq 4 + 0 = 4$$

If I set $h(x_i)=3$, $h(x_j)=1$, $h(x_T)=0$. then $h(x_i) > 1 + h(x_j)$

$h(x_j) \leq 4+0=4$

however $h(x_i) \leq dist(x_i, x_T)=3$

$h(x_j) \leq dist(x_j, x_T)=4$

Thus $h(x_i)=3$, $h(x_j)=1$, $h(x_T)=0$ is admissible but not consistent.

(e) Assume i is expanded and h is consistent, then $f_i = g_i + h_i \leq f_j \ \forall j \in OPEN$.

Suppose g_i can be improved again, i.e. g_i is larger than the least cost from s to i ($g_i > dist(s, i)$)

Then, there at least exists one state j on an optimal path from s to i, s.t. $j \in OPEN$, $j \notin CLOSED$, $f_i \leq f_j$

However.

$f_i = g_i + h_i > dist(s, i) + h_i = g_j + [dist(j, i) + h_i] \geq g_j + h_j = f_j$

contradicts with $f_i \leq f_j$

Hence g_i cannot be approved.

(f) h is neither consistent, nor admissible.

h is not consistent.

For example, $x_i \in \mathbb{R}^2$, Suppose $x_i = (6, 8)^T$, $x_T = (0, 0)^T$

then $h(x_i) = 8 + 0.4 \times 6 = 10.4$.

$cost(x_i, x_T) = 10$, $h(x_T)=0$

then $h(x_i) > h(x_T) + cost(x_i, x_T)$.

Therefore, h is not consistent.

h is not admissible.

Still suppose $x_i = (6, 8)^T$, $h(x_i) = 10.4$

Then $dist(x_i, x_T) = cost(x_i, x_T) = 10 < h(x_i)$

Thus h is not admissible.

3.

(a) We can formulate this problem as follows:

State space: $\mathcal{X} = [n] = \{1, 2, \cdots, n\}$, state $x_t \in \mathcal{X}$.

Control space: $\mathcal{U} = \{1, 0\}$, where "1" means we choose to recharge, "0" means we choose not to recharge, i.e. browsing & chatting using phone.

Control input $u_t \in \mathcal{U}$.

Motion model: $x_{t+1} \sim P_f(\cdot | x_t, u_t)$

$$= \begin{cases} x_t, & \text{if } u_t = 1, \text{ with prob. } 1-q \\ 1, & \text{if } u_t = 1, \text{ with prob. } q \\ j, & \text{if } u_t = 0, \text{ with prob. } P(x_t, j) \end{cases}$$

Reward/Cost: $\ell(x_t = i, u_t = 0) = -r(i)$, $i \in [n]$

$$\ell(x_t = i, u_t = 1) = \begin{cases} qC, & \text{with prob. } q \\ 0, & \text{with prob. } 1-q \end{cases}$$

Then value function:

$$V_t(x_t) = \mathbb{E}_{x_{t+1:T}}\left[\sum_{\tau=t}^{T-1} \gamma^{\tau-t} \ell(x_\tau, u_\tau) \,\middle|\, x_t\right]$$

As $T \to \infty$, it becomes a discounted Infinite-horizon problem:

$$\min_{u \in \mathcal{U}} V(x) = \min_{u \in \mathcal{U}} \mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t \ell(x_t, u_t) \,\middle|\, x_0 = x\right]$$

$$x_{t+1} \sim P_f(\cdot | x_t, u_t) = \begin{cases} P_f(x_{t+1} | x_t, u_t = 0) = P(x_t, x_{t+1}) \\ P_f(x_t | x_t, u_t = 1) = 1 - q \\ P_f(1 | x_t, u_t = 1) = q \end{cases}$$

$$x_t \sim \mathcal{X} = \{1, 2, 3, \cdots, n\}$$

$$u_t \sim \mathcal{U} = \{1, 0\}$$

Related Bellman equation:

$$V^*(x) = \min_{u \in \mathcal{U}(x)} \{\ell(x, u) + \gamma \mathbb{E}_{x' \sim P_f(\cdot | x, u)}[V(x')]\} \quad \forall x \in \mathcal{X}$$

where $\ell(x, u)$ is defined above.

It can also be written in a compacted version:

$$V^*(x) = \min_{u \in U(x)} \{-r(x)(1-u) + qc\, u + \gamma \sum_{x' \in X} P_f(x'|x,u)\, V^*(x')\}$$

$$= \min_{u \in U(x)} \begin{cases} -r(x) + \gamma \sum_{x'=l}^{n} P(x,x')\, V^*(x') & \text{, if } u = 0,\ l \in [n] \\ qc + r[qV^*(1) + (1-q)V^*(x)] & \text{, if } u = 1 \end{cases}$$

(b) Use Value Iteration (VI) to prove $V^*(i)$ increases in i:

① Assume $t=0$, $V_0(i) = 0$, then $t=1$,

$$V_1^*(i) = \min_{u \in \{1,0\}} \{-r(i), qc\} = -r(i), \text{ since } r(i) > 0.$$

According to that $r(i)$ is decreasing as i increases.

$V_1^*(i) = -r(i)$ is increasing as i increases.

② When $t > 0$, suppose $V_t^*(i)$ increases as i increases, then we want to prove $V_{t+1}^*(i)$ also has the same property. (Mathematical Induction).

Let $Q_t(i, u) = V_t(i)$,

then $V_{t+1}^*(i) = \min\{Q_{t+1}(i, u=1), Q_{t+1}(i, u=0)\}$.

We want to prove

$$V_{t+1}^*(i+1) = \min\{Q_{t+1}(i+1, u=1), Q_{t+1}(i+1, u=0)\} < V_{t+1}^*(i) \quad (*)$$

If we can prove

$$Q_{t+1}(i+1, u=1) > Q_{t+1}(i, u=1)$$

and $Q_{t+1}(i+1, u=0) > Q_{t+1}(i, u=0)$,

then (*) is correct.

When $u=1$, based on VI:

$$V_{t+1}(i) = qc + \gamma[qV_t(1) + (1-q)V_t(i)]$$
$$V_{t+1}(i+1) = qc + \gamma[qV_t(1) + (1-q)V_t(i+1)].$$

By induction assumption, $V_t(i) < V_t(i+1)$, $1-q>0$

then $V_{t+1}(i+1) > V_{t+1}(i)$, i.e. $Q_{t+1}(i+1, u=1) > Q_{t+1}(i, u=1)$

When $u=0$, based on VI:

$$V_{t+1}(i) = -r(i) + \gamma \sum_{j=1}^{n} P(i,j) V_t(j).$$

we know $-r(i)$ is monotonously increasing over $i \in [n]$.

$$\sum_{j=1}^{n} P(i,j) V_t(j) = \underbrace{\left(\sum_{j=1}^{n} P(i,j)\right)}_{>0} V_t(1) + \sum_{j=2}^{n} \underbrace{\left(\sum_{l=j}^{n} P(i,l)\right)}_{\text{monotonously increasing}} \underbrace{(V_t(j) - V_t(j-1))}_{>0 \text{ (Induction assumption)}}$$

$\sum_{j=1}^{n} P(i,j)$: monotonously increasing (stochastic dominance)

Thus, $\sum_{j=1}^{n} P(i,j) V_t(j)$ is also monotonously increasing over $i \in [n]$.

Then $V_{t+1}(i)$ is monotonously increasing,

so $V_{t+1}(i) < V_{t+1}(i+1)$

Therefore, (*) holds and $V^*_{t+1}(i)$ increases in $i \in [n]$

Based on ① ②, the conclusion holds.

(c) $V^*(x)$

$$= \min_{u \in \mathcal{U}(x)} \begin{cases} -r(x) + \gamma \sum_{x'=1}^{n} P(x,x') V^*(x') & \text{if } u=0,\ l \in [n] \\ qc + \gamma [q V^*(1) + (1-q) V^*(x)] & \text{if } u=1 \end{cases}$$

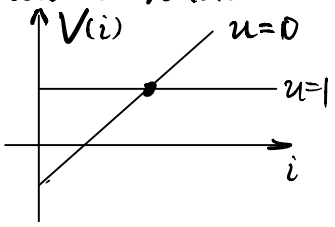
If $u=1$ is optimal, $V^*(i) = qc + \gamma [q V^*(1) + (1-q) V^*(i)]$,
then we can solve for $V^*(i)$, i.e. $V^*(i)$ is a constant. $V^*(i) = \dfrac{qc + \gamma q V^*(1)}{1 - \gamma(1-q)}$.

If $u=0$ is optimal policy,

$$V^*(i) = -r(i) + \gamma \sum_{j=1}^{n} P(i,j) V^*(j), \quad l \in [n]$$

is increasing in $i \in [n]$.

Thus, if $-r(i) + \gamma \sum_{j=1}^{n} P(i,j) V^*(j) - qc - \gamma [q V^*(1) + (1-q) V^*(i)]$
exists zero-point, then the thresh nature holds.



Otherwise, $t=1$ or n is the threshold.