

## Puterman 3.9

The official solution is in sec. 7.3.7. My solution below is different, so you can read both solution with advantage.

We follow the formulation on p. 47, sec. 3.4.1.

- Decision epochs:  $T = \{0, \dots, N\}$ ,  $N \sim \text{DPH}$ .

The decision epochs coincide with the moments, when the driver passes an available spot and has to decide whether to park or not. If there are no available spots  $N$  takes the value 0, while the largest value  $N$  can attain is the total number of spots on the road.

- States: Here  $S'$  relates to the states of the parking spots. Hence, we can define  $S' = \{A \text{ (available)}, T \text{ (taken)}\}$ , and the absorbing state  $\Delta$  refers to the process being terminated. Termination happens if the driver decides to park or we pass the last spot (no decision was taken before decision epoch  $N$ ).

## Puterman 3.9

- Actions: The driver can only take actions, when a decision epoch is reached, i.e. he finds an available spot. In this case, he has two options  $C$  (continue) or  $P$  (park). Hence  $A_s = \{C, P\}$  for  $s = A$ . If  $s \neq A$ , we do not have a decision epoch.
- Rewards: The rewards should be considered as costs in this problem. Note moreover that the cost is not dependent on the time, i.e. which decision epoch we are in. Therefore, we cannot be totally aligned with Sec. 3.4.1 (This is true for all cases with randomly occurring epochs). Instead, we let  $r$  have the subscript referring to the parking spot we make the decision at. The first spot we consider are  $Q$  units away from the restaurant, and consequently parking there would incur a cost of  $Q$ . Naturally, the cost of parking in the  $n$ 'th spot is  $|Q - n|$ . Hence,

$$r_n(s, a) = \begin{cases} 0 & , s = A, a = C \\ |Q - n| & , s = A, a = P \\ 0 & , s = \Delta \end{cases}$$

with terminal cost  $r_\tau(s) = v(s) = c$ .



## Puterman 3.9

The terminal cost is independent of the state and can be assigned a meaningful value according to the author's context. It could be the distance from a parking house, the driver knows always has available spots.

- Transition probabilities: We will only present the non-zero probabilities. Again, we let the subscript refer to which spot we are considering. Furthermore, we let  $M$  denote the total number of spots on the road.

$$p_n(j|s,a) = \begin{cases} P_{i \leftarrow n+1}, j=A, s \in \{A, T\}, a=C \\ 1 - P_{i \leftarrow n+1}, j=T, s \in \{A, T\}, a=C \\ 1, j=\Delta, s=A, a=P \text{ OR } j=\Delta, s=A, a=C \end{cases}$$

for  $n < M$ . When  $n = M$ , we are at the last spot, and we will move to  $\Delta$  no matter what and receive the terminal reward if we don't park. Hence

$$P_M(\Delta|s,a) = 1, \quad \forall s,a \in \{A, T\} \cup \{\Delta\} \times \{C, P\}.$$