1. (a) Many real-world applications. Here is one:

- Consider a gambling game, where there are different slot machines that each cost different amounts to play.
- There is an action to select a machine and pay the fee to play it.

1. (b)

Original finite-horizon value iteration:

$$V^0(s) = R(s)$$

$$V^{K+1}(s) = R(s) + \max_{a \in A} \sum_{s'} T(s,a,s') V^K(s')$$

Updated for state-action reward function:

$$V^0(s) = 0$$

$$V^{K+1}(s) = \max_{a \in A} R(s,a) + \sum_{s'} T(s,a,s') V^K(s')$$

rationale → rewards are only obtained when actions are taken and no action can be taken with zero steps to go.

1.1c) Suppose M is the original MDP.

Suppose M' is the new MDP with state reward function $R'(s)$.

<u>Key Idea:</u> we will introduce new book-keeping states in M' that will keep track of the action that was just executed.

M: State S, Actions A, Transition function T reward function $R(s,a)$.

The new state space S' of M' will contain all states in S along with a new set of states

$$\{ N_{s,a} \mid s \in S, a \in A \}$$

The transition function T' and reward function $R'(s)$ are defined as follows:

$$T'(s, a, N_{s,a}) = 1 \quad \forall\ s \in S \ \& \ a \in A$$

$$T'(N_{s,a}, a', s') = T(s, a, s')$$

$$\forall\ s \in S, a \in A, a' \in A, s' \in S$$

$$R'(s) = 0, \quad \forall\ s \in S$$

$$R'(N_{s,a}) = R(s,a), \quad \forall\ s \in S, a \in A$$

In short, we have simulated a single action in M via two actions in M', where the second action is arbitrary.

2. The state space for M' is S' whose size is $|S|^K$

$\Rightarrow$ each state in M' is a K-tuple of states in S.

Each state in S' is of the form $(s, s_1, \cdots, s_{K-1})$ where each component is a state in S.

The actions of M' are same as those of M

$$A' = A$$

Reward function of M':

$$R'(s, s_1, \cdots, s_{K-1}) = R(s)$$

Transition function of M':

$$T'((s, s_1, \cdots, s_{K-1}), a, \vec{s})$$

$$= Pr(s' \mid a, s, s_1, \cdots, s_{K-1})$$

$$\text{if } \vec{s} = (s', s, s_1, \cdots, s_{K-2})$$

$$= 0, \quad \text{otherwise}$$

No free lunch!

we were able to remove the K-order dynamics ~~for this~~ by increasing the size of state space to $|S|^K$.

3. Bellman equation for $R(s)$ case:

$$V^*(s) = R(s) + \beta \max_{a \in A} \sum_{s'} T(s,a,s') V^*(s')$$

$R(s,a)$ case:

$$V^*(s) = \max_{a \in A} R(s,a) + \beta \sum_{s'} T(s,a,s') V^*(s')$$

$R(s,a,s')$ case:

$$V^*(s) = \max_{a \in A} \sum_{s'} T(s,a,s') \left[R(s,a,s') + \beta V^*(s')\right]$$

4. (a) Suppose π is policy.

$$V_0 = V^{\pi}(s_0)$$
$$V_1 = V^{\pi}(s_1)$$

$$V_0 = R(s_0) + \beta V_1 = \beta V_1$$
$$V_1 = R(s_1) + \beta V_1 = 1 + \beta V_1$$

if $\beta = 1$,

$$V_0 = V_1$$
$$V_1 = 1 + V_1$$

System has no solution

$\Rightarrow$ policy does not have a well-defined value function.

4. (b) For $\beta = 0.9$, we get the following system:

$$V_0 = 0.9\, V_1$$
$$V_1 = 1 + 0.9\, V_1$$

Solution: $V_0 = 9$ and $V_1 = 10$