EE365: Costs and Rewards

Costs and rewards

Value iteration
Costs and rewards
Costs and rewards in a Markov chain

- associate costs (or rewards; more generally, just a function) with Markov chain \( x_0, \ldots, x_T \)

- \( g_t : \mathcal{X} \rightarrow \mathbb{R} \) is the stage cost function

- at time \( t \), we incur cost \( g_t(x) \) for being in state \( x \)

- total cost for \( T \) time periods is (random variable) \( \sum_{t=0}^{T} g_t(x_t) \)

- expected stage cost is \( \pi_t g_t \)

- expected total cost is (number)

\[
J = \mathbb{E} \sum_{t=0}^{T} g_t(x_t) = \pi_0 g_0 + \cdots + \pi_T g_T
\]
Cost evaluation by distribution propagation

\[ J = \pi_0 g_0 + \cdots + \pi_T g_T \]

- evaluate \( \pi_t \) recursively using distribution propagation
- start with \( J = \pi_0 g_0 \), then for \( t = 1, \ldots, T \),
  \[
  \begin{align*}
  \pi_t &= \pi_{t-1} P & \text{// propagate distribution forward in time} \\
  J &= J + \pi_t g_t & \text{// running sum of expected stage costs}
  \end{align*}
  \]
- requires \( n^2 T \) operations (less if \( P \) is sparse)
Value iteration
Value function

write $J$ as

$$J = \pi_0 g_0 + \cdots + \pi_T g_T$$

$$= \pi_0 g_0 + \cdots + \pi_0 P^T g_T$$

$$= \pi_0 (g_0 + P g_1 + \cdots + P^T g_T)$$

$$= \pi_0 (g_0 + P (g_1 + P g_2 + \cdots + P^{T-1} g_T))$$

$$\vdots$$

$$= \pi_0 (g_0 + P (g_1 + \cdots + P (g_{T-1} + P g_T)))$$
Value function

- define
  $$v_t = g_t + Pg_{t+1} + \cdots + P^{T-t}g_T, \quad t = 0, \ldots, T$$

- $v_t : X \rightarrow \mathbb{R}$ is value function at time $t$

- $J = \pi_0 v_0$; more generally,
  $$J = \sum_{t=0}^{s-1} \pi_t g_t + \pi_s v_s$$

- first term is expected cost over $t = 0, \ldots, s - 1$

- second term is expected cost over $t = s, \ldots, T$
Interpretation of value function

we have

\[(v_t)_i = \mathbb{E} \left( \sum_{\tau=t}^{T} g_{\tau}(x_{\tau}) \mid x_t = i \right)\]

so \(v_t\) gives expected future cost starting from each state, at time \(t\)

\(v_t\) summarizes future costs as a current cost
Recursion for value function

- from the definition of $v_t$ we have $v_T = g_T$ and

$$v_{t-1} = g_{t-1} + Pv_t, \quad t = T, \ldots, 1$$

- gives a *backward* recursion for computing $v_T, \ldots, v_0$

- called *value iteration*
Cost evaluation by value iteration

- start with $v_T = g_T$, then for $t = T, \ldots, 1$,
  \[ v_{t-1} = g_{t-1} + P v_t \]  
  // propagate value function backward in time

- let $J = \pi_0 v_0$

- requires $n^2 T$ operations (less if $P$ is sparse)

- an alternative to distribution propagation, that we will need for control
Example: Random walk

- random walk on a 2-dimensional $30 \times 30$ grid, with square obstacle
- outer boundaries are absorbing
Transition probabilities

2 different cases:

probability of staying at current state: $\frac{1}{10}$
Example: Mean time to absorption

Let $E$ be the set of absorbing states, and

$$
\tau = \min\{t > 0 \mid x_t \in E\}
$$

for $t = 0, \ldots, T$ assign costs

$$
g_t(x) = \begin{cases} 
0 & x \in E \\
1 & \text{otherwise}
\end{cases}
$$

cost $J = \mathbb{E} \min(\tau, T)$

gives mean time to absorption as $T \to \infty$
Example: Mean time to absorption

mean time to absorption as a function of initial state
Example: Mean time in each state

- pick state $j$, let

$$g_t(x) = \begin{cases} 
1 & x = j \\
0 & \text{otherwise}
\end{cases}$$

- then $J$ is the mean time spent in state $j$ during time $t \in [0, T]$
Example: Mean time in each state

plot shows the mean time spent in non-absorbing states (initial state $i = (12, 18)$)