LECTURE OUTLINE

• DP for imperfect state info
• Sufficient statistics
• Conditional state distribution as a sufficient statistic
• Finite-state systems
• Examples
REVIEW: IMPERFECT STATE INFO PROBLEM

• Instead of knowing $x_k$, we receive observations

\[ z_0 = h_0(x_0, v_0), \quad z_k = h_k(x_k, u_{k-1}, v_k), \quad k \geq 0 \]

• $I_k$: information vector available at time $k$:

\[ I_0 = z_0, \quad I_k = (z_0, z_1, \ldots, z_k, u_0, u_1, \ldots, u_{k-1}), \quad k \geq 1 \]

• Optimization over policies $\pi = \{\mu_0, \mu_1, \ldots, \mu_{N-1}\}$, where $\mu_k(I_k) \in U_k$, for all $I_k$ and $k$.

• Find a policy $\pi$ that minimizes

\[
J_\pi = \mathbb{E}_{x_0, w_k, v_k} \left\{ g_N(x_N) + \sum_{k=0}^{N-1} g_k(x_k, \mu_k(I_k), w_k) \right\}
\]

subject to the equations

\[
x_{k+1} = f_k(x_k, \mu_k(I_k), w_k), \quad k \geq 0,
\]
\[
z_0 = h_0(x_0, v_0), \quad z_k = h_k(x_k, \mu_{k-1}(I_{k-1}), v_k), \quad k \geq 1
\]
DP ALGORITHM

- DP algorithm:

\[
J_k(I_k) = \min_{u_k \in U_k} \left[ \mathbb{E}_{x_k, w_k, z_{k+1}} \left\{ g_k(x_k, u_k, w_k) + J_{k+1}(I_k, z_{k+1}, u_k) \mid I_k, u_k \right\} \right]
\]

for \( k = 0, 1, \ldots, N - 2 \), and for \( k = N - 1 \),

\[
J_{N-1}(I_{N-1}) = \min_{u_{N-1} \in U_{N-1}} \mathbb{E}_{x_{N-1}, w_{N-1}} \left[ \left\{ g_N\left(x_{N-1}, u_{N-1}, w_{N-1}\right) + g_{N-1}(x_{N-1}, u_{N-1}, w_{N-1}) \mid I_{N-1}, u_{N-1} \right\} \right]
\]

- The optimal cost \( J^* \) is given by

\[
J^* = \mathbb{E}_{z_0} \left\{ J_0(z_0) \right\}.
\]
SUFFICIENT STATISTICS

• Suppose that we can find a function $S_k(I_k)$ such that the right-hand side of the DP algorithm can be written in terms of some function $H_k$ as

$$\min_{u_k \in U_k} H_k(S_k(I_k), u_k).$$

• Such a function $S_k$ is called a sufficient statistic.

• An optimal policy obtained by the preceding minimization can be written as

$$\mu^*_k(I_k) = \overline{\mu}_k(S_k(I_k)), $$

where $\overline{\mu}_k$ is an appropriate function.

• Example of a sufficient statistic: $S_k(I_k) = I_k$

• Another important sufficient statistic

$$S_k(I_k) = P_{x_k|I_k}$$
DP ALGORITHM IN TERMS OF $P_{X_K|I_K}$

- It turns out that $P_{X_k|I_k}$ is generated recursively by a dynamic system (estimator) of the form

$$P_{X_{k+1}|I_{k+1}} = \Phi_k(P_{X_k|I_k}, u_k, z_{k+1})$$

for a suitable function $\Phi_k$.

- DP algorithm can be written as

$$\bar{J}_k(P_{X_k|I_k}) = \min_{u_k \in U_k} \left[ E \left\{ g_k(x_k, u_k, w_k) \right\} + \bar{J}_{k+1}(\Phi_k(P_{X_k|I_k}, u_k, z_{k+1})) \mid I_k, u_k \right]$$

![Diagram of DP algorithm](image)
EXAMPLE: A SEARCH PROBLEM

• At each period, decide to search or not search a site that may contain a treasure.
• If we search and a treasure is present, we find it with prob. \( \beta \) and remove it from the site.
• Treasure’s worth: \( V \). Cost of search: \( C \)
• States: treasure present & treasure not present
• Each search can be viewed as an observation of the state
• Denote

\[ p_k : \text{prob. of treasure present at the start of time } k \]

with \( p_0 \) given.

• \( p_k \) evolves at time \( k \) according to the equation

\[
p_{k+1} = \begin{cases} 
     p_k & \text{if not search,} \\
     0 & \text{if search and find treasure,} \\
     \frac{p_k(1-\beta)}{p_k(1-\beta)+1-p_k} & \text{if search and no treasure.}
\end{cases}
\]
SEARCH PROBLEM (CONTINUED)

• DP algorithm

\[ \overline{J}_k(p_k) = \max \left[ 0, -C + p_k \beta V \right. \]
\[ \left. + (1 - p_k \beta) \overline{J}_{k+1} \left( \frac{p_k (1 - \beta)}{p_k (1 - \beta) + 1 - p_k} \right) \right], \]

with \( \overline{J}_N(p_N) = 0 \).

• Can be shown by induction that the functions \( \overline{J}_k \) satisfy

\[ \overline{J}_k(p_k) = 0, \quad \text{for all } p_k \leq \frac{C}{\beta V} \]

• Furthermore, it is optimal to search at period \( k \) if and only if

\[ p_k \beta V \geq C \]

(expected reward from the next search \( \geq \) the cost of the search)
• Suppose the system is a finite-state Markov chain, with states 1, . . . , n.

• Then the conditional probability distribution $P_{x_k | I_k}$ is a vector

\[
(P(x_k = 1 | I_k), \ldots, P(x_k = n | I_k))
\]

• The DP algorithm can be executed over the $n$-dimensional simplex (state space is not expanding with increasing $k$)

• When the control and observation spaces are also finite sets the problem is called a POMDP (Partially Observed Markov Decision Problem).

• For POMDP it turns out that the cost-to-go functions $\overline{J}_k$ in the DP algorithm are piecewise linear and concave (Exercise 5.7).

• This is conceptually important. It is also useful in practice because it forms the basis for approximations.
INSTRUCTION EXAMPLE

• Teaching a student some item. Possible states are $L$: Item learned, or $\overline{L}$: Item not learned.

• Possible decisions: $T$: Terminate the instruction, or $\overline{T}$: Continue the instruction for one period and then conduct a test that indicates whether the student has learned the item.

• The test has two possible outcomes: $R$: Student gives a correct answer, or $\overline{R}$: Student gives an incorrect answer.

• Probabilistic structure

- Cost of instruction is $I$ per period
- Cost of terminating instruction; 0 if student has learned the item, and $C > 0$ if not.
INSTRUCTION EXAMPLE II

• Let \( p_k \): prob. student has learned the item given the test results so far

\[
p_k = P(x_k | I_k) = P(x_k = L \mid z_0, z_1, \ldots, z_k).
\]

• Using Bayes’ rule we can obtain

\[
p_{k+1} = \Phi(p_k, z_{k+1})
= \begin{cases}
1 - \frac{(1-t)(1-p_k)}{1-(1-t)(1-r)(1-p_k)} & \text{if } z_{k+1} = R, \\
0 & \text{if } z_{k+1} = \overline{R}.
\end{cases}
\]

• DP algorithm:

\[
\overline{J}_k(p_k) = \min \left[ (1-p_k)C, I + E \left\{ \overline{J}_{k+1}(\Phi(p_k, z_{k+1})) \right\} \right].
\]

starting with

\[
\overline{J}_{N-1}(p_{N-1}) = \min\left[ (1-p_{N-1})C, I + (1-t)(1-p_{N-1})C \right].
\]
INSTRUCTION EXAMPLE III

- Write the DP algorithm as

\[ \overline{J}_k(p_k) = \min\left[(1 - p_k)C, I + A_k(p_k)\right], \]

where

\[ A_k(p_k) = P(z_{k+1} = R \mid I_k) \overline{J}_{k+1}(\Phi(p_k, R)) \]

\[ + P(z_{k+1} = \overline{R} \mid I_k) \overline{J}_{k+1}(\Phi(p_k, \overline{R})) \]

- Can show by induction that \( A_k(p) \) are piecewise linear, concave, monotonically decreasing, with

\[ A_{k-1}(p) \leq A_k(p) \leq A_{k+1}(p), \quad \text{for all } p \in [0, 1]. \]
6.231 Dynamic Programming and Stochastic Control
Fall 2011

For information about citing these materials or our Terms of Use, visit: http://ocw.mit.edu/terms.