6.231 DYNAMIC PROGRAMMING

LECTURE 10

LECTURE OUTLINE

- Infinite horizon problems
- Stochastic shortest path (SSP) problems
- Bellman’s equation
- Dynamic programming – value iteration
- Discounted problems as special case of SSP
TYPES OF INFINITE HORIZON PROBLEMS

- Same as the basic problem, but:
  - The number of stages is infinite.
  - Stationary system and cost (except for discounting).

- **Total cost problems**: Minimize

\[
J_\pi(x_0) = \lim_{N \to \infty} E \left\{ \sum_{k=0}^{N-1} \alpha^k g(x_k, \mu_k(x_k), w_k) \right\}
\]

(if the lim exists - otherwise lim sup).
  - Stochastic shortest path (SSP) problems ($\alpha = 1$, and a termination state)
  - Discounted problems ($\alpha < 1$, bounded $g$)
  - Undiscounted, and discounted problems with unbounded $g$

- **Average cost problems**

\[
\lim_{N \to \infty} \frac{1}{N} E \left\{ \sum_{k=0}^{N-1} g(x_k, \mu_k(x_k), w_k) \right\}
\]

- **Infinite horizon characteristics**: Challenging analysis, elegance of solutions and algorithms (stationary optimal policies are likely)
PREVIEW OF INFINITE HORIZON RESULTS

• **Key issue:** The relation between the infinite and finite horizon optimal cost-to-go functions.

• For example, let $\alpha = 1$ and $J_N(x)$ denote the optimal cost of the $N$-stage problem, generated after $N$ DP iterations, starting from some $J_0$

$$J_{k+1}(x) = \min_{u \in U(x)} E \{ g(x, u, w) + J_k(f(x, u, w)) \}, \forall x$$

• Typical results for total cost problems:
  – **Convergence of value iteration to $J^*$:**

$$J^*(x) = \min_{\pi} J_\pi(x) = \lim_{N \to \infty} J_N(x), \forall x$$

  – **Bellman’s equation holds for all $x$:**

$$J^*(x) = \min_{u \in U(x)} E \{ g(x, u, w) + J^*(f(x, u, w)) \}$$

  – **Optimality condition:** If $\mu(x)$ minimizes in Bellman’s Eq., $\{\mu, \mu, \ldots\}$ is optimal.

• Bellman’s Eq. holds for all deterministic problems and “almost all” stochastic problems.

• Other results: True for SSP and discounted; exceptions for other problems.
“EASY” AND “DIFFICULT” PROBLEMS

• Easy problems (Chapter 7, Vol. I of text)
  – All of them are finite-state, finite-control
  – Bellman’s equation has unique solution
  – Optimal policies obtained from Bellman Eq.
  – Value and policy iteration algorithms apply

• Somewhat complicated problems
  – Infinite state, discounted, bounded \( g \) (contractive structure)
  – Finite-state SSP with “nearly” contractive structure
  – Bellman’s equation has unique solution, value and policy iteration work

• Difficult problems (w/ additional structure)
  – Infinite state, \( g \geq 0 \) or \( g \leq 0 \) (for all \( x, u, w \))
  – Infinite state deterministic problems
  – SSP without contractive structure

• Hugely large and/or model-free problems
  – Big state space and/or simulation model
  – Approximate DP methods

• Measure theoretic formulations (not in this course)
STOCHASTIC SHORTEST PATH PROBLEMS

• Assume finite-state system: States 1, . . . , n and special cost-free termination state t
  – Transition probabilities $p_{ij}(u)$
  – Control constraints $u \in U(i)$ (finite set)
  – Cost of policy $\pi = \{\mu_0, \mu_1, \ldots\}$ is

$$J_\pi(i) = \lim_{N \to \infty} E \left\{ \sum_{k=0}^{N-1} g(x_k, \mu_k(x_k)) \left| x_0 = i \right. \right\}$$

  – Optimal policy if $J_\pi(i) = J^*(i)$ for all $i$.
  – Special notation: For stationary policies $\pi = \{\mu, \mu, \ldots\}$, we use $J_\mu(i)$ in place of $J_\pi(i)$.

• Assumption (termination inevitable): There exists integer $m$ such that for all policies $\pi$:

$$\rho_\pi = \max_{i=1,\ldots,n} \max_{x_m \neq t} P\{x_m \neq t \mid x_0 = i, \pi\} < 1$$

• Note: We have $\rho = \max_\pi \rho_\pi < 1$, since $\rho_\pi$ depends only on the first $m$ components of $\pi$.

• Shortest path examples: Acyclic (assumption is satisfied); nonacyclic (assumption is not satisfied)
FINITENESS OF POLICY COST FUNCTIONS

- View 
  \[ \rho = \max_{\pi} \rho_\pi < 1 \]

  as an upper bound on the non-termination prob. during 1st \( m \) steps, regardless of policy used

- For any \( \pi \) and any initial state \( i \)

  \[ P\{x_{2m} \neq t \mid x_0 = i, \pi\} = P\{x_{2m} \neq t \mid x_m \neq t, x_0 = i, \pi\} \times P\{x_m \neq t \mid x_0 = i, \pi\} \leq \rho^2 \]

  and similarly

  \[ P\{x_{km} \neq t \mid x_0 = i, \pi\} \leq \rho^k, \quad i = 1, \ldots, n \]

- So \( E\{\text{Cost between times } km \text{ and } (k + 1)m - 1 \} \)

  \[ \leq m\rho^k \max_{i=1,\ldots,n} \max_{u \in U(i)} |g(i, u)| \]

  and

  \[ |J_\pi(i)| \leq \sum_{k=0}^{\infty} m\rho^k \max_{i=1,\ldots,n} \max_{u \in U(i)} |g(i, u)| = \frac{m}{1 - \rho} \max_{i=1,\ldots,n} \max_{u \in U(i)} |g(i, u)| \]
MAIN RESULT

• Given any initial conditions $J_0(1), \ldots, J_0(n)$, the sequence $J_k(i)$ generated by value iteration,

$$J_{k+1}(i) = \min_{u \in U(i)} \left[ g(i, u) + \sum_{j=1}^{n} p_{ij}(u) J_k(j) \right], \quad \forall \ i$$

converges to the optimal cost $J^*(i)$ for each $i$.

• Bellman’s equation has $J^*(i)$ as unique solution:

$$J^*(i) = \min_{u \in U(i)} \left[ g(i, u) + \sum_{j=1}^{n} p_{ij}(u) J^*(j) \right], \quad \forall \ i$$

$$J^*(t) = 0$$

• A stationary policy $\mu$ is optimal if and only if for every state $i$, $\mu(i)$ attains the minimum in Bellman’s equation.

• Key proof idea: The “tail” of the cost series,

$$\sum_{k=mK}^{\infty} E \left\{ g(x_k, \mu_k(x_k)) \right\}$$

vanishes as $K$ increases to $\infty$. 
OUTLINE OF PROOF THAT $J_N \to J^*$

- Assume for simplicity that $J_0(i) = 0$ for all $i$. For any $K \geq 1$, write the cost of any policy $\pi$ as

$$J_\pi(x_0) = \sum_{k=0}^{mK-1} E \left\{ g(x_k, \mu_k(x_k)) \right\} + \sum_{k=mK}^{\infty} E \left\{ g(x_k, \mu_k(x_k)) \right\}$$

$$\leq \sum_{k=0}^{mK-1} E \left\{ g(x_k, \mu_k(x_k)) \right\} + \sum_{k=K}^{\infty} \rho^k m \max_{i,u} |g(i, u)|$$

Take the minimum of both sides over $\pi$ to obtain

$$J^*(x_0) \leq J_{mK}(x_0) + \frac{\rho^K}{1 - \rho} m \max_{i,u} |g(i, u)|.$$  

Similarly, we have

$$J_{mK}(x_0) - \frac{\rho^K}{1 - \rho} m \max_{i,u} |g(i, u)| \leq J^*(x_0).$$

It follows that $\lim_{K \to \infty} J_{mK}(x_0) = J^*(x_0)$.

- $J_{mK}(x_0)$ and $J_{mK+k}(x_0)$ converge to the same limit for $k < m$ (since $k$ extra steps far into the future don’t matter), so $J_N(x_0) \to J^*(x_0)$.

- Similarly, $J_0 \neq 0$ does not matter.
EXAMPLE

- Minimizing the $E\{\text{Time to Termination}\}$: Let

$$g(i,u) = 1, \quad \forall \ i = 1, \ldots, n, \ u \in U(i)$$

- Under our assumptions, the costs $J^*(i)$ uniquely solve Bellman’s equation, which has the form

$$J^*(i) = \min_{u \in U(i)} \left[ 1 + \sum_{j=1}^{n} p_{ij}(u) J^*(j) \right], \quad i = 1, \ldots, n$$

- In the special case where there is only one control at each state, $J^*(i)$ is the mean first passage time from $i$ to $t$. These times, denoted $m_i$, are the unique solution of the classical equations

$$m_i = 1 + \sum_{j=1}^{n} p_{ij} m_j, \quad i = 1, \ldots, n,$$

which are seen to be a form of Bellman’s equation