6.231 Dynamic Programming and Stochastic Control
Fall 2008

6.231 DYNAMIC PROGRAMMING

LECTURE 17

LECTURE OUTLINE

- We start a four-lecture sequence on advanced infinite horizon DP
- We allow infinite state space, so the stochastic shortest path framework cannot be used any more
- The discounted problem is the proper starting point for this analysis
- The central mathematical structure is that the DP mapping is a contraction mapping (instead of existence of a termination state)

DISCOUNTED PROBLEMS W/ BOUNDED COST

- Stationary system with arbitrary state space

$$x_{k+1} = f(x_k, u_k, w_k), \qquad k = 0, 1, \ldots$$

- Cost of a policy $\pi = \{\mu_0, \mu_1, \ldots\}$

$$J_\pi(x_0) = \lim_{N\to\infty} \mathop{E}_{\substack{w_k \\ k=0,1,\ldots}} \left\{ \sum_{k=0}^{N-1} \alpha^k g\big(x_k, \mu_k(x_k), w_k\big) \right\}$$

with $\alpha < 1$, and for some M, we have $|g(x, u, w)| \leq M$ for all (x, u, w)

- Shorthand notation for DP mappings (operate on functions of state to produce other functions)

$$(TJ)(x) = \min_{u\in U(x)} \mathop{E}_{w} \Big\{ g(x, u, w) + \alpha J\big(f(x, u, w)\big) \Big\}, \ \forall\, x$$

TJ is the optimal cost function for the one-stage problem with stage cost g and terminal cost αJ.

- For any stationary policy μ

$$(T_\mu J)(x) = \mathop{E}_{w} \Big\{ g\big(x, \mu(x), w\big) + \alpha J\big(f(x, \mu(x), w)\big) \Big\}, \ \forall\, x$$

“SHORTHAND” THEORY – A SUMMARY

- Cost function expressions [with $J_0(x) \equiv 0$]

$$J_\pi(x) = \lim_{k\to\infty} (T_{\mu_0} T_{\mu_1} \cdots T_{\mu_k} J_0)(x), \quad J_\mu(x) = \lim_{k\to\infty} (T_\mu^k J_0)(x)$$

- Bellman's equation: $J^* = TJ^*, \quad J_\mu = T_\mu J_\mu$
- Optimality condition:

$$\mu\text{: optimal} \quad \Longleftrightarrow \quad T_\mu J^* = TJ^*$$

- Value iteration: For any (bounded) J and all x,

$$J^*(x) = \lim_{k\to\infty} (T^k J)(x)$$

- Policy iteration: Given μ^k,
  - Policy evaluation: Find J_{μ^k} by solving

$$J_{\mu^k} = T_{\mu^k} J_{\mu^k}$$

  - Policy improvement: Find μ^{k+1} such that

$$T_{\mu^{k+1}} J_{\mu^k} = TJ_{\mu^k}$$

TWO KEY PROPERTIES

- **Monotonicity property:** For any functions J and J' such that $J(x) \le J'(x)$ for all x, and any μ

$$(TJ)(x) \le (TJ')(x), \qquad \forall\ x,$$

$$(T_\mu J)(x) \le (T_\mu J')(x), \qquad \forall\ x.$$

- **Additivity property:** For any J, any scalar r, and any μ

$$\big(T(J + re)\big)(x) = (TJ)(x) + \alpha r, \qquad \forall\ x,$$

$$\big(T_\mu(J + re)\big)(x) = (T_\mu J)(x) + \alpha r, \qquad \forall\ x,$$

where e is the unit function $[e(x) \equiv 1]$.

CONVERGENCE OF VALUE ITERATION

- If $J_0 \equiv 0$,

$$J^*(x) = \lim_{N\to\infty} (T^N J_0)(x), \qquad \text{for all } x$$

Proof: For any initial state x_0, and policy $\pi = \{\mu_0, \mu_1, \ldots\}$,

$$\begin{aligned} J_\pi(x_0) &= E\left\{\sum_{k=0}^{\infty} \alpha^k g\big(x_k, \mu_k(x_k), w_k\big)\right\} \\ &= E\left\{\sum_{k=0}^{N-1} \alpha^k g\big(x_k, \mu_k(x_k), w_k\big)\right\} \\ &\quad + E\left\{\sum_{k=N}^{\infty} \alpha^k g\big(x_k, \mu_k(x_k), w_k\big)\right\} \end{aligned}$$

The tail portion satisfies

$$\left| E\left\{\sum_{k=N}^{\infty} \alpha^k g\big(x_k, \mu_k(x_k), w_k\big)\right\}\right| \le \frac{\alpha^N M}{1-\alpha},$$

where $M \ge |g(x, u, w)|$. Take the min over π of both sides. **Q.E.D.**

BELLMAN'S EQUATION

- The optimal cost function J^* satisfies Bellman's Eq., i.e. $J^* = T(J^*)$.

Proof: For all x and N,

$$J^*(x) - \frac{\alpha^N M}{1-\alpha} \leq (T^N J_0)(x) \leq J^*(x) + \frac{\alpha^N M}{1-\alpha},$$

where $J_0(x) \equiv 0$ and $M \geq |g(x,u,w)|$. Applying T to this relation, and using Monotonicity and Additivity,

$$\begin{aligned}(TJ^*)(x) - \frac{\alpha^{N+1} M}{1-\alpha} &\leq (T^{N+1} J_0)(x) \\ &\leq (TJ^*)(x) + \frac{\alpha^{N+1} M}{1-\alpha}\end{aligned}$$

Taking the limit as $N \to \infty$ and using the fact

$$\lim_{N\to\infty} (T^{N+1} J_0)(x) = J^*(x)$$

we obtain $J^* = TJ^*$. **Q.E.D.**

THE CONTRACTION PROPERTY

- **Contraction property:** For any bounded functions J and J', and any μ,

$$\max_x \bigl|(TJ)(x) - (TJ')(x)\bigr| \leq \alpha \max_x \bigl|J(x) - J'(x)\bigr|,$$

$$\max_x \bigl|(T_\mu J)(x) - (T_\mu J')(x)\bigr| \leq \alpha \max_x \bigl|J(x) - J'(x)\bigr|.$$

Proof: Denote $c = \max_{x \in S} \bigl|J(x) - J'(x)\bigr|$. Then

$$J(x) - c \leq J'(x) \leq J(x) + c, \qquad \forall\ x$$

Apply T to both sides, and use the Monotonicity and Additivity properties:

$$(TJ)(x) - \alpha c \leq (TJ')(x) \leq (TJ)(x) + \alpha c, \qquad \forall\ x$$

Hence

$$\bigl|(TJ)(x) - (TJ')(x)\bigr| \leq \alpha c, \qquad \forall\ x.$$

Q.E.D.

IMPLICATIONS OF CONTRACTION PROPERTY

- Bellman's equation $J = TJ$ has a unique solution, namely J^*, and for any bounded J, we have

$$\lim_{k\to\infty} (T^k J)(x) = J^*(x), \qquad \forall\, x$$

Proof: Use

$$\begin{aligned}\max_x \big|(T^k J)(x) - J^*(x)\big| &\le \max_x \big|(T^k J)(x) - (T^k J^*)(x)\big| \\ &\le \alpha^k \max_x \big|J(x) - J^*(x)\big|\end{aligned}$$

- **Convergence rate:** For all k,

$$\max_x \big|(T^k J)(x) - J^*(x)\big| \le \alpha^k \max_x \big|J(x) - J^*(x)\big|$$

- Also, for each stationary μ, J_μ is the unique solution of $J = T_\mu J$ and

$$\lim_{k\to\infty} (T_\mu^k J)(x) = J_\mu(x), \qquad \forall\, x,$$

for any bounded J.

NEC. AND SUFFICIENT OPT. CONDITION

- A stationary policy μ is optimal if and only if $\mu(x)$ attains the minimum in Bellman's equation for each x; i.e.,

$$TJ^* = T_\mu J^*.$$

Proof: If $TJ^* = T_\mu J^*$, then using Bellman's equation ($J^* = TJ^*$), we have

$$J^* = T_\mu J^*,$$

so by uniqueness of the fixed point of T_μ, we obtain $J^* = J_\mu$; i.e., μ is optimal.

- Conversely, if the stationary policy μ is optimal, we have $J^* = J_\mu$, so

$$J^* = T_\mu J^*.$$

Combining this with Bellman's equation ($J^* = TJ^*$), we obtain $TJ^* = T_\mu J^*$. **Q.E.D.**

COMPUTATIONAL METHODS

- Value iteration and variants
  - Gauss-Seidel version
  - Approximate value iteration
- Policy iteration and variants
  - Combination with value iteration
  - Modified policy iteration
  - Asynchronous policy iteration
- Linear programming

$$\text{maximize} \quad \sum_{i=1}^{n} J(i)$$

$$\text{subject to} \quad J(i) \leq g(i,u) + \alpha \sum_{j=1}^{n} p_{ij}(u) J(j), \quad \forall\ (i,u)$$

- Approximate linear programming: use in place of $J(i)$ a low-dim. basis function representation

$$\tilde{J}(i,r) = \sum_{k=1}^{m} r_k w_k(i)$$

and low-dim. LP (with many constraints)