6.231 Dynamic Programming and Stochastic Control
Fall 2008

6.231 DYNAMIC PROGRAMMING

LECTURE 18

LECTURE OUTLINE

- One-step lookahead and rollout for discounted problems
- Approximate policy iteration: Infinite state space
- Contraction mappings in DP
- Discounted problems: Countable state space with unbounded costs

ONE-STEP LOOKAHEAD POLICIES

- At state i use the control $\overline{\mu}(i)$ that attains the minimum in

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

where $\tilde{J}$ is some approximation to J^*.

- Assume that $\hat{J} \leq \tilde{J} + \delta e$, for some δ, where

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

Then

$$J_{\overline{\mu}} \leq \hat{J} + \frac{\alpha\delta}{1-\alpha} e \leq \tilde{J} + \frac{\delta}{1-\alpha} e.$$

- Assume that $J^* - \epsilon e \leq \tilde{J} \leq J^* + \epsilon e$, for some ϵ. Then

$$J_{\overline{\mu}} \leq J^* + \frac{2\alpha\epsilon}{1-\alpha} e.$$

APPLICATION TO ROLLOUT POLICIES

- Let $\mu_1, \ldots, \mu_M$ be stationary policies, and let

$$\tilde{J}(i) = \min\big\{J_{\mu_1(i)}, \ldots, J_{\mu_M(i)}\big\}, \qquad \forall\ i.$$

- Then, for all i, and $m = 1, \ldots, M$, we have

$$\begin{aligned}\hat{J}(i) &= \min_{u \in U(i)} \left[g(i,u) + \alpha \sum_{j=1}^{n} p_{ij}(u) \tilde{J}(j) \right] \\ &\le \min_{u \in U(i)} \left[g(i,u) + \alpha \sum_{j=1}^{n} p_{ij}(u) \tilde{J}_{\mu_m}(j) \right] \\ &\le J_{\mu_m}(i)\end{aligned}$$

- Taking minimum over m,

$$\hat{J}(i) \le \tilde{J}(i), \qquad \forall\ i.$$

- Using the preceding slide result with $\delta = 0$,

$$J_{\overline{\mu}}(i) \le \tilde{J}(i) = \min\big\{J_{\mu_1(i)}, \ldots, J_{\mu_M(i)}\big\}, \qquad \forall\ i,$$

i.e., the rollout policy $\overline{\mu}$ improves over each μ_m.

APPROXIMATE POLICY ITERATION

- Suppose that the policy evaluation is approximate, according to,

$$\max_x |J_k(x) - J_{\mu^k}(x)| \le \delta, \qquad k = 0, 1, \ldots$$

and policy improvement is approximate, according to,

$$\max_x |(T_{\mu^{k+1}} J_k)(x) - (T J_k)(x)| \le \epsilon, \qquad k = 0, 1, \ldots$$

where δ and ϵ are some positive scalars.

- **Error Bound:** The sequence $\{\mu^k\}$ generated by approximate policy iteration satisfies

$$\limsup_{k\to\infty} \max_{x\in S} \big(J_{\mu^k}(x) - J^*(x)\big) \le \frac{\epsilon + 2\alpha\delta}{(1-\alpha)^2}$$

- Typical practical behavior: The method makes steady progress up to a point and then the iterates J_{μ^k} oscillate within a neighborhood of J^*.

CONTRACTION MAPPINGS

- Given a real vector space Y with a norm $\|\cdot\|$ (i.e., $\|y\| \geq 0$ for all $y \in Y$, $\|y\| = 0$ if and only if $y = 0$, and $\|y + z\| \leq \|y\| + \|z\|$ for all $y, z \in Y$)

- A function $F : Y \mapsto Y$ is said to be a *contraction mapping* if for some $\rho \in (0, 1)$, we have

$$\|F(y) - F(z)\| \leq \rho\|y - z\|, \qquad \text{for all } y, z \in Y.$$

ρ is called the *modulus of contraction* of F.

- For $m > 1$, we say that F is an *m-stage contraction* if F^m is a contraction.

- Important example: Let S be a set (e.g., state space in DP), $v : S \mapsto \Re$ be a positive-valued function. Let $B(S)$ be the set of all functions $J : S \mapsto \Re$ such that $J(s)/v(s)$ is bounded over s.

- We define a norm on $B(S)$, called the *weighted sup-norm*, by

$$\|J\| = \max_{s \in S} \frac{|J(s)|}{v(s)}.$$

- Important special case: The discounted problem mappings T and T_μ [for $v(s) \equiv 1$, $\rho = \alpha$].

CONTRACTION MAPPING FIXED-POINT TH.

- **Contraction Mapping Fixed-Point Theorem:** If $F : B(S) \mapsto B(S)$ is a contraction with modulus $\rho \in (0,1)$, then there exists a unique $J^* \in B(S)$ such that

$$J^* = FJ^*.$$

Furthermore, if J is any function in $B(S)$, then $\{F^k J\}$ converges to J^* and we have

$$\|F^k J - J^*\| \leq \rho^k \|J - J^*\|, \qquad k = 1, 2, \ldots.$$

- Similar result if F is an m-stage contraction mapping.

- This is a special case of a general result for contraction mappings $F : Y \mapsto Y$ over normed vector spaces Y that are *complete*: every sequence $\{y_k\}$ that is Cauchy (satisfies $\|y_m - y_n\| \to 0$ as $m, n \to \infty$) converges.

- The space $B(S)$ is complete (see the text for a proof).

A DP-LIKE CONTRACTION MAPPING I

- Let $S = \{1, 2, \ldots\}$, and let $F : B(S) \mapsto B(S)$ be a linear mapping of the form

$$(FJ)(i) = b(i) + \sum_{j \in S} a(i,j)\, J(j), \qquad \forall\ i$$

where $b(i)$ and $a(i,j)$ are some scalars. Then F is a contraction with modulus ρ if

$$\frac{\sum_{j \in S} |a(i,j)|\, v(j)}{v(i)} \le \rho, \qquad \forall\ i$$

- Let $F : B(S) \mapsto B(S)$ be a mapping of the form

$$(FJ)(i) = \min_{\mu \in M} (F_\mu J)(i), \qquad \forall\ i$$

where M is parameter set, and for each $\mu \in M$, F_μ is a contraction mapping from $B(S)$ to $B(S)$ with modulus ρ. Then F is a contraction mapping with modulus ρ.

A DP-LIKE CONTRACTION MAPPING II

- Let $S = \{1, 2, \ldots\}$, let M be a parameter set, and for each $\mu \in M$, let

$$(F_\mu J)(i) = b(i,\mu) + \sum_{j \in S} a(i,j,\mu)\, J(j), \qquad \forall\ i$$

- We have $F_\mu J \in B(S)$ for all $J \in B(S)$ provided $b_\mu \in B(S)$ and $V_\mu \in B(S)$, where

$$b_\mu = \big\{b(1,\mu), b(2,\mu), \ldots\big\},\ \ V_\mu = \big\{V(1,\mu), V(2,\mu), \ldots\big\},$$

$$V(i,\mu) = \sum_{j \in S} \big|a(i,j,\mu)\big|\, v(j), \qquad \forall\ i$$

- Consider the mapping F

$$(FJ)(i) = \min_{\mu \in M} (F_\mu J)(i), \qquad \forall\ i$$

We have $FJ \in B(S)$ for all $J \in B(S)$, provided $b \in B(S)$ and $V \in B(S)$, where

$$b = \big\{b(1), b(2), \ldots\big\}, \qquad V = \big\{V(1), V(2), \ldots\big\},$$

with $b(i) = \max_{\mu \in M} b(i,\mu)$ and $V(i) = \max_{\mu \in M} V(i,\mu)$.

DISCOUNTED DP - UNBOUNDED COST I

- State space $S = \{1, 2, \ldots\}$, transition probabilities $p_{ij}(u)$, cost $g(i, u)$.
- Weighted sup-norm
$$\|J\| = \max_{i \in S} \frac{|J(i)|}{v_i}$$
on $B(S)$: sequences $\{J(i)\}$ such that $\|J\| < \infty$.
- Assumptions:

(a) $G = \{G(1), G(2), \ldots\} \in B(S)$, where
$$G(i) = \max_{u \in U(i)} \big|g(i, u)\big|, \qquad \forall\ i$$

(b) $V = \{V(1), V(2), \ldots\} \in B(S)$, where
$$V(i) = \max_{u \in U(i)} \sum_{j \in S} p_{ij}(u)\, v_j, \qquad \forall\ i$$

(c) There exists an integer $m \geq 1$ and a scalar $\rho \in (0, 1)$ such that for every policy π,
$$\alpha^m \frac{\sum_{j \in S} P(x_m = j \mid x_0 = i, \pi)\, v_j}{v_i} \leq \rho, \qquad \forall\ i$$

DISCOUNTED DP - UNBOUNDED COST II

- Example: Let $v_i = i$ for all $i = 1, 2, \ldots$
- Assumption (a) is satisfied if the maximum expected absolute cost per stage at state i grows no faster than linearly with i.
- Assumption (b) states that the maximum expected next state following state i,

$$\max_{u \in U(i)} E\{j \mid i, u\},$$

also grows no faster than linearly with i.

- Assumption (c) is satisfied if

$$\alpha^m \sum_{j \in S} P(x_m = j \mid x_0 = i, \pi)\, j \le \rho\, i, \qquad \forall\ i$$

It requires that for all π, the expected value of the state obtained m stages after reaching state i is no more than $\alpha^{-m} \rho\, i$.

- If there is bounded upward expected change of the state starting at i, there exists m sufficiently large so that Assumption (c) is satisfied.

DISCOUNTED DP - UNBOUNDED COST III

- Consider the DP mappings T_μ and T,

$$(T_\mu J)(i) = g\big(i,\mu(i)\big) + \alpha \sum_{j\in S} p_{ij}\big(\mu(i)\big) J(j), \qquad \forall\, i,$$

$$(TJ)(i) = \min_{u\in U(i)} \left[g(i,u) + \alpha \sum_{j\in S} p_{ij}(u) J(j) \right], \ \forall\, i$$

- **Proposition:** Under the earlier assumptions, T and T_μ map $B(S)$ into $B(S)$, and are m-stage contraction mappings with modulus ρ.

- The m-stage contraction properties can be used to essentially replicate the analysis for the case of bounded cost, and to show the standard results:
  - The value iteration method $J_{k+1} = TJ_k$ converges to the unique solution J^* of Bellman's equation $J = TJ$.
  - The unique solution J^* of Bellman's equation is the optimal cost function.
  - A stationary policy μ is optimal if and only if $T_\mu J^* = TJ^*$.