Dynamic Programming Lecture #15

Outline:

- Discounted Problems
- Contraction Mapping Proof

Discounted Problem Setup

- System: Controlled finite-state Markov chain with transition probabilities $p_{ij}(u)$.
- Cost:

$$\min_{\pi} \lim_{N \to \infty} E \left\{ \sum_{k=0}^{N-1} \alpha^k g(x_k, \mu_k(x_k)) \right\}$$

- Analysis is (almost) exactly same as before. Main idea:

$$\sum_{k=0}^{\infty} \alpha^k g \approx \sum_{k=0}^{N-1} \alpha^k g \pm \sum_{k=N}^{\infty} \alpha^k G \approx \sum_{k=0}^{N-1} \alpha^k g \pm G \frac{\alpha^N}{1-\alpha}$$

  Future penalties explicitly discounted

- Immediate vs future trade-off:

$$\begin{aligned} J^*(x_0) &= E \left\{ \sum_{k=0}^{\infty} \alpha^k g(x_k, \mu^*(x_k)) \right\} \\ &= E \left\{ g(x_0, \mu^*(x_0)) + \sum_{k=1}^{\infty} \alpha^k g(x_k, \mu^*(x_k)) \right\} \\ &= E \left\{ g(x_0, \mu^*(x_0)) + \alpha \sum_{k=0}^{\infty} g(x_{k+1}, \mu^*(x_{k+1})) \right\} \\ &= E \left\{ g(x_0, \mu^*(x_0)) + \alpha J(x_1) \right\} \end{aligned}$$

- Suggest Bellman equation:

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

Main Results

- Value iteration: Converges to J^* for any J_0.

$$J_{k+1}(i) = \min_{u \in U(i)} g(i,u) + \underbrace{\alpha}_{\text{new term}} \sum_{j=1}^{n} p_{ij}(u) J_k(j)$$

- J^* unique solution to Bellman equation.

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

- μ-specific value iteration & Bellman equation.

$$J_{k+1}(i) = g(i, \mu(k)) + \alpha \sum_{j=1}^{n} p_{ij}(\mu(i)) J_k(j)$$

$$J_\mu(i) = g(i, \mu(i)) + \alpha \sum_{j=1}^{n} p_{ij}(\mu(i)) J_\mu(j)$$

- Stationary policy μ is optimal $\Leftrightarrow$ it is minimizer in Bellman equation.
- Policy iteration terminates with optimal policy.

$$\mu^+(i) = \arg \min_{u \in U(i)} g(i,u) + \alpha \sum_{j=1}^{n} p_{ij}(u) J_\mu(j)$$

Contraction Mapping Theorem

- Let $\|\cdot\|$ be a norm on $\mathcal{R}^n$, e.g.,

$$\|x\| = \left(\sum_{i=1}^{n} x_i^2 \right)^{1/2}$$

  or

$$\|x\| = \max_i |x_i|$$

- The mapping $T : S \to S$ on the closed set S is a CONTRACTION if for some $\rho < 1$,

$$\|Ts - Ts'\| \leq \rho \|s - s'\|$$

- Notation: Write Ts rather than $T(s)$.
- THEOREM:
  1. There exists a unique $s^* \in S$ such that

$$s^* = Ts^*$$

  2. The iterations

$$s_{k+1} = Ts_k$$

     converge to s^* for any initial $s_0 \in S$ and satisfy the error bound

$$\|s_k - s^*\| \leq \rho^k \|s_0 - s^*\|$$

- Contraction mapping arguments are very versatile:
  - Applicable to more general spaces than $\mathcal{R}^n$.
  - Unique solution of $x = f(x)$ in case $|\nabla f(x)| \leq \rho < 1$.
  - Unique solution of $\frac{dx}{dt} = f(x)$ in case $|f(x_a) - f(x_b)| \leq L |x_a - x_b|$.

Contraction Mapping Proof

- Inspect

$$\left\|T^2 s - Ts\right\| = \|T(Ts) - T(s)\| \leq \rho \|Ts - s\|$$

- Likewise

$$\left\|T^3 s - T^2 s\right\| = \left\|T(T^2 s) - T(Ts)\right\| \leq \rho \left\|T^2 s - Ts\right\| \leq \rho^2 \|Ts - s\|$$

- In general

$$\left\|T^{n+1} s - T^n s\right\| \leq \rho^n \|Ts - s\|$$

- Inspect $\|T^n s - T^m s\|$:

$$T^n s - T^m s = T^n s - T^{n-1} s + T^{n-1} s ... - T^{m+1} s + T^{m+1} s - T^m s$$

$$\Rightarrow$$

$$\|T^n s - T^m s\| \leq (\rho^{n-1} + \rho^{n-2} + ... + \rho^m) \|Ts - s\|$$

So starting from s_0,

$$\|s_n - s_m\| = \|T^n s - T^m s\| \leq \frac{1}{1-\rho} \rho^m \|s_1 - s_0\|$$

- Implication: s_k is a Cauchy sequence and hence convergent to some s^*.

Contraction Mapping Proof, cont.

- To show that s^* is a fixed point:

$$\begin{aligned}\|Ts^* - s^*\| &\leq \|Ts^* - s_n\| + \|s_n - s^*\| \\ &= \|Ts^* - Ts_{n-1}\| + \|s_n - s^*\| \\ &\leq \rho \|s^* - s_{n-1}\| + \|s_n - s^*\| \\ &\to 0\end{aligned}$$

- To show that s^* is unique, suppose s_a^* and s_b^* are fixed points. Then

$$\|s_a^* - s_b^*\| = \|Ts_a^* - Ts_b^*\| \leq \rho \|s_a^* - s_b^*\| \Rightarrow \|s_a^* - s_b^*\| = 0$$

- To show the error bounds,

$$\|T^n s_0 - s^*\| = \|T^n s_0 - Ts^*\| \leq \rho \|s_{n-1} - s^*\| = \rho \left\|T^{n-1} s_0 - s^*\right\|$$

  Repeating recursively results in

$$\|s_n - s^*\| \leq \rho^n \|s_0 - s^*\|$$

Connection to DP

- Define $T : \mathcal{R}^n \to \mathcal{R}^n$ by

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

- Notation: For vectors x and y, $x \leq y$ denotes element-by-element $x(i) \leq y(i)$
- Monotonicity: $J \leq J' \Rightarrow TJ \leq TJ'$.
- Proof:

$$\begin{aligned}(TJ)(i) &= \min_{u \in U(i)} g(i,u) + \alpha \sum_{j=1}^{n} p_{ij}(u) J(j) \\ &\leq \min_{u \in U(i)} g(i,u) + \alpha \sum_{j=1}^{n} p_{ij}(u) J'(j) \\ &= (TJ')(i)\end{aligned}$$

- FACT: T is a contraction under the $\max$-norm.

Connection to DP, cont

- Want to show that for some $\rho < 1$,

$$\|TJ - TJ'\| \leq \rho \, \|J - J'\|$$

- Proof: Let

$$c = \|J - J'\|$$

Then

$$J - c \leq J' \leq J + c$$

By monotonicity:

$$T(J - c) \leq TJ' \leq T(J + c)$$

Inspect:

$$\begin{aligned} (T(J-c))(i) &= \min_{u \in U(i)} g(i,u) + \alpha \sum_{j=1}^{n} p_{ij}(u)(J(j) - c) \\ &= \min_{u \in U(i)} g(i,u) + \alpha \sum_{j=1}^{n} p_{ij}(u)J(j) - \alpha \sum_{j=1}^{n} p_{ij}(u)c \\ &= TJ - \alpha c \end{aligned}$$

Likewise,

$$T(J + c) = TJ + \alpha c$$

Therefore,

$$TJ - \alpha c \leq TJ' \leq TJ + \alpha c$$

or

$$\|TJ - TJ'\| \leq \alpha c$$

$$\Rightarrow$$

$$\|TJ - TJ'\| \leq \alpha \, \|J - J'\|$$

Implications for DP

- Value iterations $J_{k+1} = TJ_k$, i.e.,

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

  converge to J^* at an exponential rate.

- J^* is the unique solution to $TJ^* = J^*$, i.e.,

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

- Similar results for μ-specific value iterations $J_{k+1} = T_\mu J_k$, i.e.,

$$J_{k+1}(i) = g(i,\mu(i)) + \sum_{j=1}^{n} p_{ij}(\mu(i)) J_k(j)$$

  and μ-specific Bellman equation $J_\mu = T_\mu J_\mu$, i.e.,

$$J_\mu(i) = g(i,\mu(i)) + \sum_{j=1}^{n} p_{ij}(\mu(i)) J_\mu(j)$$