EE 223: Stochastic Estimation and Control **Spring 2007**

Lecture 19 — March 22

Lecturer: Venkat Anantharam *Scribe: Jiening Zhan*

Review

Recall that we are considering the discounted infinite horizon problem with state evolution:

$$x_{k+1} = f(x_k, u_k, w_k), \quad k \geq 0 \ .$$

The aim was to minimize (informally):

$$E\left[\sum_{k=0}^{\infty} \alpha^k g(X_k, U_k, w_k)\right] \ , \tag{19.1}$$

where $(w_k, k \geq 0)$ are i.i.d. random variables.

To solve this problem, we defined a mapping T from functions on $\mathcal{X}$ to functions on $\mathcal{X}$ given by:

$$(TJ)(x) = \min_u E\left[g(x,u,w) + \alpha J(f(x,u,w))\right] \tag{19.2}$$

The above equation is known as Bellman's equation. We will look at this mapping in the special case of a finite state controlled Markov chain with finite control space. There, we have $P(u) = [P_{ij}(u)]$ and $g(i,u,w) = g(i,u)$, $i \in \mathcal{X}, u \in U$. Bellman's equation becomes:

$$(TJ)(i) = \min_u \left[g(i,u) + \alpha \sum_{j \in \mathcal{X}} P_{ij}(u) J(j)\right] \tag{19.3}$$

We showed that this mapping has a unique fixed point (we did this under the assumption of bounded cost: $|g(x,u,w)| \leq M \ \forall (x,u,w)$). If J^* denotes the optimal overall cost then $TJ^* = J^*$. Furthermore, any minimizing $\mu = \mu(x)$ in Bellman's equation $TJ^* = J^*$ defines an optimal stationary Markov strategy.

Contraction Mapping

Definition 1. *A metric space (M,d) is a set M with a notion of distance:*

(a) $d(x,y) \geq 0$ *with* $d(x,y) = 0$ *iff* $x = y$

(b) $d(x,y) = d(y,x) \ \forall x,y \in M$

(c) $d(x,y) + d(y,z) \geq d(y,z) \ \forall x,y,z \in M$

A metric space (M, d) is called complete if every Cauchy sequence in M has a limit. i.e if $z_n \in M$ and $\forall \epsilon \geq 0$, $\exists N(\epsilon)$ s.t $d(z_n, z_m) < \epsilon \; \forall n, m \geq N(\epsilon)$ then $\exists z \in M$ with $d(z_n, z) \to 0$ as $n \to \infty$.

Definition 2. *A mapping $T : (M, d) \mapsto (M, d)$ is called a contraction mapping if for some $0 \leq \beta < 1$ and $\forall z_1, z_2 \in M$:*

$$d(Tz_1, Tz_2) \leq \beta d(z_1, z_2) \; . \tag{19.4}$$

The condition (19.4) is called a Lipshitz condition.

Theorem: Every contraction mapping on a complete metric space has a unique fixed point

To get a feeling for what this theorem says, consider $(M, d) = (\mathbb{R}$, Euclidean), and the function $f(x) = e^x$ as shown in Figure 1. ($\mathbb{R}$, Euclidean) is a complete metric space, but this function has no fixed point. Note that this function is not a contraction mapping on ($\mathbb{R}$, Euclidean). Since the values of the function are able to change faster than the change in the argument, the function is able to "escape". If there were some $\beta < 1$ such that the Lipshitz condition (19.4) held, then the argument would "catch up with the values" and there would be a fixed point. The contraction mapping theorem applies much more generally than to real valued functions on the real line, but this example displays some of the underlying intuition. We now prove the contraction mapping theorem.

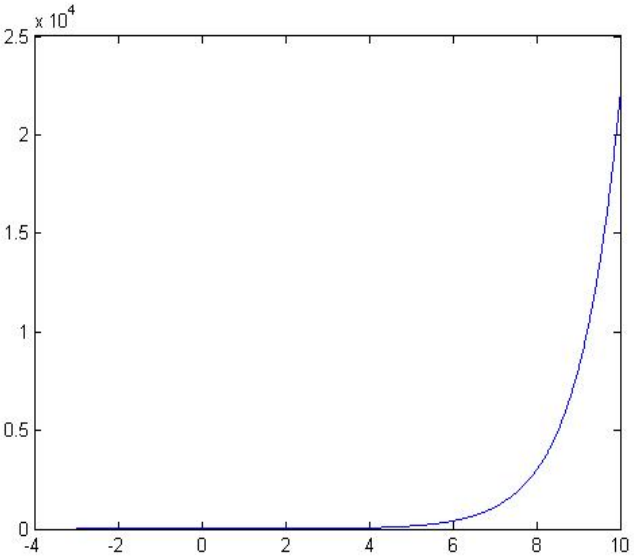

Figure 19.1. Exponential Function

Proof. Take any $z \in M$. Consider the sequence $(T^k z, k \geq 0)$. We have $d(T^{k+1}z, T^k z) \leq \beta d(T^k z, T^{k-1} z) \; \forall k \geq 1$. Since $\beta < 1$, this implies that $(T^k z, k \geq 0)$ is Cauchy. Hence it has a limit (by the completeness assumption). Call the limit z^*. Since $\lim_{k \to \infty} T^k z$ is also z^*, we have $Tz^* = z^*$.

If $\tilde{z}$ were any other fixed point i.e $T\tilde{z} = \tilde{z}$, then

$$d(z^*, \tilde{z}) = d(Tz^*, T\tilde{z}) \leq \beta d(z^*, \tilde{z}) \ .$$

This is only possible if $d(z^*, \tilde{z}) = 0$ i.e $z^* = \tilde{z}$.

From the contraction mapping theorem, it follows that Bellman's equation has a unique fixed point. This is because Bellman's equation defines a contraction mapping on the metric space comprised of all bounded functions on $\mathcal{X}$ with the L^∞ norm. To see that T is a contraction mapping, take two bounded functions J_1 and J_2 on $\mathcal{X}$. Let u_2^* achieve the minimum in the definition of $TJ_2(x)$ (it is not really necessary to assume that such a minimizer exists, since we could be more careful via the use of an approximation argument, but we prefer to make this assumption for simplicity). We have:

$$\begin{aligned} TJ_1(x) - TJ_2(x) &= \min_u E[g(x,u,w) + \alpha J_1(f(x,u,w))] - E[g(x,u_2^*,w) + \alpha J_2(f(x,u_2^*,w))] \quad (19.5) \\ &\leq E[g(x,u_2^*,w) + \alpha J_1(f(x,u_2^*,w))] - E[g(x,u_2^*,w) + \alpha J_2(f(x,u_2^*,w))] \\ &= \alpha E[J_1(f(x,u_2^*,w)) - J_2(f(x,u_2^*,w))] \\ &\leq \alpha \left\| J_1 - J_2 \right\|_\infty \end{aligned}$$

where $\left\| J_1 - J_2 \right\|_\infty = sup_{x \in X} |J_1(x) - J_2(x)|$. A similar argument in the other direction gives $TJ_2(x) - TJ_1(x) \leq \alpha \left\| J_1 - J_2 \right\|_\infty$.

Value Iteration

Value Iteration can be used to solve Bellman's equation.

(a) Start with some

$$\begin{bmatrix} J(1) \\ J(2) \\ . \\ . \\ . \\ J(d) \end{bmatrix}$$

(b) Find

$$\begin{bmatrix} TJ(1) \\ TJ(2) \\ . \\ . \\ . \\ TJ(d) \end{bmatrix}$$

where $TJ(i) = \min_u \left\{ g(i,u) + \alpha \sum_j P_{ij}(u) J(j) \right\}$

(c) If $TJ = J$, then $J = J^*$ and μ for which $TJ = T_\mu J$ is an optimal stationary Markov strategy. If $TJ \neq J$, then replace J by TJ and repeat.

Note. Let $\mu : \mathcal{X} \mapsto \mathcal{U}$ be s.t $TJ = T_\mu J \triangleq g_\mu + \alpha P_\mu J$, where

$$T_\mu \triangleq \begin{bmatrix} T_\mu J(1) \\ T_\mu J(2) \\ . \\ . \\ . \\ T_\mu J(d) \end{bmatrix}$$

$$g_\mu \triangleq \begin{bmatrix} g(1, \mu(1)) \\ g(2, \mu(2)) \\ . \\ . \\ . \\ g(d, \mu(d)) \end{bmatrix}$$

$$P_\mu \triangleq \left[\ P_{ij}(\mu(i))\ \right]$$

So $(T_\mu J)(i)$ is defined as $g(i, \mu(i)) + \alpha \sum_j P_{ij}(\mu(i))J(j)$

Let $\mu^{(k)}, k \geq 0$ be s.t $\mu^{(k)}$ is a minimizer at the kth step. For example:

current step	next step	minimizer
$J = J^{(0)}$	$TJ^{(0)}$	$\mu^{(0)}$
$TJ^{(0)} = J^{(1)}$	$T^2J^{(0)}$	$\mu^{(1)}$
$T^2J^{(0)} = J^{(2)}$	$T^3J^{(0)}$	$\mu^{(2)}$
.	.	.
.	.	.
.	.	.
$T^kJ^{(0)} = J^{(k)}$	$T^{(k+1)}J^{(0)}$	$\mu^{(k)}$

Then (proved last time), there is a finite K s.t $\mu^{(k)}$ is an optimal strategy $\forall k \geq K$. But $\mu^{(k)}$ may oscillate infinitely often between optimal strategies as shown in the following example.
Let $\mathcal{X} = \{1, 2, 3\}$, $\mu = \{a, b\}$,

$$P_{ij}(a) = \begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ 1 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix}.$$

$$P_{ij}(b) = \begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 1 \\ \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix}.$$

Note that the transition probabilities from states 1 and 3 do not depend on the control choice. Define costs:

$$\begin{aligned} g(1,a) = g(1,b) &= 10 \\ g(2,a) = g(2,b) &= 0 \\ g(3,a) = g(3,b) &= 10 \end{aligned} \tag{19.6}$$

Let $0 < \alpha < 1$ be the discount factor. It follows that

$$\begin{aligned} TJ(1) &= 10 + \frac{1}{2}\alpha J(2) + \frac{1}{2}\alpha J(3) \\ TJ(2) &= \min\{\alpha J(1), \alpha J(2)\} \\ TJ(3) &= 10 + \frac{1}{2}\alpha J(2) + \frac{1}{2}\alpha J(1) \end{aligned} \tag{19.7}$$

If $J(1) > J(3)$, then $TJ(1) < TJ(3)$. So if we started with $J^{(0)}$ where $J^{(0)}(1) \neq J^{(0)}(3)$, then during value iteration, the control choice will flip flop at state 2 infinitely often.

Policy Iteration

(This relies on the fact that there are only finitely many stationary strategies since $\mathcal{X}$ and $\mathcal{U}$ are finite).

(a) Start with a stationary policy $\mu^{(0)}$ and solve $T_{\mu^{(0)}} J^*_{\mu^{(0)}} = J^*_{\mu^{(0)}}$.

(b) Check if $J^*_{\mu^{(0)}}$ is a fixed point of Bellman's equation, i.e whether $TJ^*_{\mu^{(0)}} = J^*_{\mu^{(0)}}$. If so, then $J^*_{\mu^{(0)}} = J^*$ (overall optimal cost) and $\mu^{(0)}$ defines an optimal stationary Markov strategy. If not, then the process of checking gives $\mu^{(1)}$ where:

$$TJ^*_{\mu^{(0)}} = T_{\mu^{(1)}} J^*_{\mu^{(0)}} < T_{\mu^{(0)}} J^*_{\mu^{(0)}} = J^*_{\mu^{(0)}} \ , \tag{19.8}$$

(where the strictness of inequality is in least at one coordinate and $\leq$ everywhere).

(c) Replace $\mu^{(0)}$ by $\mu^{(1)}$ and iterate.

Note that applying $T^k_{\mu^{(1)}}$ to the inequality $T_{\mu^{(1)}} J^*_{\mu^{(0)}} < J^*_{\mu^{(0)}}$ we conclude, via monotonicity of $T_{\mu^{(1)}}$, that $J^*_{\mu^{(1)}} < J^*_{\mu^{(0)}}$ (strictness at least one coordinate and $\leq$ everywhere). This process will therefore find an optimal strategy within $|\mathcal{X}|^{|\mathcal{U}|}$ steps, because there are only this many stationary strategies.

Aside: For a stationary Markov strategy μ, J^*_μ is the unique fixed point of the mapping T_μ (which exists and is unique because T_μ is a contraction mapping). Recall that

$$(T_\mu J)(i) = g(i, \mu(i)) + \alpha \sum_j P_{ij}(u(i)) J(j) \ . \tag{19.9}$$

In vector notation, this reads:

$$T_\mu J = g_\mu + \alpha P_\mu J \ . \tag{19.10}$$

So

$$J_\mu^* = (I - \alpha P_\mu)^{-1} g_\mu \ . \tag{19.11}$$