Last Time

Value Function - Expectation of future Rewards starting at s

$$V^*(s) = \max_a \left(R(s,a) + \gamma \mathop{E}_{s' \sim T(s,a)} \left[V^*(s') \right] \right)$$

Two Algorithms

Policy Iteration

Policy Evalution V^π

Policy Improvement

Value Iteration

$V' = B[V]$

$$B[V](s) = \max_a \left(R(s,a) + \gamma \underline{E}[V(s')] \right)$$

$\underline{E}[V(s')] = \sum_{s'} T(s'|s,a) V(s')$

Today

V.I. converges

V^* unique

Julia perf + debugging

Q value

$$Q^\pi(s,a) = R(s,a) + \gamma E\left[V^\pi(s') \right]$$

$$V^\pi(s) = \max_a Q^\pi(s,a)$$

$$\pi^*(s) = \arg\max_a Q^\pi(s,a)$$

~~$\arg\max_a R(s,a)$~~ Myopic

$M = \mathbb{R}^{|S|}$

Theorem Let $\{V_k\}_{k=1}^{\infty}$ be a sequence of value fcns for a discrete MDP, generated by applying B. If $\gamma < 1$, then $\lim_{k\to\infty} V_k = V^*$.

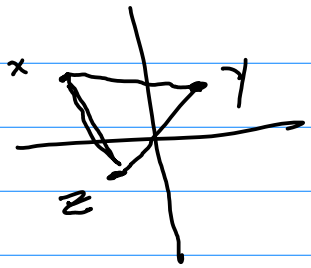
$V_{k+1} = B[V_k]$

Def Let M be a set. A metric on M is a function $d: M \times M \to [0, \infty)$ that satisfies

i) $d(x,y) = 0$ iff $x = y$

ii) $d(x,y) = d(y,x)$ $\forall$ (for all) $x, y \in M$ (in)

iii) $d(x,z) \le d(x,y) + d(y,z)$ $\forall\ x, y, z \in M$



Def A contraction mapping on set M is a function $f: M \to M$ satisfying

$$d(f(x), f(y)) \le c\, d(x,y)$$

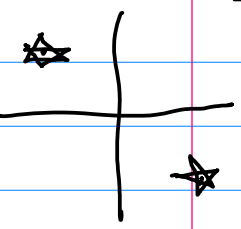
for some c, $0 \le c < 1$ and all $x, y \in M$

Def x^* is said to be a fixed point of f if $f(x^*) = x^*$

Banach's Theorem If f is a contraction mapping on (M, d), then

i) f has a single, unique fixed point

ii) If $\{x_n\}$ is a sequence defined by $x_{k+1} = f(x_k)$ then $\lim_{k\to\infty} x_k = x^*$



$\|V\|_1 = \text{sum } V$

$\|V\|_2 = \sqrt{\sum V_i^2}$

$\|V\|_n = \sqrt[n]{\sum |V_i|^n}$ $\quad \|V\|_\infty = \max_i |V_i|$

Prove that

$\|V_1 - V_2\|_\infty$ is a metric

ii) $\|V_1 - V_2\|_\infty = \|V_2 - V_1\|_\infty$

iii) $\max_i |x_i - z_i| = \max |x - y + y - z|$ ← triangle inequality

$d(x,z)$ $\quad |x_i - y_i + y_i - z_i| \le |x_i - y_i| + |y_i - z_i|$

$\max |x - y + y - z| \le \max (|x-y| + |y-z|)$

$\le \max |x-y| + \max |y-z|$

$d(x,y) + d(y,z)$

Lemma 1 $\|V_1 - V_2\|_\infty$ is a metric on $\mathbb{R}^{|S|}$

$$\|B[V_1] - B[V_2]\|_\infty = \max_{s \in S} |B[V_1](s) - B[V_2](s)|$$

$$= \max_{s \in S} \left| \max_{a \in A} \left(R(s,a) + \gamma \sum_{s'} T(s'|s,a) V_1(s')\right) - \max_a \left(R(s,a) + \gamma \sum_{s'} T(s'|s,a) V_2(s')\right)\right|$$

$$\le \max_{s \in S} \left| \max_a \left(R(s,a) + \gamma \sum_{s'} T(s'|s,a) V_1(s') - \left(R(s,a) + \gamma \sum T(s'|s,a) V_2(s')\right)\right)\right|$$

$$= \max_{\substack{s \in S \\ a \in A}} \gamma \left| \sum_{s'} T(s'|s,a) \left(V_1(s') - V_2(s')\right)\right|$$

$$\le \max_{\substack{s \in S \\ a \in A}} \gamma \sum_{s'} T(s'|s,a) \left|V_1(s') - V_2(s')\right|$$

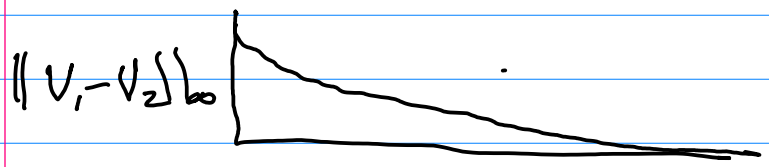
$$\le \max_{\substack{s \in S \\ a \in A}} \gamma \sum_{s'} T(s'|s,a) \|V_1 - V_2\|_\infty$$

$$= \gamma \|V_1 - V_2\|_\infty \max_{s,a} \sum_{s'} T(s'|s,a)$$

$$\leq \gamma \|V_1 - V_2\|_\infty$$

Lemma 2 B is a contraction mapping on $\mathbb{R}^{|S|}$

By L1, L2, Banach's Theorem, we have proved the original Theorem □



$S = \{1, 2, 3\}$ $|S| = 3$ "cardinality" size of S

$\mathbb{R}^{|S|}$ = 3-dimensional vector space

$V_1 = [0.0, 0.1, 0.2]$

$V_2 = [0.1, 0.3, 0.5]$

Breakout Rooms

$S = \{1, 2, 3\}$ $A = \{1, -1\}$

desired

$$T(s'|s,a) = \begin{cases} 0.7 & \text{if } s' = \text{clamp}(s+a, 1, 3) \\ 0.3 & \text{if } s' = \text{clamp}(s-a, 1, 3) \\ 0 & \text{o.w.} \end{cases}$$

opposite

$V^*(3) = 98$

$R(2, a) = 0$

$R(1, a) = 1$

$\gamma = 0.99$

Optimal policy? $\pi^*(s) = \rightarrow$

$V_1 = 96.67$

$V_2 = 96.62$

$V_1 = 95$

$V_2 = 95$

$R(1,a) + \gamma E(V(s'))$

$V_1 = 1 + \gamma(0.7V_1 + 0.3V_2)$ $\quad$ $1 + \gamma(0.7V_2 + 0.3V_1)$

$V_2 = 0 + \gamma(0.7V_1 + 0.3 \cdot 98)$ $\quad$ $0 + \gamma(0.7 \cdot 98 + 0.3V_1)$