Last Time: ~~I~~ Value Function

How can we find optimal policies?

- Policy Iteration
  - ↳ 2 steps
    1. Evaluate Policy
    2. Policy Improvement
- Value Iteration ← Today

Does Value Iteration converge?

$$V^*(s) = \max_a \left(R(s,a) + \gamma E\left(V^*(s') \mid s' \sim T(s,a)\right]\right)$$

Bellman's Equation

Value Iteration

$V_0 = \text{rand}$ (or zeros)

while $\|V_k - V_{k-1}\|_\infty > \varepsilon$ ← infinity norm

$\quad V_{k+1} = B[V_k]$ ← Bellman Operator

$\quad k = k+1$

V_k

$\|x\|_\infty = \text{maximum } |x|$

$\|x\|_1 = |x_1| + |x_2| + |x_3| \ldots$

$\|x\|_2 = \sqrt{x_1^2 + x_2^2 + x_3^2}$

$\|x\|_3 = \sqrt[3]{|x_1|^3 + |x_2|^3 + \ldots}$

$\vdots$

$\|x\|_\infty = \text{maximum } |x|$

$$B[V](s) = \max_a \left(R(s,a) + \gamma E[V(s') \mid s' \sim T(s,a)]\right]$$

will this converge?

Theorem: Let $\{V_k\}_k$ be a sequence of value functions for a discrete MDP calculated with $V_{k+1} = B[V_k]$. If $\gamma < 1$, then

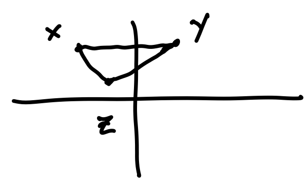
$$\lim_{k \to \infty} V_k = V^*$$

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

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

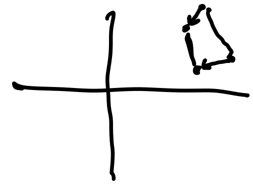
ii) $d(x,y) = d(y,x) \quad \forall\, x,y \in M$

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



Def. A <u>contraction mapping</u> on (M, d) is a function f

~~Ron~~ $f: M \to M$ that satisfies



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

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

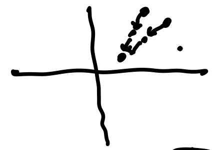
Def x^* is said to be a <u>fixed point</u> of g if $g(x^*) = x^*$

<u>Banach's Theorem</u>: If f is a contraction mapping on (M, d), then

i) f has a single, unique fixed point, x^*

ii) If $\{x_n\}$ is a sequenced defined by $x_{k+1} = f(x_k)$

then $\lim_{n \to \infty} x_n = x^*$



Prove that 1. $d(V_1, V_2) = \|V_1 - V_2\|_\infty$ is a metric on $\mathbb{R}^{|S|}$

2. B is a contraction mapping

$\|x - y\|_\infty = \max|x - y|$

i) $\max|x - y| = 0$ iff $x = y$

ii) $|x - y| = |-(x - y)| = |y - x|$

$\therefore \max|x - y| = \max|y - x|$

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

$\vdots$

iii) $\max|x - z| = \max|x - y + y - z|$

$|x_i - y_i + y_i - z_i| \le |x_i - y_i| + |y_i - z_i| \quad \forall i$

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

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

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

~~B~~

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

$|\max(x)| \le \max|x|$ ~~true~~ trust me

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

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

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

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

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

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

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

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

$$\sum_a P(a|b) = 1$$

$$\|B[V_1] - B[V_2]\|_\infty \le \gamma \|V_1 - V_2\|_\infty \quad \square$$

T~~heorem~~ ~~Value Iteration~~ c

By Lemma 1 and 2 Theorem 1 is proven $\square$

Convergence rate related to γ (S, A, R, T, γ)

If $\gamma = 0$, $V^*(s) = \max_a R(s,a)$

Finite Time

objective $\sum_{t=0}^{N} R(s_t, a_t)$

start at end $V_N(s) = \max_a R(s,a)$

work backwards $V_{k-1}(s) = \max_a \left(R(s,a + \sum_{s'} T(s'|s,a) \; V_k(s')\right)$

Continuous States ?

$$S = \mathbb{R}^n \qquad A = \mathbb{R}^m \qquad s' \sim \mathcal{N}(As + Ba, \Sigma)$$

Linea

$$R(s,a) = s^T Q s + a^T R a$$

LQR

Finite Time Case

$$V_n(s) = \max_a \left(R(s,a) + \int T(s' \mid s,a)\, V_{n+1}(s')\, ds' \right)$$

$$= \max_a \left(s^T Q s + a^T R a + \int \mathcal{N}(s' \mid As + Ba, \Sigma)\, V_{n+1}(s')\, ds' \right.$$

$$V_n(s) = s^T P_n s + q_n$$

$$= \max_a \left(s^T Q s + a^T R a + \int \mathcal{N}(s' \mid As + Ba, \Sigma)\left(s'^T P_{n+1} s' + q_{n+1}\right) ds' \right.$$

integral

$$\mathrm{Tr}(\Sigma P_{n+1}) + (As + Ba)^T P_{n+1} (As + Ba)$$

$$= q_{n+1} + s^T Q s + \mathrm{Tr}(\Sigma P_{n+1}) + \max_a \left(a^T R a + (As + Ba)^T P_{n+1} (As + Ba) \right)$$

find derivative, set $= 0$

$$a^* = -(B^T P_{n+1} B + R)^{-1} B P_{n-1} A s$$

substitute

$$V_n(s) = s^T P_n s + q_n$$

where

$$P_n = A^T P_{n+1} A - A^T P_{n+1} B (B^T B + R)^{-1} B^T P_{n+1} A + Q$$

$$q_n = q_{n+1} + \mathrm{Tr}(\Sigma P_{n+1})$$

$$P_N \rightarrow P_{N-1} \rightarrow P_{N-2} \text{ — } \cdots \; P_0$$

$$\pi_n(s) = -(B^T P_{n+1} B + R)^{-1} B P_{n+1} A s$$

does not depend on Σ