Last Time

Last Time

$\max_a R(s,a)$ ← Myopic

$\max \underbrace{R(s,a) + \gamma E[U(s')]}_{Q(s,a)}$

$\max_a Q^*(s,a)$

- How do we reason about the **future consequences** of actions in an MDP?

Last Time

- How do we reason about the **future consequences** of actions in an MDP?
- What are the basic **algorithms for solving MDPs**?

Policy Iteration
Evaluate policy
Improve policy

Value Iteration
Improve V estimate

Guiding Questions

Guiding Questions

- Does value iteration always converge?
- Is the value function unique?

Value Iteration: The Bellman Operator

Value Iteration: The Bellman Operator

Algorithm: Value Iteration

while $\|V - V'\|_\infty < \epsilon$

$V \leftarrow V'$

$V' \leftarrow B[V]$

return V'

Value Iteration: The Bellman Operator

Algorithm: Value Iteration

while $\|V - V'\|_\infty < \epsilon$

$V \leftarrow V'$

$V' \leftarrow B[V]$

return V'

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

Value Iteration Convergence

Value Iteration Convergence

Theorem 1: Let $\{V_1, \ldots, V_\infty\}$ be a sequence of value functions for a discrete MDP generated by the recurrence $V_{k+1} = B[V_k]$. If $\gamma < 1$, then $\lim_{k\to\infty} V_k = V^*$.

Metrics

Metrics

Definition: Let M be a set. A *metric* on M is a function $d : M \times M \to [0, \infty)$ which satisfies the following three conditions for all $x, y, z \in M$:

Metrics

<u>Definition</u>: Let M be a set. A *metric* on M is a function $d : M \times M \to [0, \infty)$ which satisfies the following three conditions for all $x, y, z \in M$:

1. $d(x, y) = 0$ if and only if $x = y$

Metrics

Definition: Let M be a set. A *metric* on M is a function $d : M \times M \to [0, \infty)$ which satisfies the following three conditions for all $x, y, z \in M$:

1. $d(x, y) = 0$ if and only if $x = y$
2. $d(x, y) = d(y, x)$

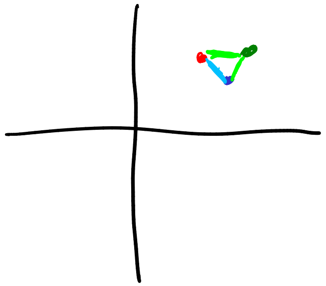
Metrics

<u>Definition</u>: Let M be a set. A *metric* on M is a function $d : M \times M \to [0, \infty)$ which satisfies the following three conditions for all $x, y, z \in M$:

1. $d(x, y) = 0$ if and only if $x = y$
2. $d(x, y) = d(y, x)$
3. $d(x, y) \leq d(x, z) + d(z, y)$

• x
• y
• z

$\|x - y\|_2$



Contraction Mappings

Contraction Mappings

<u>Definition</u>: A *contraction mapping* on metric space (M, d) is a function $f : M \to M$ satisfying

$$d(f(x), f(y)) \leq \alpha\, d(x, y)$$

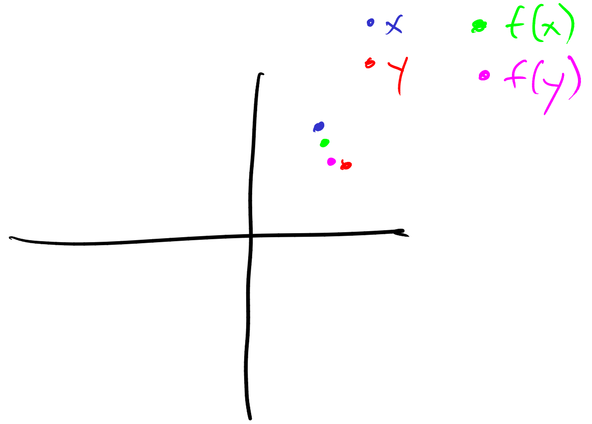
for some α, $0 \leq \alpha \leq 1$ and all x and y in M.

Contraction Mappings

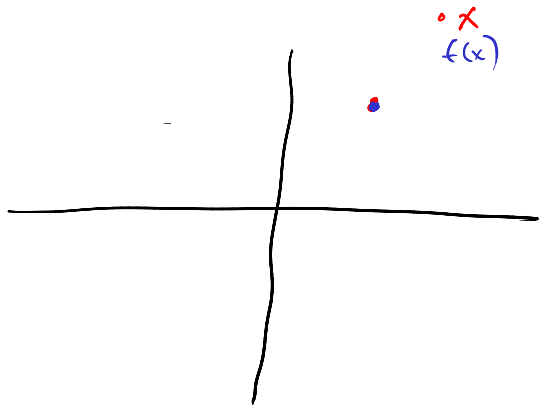
<u>Definition</u>: A *contraction mapping* on metric space (M, d) is a function $f : M \to M$ satisfying

$$d(f(x), f(y)) \leq \alpha\, d(x, y)$$

for some α, $0 \leq \alpha \leq 1$ and all x and y in M.



<u>Definition</u>: x^* is said to be a *fixed point* of f if $f(x^*) = x^*$.



Contraction Mappings

<u>Definition</u>: A *contraction mapping* on metric space (M, d) is a function $f : M \to M$ satisfying

$$d(f(x), f(y)) \leq \alpha \, d(x, y)$$

for some α, $0 \leq \alpha \leq 1$ and all x and y in M.

<u>Definition</u>: x^* is said to be a *fixed point* of f if $f(x^*) = x^*$.

Script: contraction_mapping.jl

Banach's Theorem

Banach's Theorem

Theorem (Banach): If f is a contraction mapping on metric space (M, d), then

1. f has a single, unique fixed point x^*.
2. If $\{x_k\}$ is a sequence defined by $x_{k+1} = f(x_k)$, then $\lim_{k \to \infty} x_k = x^*$.

Max Norm

Max Norm

Lemma 1: $\left(\mathbb{R}^{|S|}, \|\cdot\|_\infty\right)$ is a metric space.

Max Norm

Lemma 1: $\left(\mathbb{R}^{|S|}, \|\cdot\|_\infty\right)$ is a metric space.

Definition: Let M be a set. A *metric* on M is a function $d : M \times M \to [0, \infty)$ which satisfies the following three conditions for all $x, y, z \in M$:

1. $d(x, y) = 0$ if and only if $x = y$
2. $d(x, y) = d(y, x)$
3. $d(x, y) \leq d(x, z) + d(z, y)$

Max Norm

Lemma 1: $\left(\mathbb{R}^{|S|}, \|\cdot\|_\infty\right)$ is a metric space.

Definition: Let M be a set. A *metric* on M is a function $d : M \times M \to [0, \infty)$ which satisfies the following three conditions for all $x, y, z \in M$:

1. $d(x, y) = 0$ if and only if $x = y$
2. $d(x, y) = d(y, x)$
3. $d(x, y) \leq d(x, z) + d(z, y)$

Proof:

$\|x\|_n = \left(\sum |x_i|^n\right)^{1/n}$

$\|x-y\|_1$

Max Norm

<u>Lemma 1</u>: $(\mathbb{R}^{|S|}, \|\cdot\|_\infty)$ is a metric space.

<u>Definition</u>: Let M be a set. A *metric* on M is a function $d : M \times M \to [0, \infty)$ which satisfies the following three conditions for all $x, y, z \in M$:

1. $d(x, y) = 0$ if and only if $x = y$
2. $d(x, y) = d(y, x)$
3. $d(x, y) \leq d(x, z) + d(z, y)$

Proof:

Note: $\|x - y\|_\infty = \max_i |x_i - y_i|$

Max Norm

Lemma 1: $(\mathbb{R}^{|S|}, \|\cdot\|_\infty)$ is a metric space.

Definition: Let M be a set. A *metric* on M is a function $d : M \times M \to [0, \infty)$ which satisfies the following three conditions for all $x, y, z \in M$:

1. $d(x,y) = 0$ if and only if $x = y$
2. $d(x,y) = d(y,x)$
3. $d(x,y) \leq d(x,z) + d(z,y)$

Proof: Note: $\|x - y\|_\infty = \max_i |x_i - y_i|$

1. $\max |x - y| = 0 \text{ iff } x_i = y_i \quad \forall i$

Max Norm

Lemma 1: $\left(\mathbb{R}^{|S|}, \|\cdot\|_\infty\right)$ is a metric space.

Definition: Let M be a set. A *metric* on M is a function $d : M \times M \to [0, \infty)$ which satisfies the following three conditions for all $x, y, z \in M$:

1. $d(x, y) = 0$ if and only if $x = y$
2. $d(x, y) = d(y, x)$
3. $d(x, y) \leq d(x, z) + d(z, y)$

Proof: Note: $\|x - y\|_\infty = \max_i |x_i - y_i|$

1. $\max |x - y| = 0 \text{ iff } x_i = y_i \quad \forall i$

2. $|x - y| = |-(x - y)| = |y - x|$

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

Max Norm

Lemma 1: $\left(\mathbb{R}^{|S|}, \|\cdot\|_\infty\right)$ is a metric space.

Definition: Let M be a set. A *metric* on M is a function $d : M \times M \to [0, \infty)$ which satisfies the following three conditions for all $x, y, z \in M$:

1. $d(x, y) = 0$ if and only if $x = y$
2. $d(x, y) = d(y, x)$
3. $d(x, y) \leq d(x, z) + d(z, y)$

Proof: Note: $\|x - y\|_\infty = \max_i |x_i - y_i|$

1. $\max |x - y| = 0 \text{ iff } x_i = y_i \quad \forall i$

2. $|x - y| = |-(x - y)| = |y - x|$

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

3. $\max |x - z| = \max |x - y + y - z|$

Assumption: $|S| < \infty$

Max Norm

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

$\mathbb{R}^{|S|}$ = space of 3D vectors

Lemma 1: $(\mathbb{R}^{|S|}, \|\cdot\|_\infty)$ is a metric space.

Definition: Let M be a set. A *metric* on M is a function $d : M \times M \to [0, \infty)$ which satisfies the following three conditions for all $x, y, z \in M$:

1. $d(x, y) = 0$ if and only if $x = y$
2. $d(x, y) = d(y, x)$
3. $d(x, y) \leq d(x, z) + d(z, y)$

Proof: Note: $\|x - y\|_\infty = \max_i |x_i - y_i|$

1. $\max |x - y| = 0 \text{ iff } x_i = y_i \quad \forall i$

2. $|x - y| = |-(x - y)| = |y - x|$

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

3. $\max |x - z| = \max |x - y + y - z|$

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

Max Norm

<u>Lemma 1</u>: $\left(\mathbb{R}^{|S|}, \|\cdot\|_\infty\right)$ is a metric space.

<u>Definition</u>: Let M be a set. A *metric* on M is a function $d : M \times M \to [0, \infty)$ which satisfies the following three conditions for all $x, y, z \in M$:

1. $d(x,y) = 0$ if and only if $x = y$
2. $d(x,y) = d(y,x)$
3. $d(x,y) \leq d(x,z) + d(z,y)$

Proof: Note: $\|x-y\|_\infty = \max_i |x_i - y_i|$

1. $\max|x-y| = 0 \text{ iff } x_i = y_i \quad \forall i$

2. $|x-y| = |-(x-y)| = |y-x|$
   $\therefore \quad \max|x-y| = max|y-x|$

3. $\max|x-z| = \max|x-y+y-z|$
   $\leq \max(|x-y| + |y-z|)$
   $\leq \max|x-y| + \max|y-z|$

Bellman Operator Contraction

Bellman Operator Contraction

Lemma 2: B is a γ contraction mapping on $(\mathbb{R}^{|S|}, \|\cdot\|_\infty)$.

Bellman Operator Contraction

Lemma 2: B is a γ contraction mapping on $(\mathbb{R}^{|S|}, \|\cdot\|_\infty)$.

Proof

Bellman Operator Contraction

Lemma 2: B is a γ contraction mapping on $(\mathbb{R}^{|S|}, \|\cdot\|_\infty)$.

Proof

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

Bellman Operator Contraction

<u>Lemma 2</u>: B is a γ contraction mapping on $(\mathbb{R}^{|S|}, \|\cdot\|_\infty)$.

Proof

$$\begin{aligned}\|B[V_1] - B[V_2]\|_\infty &= \max_{s\in S} |B[V_1](s) - B[V_2](s)| \\ &= \max_{s\in S} \left| \max_{a\in A} \left(R(s,a) + \gamma \sum_{s'\in S} T(s'|s,a) V_1(s') \right) - \max_{a\in A} \left(R(s,a) + \gamma \sum_{s'\in S} T(s'|s,a) V_2(s') \right) \right|\end{aligned}$$

Bellman Operator Contraction

<u>Lemma 2</u>: B is a γ contraction mapping on $(\mathbb{R}^{|S|}, \|\cdot\|_\infty)$.

Proof

$$\begin{aligned}
\|B[V_1] - B[V_2]\|_\infty &= \max_{s\in S} |B[V_1](s) - B[V_2](s)| \\
&= \max_{s\in S} \left| \max_{a\in A} \left(R(s,a) + \gamma \sum_{s'\in S} T(s'|s,a)V_1(s') \right) - \max_{a\in A} \left(R(s,a) + \gamma \sum_{s'\in S} T(s'|s,a)V_2(s') \right) \right| \\
&\leq \max_{s\in S} \left| \max_{a\in A} \left(R(s,a) + \gamma \sum_{s'\in S} T(s'|s,a)V_1(s') - R(s,a) - \gamma \sum_{s'\in S} T(s'|s,a)V_2(s') \right) \right|
\end{aligned}$$

Bellman Operator Contraction

$$\begin{bmatrix} 3 \\ -4 \end{bmatrix}$$

Lemma 2: B is a γ contraction mapping on $(\mathbb{R}^{|S|}, \|\cdot\|_\infty)$.

Proof

$$\begin{aligned}
\|B[V_1] - B[V_2]\|_\infty &= \max_{s \in S} |B[V_1](s) - B[V_2](s)| \\
&= \max_{s \in S} \left| \max_{a \in A} \left(R(s,a) + \gamma \sum_{s' \in S} T(s'|s,a) V_1(s') \right) - \max_{a \in A} \left(R(s,a) + \gamma \sum_{s' \in S} T(s'|s,a) V_2(s') \right) \right| \\
&\leq \max_{s \in S} \left| \max_{a \in A} \left(R(s,a) + \gamma \sum_{s' \in S} T(s'|s,a) V_1(s') - R(s,a) - \gamma \sum_{s' \in S} T(s'|s,a) V_2(s') \right) \right|
\end{aligned}$$

$$|\max(x)| \leq \max |x|$$

Bellman Operator Contraction

Lemma 2: B is a γ contraction mapping on $(\mathbb{R}^{|S|}, \|\cdot\|_\infty)$.

Proof

$$
\begin{aligned}
\|B[V_1] - B[V_2]\|_\infty &= \max_{s \in S} |B[V_1](s) - B[V_2](s)| \\
&= \max_{s \in S} \left| \max_{a \in A} \left(R(s,a) + \gamma \sum_{s' \in S} T(s'|s,a) V_1(s') \right) - \max_{a \in A} \left(R(s,a) + \gamma \sum_{s' \in S} T(s'|s,a) V_2(s') \right) \right| \\
&\le \max_{s \in S} \left| \max_{a \in A} \left(\cancel{R(s,a)} + \gamma \sum_{s' \in S} T(s'|s,a) V_1(s') \cancel{- R(s,a)} - \gamma \sum_{s' \in S} T(s'|s,a) V_2(s') \right) \right| \\
&\le \max_{s \in S, a \in A} \left| \gamma \sum_{s' \in S} T(s'|s,a) \left(V_1(s') - V_2(s') \right) \right|
\end{aligned}
$$

$|\max(x)| \le \max |x|$

Bellman Operator Contraction

<u>Lemma 2</u>: B is a γ contraction mapping on $(\mathbb{R}^{|S|}, \|\cdot\|_\infty)$.

Proof

$$\begin{aligned}
\|B[V_1] - B[V_2]\|_\infty &= \max_{s\in S} |B[V_1](s) - B[V_2](s)| \\
&= \max_{s\in S} \left| \max_{a\in A} \left(R(s,a) + \gamma \sum_{s'\in S} T(s'|s,a) V_1(s') \right) - \max_{a\in A} \left(R(s,a) + \gamma \sum_{s'\in S} T(s'|s,a) V_2(s') \right) \right| \\
&\leq \max_{s\in S} \left| \max_{a\in A} \left(R(s,a) + \gamma \sum_{s'\in S} T(s'|s,a) V_1(s') - R(s,a) - \gamma \sum_{s'\in S} T(s'|s,a) V_2(s') \right) \right| \\
&\leq \max_{s\in S, a\in A} \left| \gamma \sum_{s'\in S} T(s'|s,a) \left(V_1(s') - V_2(s') \right) \right| \qquad |\max(x)| \leq \max |x| \\
&\leq \max_{s\in S, a\in A} \gamma \sum_{s'\in S} T(s'|s,a) \left| V_1(s') - V_2(s') \right|
\end{aligned}$$

Bellman Operator Contraction

Lemma 2: B is a γ contraction mapping on $(\mathbb{R}^{|S|}, \|\cdot\|_\infty)$.

Proof

$$
\begin{aligned}
\|B[V_1] - B[V_2]\|_\infty &= \max_{s \in S} |B[V_1](s) - B[V_2](s)| \\
&= \max_{s \in S} \left| \max_{a \in A} \left(R(s,a) + \gamma \sum_{s' \in S} T(s'|s,a) V_1(s') \right) - \max_{a \in A} \left(R(s,a) + \gamma \sum_{s' \in S} T(s'|s,a) V_2(s') \right) \right| \\
&\leq \max_{s \in S} \left| \max_{a \in A} \left(R(s,a) + \gamma \sum_{s' \in S} T(s'|s,a) V_1(s') - R(s,a) - \gamma \sum_{s' \in S} T(s'|s,a) V_2(s') \right) \right| \\
&\leq \max_{s \in S, a \in A} \left| \gamma \sum_{s' \in S} T(s'|s,a) \left(V_1(s') - V_2(s') \right) \right| \qquad |\max(x)| \leq \max |x| \\
&\leq \max_{s \in S, a \in A} \gamma \sum_{s' \in S} T(s'|s,a) \left| V_1(s') - V_2(s') \right| \\
&\leq \max_{s \in S, a \in A} \gamma \sum_{s' \in S} T(s'|s,a) \|V_1 - V_2\|_\infty
\end{aligned}
$$

Bellman Operator Contraction

Lemma 2: B is a γ contraction mapping on $(\mathbb{R}^{|S|}, \|\cdot\|_\infty)$.

Proof

$$
\begin{aligned}
\|B[V_1] - B[V_2]\|_\infty &= \max_{s\in S} |B[V_1](s) - B[V_2](s)| \\
&= \max_{s\in S} \left| \max_{a\in A} \left(R(s,a) + \gamma \sum_{s'\in S} T(s'|s,a) V_1(s') \right) - \max_{a\in A} \left(R(s,a) + \gamma \sum_{s'\in S} T(s'|s,a) V_2(s') \right) \right| \\
&\le \max_{s\in S} \left| \max_{a\in A} \left(R(s,a) + \gamma \sum_{s'\in S} T(s'|s,a) V_1(s') - R(s,a) - \gamma \sum_{s'\in S} T(s'|s,a) V_2(s') \right) \right| \\
&\le \max_{s\in S, a\in A} \left| \gamma \sum_{s'\in S} T(s'|s,a) \left(V_1(s') - V_2(s') \right) \right| \qquad |\max(x)| \le \max |x| \\
&\le \max_{s\in S, a\in A} \gamma \sum_{s'\in S} T(s'|s,a) \left| V_1(s') - V_2(s') \right| \\
&\le \max_{s\in S, a\in A} \gamma \sum_{s'\in S} T(s'|s,a) \underline{\|V_1 - V_2\|_\infty} \\
&= \gamma \|V_1 - V_2\|_\infty \max_{s\in S, a\in A} \sum_{s'\in S} T(s'|s,a)
\end{aligned}
$$

Bellman Operator Contraction

Lemma 2: B is a γ contraction mapping on $(\mathbb{R}^{|S|}, \|\cdot\|_\infty)$.

Proof

$$
\begin{aligned}
\|B[V_1] - B[V_2]\|_\infty &= \max_{s\in S} |B[V_1](s) - B[V_2](s)| \\
&= \max_{s\in S} \left| \max_{a\in A} \left(R(s,a) + \gamma \sum_{s'\in S} T(s'|s,a) V_1(s') \right) - \max_{a\in A} \left(R(s,a) + \gamma \sum_{s'\in S} T(s'|s,a) V_2(s') \right) \right| \\
&\le \max_{s\in S} \left| \max_{a\in A} \left(R(s,a) + \gamma \sum_{s'\in S} T(s'|s,a) V_1(s') - R(s,a) - \gamma \sum_{s'\in S} T(s'|s,a) V_2(s') \right) \right| \\
&\le \max_{s\in S, a\in A} \left| \gamma \sum_{s'\in S} T(s'|s,a) \left(V_1(s') - V_2(s') \right) \right| \\
&\le \max_{s\in S, a\in A} \gamma \sum_{s'\in S} T(s'|s,a) |V_1(s') - V_2(s')| \\
&\le \max_{s\in S, a\in A} \gamma \sum_{s'\in S} T(s'|s,a) \|V_1 - V_2\|_\infty \\
&= \gamma \|V_1 - V_2\|_\infty \cancel{\max_{s\in S, a\in A} \sum_{s'\in S} T(s'|s,a)} \\
&= \gamma \|V_1 - V_2\|_\infty
\end{aligned}
$$

$|\max(x)| \le \max |x|$

Value Iteration Convergence

Value Iteration Convergence

Theorem 1: Let $\{V_1, \ldots, V_\infty\}$ be a sequence of value functions for a discrete MDP generated by the recurrence $V_{k+1} = B[V_k]$. If $\gamma < 1$, then $\lim_{k \to \infty} V_k = V^*$.

Value Iteration Convergence

Theorem 1: Let $\{V_1, \ldots, V_\infty\}$ be a sequence of value functions for a discrete MDP generated by the recurrence $V_{k+1} = B[V_k]$. If $\gamma < 1$, then $\lim_{k \to \infty} V_k = V^*$.

Lemma 2: B is a γ contraction mapping on $(\mathbb{R}^{|S|}, \|\cdot\|_\infty)$.

Value Iteration Convergence

Theorem 1: Let $\{V_1, \ldots, V_\infty\}$ be a sequence of value functions for a discrete MDP generated by the recurrence $V_{k+1} = B[V_k]$. If $\gamma < 1$, then $\lim_{k\to\infty} V_k = V^*$.

Lemma 2: B is a γ contraction mapping on $(\mathbb{R}^{|S|}, \|\cdot\|_\infty)$.

Theorem (Banach): If f is a contraction mapping on metric space (M, d), then

1. f has a single, unique fixed point x^*.
2. If $\{x_k\}$ is a sequence defined by $x_{k+1} = f(x_k)$, then $\lim_{k\to\infty} x_k = x^*$.

Value Iteration Convergence

Theorem 1: Let $\{V_1, \ldots, V_\infty\}$ be a sequence of value functions for a discrete MDP generated by the recurrence $V_{k+1} = B[V_k]$. If $\gamma < 1$, then $\lim_{k\to\infty} V_k = V^*$.

Proof:

Lemma 2: B is a γ contraction mapping on $(\mathbb{R}^{|S|}, \|\cdot\|_\infty)$.

Theorem (Banach): If f is a contraction mapping on metric space (M, d), then

1. f has a single, unique fixed point x^*.
2. If $\{x_k\}$ is a sequence defined by $x_{k+1} = f(x_k)$, then $\lim_{k\to\infty} x_k = x^*$.

Value Iteration Convergence

Theorem 1: Let $\{V_1, \ldots, V_\infty\}$ be a sequence of value functions for a discrete MDP generated by the recurrence $V_{k+1} = B[V_k]$. If $\gamma < 1$, then $\lim_{k\to\infty} V_k = V^*$.

Proof:

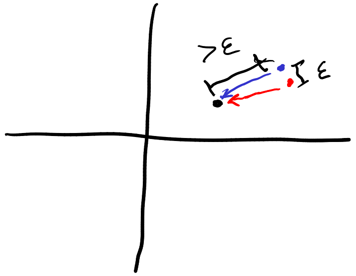
Lemma 2: B is a γ contraction mapping on $(\mathbb{R}^{|S|}, \|\cdot\|_\infty)$.

Theorem (Banach): If f is a contraction mapping on metric space (M, d), then

1. f has a single, unique fixed point x^*.
2. If $\{x_k\}$ is a sequence defined by $x_{k+1} = f(x_k)$, then $\lim_{k\to\infty} x_k = x^*$.

By Lemma 2 and Banach's theorem (part 2), repeated application of the Bellman operator always has a fixed point limit, $\hat{V}$.

Value Iteration Convergence



<u>Theorem 1</u>: Let $\{V_1, \ldots, V_\infty\}$ be a sequence of value functions for a discrete MDP generated by the recurrence $V_{k+1} = B[V_k]$. If $\gamma < 1$, then $\lim_{k \to \infty} V_k = V^*$.

<u>Lemma 2</u>: B is a γ contraction mapping on $(\mathbb{R}^{|S|}, \|\cdot\|_\infty)$.

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

1. f has a single, unique fixed point x^*.
2. If $\{x_k\}$ is a sequence defined by $x_{k+1} = f(x_k)$, then $\lim_{k \to \infty} x_k = x^*$.

Proof:

By Lemma 2 and Banach's theorem (part 2), repeated application of the Bellman operator always has a fixed point limit, $\hat{V}$.

By Banach's theorem (part 1), $\hat{V} = B[\hat{V}]$. Since $\hat{V}$ satisfies Bellman's equation, it is optimal and $\hat{V} = V^*$.

Does Policy Iteration Converge?

Does Policy Iteration Converge?

Theorem: Policy iteration converges to an optimal policy for a finite MDP in finite time.

Does Policy Iteration Converge?

Theorem: Policy iteration converges to an optimal policy for a finite MDP in finite time.

Proof (sketch):

Does Policy Iteration Converge?

Theorem: Policy iteration converges to an optimal policy for a finite MDP in finite time.

Proof (sketch):

1. The policy will either improve or stay the same at each iteration

Does Policy Iteration Converge?

Theorem: Policy iteration converges to an optimal policy for a finite MDP in finite time.

Proof (sketch):

1. The policy will either improve or stay the same at each iteration
2. The policy will stay the same if and only if $V^{\pi} = V^{*}$

Does Policy Iteration Converge?

Theorem: Policy iteration converges to an optimal policy for a finite MDP in finite time.

Proof (sketch):

1. The policy will either improve or stay the same at each iteration
2. The policy will stay the same if and only if $V^{\pi} = V^{*}$
3. There are a finite number of possible policies

Does Policy Iteration Converge?

Theorem: Policy iteration converges to an optimal policy for a finite MDP in finite time.

Proof (sketch):

1. The policy will either improve or stay the same at each iteration
2. The policy will stay the same if and only if $V^{\pi} = V^{*}$
3. There are a finite number of possible policies
4. By (1), (2), and (3), the policy will improve until it finds the optimal policy, and it will always find the optimal policy.

Guiding Questions

Guiding Questions

- Does value iteration always converge?
- Is the value function unique?

$S = \{1, 0.8 \ldots 0\}$

Break

Conservation MDP

$S = \{0, 1\}$ ← extinct (0)

$A = \{p, n\}$

$R(s,a) = R(s) + R(a)$

$R(1) = 1$

$R(0) = -1$

$R(p) = -1$

$R(n) = 1$

$$T^p = \begin{array}{c|cc} & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0.1 & 0.9 \end{array}$$

$$T^n = \begin{bmatrix} 1 & 0 \\ 0.5 & 0.5 \end{bmatrix}$$

$\gamma = 0.9$

Find V^*, π^*

$V^*(0) = \max_a \left(R(s,a) + \gamma E[V^*(s')]\right)$ ← Bellman's Optimality Equation

$\max_a \left(R(s,a) + 0.9\, V^*(0)\right)$ (a → n)

$V^*(0) = 0 + 0.9\, V^*(0)$

$(1 - 0.9) V^*(0) = 0$

$V^*(0) = 0$

$\pi^*(0) = n$

Two options: $\pi_p(1) = p$ $\quad \pi_n(1) = n$

$V^{\pi_p}(1) = R(1,p) + \gamma\left(0.1\, \cancel{V^{\pi_p}(0)} + 0.9\, V^{\pi_p}(1)\right)$ ← Policy Eval. Bellman's Exp. Eqn.

$0 + \cancel{\gamma}\left(0.9\, V^{\pi_p}(1)\right)$ ($\gamma \to 0.9$)

$(1 - 0.81) V^{\pi_p}(1) = 0$

$V^{\pi_p}(0) = 0$

$V^{\pi_n}(1) = R(1,n) + \gamma\left(\cancel{0.5\, V^{\pi_n}(0)} + 0.5\, V^{\pi_n}(1)\right)$ ($\to 0$)

$2 + \gamma\left(0.5\, V^{\pi_n}(1)\right)$

$(1 - 0.45) V^{\pi_n}(1) = 2$

$V^{\pi_n}(1) = 3.64 = V^*(1)$

$\pi^*(1) = n$