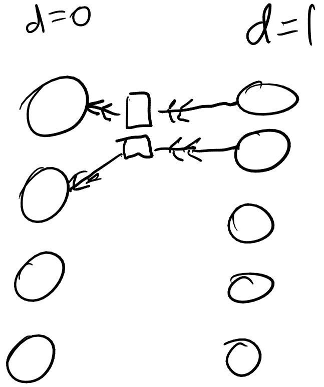
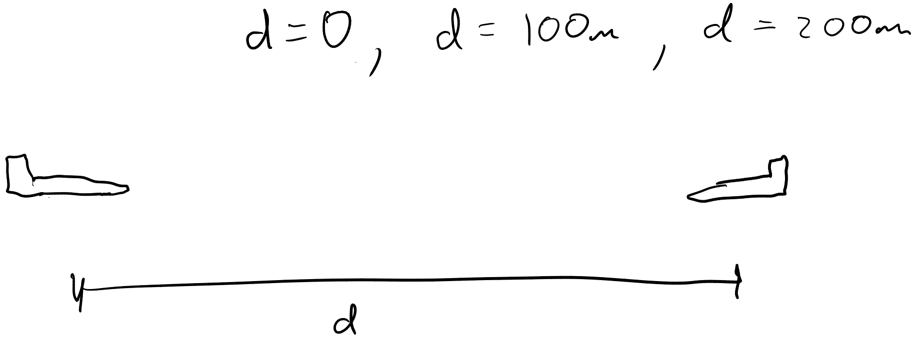




Value Iteration Convergence

Review

$Q^{\pi}(s,a)$

$V^{\pi}(s)$ $U^{\pi}(s)$

Review

Value function

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

Review

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

Offline   Value Iteration
Policy Iteration

Online   MCTS $\leftarrow$ insensitive $|S|$
FSSS

Guiding Questions

Guiding Questions

- Does value iteration always converge?
- Is the optimal value function unique?
- Can there be multiple optimal policies?
- Is there always a deterministic optimal policy?

Value Iteration: The Bellman Operator

Value Iteration: The Bellman Operator

Algorithm: Value Iteration

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

$\quad V \leftarrow V'$

$\quad V' \leftarrow B[V]$

return V'

Value Iteration: The Bellman Operator

Algorithm: Value Iteration

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

$V \leftarrow V'$

$V' \leftarrow B[V]$

return V'

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

Value Iteration Convergence

Value Iteration Convergence

Theorem 1: Let $\{V_1, \dots, V_\infty\}$ 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

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

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 $x, y, z \in M$:

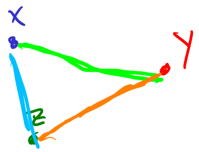
1. $d(x, y) = 0$ if and only if $x = y$
2. $d(x, y) = d(y, x)$

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. $d(x, y) \leq d(x, z) + d(z, y)$

$M = \mathbb{R}^2$



$d(x,y) = \|x - y\|$

Contraction Mappings

Contraction Mappings

Definition: 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 $f(x^*) = x^*$.

$$f(x) = \begin{bmatrix} \frac{x_2}{2} + 1 \\ \frac{x_1}{2} + \frac{1}{2} \end{bmatrix}$$

Script: contraction_mapping.jl

Banach's Theorem

Banach's Theorem

If we can prove
$d(f(x), f(y)) \leq \alpha \cdot d(x,y)$
then
convergence to fixed point

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 $\lim_{k\to\infty} x_k = x^*$.

B

Max Norm

Max Norm

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

$V \in \mathbb{R}^{|S|}$ ← size of state space

Max Norm

$\|x\|_\infty = \max |x_i|$

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. $d(x, y) \leq d(x, z) + d(z, y)$

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:

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: $\|x - y\|_\infty = \mathbf{max}_i \, |x_i - y_i|$

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

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

$\therefore \quad \max_i |x_i - y_i| = max|y - x|$

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. $d(x, y) \leq d(x, z) + d(z, y)$

x, z

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

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

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

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

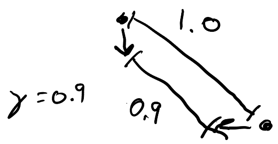
3. $\max|x - z| = \max|x - y + y - z|$

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

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

Bellman Operator Contraction

Bellman Operator Contraction



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

Bellman Operator Contraction

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

Proof

Bellman Operator Contraction

<u>Lemma 2</u>: 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

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

Proof

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

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

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

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

Proof

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

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

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

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

Proof

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

Bellman Operator Contraction

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

Proof

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

Bellman Operator Contraction

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

Proof

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

Bellman Operator Contraction

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

Proof

$$
\begin{aligned}
\|B[V_1] - B[V_2]\|_\infty &= \max_{s\in S} |B[V_1](s) - B[V_2](s)| \\
&= \max_{s\in S} \left| \max_{a\in A} \left(R(s,a) + \gamma \sum_{s'\in S} T(s'|s,a) V_1(s') \right) - \max_{a\in A} \left(R(s,a) + \gamma \sum_{s'\in S} T(s'|s,a) V_2(s') \right) \right| \\
&\leq \max_{s\in S} \left| \max_{a\in A} \left(R(s,a) + \gamma \sum_{s'\in S} T(s'|s,a) V_1(s') - R(s,a) - \gamma \sum_{s'\in S} T(s'|s,a) V_2(s') \right) \right| \\
&\leq \max_{s\in S, a\in A} \left| \gamma \sum_{s'\in S} T(s'|s,a) \left(V_1(s') - V_2(s') \right) \right| \qquad |\max(x)| \leq \max |x| \\
&\leq \max_{s\in S, a\in A} \gamma \sum_{s'\in S} T(s'|s,a) \left| V_1(s') - V_2(s') \right| \\
&\leq \max_{s\in S, a\in A} \gamma \sum_{s'\in S} T(s'|s,a) \|V_1 - V_2\|_\infty \\
&= \gamma \|V_1 - V_2\|_\infty \max_{s\in S, a\in A} \sum_{s'\in S} T(s'|s,a) \quad = 1 \\
&= \gamma \|V_1 - V_2\|_\infty
\end{aligned}
$$

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

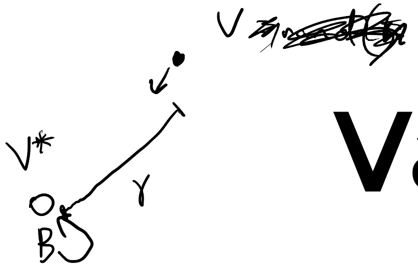
Lemma 2: B is a γ contraction mapping on $(\mathbb{R}^{|S|}, \|\cdot\|_\infty)$.

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 $\lim_{k\to\infty} x_k = x^*$.

Proof:

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

$\|V_k - V^*\| < (\bar{V} - \underline{V})\gamma^k$

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, $\hat{V}$.

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.

Is there always a deterministic optimal policy?

<u>Thm</u>

For any finite MDP, there is always at least 1 deterministic optimal policy

Suppose $\pi^*(a|s)$ is a nondeterministic optimal policy. Then

$$V^*(s) = \sum_{a \in \text{support}(\pi^*(a|s))} \pi^*(a|s) \underbrace{\left(R(s,a) + \gamma \sum_{s' \in S} T(s'|s,a) V^*(s') \right)}_{Q^*(s,a)}$$

Lemma 3: If $a' \in \text{support}(\pi^*(a|s))$ then $Q^*(s,a') = V^*(s)$

Proof: (by contradiction) w.l.o.g. suppose that $Q^*(s,a^1) > Q^*(s,a^2)$ and $\text{support}(\pi(a|s)) = \{a^1, a^2\}$

Let π' be same as π^* except that $\pi'(s) = a^1$

$$V^{\pi'}(s) = Q^*(s,a^1) > \pi^*(a^1|s)\, Q^*(s,a^1) + \pi^*(a^2|s)\, Q^*(s,a^2) = V^*(s)$$

Let π'' be a deterministic policy with $\pi''(s) \in \text{support}(\pi^*(a|s))$.

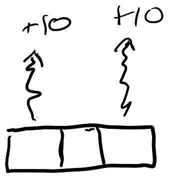
By Lemma 3, $Q^*(s,\pi''(s)) = V^*(s)$ at every state so π'' is optimal.

Guiding Questions

Guiding Questions

B is a contraction mapping

- Does value iteration always converge? Yes
- Is the value function unique? Yes
- Can there be multiple optimal policies? Yes
- Is there always a deterministic optimal policy? Yes



Break

Conservation MDP