# Value Iteration Convergence

# Review

# Review

- 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**?

# Guiding Questions

# Guiding Questions

- Does value iteration always converge?
- Is the 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 $\|V - V'\|_\infty > \epsilon$

$V \leftarrow V'$

$V' \leftarrow B[V]$

return $V'$

# Value Iteration: The Bellman Operator

<u>Algorithm: Value Iteration</u>

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

$\quad V \leftarrow V'$

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

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 $\{V_1, \ldots, V_\infty\}$ 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

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 $x = y$

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

# 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 $\alpha$, $0 \leq \alpha \leq 1$ and all $x$ and $y$ in $M$.

# 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 $\alpha$, $0 \leq \alpha \leq 1$ and all $x$ and $y$ in $M$.

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

# 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 $\alpha$, $0 \leq \alpha \leq 1$ and all $x$ and $y$ in $M$.

Definition: $x^*$ 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

<u>Lemma 1</u>: $\left(\mathbb{R}^{|S|}, \|\cdot\|_\infty\right)$ 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: $(\mathbb{R}^{|S|}, \|\cdot\|_\infty)$ 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: $(\mathbb{R}^{|S|}, \|\cdot\|_\infty)$ 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: $\|x - y\|_\infty = \max_i |x_i - y_i|$

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

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

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

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

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

$\therefore \quad \mathbf{max} \, |x - y| = max|y - x|$

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

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

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

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

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

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

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

# Bellman Operator Contraction

# Bellman Operator Contraction

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

# Bellman Operator Contraction

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

Proof

# Bellman Operator Contraction

Lemma 2: $B$ is a $\gamma$ 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 $\gamma$ 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

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

Proof

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

# Bellman Operator Contraction

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

Proof

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

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

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

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

# Bellman Operator Contraction

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

Proof

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

# Bellman Operator Contraction

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

Proof

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

# Bellman Operator Contraction

<u>Lemma 2</u>: $B$ is a $\gamma$ 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

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

Proof

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

# 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 $\{V_1, \ldots, V_\infty\}$ 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 $\gamma$ 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 $\gamma$ 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 $\gamma$ 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 $\gamma$ 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

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 $\gamma$ 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.

# Properties of optimal MDP solutions

- Does every MDP have a unique optimal value function, $V^*$?
- Does every MDP have a unique optimal policy, $\pi^*$?
- Does every MDP have a *deterministic* optimal policy?

# Properties of optimal MDP solutions

- Does every MDP have a unique optimal value function, $V^*$?
- Does every MDP have a unique optimal policy, $\pi^*$?
- Does every MDP have a *deterministic* optimal policy?

Justification

- Suppose that $\tilde{\pi}$ is optimal and, for some $s$, $\tilde{\pi}(a^1 \mid s) > 0$, $\tilde{\pi}(a^2 \mid s) > 0$, and $\tilde{\pi}(a^1 \mid s) + \tilde{\pi}(a^2 \mid s) = 1$.
- Then $Q^*(s, a^1) = Q^*(s, a^2) = V^*(s)$. If this were not true, then $\tilde{\pi}$ would not be optimal.
- As a consequence, a deterministic policy $\tilde{\pi}'$ with $\tilde{\pi}'(s) = a^1$ is also optimal!

# Guiding Questions

# Guiding Questions

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

# Break

Conservation MDP