CS 747, Autumn 2022: Lecture 8

Shivaram Kalyanakrishnan

Department of Computer Science and Engineering
Indian Institute of Technology Bombay

Autumn 2022

Markov Decision Problems

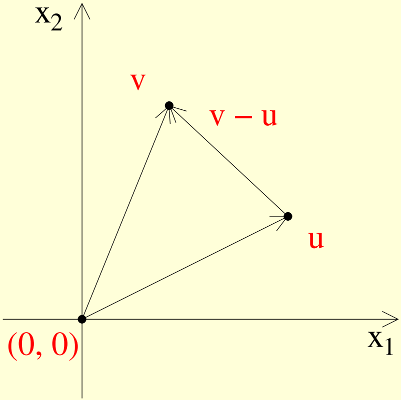
1. Banach's fixed-point theorem
2. Bellman optimality operator
3. Value iteration

Markov Decision Problems

1. Banach's fixed-point theorem
2. Bellman optimality operator
3. Value iteration

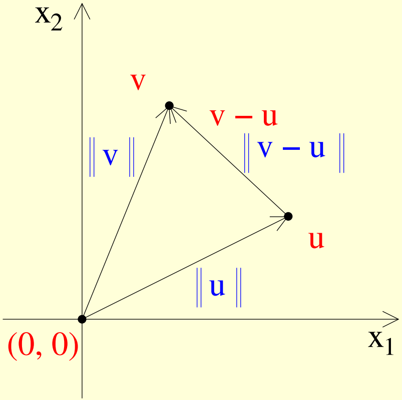
Complete, Normed Vector Spaces

- A vector space X has objects called vectors that can be added and scaled.



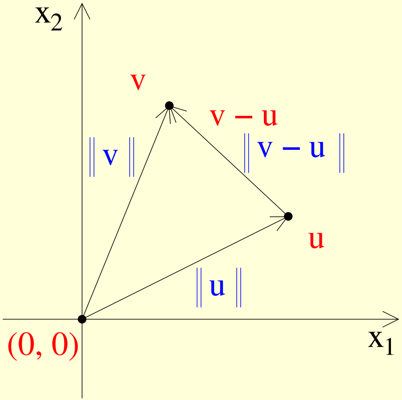
Complete, Normed Vector Spaces

- A vector space X has objects called vectors that can be added and scaled.
- A norm $\|\cdot\|$ associates a length with each vector (and satisfies some conditions).



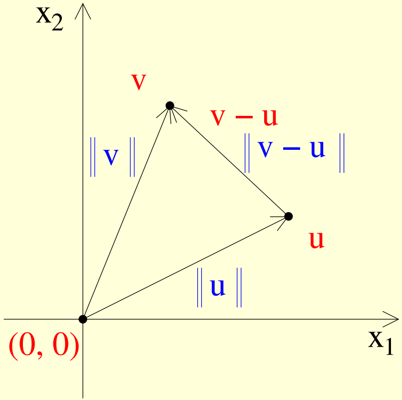
Complete, Normed Vector Spaces

- A vector space X has objects called vectors that can be added and scaled.
- A norm $\|\cdot\|$ associates a length with each vector (and satisfies some conditions).
- A complete, normed vector space $(X, \|\cdot\|)$ is one in which every Cauchy sequence has a limit in X.



Complete, Normed Vector Spaces

- A vector space X has objects called vectors that can be added and scaled.
- A norm $\|\cdot\|$ associates a length with each vector (and satisfies some conditions).
- A complete, normed vector space $(X, \|\cdot\|)$ is one in which every Cauchy sequence has a limit in X.



- A complete, normed vector space is called a Banach space.

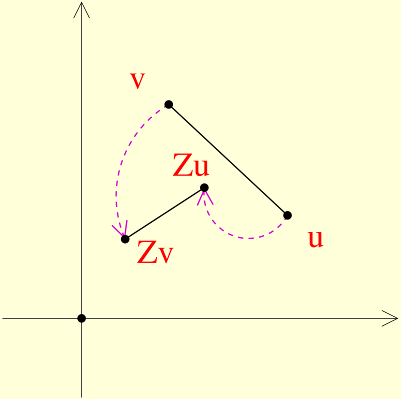
Two Definitions

- Let $(X, \|\cdot\|)$ be a normed vector space, and let $0 \leq \ell < 1$.

Two Definitions

- Let $(X, \|\cdot\|)$ be a normed vector space, and let $0 \leq \ell < 1$.

- **Contraction mapping.** A mapping $Z : X \to X$ is called a contraction mapping with contraction factor ℓ if $\forall u, v \in X$,
$$\|Zv - Zu\| \leq \ell \|v - u\|.$$

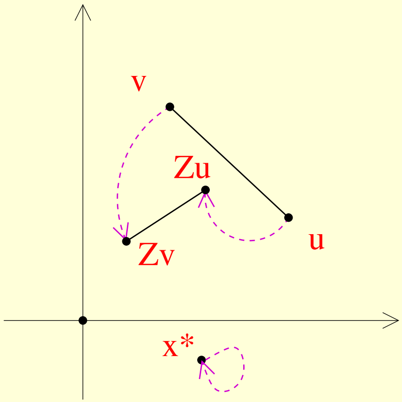


Two Definitions

- Let $(X, \|\cdot\|)$ be a normed vector space, and let $0 \leq \ell < 1$.
- **Contraction mapping.** A mapping $Z : X \to X$ is called a contraction mapping with contraction factor ℓ if $\forall u, v \in X$,

$$\|Zv - Zu\| \leq \ell \|v - u\|.$$

- **Fixed-point.** $x^\star \in X$ is called a fixed-point of Z if $Zx^\star = x^\star$.



Banach's Fixed-point Theorem

(Adapted from Szepesvári, 2009 (see Appendix A.1).)

Let $(X, \|\cdot\|)$ be a Banach space, and let $Z : X \to X$ be a contraction mapping with contraction factor $\ell \in [0, 1)$. Then:

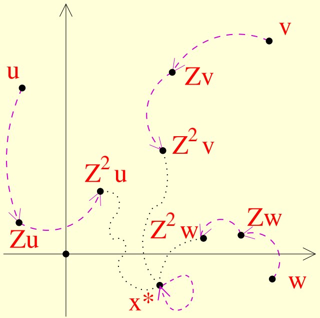
1. Z has a unique fixed point $x^\star \in X$.
2. For $x \in X, m \geq 0$: $\|Z^m x - x^\star\| \leq \ell^m \|x - x^\star\|$.

Banach's Fixed-point Theorem

(Adapted from Szepesvári, 2009 (see Appendix A.1).)

Let $(X, \|\cdot\|)$ be a Banach space, and let $Z : X \to X$ be a contraction mapping with contraction factor $\ell \in [0, 1)$. Then:

1. Z has a unique fixed point $x^\star \in X$.
2. For $x \in X, m \geq 0$: $\|Z^m x - x^\star\| \leq \ell^m \|x - x^\star\|$.



Markov Decision Problems

1. Banach's fixed-point theorem
2. Bellman optimality operator
3. Value iteration

Bellman Optimality Operator

- Take $S = \{s_1, s_2, \ldots, s_n\}$. A function $F : S \to \mathbb{R}$ is equivalently a point in $\mathbb{R}^n$.

Bellman Optimality Operator

- Take $S = \{s_1, s_2, \ldots, s_n\}$. A function $F : S \to \mathbb{R}$ is equivalently a point in $\mathbb{R}^n$.
- The Bellman optimality operator $B^\star : \mathbb{R}^n \to \mathbb{R}^n$ for MDP (S, A, T, R, γ) is defined as follows. For $F \in \mathbb{R}^n$, $s \in S$:

$$(B^\star(F))(s) \stackrel{\text{def}}{=} \max_{a \in A} \sum_{s' \in S} T(s, a, s')\{R(s, a, s') + \gamma F(s')\}.$$

Bellman Optimality Operator

- Take $S = \{s_1, s_2, \ldots, s_n\}$. A function $F : S \to \mathbb{R}$ is equivalently a point in $\mathbb{R}^n$.
- The Bellman optimality operator $B^\star : \mathbb{R}^n \to \mathbb{R}^n$ for MDP (S, A, T, R, γ) is defined as follows. For $F \in \mathbb{R}^n$, $s \in S$:

$$(B^\star(F))(s) \stackrel{\text{def}}{=} \max_{a \in A} \sum_{s' \in S} T(s, a, s')\{R(s, a, s') + \gamma F(s')\}.$$

- Recall that the max norm $\|\cdot\|_\infty$ of $F = (f_1, f_2, \ldots, f_n) \in \mathbb{R}^n$ is

$$\|F\|_\infty = \max\{|f_1|, |f_2|, \ldots, |f_n|\}.$$

Bellman Optimality Operator

- Take $S = \{s_1, s_2, \dots, s_n\}$. A function $F : S \to \mathbb{R}$ is equivalently a point in $\mathbb{R}^n$.
- The Bellman optimality operator $B^\star : \mathbb{R}^n \to \mathbb{R}^n$ for MDP (S, A, T, R, γ) is defined as follows. For $F \in \mathbb{R}^n$, $s \in S$:

$$(B^\star(F))(s) \stackrel{\text{def}}{=} \max_{a \in A} \sum_{s' \in S} T(s, a, s')\{R(s, a, s') + \gamma F(s')\}.$$

- Recall that the max norm $\|\cdot\|_\infty$ of $F = (f_1, f_2, \dots, f_n) \in \mathbb{R}^n$ is

$$\|F\|_\infty = \max\{|f_1|, |f_2|, \dots, |f_n|\}.$$

- It is an established result that $(\mathbb{R}^n, \|\cdot\|_\infty)$ is a Banach space.

Bellman Optimality Operator

- Take $S = \{s_1, s_2, \ldots, s_n\}$. A function $F : S \to \mathbb{R}$ is equivalently a point in $\mathbb{R}^n$.
- The Bellman optimality operator $B^\star : \mathbb{R}^n \to \mathbb{R}^n$ for MDP (S, A, T, R, γ) is defined as follows. For $F \in \mathbb{R}^n$, $s \in S$:

$$(B^\star(F))(s) \stackrel{\text{def}}{=} \max_{a \in A} \sum_{s' \in S} T(s, a, s')\{R(s, a, s') + \gamma F(s')\}.$$

- Recall that the max norm $\|\cdot\|_\infty$ of $F = (f_1, f_2, \ldots, f_n) \in \mathbb{R}^n$ is

$$\|F\|_\infty = \max\{|f_1|, |f_2|, \ldots, |f_n|\}.$$

- It is an established result that $(\mathbb{R}^n, \|\cdot\|_\infty)$ is a Banach space.

> **Fact.** $B^\star$ is a contraction mapping in the $(\mathbb{R}^n, \|\cdot\|_\infty)$ Banach space with contraction factor γ.

Proof that $B^\star$ is a Contraction Mapping

We use: $|\max_a f(a) - \max_a g(a)| \leq \max_a |f(a) - g(a)|$.

Proof that $B^\star$ is a Contraction Mapping

We use: $|\max_a f(a) - \max_a g(a)| \leq \max_a |f(a) - g(a)|$.

$$\|B^\star(F) - B^\star(G)\|_\infty$$

Proof that $B^\star$ is a Contraction Mapping

We use: $|\max_a f(a) - \max_a g(a)| \leq \max_a |f(a) - g(a)|$.

$$\|B^\star(F) - B^\star(G)\|_\infty = \max_{s \in S} |(B^\star(F))(s) - (B^\star(G))(s)|$$

Proof that $B^\star$ is a Contraction Mapping

We use: $|\max_a f(a) - \max_a g(a)| \leq \max_a |f(a) - g(a)|$.

$$\|B^\star(F) - B^\star(G)\|_\infty = \max_{s \in S} |(B^\star(F))(s) - (B^\star(G))(s)|$$

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

$$\left. \max_{a \in A} \sum_{s' \in S} T(s, a, s')\{R(s, a, s') + \gamma G(s')\} \right|$$

Proof that $B^\star$ is a Contraction Mapping

We use: $|\max_a f(a) - \max_a g(a)| \leq \max_a |f(a) - g(a)|$.

$$
\begin{aligned}
\|B^\star(F) - B^\star(G)\|_\infty &= \max_{s \in S} |(B^\star(F))(s) - (B^\star(G))(s)| \\
&= \max_{s \in S} \left| \max_{a \in A} \sum_{s' \in S} T(s, a, s')\{R(s, a, s') + \gamma F(s')\} - \right. \\
&\qquad \left. \max_{a \in A} \sum_{s' \in S} T(s, a, s')\{R(s, a, s') + \gamma G(s')\} \right| \\
&\leq \gamma \max_{(s,a) \in S \times A} \left| \sum_{s' \in S} T(s, a, s')\{F(s') - G(s')\} \right|
\end{aligned}
$$

Proof that $B^\star$ is a Contraction Mapping

We use: $|\max_a f(a) - \max_a g(a)| \leq \max_a |f(a) - g(a)|$.

$$\begin{aligned}
\|B^\star(F) - B^\star(G)\|_\infty &= \max_{s \in S} |(B^\star(F))(s) - (B^\star(G))(s)| \\
&= \max_{s \in S} \left| \max_{a \in A} \sum_{s' \in S} T(s, a, s')\{R(s, a, s') + \gamma F(s')\} - \right. \\
&\qquad \left. \max_{a \in A} \sum_{s' \in S} T(s, a, s')\{R(s, a, s') + \gamma G(s')\} \right| \\
&\leq \gamma \max_{(s,a) \in S \times A} \left| \sum_{s' \in S} T(s, a, s')\{F(s') - G(s')\} \right| \\
&\leq \gamma \max_{(s,a) \in S \times A} \sum_{s' \in S} T(s, a, s')|F(s') - G(s')|
\end{aligned}$$

Proof that $B^\star$ is a Contraction Mapping

We use: $|\max_a f(a) - \max_a g(a)| \leq \max_a |f(a) - g(a)|$.

$$\begin{aligned}
\|B^\star(F) - B^\star(G)\|_\infty &= \max_{s \in S} |(B^\star(F))(s) - (B^\star(G))(s)| \\
&= \max_{s \in S} \Bigg| \max_{a \in A} \sum_{s' \in S} T(s, a, s')\{R(s, a, s') + \gamma F(s')\} - \\
&\qquad \max_{a \in A} \sum_{s' \in S} T(s, a, s')\{R(s, a, s') + \gamma G(s')\} \Bigg| \\
&\leq \gamma \max_{(s,a) \in S \times A} \left| \sum_{s' \in S} T(s, a, s')\{F(s') - G(s')\} \right| \\
&\leq \gamma \max_{(s,a) \in S \times A} \sum_{s' \in S} T(s, a, s') |F(s') - G(s')| \\
&\leq \gamma \max_{(s,a) \in S \times A} \sum_{s' \in S} T(s, a, s') \|F - G\|_\infty
\end{aligned}$$

Proof that $B^\star$ is a Contraction Mapping

We use: $|\max_a f(a) - \max_a g(a)| \leq \max_a |f(a) - g(a)|$.

$$
\begin{aligned}
\|B^\star(F) - B^\star(G)\|_\infty &= \max_{s \in S} |(B^\star(F))(s) - (B^\star(G))(s)| \\
&= \max_{s \in S} \left| \max_{a \in A} \sum_{s' \in S} T(s, a, s')\{R(s, a, s') + \gamma F(s')\} - \right. \\
&\qquad \left. \max_{a \in A} \sum_{s' \in S} T(s, a, s')\{R(s, a, s') + \gamma G(s')\} \right| \\
&\leq \gamma \max_{(s,a) \in S \times A} \left| \sum_{s' \in S} T(s, a, s')\{F(s') - G(s')\} \right| \\
&\leq \gamma \max_{(s,a) \in S \times A} \sum_{s' \in S} T(s, a, s')|F(s') - G(s')| \\
&\leq \gamma \max_{(s,a) \in S \times A} \sum_{s' \in S} T(s, a, s')\|F - G\|_\infty = \gamma \|F - G\|_\infty.
\end{aligned}
$$

The Fixed-point of $B^\star$

- Banach's Fixed-point Theorem implies there is a unique fixed point for $B^\star$.

The Fixed-point of $B^\star$

- Banach's Fixed-point Theorem implies there is a unique fixed point for $B^\star$.
- Denote the fixed point $V^\star : S \to \mathbb{R}$. Note that $B^\star(V^\star) = V^\star$. In other words, for $s \in S$:

$$V^\star(s) = \max_{a \in A} \sum_{s' \in S} T(s, a, s') \{R(s, a, s') + \gamma V^\star(s')\}.$$

The Fixed-point of $B^\star$

- Banach's Fixed-point Theorem implies there is a unique fixed point for $B^\star$.
- Denote the fixed point $V^\star : S \to \mathbb{R}$. Note that $B^\star(V^\star) = V^\star$. In other words, for $s \in S$:

$$V^\star(s) = \max_{a \in A} \sum_{s' \in S} T(s, a, s') \{R(s, a, s') + \gamma V^\star(s')\}.$$

- These are the Bellman optimality equations for MDP (S, A, T, R, γ).

The Fixed-point of $B^\star$

- Banach's Fixed-point Theorem implies there is a unique fixed point for $B^\star$.
- Denote the fixed point $V^\star : S \to \mathbb{R}$. Note that $B^\star(V^\star) = V^\star$. In other words, for $s \in S$:

$$V^\star(s) = \max_{a \in A} \sum_{s' \in S} T(s, a, s') \{R(s, a, s') + \gamma V^\star(s')\}.$$

- These are the Bellman optimality equations for MDP (S, A, T, R, γ).

  n equations, n unknowns, but non-linear!

The Fixed-point of $B^\star$

- Banach's Fixed-point Theorem implies there is a unique fixed point for $B^\star$.
- Denote the fixed point $V^\star : S \to \mathbb{R}$. Note that $B^\star(V^\star) = V^\star$. In other words, for $s \in S$:

$$V^\star(s) = \max_{a \in A} \sum_{s' \in S} T(s, a, s') \{R(s, a, s') + \gamma V^\star(s')\}.$$

- These are the Bellman optimality equations for MDP (S, A, T, R, γ).

  n equations, n unknowns, but non-linear!
- Value iteration, linear programming, and policy iteration are three distinct families of algorithms to compute $V^\star$.

The Fixed-point of $B^\star$

- Banach's Fixed-point Theorem implies there is a unique fixed point for $B^\star$.
- Denote the fixed point $V^\star : S \to \mathbb{R}$. Note that $B^\star(V^\star) = V^\star$. In other words, for $s \in S$:

$$\boxed{V^\star(s) = \max_{a \in A} \sum_{s' \in S} T(s, a, s') \{R(s, a, s') + \gamma V^\star(s')\}.}$$

- These are the Bellman optimality equations for MDP (S, A, T, R, γ).

  n equations, n unknowns, but non-linear!
- Value iteration, linear programming, and policy iteration are three distinct families of algorithms to compute $V^\star$.
- **Fact.** $V^\star$ is the value function of every policy $\pi^\star : S \to A$ such that for $s \in S$:

$$\pi^\star(s) = \text{argmax}_{a \in A} \sum_{s' \in S} T(s, a, s') \{R(s, a, s') + \gamma V^*(s')\}.$$

The Fixed-point of $B^\star$

- Banach's Fixed-point Theorem implies there is a unique fixed point for $B^\star$.
- Denote the fixed point $V^\star : S \to \mathbb{R}$. Note that $B^\star(V^\star) = V^\star$. In other words, for $s \in S$:

$$V^\star(s) = \max_{a \in A} \sum_{s' \in S} T(s, a, s') \{R(s, a, s') + \gamma V^\star(s')\}.$$

- These are the Bellman optimality equations for MDP (S, A, T, R, γ).

  n equations, n unknowns, but non-linear!
- Value iteration, linear programming, and policy iteration are three distinct families of algorithms to compute $V^\star$.
- **Fact.** $V^\star$ is the value function of every policy $\pi^\star : S \to A$ such that for $s \in S$:

$$\pi^\star(s) = \text{argmax}_{a \in A} \sum_{s' \in S} T(s, a, s') \{R(s, a, s') + \gamma V^*(s')\}.$$

- We shall prove next week that every such policy $\pi^\star$ is an optimal policy. Hence $V^\star$ is the optimal value function.

Markov Decision Problems

1. Banach's fixed-point theorem

2. Bellman optimality operator

3. Value iteration

Value Iteration

- Iterative approach to compute $V^{\star}$.

Value Iteration

- Iterative approach to compute $V^\star$.
- $V_0 \xrightarrow{B^\star} V_1 \xrightarrow{B^\star} V_2 \xrightarrow{B^\star} \ldots$.

Value Iteration

- Iterative approach to compute $V^\star$.
- $V_0 \xrightarrow{B^\star} V_1 \xrightarrow{B^\star} V_2 \xrightarrow{B^\star} \ldots$.

$V_0 \leftarrow$ Arbitrary, element-wise bounded, n-length vector.
$t \leftarrow 0$.
Repeat:
  For $s \in S$**:**
    $V_{t+1}(s) \leftarrow \max_{a \in A} \sum_{s' \in S} T(s, a, s') \left(R(s, a, s') + \gamma V_t(s')\right)$.
    $t \leftarrow t + 1$.
Until $V_t \approx V_{t-1}$ (up to machine precision).

Value Iteration

- Iterative approach to compute $V^\star$.
- $V_0 \xrightarrow{B^\star} V_1 \xrightarrow{B^\star} V_2 \xrightarrow{B^\star} \ldots$.

$V_0 \leftarrow$ Arbitrary, element-wise bounded, n-length vector.
$t \leftarrow 0$.
Repeat:
  For $s \in S$**:**
    $V_{t+1}(s) \leftarrow \max_{a \in A} \sum_{s' \in S} T(s, a, s')\,(R(s, a, s') + \gamma V_t(s'))$.
    $t \leftarrow t + 1$.
Until $V_t \approx V_{t-1}$ (up to machine precision).

- Popular; easy to implement; quick to converge in practice.

Markov Decision Problems

1. Banach's fixed-point theorem
2. Bellman optimality operator
3. Value iteration

Markov Decision Problems

1. Banach's fixed-point theorem
2. Bellman optimality operator
3. Value iteration

Next class: MDP planning through linear programming.