CS 747, Autumn 2020: Week 5, Lecture 1

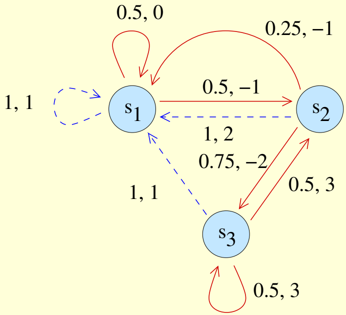
Shivaram Kalyanakrishnan

Department of Computer Science and Engineering
Indian Institute of Technology Bombay

Autumn 2020

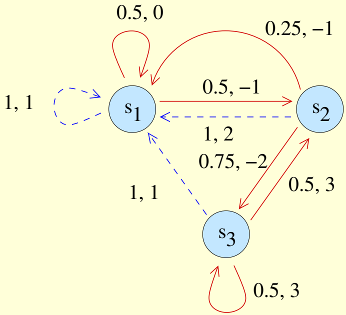
Summary of Previous Lecture

1. Definitions
   - MDP (S, A, T, R, γ)
   - Policy (π)
   - Value Function (V^{π})
2. MDP planning
3. Alternative formulations
4. Applications
5. Policy Evaluation



Summary of Previous Lecture

1. Definitions
   - MDP (S, A, T, R, γ)
   - Policy (π)
   - Value Function (V^{π})
2. MDP planning
3. Alternative formulations
4. Applications
5. Policy Evaluation



What is coming up this week?

Markov Decision Problems

1. Bellman optimality
   - Banach's fixed-point theorem
   - Bellman optimality operator

2. Value Iteration

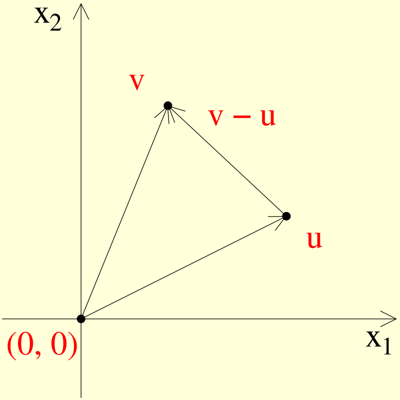
3. Linear Programming formulation
   - Review of LP
   - MDP Planning as LP

Markov Decision Problems

1. Bellman optimality
   - Banach's fixed-point theorem
   - Bellman optimality operator
2. Value Iteration
3. Linear Programming formulation
   - Review of LP
   - MDP Planning as LP

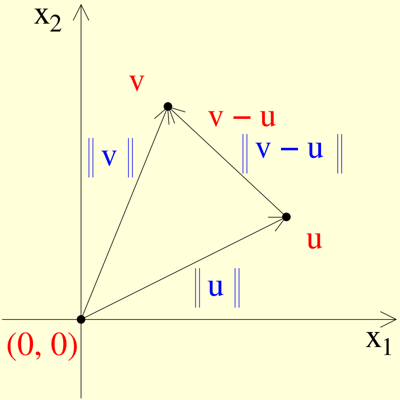
Complete, Normed Vector Spaces

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



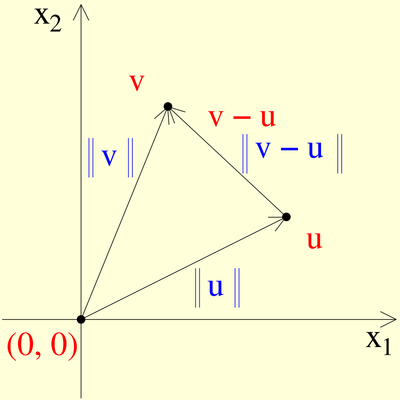
Complete, Normed Vector Spaces

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



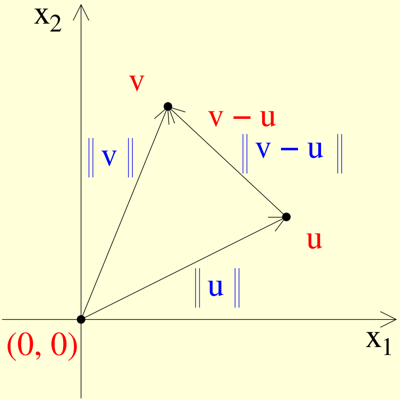
Complete, Normed Vector Spaces

- A vector space X has objects called vectors that can be added and scaled.
- A norm $\|\cdot\|$ associates a length which 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 which 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.

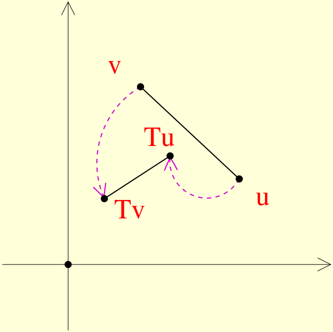
Two Definitions

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

Two Definitions

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

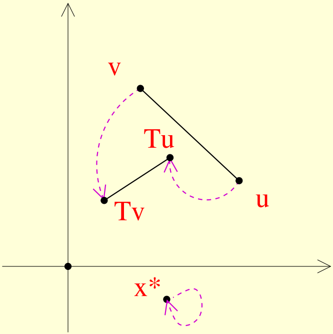
$$\|Tv - Tu\| \leq L\|v - u\|.$$



Two Definitions

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

$$\|Tv - Tu\| \leq L\|v - u\|.$$



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

Banach's Fixed-point Theorem

(Adapted from Szepesvári, 2010 (see Appendix A.1).)

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

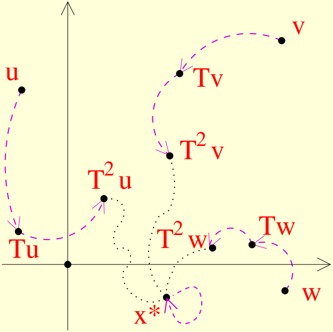
1. T has a unique fixed point $x^\star \in X$.
2. For $x \in X, m \geq 0$: $\|T^m x - x^\star\| \leq L^m \|x - x^\star\|$.

Banach's Fixed-point Theorem

(Adapted from Szepesvári, 2010 (see Appendix A.1).)

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

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



Bellman Optimality Operator

- The Bellman optimality operator $B^\star : (S \to \mathbb{R}) \to (S \to \mathbb{R})$ for an MDP (S, A, T, R, γ) is defined as follows.
  For $F : S \to \mathbb{R}$ and $s \in S$:

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

Bellman Optimality Operator

- The Bellman optimality operator $B^\star : (S \to \mathbb{R}) \to (S \to \mathbb{R})$ for an MDP (S, A, T, R, γ) is defined as follows.
  For $F : S \to \mathbb{R}$ and $s \in S$:

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

- Since $S = \{s_1, s_2, \ldots, s_n\}$, we may equivalently view $B^\star$ as a mapping from $\mathbb{R}^n$ to $\mathbb{R}^n$.

Bellman Optimality Operator

- The Bellman optimality operator $B^\star : (S \to \mathbb{R}) \to (S \to \mathbb{R})$ for an MDP (S, A, T, R, γ) is defined as follows.
  For $F : S \to \mathbb{R}$ and $s \in S$:

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

- Since $S = \{s_1, s_2, \ldots, s_n\}$, we may equivalently view $B^\star$ as a mapping from $\mathbb{R}^n$ to $\mathbb{R}^n$.
- 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

- The Bellman optimality operator $B^\star : (S \to \mathbb{R}) \to (S \to \mathbb{R})$ for an MDP (S, A, T, R, γ) is defined as follows.
  For $F : S \to \mathbb{R}$ and $s \in S$:

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

- Since $S = \{s_1, s_2, \ldots, s_n\}$, we may equivalently view $B^\star$ as a mapping from $\mathbb{R}^n$ to $\mathbb{R}^n$.
- 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.

Bellman Optimality Operator

- The Bellman optimality operator $B^\star : (S \to \mathbb{R}) \to (S \to \mathbb{R})$ for an MDP (S, A, T, R, γ) is defined as follows.
  For $F : S \to \mathbb{R}$ and $s \in S$:
  $$(B^\star(F))(s) \stackrel{\text{def}}{=} \max_{a \in A} \sum_{s' \in S} T(s, a, s')\{R(s, a, s') + \gamma F(s')\}.$$
- Since $S = \{s_1, s_2, \dots, s_n\}$, we may equivalently view $B^\star$ as a mapping from $\mathbb{R}^n$ to $\mathbb{R}^n$.
- 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.

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

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

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

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

$$\leq \gamma \max_{s \in S} \max_{a \in A} \left| \sum_{s' \in S} T(s, a, s')\{F(s') - G(s')\} \right|$$

$$\leq \gamma \max_{s \in S} \max_{a \in A} \sum_{s' \in S} T(s, a, s') |F(s') - G(s')|$$

$$\leq \gamma \max_{s \in S} \max_{a \in A} \sum_{s' \in S} T(s, a, s') \|F - G\|_\infty = \gamma \|F - G\|_\infty.$$

The Fixed-point of $B^{\star}$

- By Banach's Fixed-point Theorem, it follows that there is a unique fixed point for $B^{\star}$.

The Fixed-point of $B^{\star}$

- By Banach's Fixed-point Theorem, it follows that 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 $s \in S$:

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

The Fixed-point of $B^\star$

- By Banach's Fixed-point Theorem, it follows that 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 $s \in S$:

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

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

The Fixed-point of $B^\star$

- By Banach's Fixed-point Theorem, it follows that 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 $s \in S$:

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

- These are the Bellman optimality equations for MDP (S, A, T, R, γ). n equations, n unknowns, but non-linear!

The Fixed-point of $B^\star$

- By Banach's Fixed-point Theorem, it follows that 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 $s \in S$:

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

- 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 $B^\star$

- By Banach's Fixed-point Theorem, it follows that 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 $s \in S$:

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

- 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 $\pi^\star : S \to A$ that satisfies, for all $s \in S$:

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

The Fixed-point of $B^\star$

- By Banach's Fixed-point Theorem, it follows that 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 $s \in S$:

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

- 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 $\pi^\star : S \to A$ that satisfies, for all $s \in S$:

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

- 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. Bellman optimality
   - Banach's fixed-point theorem
   - Bellman optimality operator

2. Value Iteration

3. Linear Programming formulation
   - Review of LP
   - MDP Planning as LP

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 $s \in S$**:**
    $V_{t+1}(s) \leftarrow \max_{a \in A} \sum_{s' \in S} T(s, a, s') \left(R(s, a, s') + \gamma V_t(s')\right)$.
    $t \leftarrow t + 1$.
Until $V_t \approx V_{t-1}$ (up to machine precision).

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

Relationship of $V^\star$, $Q^\star$, $\pi^\star$

- Say we are working with MDP (S, A, T, R, γ).

Relationship of $V^\star$, $Q^\star$, $\pi^\star$

- Say we are working with MDP (S, A, T, R, γ).

- Suppose you have computed $V^\star$. How to get $Q^\star$?

Relationship of $V^\star$, $Q^\star$, $\pi^\star$

- Say we are working with MDP (S, A, T, R, γ).
- Suppose you have computed $V^\star$. How to get $Q^\star$?
  For $s \in S, a \in A$:
  $$Q^\star(s, a) = \sum_{s' \in S} T(s, a, s')\{R(s, a, s') + \gamma V^\star(s')\}.$$
- Now that you have $Q^\star$, how to get $\pi^\star$?

Relationship of $V^\star$, $Q^\star$, $\pi^\star$

- Say we are working with MDP (S, A, T, R, γ).
- Suppose you have computed $V^\star$. How to get $Q^\star$?
  For $s \in S, a \in A$:
  $$Q^\star(s, a) = \sum_{s' \in S} T(s, a, s')\{R(s, a, s') + \gamma V^\star(s')\}.$$
- Now that you have $Q^\star$, how to get $\pi^\star$?
  For $s \in S$,
  $$\pi^\star(s) = \operatorname*{argmax}_{a \in A} Q^\star(s, a).$$

Relationship of $V^\star$, $Q^\star$, $\pi^\star$

- Say we are working with MDP (S, A, T, R, γ).

- Suppose you have computed $V^\star$. How to get $Q^\star$?
  For $s \in S, a \in A$:
  $$Q^\star(s, a) = \sum_{s' \in S} T(s, a, s')\{R(s, a, s') + \gamma V^\star(s')\}.$$

- Now that you have $Q^\star$, how to get $\pi^\star$?
  For $s \in S$,
  $$\pi^\star(s) = \operatorname*{argmax}_{a \in A} Q^\star(s, a).$$

- Suppose you have computed $\pi^\star$. How to get $V^\star$?

Relationship of $V^\star$, $Q^\star$, $\pi^\star$

- Say we are working with MDP (S, A, T, R, γ).

- Suppose you have computed $V^\star$. How to get $Q^\star$?
  For $s \in S, a \in A$:

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

- Now that you have $Q^\star$, how to get $\pi^\star$?
  For $s \in S$,

$$\pi^\star(s) = \operatorname*{argmax}_{a \in A} Q^\star(s, a).$$

- Suppose you have computed $\pi^\star$. How to get $V^\star$?
  Solve Bellman equations for $\pi^\star$!

Markov Decision Problems

1. Bellman optimality
   - Banach's fixed-point theorem
   - Bellman optimality operator
2. Value Iteration
3. Linear Programming formulation
   - Review of LP
   - MDP Planning as LP

Linear Programming

- To solve for real-valued variables $x_1, x_2, \ldots, x_m$ such that
  - a given linear function of the variables is maximised, while
  - given linear constraints on the variables are satisfied.

Linear Programming

- To solve for real-valued variables $x_1, x_2, \ldots, x_m$ such that
  - a given linear function of the variables is maximised, while
  - given linear constraints on the variables are satisfied.

Maximise $x_1 + 2x_2$ //Objective function
subject to: //Constraints

$$
\begin{aligned}
x_1 + x_2 &\leq 9, && \text{(C1)}\\
4x_1 - 13x_2 &\leq -75, && \text{(C2)}\\
x_1 &\leq 5. && \text{(C3)}
\end{aligned}
$$

Linear Programming

- To solve for real-valued variables $x_1, x_2, \ldots, x_m$ such that
  - a given linear function of the variables is maximised, while
  - given linear constraints on the variables are satisfied.

Maximise $x_1 + 2x_2$ //Objective function
subject to: //Constraints

$$x_1 + x_2 \leq 9, \qquad \text{(C1)}$$
$$4x_1 - 13x_2 \leq -75, \qquad \text{(C2)}$$
$$x_1 \leq 5. \qquad \text{(C3)}$$

- Well-studied problem with wide-ranging applications in mathematics, engineering.

Linear Programming

- To solve for real-valued variables $x_1, x_2, \ldots, x_m$ such that
  - a given linear function of the variables is maximised, while
  - given linear constraints on the variables are satisfied.

Maximise $x_1 + 2x_2$ //Objective function
subject to: //Constraints

$$x_1 + x_2 \leq 9, \quad \text{(C1)}$$
$$4x_1 - 13x_2 \leq -75, \quad \text{(C2)}$$
$$x_1 \leq 5. \quad \text{(C3)}$$

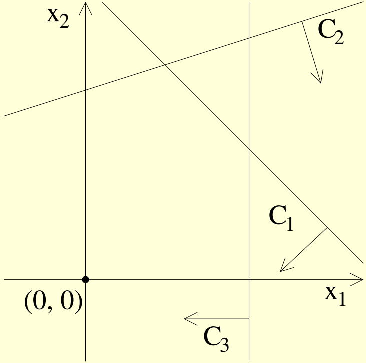
- Well-studied problem with wide-ranging applications in mathematics, engineering.
- Today's solvers (commercial, as well as open source) can handle LPs with millions of variables.

Solving a Linear Program

- **Step 1**: Identify the feasible set, which contains all the points satisfying the constraints. Might be empty, but otherwise will be convex.

Maximise $x_1 + 2x_2$
subject to:

$$\begin{aligned} x_1 + x_2 &\leq 9, & \text{(C1)} \\ 4x_1 - 13x_2 &\leq -75, & \text{(C2)} \\ x_1 &\leq 5. & \text{(C3)} \end{aligned}$$

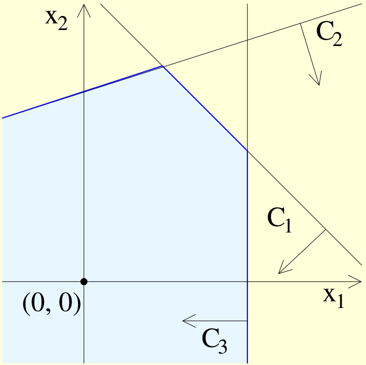


Solving a Linear Program

- **Step 1**: Identify the feasible set, which contains all the points satisfying the constraints. Might be empty, but otherwise will be convex.

Maximise $x_1 + 2x_2$
subject to:

$$\begin{aligned} x_1 + x_2 &\leq 9, && \text{(C1)}\\ 4x_1 - 13x_2 &\leq -75, && \text{(C2)}\\ x_1 &\leq 5. && \text{(C3)} \end{aligned}$$

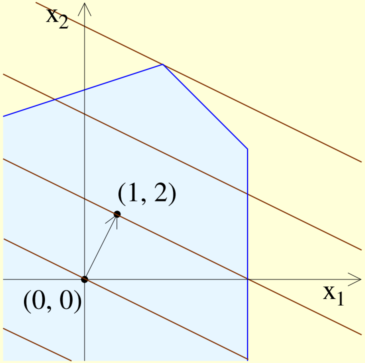


Solving a Linear Program

- **Step 1**: Identify the feasible set, which contains all the points satisfying the constraints. Might be empty, but otherwise will be convex.
- **Step 2**: Identify points within the feasible set that maximise the objective. Usually a single point.

Maximise $x_1 + 2x_2$
subject to:

$$\begin{aligned} x_1 + x_2 &\leq 9, & \text{(C1)} \\ 4x_1 - 13x_2 &\leq -75, & \text{(C2)} \\ x_1 &\leq 5. & \text{(C3)} \end{aligned}$$



Bellman Optimality Equations as an LP

- Bellman optimality equations: for $s \in S$,

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

Bellman Optimality Equations as an LP

- Bellman optimality equations: for $s \in S$,

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

- Let us create n variables $V(s_1), V(s_2), \ldots, V(s_n)$, and attempt to create an LP whose unique solution is $V^\star$.

Bellman Optimality Equations as an LP

- Bellman optimality equations: for $s \in S$,

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

- Let us create n variables $V(s_1), V(s_2), \dots, V(s_n)$, and attempt to create an LP whose unique solution is $V^\star$.
- Although the Bellman optimality equations are non-linear, we can easily create linear constraints. For $s \in S, a \in A$:

$$\boxed{V(s) \geq \sum_{s' \in S} T(s, a, s')\{R(s, a, s') + \gamma V(s')\}.}$$

Bellman Optimality Equations as an LP

- Bellman optimality equations: for $s \in S$,

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

- Let us create n variables $V(s_1), V(s_2), \ldots, V(s_n)$, and attempt to create an LP whose unique solution is $V^\star$.
- Although the Bellman optimality equations are non-linear, we can easily create linear constraints. For $s \in S, a \in A$:

$$\boxed{V(s) \geq \sum_{s' \in S} T(s, a, s')\{R(s, a, s') + \gamma V(s')\}.}$$

- These are nk linear constraints.

Bellman Optimality Equations as an LP

- Bellman optimality equations: for $s \in S$,

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

- Let us create n variables $V(s_1), V(s_2), \dots, V(s_n)$, and attempt to create an LP whose unique solution is $V^\star$.
- Although the Bellman optimality equations are non-linear, we can easily create linear constraints. For $s \in S, a \in A$:

$$\boxed{V(s) \geq \sum_{s' \in S} T(s, a, s')\{R(s, a, s') + \gamma V(s')\}.}$$

- These are *nk* linear constraints.
- Observe that $V^\star$ is in the feasible set.

Bellman Optimality Equations as an LP

- Bellman optimality equations: for $s \in S$,

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

- Let us create n variables $V(s_1), V(s_2), \ldots, V(s_n)$, and attempt to create an LP whose unique solution is $V^\star$.
- Although the Bellman optimality equations are non-linear, we can easily create linear constraints. For $s \in S, a \in A$:

$$\boxed{V(s) \geq \sum_{s' \in S} T(s, a, s')\{R(s, a, s') + \gamma V(s')\}.}$$

- These are *nk* linear constraints.
- Observe that $V^\star$ is in the feasible set.

Can we construct an objective function for which $V^\star$ is the sole optimiser?

Vector Comparison

- For $X : S \to \mathbb{R}$ and $Y : S \to \mathbb{R}$, we define

$$X \succeq Y \iff \forall s \in S : X(s) \geq Y(s),$$
$$X \succ Y \iff X \succeq Y \text{ and } \exists s \in S : X(s) > Y(s).$$

Vector Comparison

- For $X : S \to \mathbb{R}$ and $Y : S \to \mathbb{R}$, we define

$$X \succeq Y \iff \forall s \in S : X(s) \geq Y(s),$$
$$X \succ Y \iff X \succeq Y \text{ and } \exists s \in S : X(s) > Y(s).$$

- For policies $\pi_1, \pi_2 \in \Pi$, we define

$$\pi_1 \succeq \pi_2 \iff V^{\pi_1} \succeq V^{\pi_2},$$
$$\pi_1 \succ \pi_2 \iff V^{\pi_1} \succ V^{\pi_2}.$$

Vector Comparison

- For $X : S \to \mathbb{R}$ and $Y : S \to \mathbb{R}$, we define

$$X \succeq Y \iff \forall s \in S : X(s) \geq Y(s),$$
$$X \succ Y \iff X \succeq Y \text{ and } \exists s \in S : X(s) > Y(s).$$

- For policies $\pi_1, \pi_2 \in \Pi$, we define

$$\pi_1 \succeq \pi_2 \iff V^{\pi_1} \succeq V^{\pi_2},$$
$$\pi_1 \succ \pi_2 \iff V^{\pi_1} \succ V^{\pi_2}.$$

- Note that we can have incomparable policies $\pi_1, \pi_2 \in \Pi$: that is, neither $\pi_1 \succeq \pi_2$ nor $\pi_2 \succeq \pi_1$.

Vector Comparison

- For $X : S \to \mathbb{R}$ and $Y : S \to \mathbb{R}$, we define

$$X \succeq Y \iff \forall s \in S : X(s) \geq Y(s),$$
$$X \succ Y \iff X \succeq Y \text{ and } \exists s \in S : X(s) > Y(s).$$

- For policies $\pi_1, \pi_2 \in \Pi$, we define

$$\pi_1 \succeq \pi_2 \iff V^{\pi_1} \succeq V^{\pi_2},$$
$$\pi_1 \succ \pi_2 \iff V^{\pi_1} \succ V^{\pi_2}.$$

- Note that we can have incomparable policies $\pi_1, \pi_2 \in \Pi$: that is, neither $\pi_1 \succeq \pi_2$ nor $\pi_2 \succeq \pi_1$.

- Also note that if $\pi_1 \succeq \pi_2$ and $\pi_2 \succeq \pi_1$, then $V^{\pi_1} = V^{\pi_2}$.

$B^\star$ Preserves $\succeq$

- **Fact.** For $X : S \to \mathbb{R}$ and $Y : S \to \mathbb{R}$,
$$X \succeq Y \implies B^\star(X) \succeq B^\star(Y).$$

$B^\star$ Preserves $\succeq$

- **Fact.** For $X : S \to \mathbb{R}$ and $Y : S \to \mathbb{R}$,
$$X \succeq Y \implies B^\star(X) \succeq B^\star(Y).$$
As proof it suffices to show that if $X \succeq Y$, then for $s \in S$,
$$(B^\star(X))(s) - (B^\star(Y))(s) \geq 0.$$

$B^\star$ Preserves $\succeq$

- **Fact.** For $X : S \to \mathbb{R}$ and $Y : S \to \mathbb{R}$,
$$X \succeq Y \implies B^\star(X) \succeq B^\star(Y).$$
As proof it suffices to show that if $X \succeq Y$, then for $s \in S$,
$$(B^\star(X))(s) - (B^\star(Y))(s) \geq 0.$$

We use: $\max_a f(a) - \max_a g(a) \geq \min_a (f(a) - g(a))$.

$B^\star$ Preserves $\succeq$

- **Fact.** For $X : S \to \mathbb{R}$ and $Y : S \to \mathbb{R}$,
$$X \succeq Y \implies B^\star(X) \succeq B^\star(Y).$$
As proof it suffices to show that if $X \succeq Y$, then for $s \in S$,
$$(B^\star(X))(s) - (B^\star(Y))(s) \geq 0.$$

We use: $\max_a f(a) - \max_a g(a) \geq \min_a(f(a) - g(a))$.

$$\begin{aligned}
&(B^\star(X))(s) - (B^\star(Y))(s) \\
&= \max_{a \in A} \sum_{s' \in S} T(s, a, s')\{R(s, a, s') + \gamma X(s')\} - \\
&\qquad \max_{a \in A} \sum_{s' \in S} T(s, a, s')\{R(s, a, s') + \gamma Y(s')\} \\
&\geq \gamma \min_{a \in A} \sum_{s' \in S} T(s, a, s')\{X(s') - Y(s')\} \geq 0.
\end{aligned}$$

Examining the Feasible Set of our LP

- Each $V : S \to \mathbb{R}$ in our feasible set satisfies $V \succeq B^{\star}(V)$.

Examining the Feasible Set of our LP

- Each $V : S \to \mathbb{R}$ in our feasible set satisfies $V \succeq B^\star(V)$.
- Since $B^\star$ preserves $\succeq$, we get

$$\begin{aligned} V &\succeq B^\star(V) \\ \implies B^\star(V) &\succeq (B^\star)^2(V) \\ \implies (B^\star)^2(V) &\succeq (B^\star)^3(V) \\ &\vdots \end{aligned}$$

Examining the Feasible Set of our LP

- Each $V : S \to \mathbb{R}$ in our feasible set satisfies $V \succeq B^\star(V)$.
- Since $B^\star$ preserves $\succeq$, we get

$$
\begin{aligned}
& V \succeq B^\star(V) \\
\implies & B^\star(V) \succeq (B^\star)^2(V) \\
\implies & (B^\star)^2(V) \succeq (B^\star)^3(V) \\
& \vdots
\end{aligned}
$$

- By implication and by Banach's Fixed-point Theorem,
$$V \succeq \lim_{l \to \infty} (B^\star)^l(V) = V^\star.$$

Examining the Feasible Set of our LP

- Each $V: S \to \mathbb{R}$ in our feasible set satisfies $V \succeq B^\star(V)$.
- Since $B^\star$ preserves $\succeq$, we get

$$\begin{aligned} & V \succeq B^\star(V) \\ \implies & B^\star(V) \succeq (B^\star)^2(V) \\ \implies & (B^\star)^2(V) \succeq (B^\star)^3(V) \\ & \vdots \end{aligned}$$

- By implication and by Banach's Fixed-point Theorem,
$$V \succeq \lim_{l \to \infty} (B^\star)^l(V) = V^\star.$$
- We "linearise" this result: for $V: S \to R$ in the feasible set.
$$\sum_{s \in S} V(s) \geq \sum_{s \in S} V^\star(s).$$

Linear Programming Formulation

Maximise $\left(-\sum_{s \in S} V(s)\right)$

subject to

$$V(s) \geq \sum_{s' \in S} T(s, a, s')\{R(s, a, s') + \gamma V(s')\}, \forall s \in S, a \in A.$$

- This LP has n variables, nk constraints.

Linear Programming Formulation

$$\text{Maximise} \left(- \sum_{s \in S} V(s) \right)$$

subject to

$$V(s) \geq \sum_{s' \in S} T(s, a, s') \{ R(s, a, s') + \gamma V(s') \}, \forall s \in S, a \in A.$$

- This LP has n variables, nk constraints.
- There is also a *dual* LP formulation with nk variables and n constraints. See Littman et al. (1995) if interested.

Markov Decision Problems

1. Bellman optimality
   - Banach's fixed-point theorem
   - Bellman optimality operator

2. Value Iteration

3. Linear Programming formulation
   - Review of LP
   - MDP Planning as LP

Markov Decision Problems

1. Bellman optimality
   - Banach's fixed-point theorem
   - Bellman optimality operator
2. Value Iteration
3. Linear Programming formulation
   - Review of LP
   - MDP Planning as LP

Next week: Policy Iteration, proof of optimality of $\pi^\star$.