CS 747, Autumn 2023: Lecture 9

Shivaram Kalyanakrishnan

Department of Computer Science and Engineering
Indian Institute of Technology Bombay

Autumn 2023

Markov Decision Problems

1. Review of linear programming
2. MDP planning through linear programming

Markov Decision Problems

1. Review of linear programming
2. MDP planning through linear programming

Linear Programming

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

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

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

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 ("$\leq$") 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.

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 ("$\leq$") on the variables are satisfied.

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

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

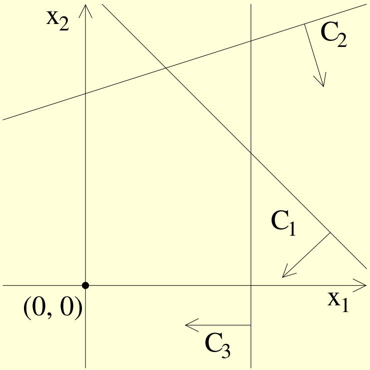
- Well-studied problem with wide-ranging applications in mathematics, engineering.
- Today's solvers (commercial, as well as open source) can handle LPs with millions of variables.

Conceptual Steps towards 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:

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

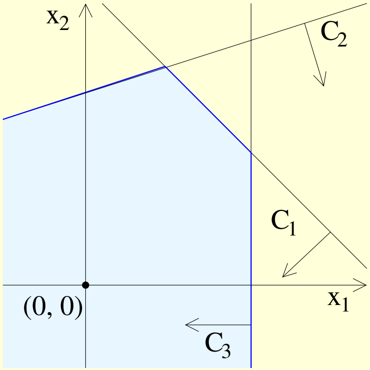


Conceptual Steps towards Solving a Linear Program

- **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:

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

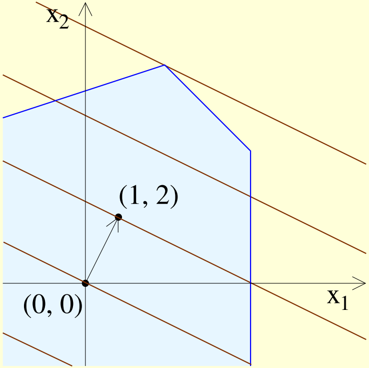


Conceptual Steps towards Solving a Linear Program

- **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:

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



Actually Solving a Linear Program

- Common approaches: Simplex, interior-point methods.

Actually Solving a Linear Program

- Common approaches: Simplex, interior-point methods.

- LP with d variables, m constraints, B-bit representation of floats.
  - Can be solved in poly(d, m, B) operations.
  - Can be solved in poly$(d, m) \cdot e^{O(\sqrt{d \log(m)})}$ expected "real RAM" operations.

Actually Solving a Linear Program

- Common approaches: Simplex, interior-point methods.

- LP with d variables, m constraints, B-bit representation of floats.
  - Can be solved in poly(d, m, B) operations.
  - Can be solved in poly$(d, m) \cdot e^{O(\sqrt{d \log(m)})}$ expected "real RAM" operations.

- Modern LP solvers can solve LPs with thousands/millions of variables/constraints in reasonable time (hours/days).

Actually Solving a Linear Program

- Common approaches: Simplex, interior-point methods.

- LP with d variables, m constraints, B-bit representation of floats.
  - Can be solved in poly(d, m, B) operations.
  - Can be solved in poly(d, m) $\cdot\, e^{O(\sqrt{d \log(m)})}$ expected "real RAM" operations.

- Modern LP solvers can solve LPs with thousands/millions of variables/constraints in reasonable time (hours/days).

- Most engineers' focus is on formulating, rather than solving, LP.

Markov Decision Problems

1. Review of linear programming
2. MDP planning through linear programming

Bellman Optimality Equations as an LP

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

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

Bellman Optimality Equations as an LP

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

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

- 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 $V^\star$.

Bellman Optimality Equations as an LP

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

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

- 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 $s \in S$, $a \in A$:

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

Bellman Optimality Equations as an LP

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

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

- 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 $s \in S$, $a \in A$:

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

- These are nk linear constraints.

Bellman Optimality Equations as an LP

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

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

- 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 $s \in S, a \in A$:

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

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

Bellman Optimality Equations as an LP

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

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

- 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 $s \in S, a \in A$:

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

- 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 $Y : S \to \mathbb{R}$ (equivalently $X, Y \in \mathbb{R}^n$), 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 $Y : S \to \mathbb{R}$ (equivalently $X, Y \in \mathbb{R}^n$), 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 $Y : S \to \mathbb{R}$ (equivalently $X, Y \in \mathbb{R}^n$), 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 $Y : S \to \mathbb{R}$ (equivalently $X, Y \in \mathbb{R}^n$), 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: $\max_a f(a) - \max_a g(a) \geq \min_a(f(a) - g(a))$.

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

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,

$$V \succeq \lim_{l \to \infty} (B^\star)^l(V) = V^\star.$$

- For all $V \neq V^\star$ in the feasible set, $V \succ V^\star$.

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,

$$V \succeq \lim_{l \to \infty} (B^\star)^l(V) = V^\star.$$

- For all $V \neq V^\star$ in the feasible set, $V \succ V^\star$. By implication:

$$\sum_{s \in S} V(s) > \sum_{s \in S} V^\star(s).$$

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

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. $V^{\star}$ is the unique solution.

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. $V^\star$ is the unique solution.
- There is also a *dual* LP formulation with nk variables and n 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. $V^\star$ is the unique solution.
- There is also a *dual* LP formulation with nk variables and n constraints. The dual's solution immediately yields $\pi^\star$.

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. $V^\star$ is the unique solution.
- There is also a *dual* LP formulation with nk variables and n constraints. The dual's solution immediately yields $\pi^\star$. See Littman et al. (1995) if interested.

Markov Decision Problems

1. Review of linear programming
2. MDP planning through linear programming

Markov Decision Problems

1. Review of linear programming

2. MDP planning through linear programming

Next class: policy iteration.