Planning and Optimization

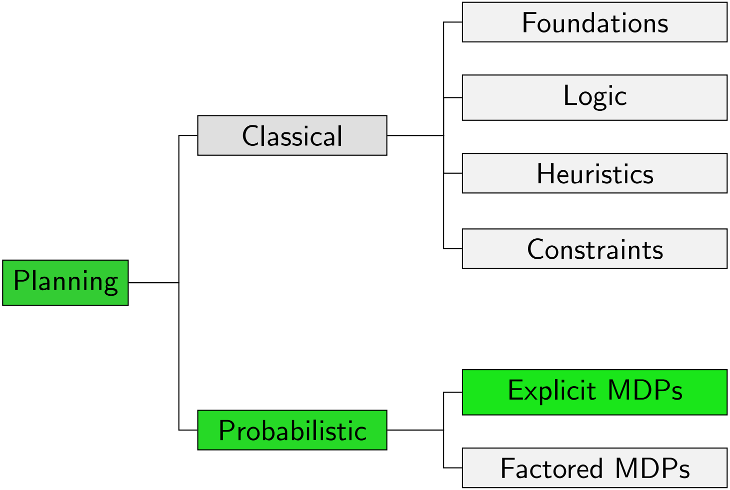
F2. Bellman Equation & Linear Programming

Malte Helmert and Thomas Keller

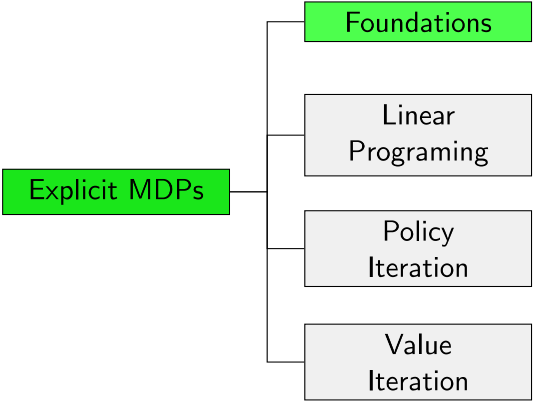
Universität Basel

November 27, 2019

Content of this Course



Content of this Course: Explicit MDPs



Introduction

Quality of Solutions

- Solution in classical planning: plan
- Optimality criterion of a solution in classical planning: minimize plan cost

Quality of Solutions

- Solution in classical planning: plan
- Optimality criterion of a solution in classical planning: minimize plan cost

- Solution in probabilistic planning: policy
- What is the optimality criterion of a solution in probabilistic planning?

Example: Swiss Lotto

Example (Swiss Lotto)

What is the expected payoff of placing one bet in Swiss Lotto for a cost of $CHF2.50$ with (simplified) payouts and probabilities:

$$\begin{aligned}
\text{CHF } 30.000.000 \text{ with prob. } 1/31474716 &\qquad (6+1)\\
\text{CHF } 1.000.000 \text{ with prob. } 1/5245786 &\qquad (6)\\
\text{CHF } 5.000 \text{ with prob. } 1/850668 &\qquad (5)\\
\text{CHF } 50 \text{ with prob. } 1/111930 &\qquad (4)\\
\text{CHF } 10 \text{ with prob. } 1/11480 &\qquad (3)
\end{aligned}$$

Example: Swiss Lotto

Example (Swiss Lotto)

What is the expected payoff of placing one bet in Swiss Lotto for a cost of $CHF 2.50$ with (simplified) payouts and probabilities:

$$
\begin{aligned}
\text{CHF } 30.000.000 \text{ with prob. } & 1/31474716 & (6+1) \\
\text{CHF } 1.000.000 \text{ with prob. } & 1/5245786 & (6) \\
\text{CHF } 5.000 \text{ with prob. } & 1/850668 & (5) \\
\text{CHF } 50 \text{ with prob. } & 1/111930 & (4) \\
\text{CHF } 10 \text{ with prob. } & 1/11480 & (3)
\end{aligned}
$$

Solution: $$\frac{30000000}{31474716} + \frac{1000000}{5245786} + \frac{5000}{850668} + \frac{50}{111930} + \frac{10}{11480} - 2.5 \approx -1.35.$$

Expected Values under Uncertainty

Definition (Expected Value of a Random Variable)

Let X be a random variable with a finite number of outcomes $d_1, \dots, d_n \in \mathbb{R}$, and let d_i happen with probability $p_i \in [0, 1]$ (for $i = 1, \dots n$) s.t. $\sum_{i=1}^{n} p_i = 1$.

The expected value of X is $\mathbb{E}[X] = \sum_{i=1}^{n}(p_i \cdot d_i)$.

Bellman Equation

Value Functions for MDPs

Definition (Value Functions for MDPs)

Let $\mathcal{T} = \langle S, L, c, T, s_0, \gamma \rangle$ be an MDP and π be an executable policy for $\mathcal{T}$.

The state-value $V_\pi(s)$ of s under π is defined as

$$V_\pi(s) := Q_\pi(s, \pi(s))$$

where the action-value $Q_\pi(s, \ell)$ of s and ℓ under π is defined as

$$Q_\pi(s, \ell) := R(s, \ell) + \gamma \cdot \sum_{s' \in \mathrm{succ}(s,\ell)} T(s, \ell, s') \cdot V_\pi(s').$$

The state-value $V_\pi(s)$ describes the expected reward of applying π in MDP $\mathcal{T}$, starting from s.

Bellman Equation in MDPs

Definition (Bellman Equation in MDPs)

Let $\mathcal{T} = \langle S, L, c, T, s_0, \gamma \rangle$ be an MDP.

The Bellman equation for a state s of $\mathcal{T}$ is the set of equations that describes $V_\star(s)$, where

$$V_\star(s) := \max_{\ell \in L(s)} Q_\star(s, \ell)$$

$$Q_\star(s, \ell) := R(s, \ell) + \gamma \cdot \sum_{s' \in \mathrm{succ}(s,\ell)} T(s, \ell, s') \cdot V_\star(s').$$

The solution $V_\star(s)$ of the Bellman equation describes the maximal expected reward that can be achieved from state s in MDP $\mathcal{T}$.

Optimal Policy in MDPs

What is the policy that achieves the maximal expected reward?

Definition (Optimal Policy in MDPs)

Let $\mathcal{T} = \langle S, L, c, T, s_0, \gamma \rangle$ be an MDP.
A policy π is an optimal policy if $\pi(s) \in \arg\max_{\ell \in L(s)} Q_\star(s, \ell)$ for all $s \in S$ and the expected reward of π in $\mathcal{T}$ is $V_\star(s_0)$.

W.l.o.g., we assume the optimal policy is unique and written as $\pi^\star$.

Value Functions for SSPs

Definition (Value Functions for SSPs)

Let $\mathcal{T} = \langle S, L, c, T, s_0, S_\star \rangle$ be an SSP and π be an executable policy for $\mathcal{T}$.

The state-value $V_\pi(s)$ of s under π is defined as

$$V_\pi(s) := \begin{cases} 0 & \text{if } s \in S_\star \\ Q_\pi(s, \pi(s)) & \text{otherwise,} \end{cases}$$

where the action-value $Q_\pi(s, \ell)$ of s and ℓ under π is defined as

$$Q_\pi(s, \ell) := c(\ell) + \sum_{s' \in \mathrm{succ}(s,\ell)} T(s, \ell, s') \cdot V_\pi(s').$$

The state-value $V_\pi(s)$ describes the expected cost of applying π in SSP $\mathcal{T}$, starting from s.

Bellman Equation in SSPs

Definition (Bellman Equation in SSPs)

Let $\mathcal{T} = \langle S, L, c, T, s_0, S_\star \rangle$ be an SSP.

The Bellman equation for a state s of $\mathcal{T}$ is the set of equations that describes $V_\star(s)$, where

$$V_\star(s) := \min_{\ell \in L(s)} Q_\star(s, \ell)$$

$$Q_\star(s, \ell) := c(\ell) + \sum_{s' \in \mathrm{succ}(s,\ell)} T(s, \ell, s') \cdot V_\star(s').$$

The solution $V_\star(s)$ of the Bellman equation describes the minimal expected cost that can be achieved from state s in SSP $\mathcal{T}$.

Optimal Policy in SSPs

What is the policy that achieves the minimal expected cost?

Definition (Optimal Policy in SSPs)

Let $\mathcal{T} = \langle S, L, c, T, s_0, S_\star \rangle$ be an SSP.
A policy π is an optimal policy if $\pi(s) \in \arg\min_{\ell \in L(s)} Q_\star(s, \ell)$ for all $s \in S$ and the expected cost of π in $\mathcal{T}$ is $V_\star(s_0)$.

W.l.o.g., we assume the optimal policy is unique and written as $\pi^\star$.

Proper SSP Policy

Definition (Proper SSP Policy)

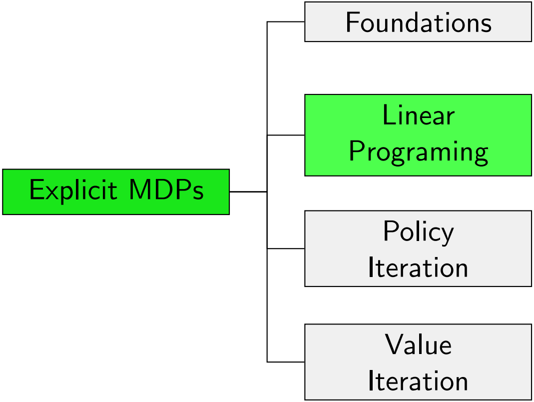
Let $\mathcal{T} = \langle S, L, c, T, s_0, S_\star \rangle$ be an SSP and π be an executable policy for $\mathcal{T}$. π is proper if it reaches a goal state from each reachable state with probability 1, i.e. if

$$\sum_{s \xrightarrow{p_1:\ell_1} s', \ldots, s'' \xrightarrow{p_n:\ell_n} s_\star} \prod_{i=1}^{n} p_i = 1$$

for all states $s \in S_\pi(s)$.

Linear Programming

Content of this Course: Explicit MDPs



Linear Programming for SSPs

- Bellman equation gives set of equations that describes expected cost for each state
- there are $|S|$ variables and $|S|$ equations (assuming $Q_\star$ is replaced in $V_\star$ with corresponding equation)
- If we solve these equations, we have solved the SSP
- Problem: how can we deal with the minimization?

$\Rightarrow$ We have solved the "same" problem before with the help of an LP solver

Reminder: LP for Shortest Path in State Space

Variables

Non-negative variable Distance_s for each state s

Objective

Maximize Distance_{s_0}

Subject to

$$
\begin{aligned}
\text{Distance}_{s_\star} &= 0 && \text{for all goal states } s_\star \\
\text{Distance}_s &\leq \text{Distance}_{s'} + c(\ell) && \text{for all transitions } s \xrightarrow{\ell} s'
\end{aligned}
$$

LP for Expected Cost in SSP

Variables

Non-negative variable ExpCost_s for each state s

Objective

Maximize ExpCost_{s_0}

Subject to

$$\text{ExpCost}_{s_\star} = 0 \quad \text{for all goal states } s_\star$$

$$\text{ExpCost}_s \leq \Big(\sum_{s' \in S} T(s, \ell, s') \cdot \text{ExpCost}_{s'}\Big) + c(\ell)$$

$$\text{for all } s \in S \text{ and } \ell \in L(s)$$

LP for Expected Reward in MDP

Variables

Non-negative variable ExpReward_s for each state s

Objective

Minimize ExpReward_{s_0}

Subject to

$$\text{ExpReward}_s \geq (\gamma \cdot \sum_{s' \in S} T(s, \ell, s')\text{ExpReward}_{s'}) + R(s, \ell)$$

for all $s \in S$ and $\ell \in L(s)$

Complexity of Probabilistic Planning

- optimal solution for MDPs or SSPs can be computed with LP solver
- requires $|S|$ variables and $|S| \cdot |L|$ constraints
- we know that LPs can be solved in polynomial time
- $\Rightarrow$ solving MDPs or SSPs is a polynomial time problem

How does this relate to the complexity result for classical planning?

Complexity of Probabilistic Planning

- optimal solution for MDPs or SSPs can be computed with LP solver
- requires $|S|$ variables and $|S| \cdot |L|$ constraints
- we know that LPs can be solved in polynomial time
- $\Rightarrow$ solving MDPs or SSPs is a polynomial time problem

How does this relate to the complexity result for classical planning?

Solving MDPs or SSPs is polynomial in $|S| \cdot |L|$

Summary

Summary

- State-value of a policy describes expected reward (cost) of following that policy
- Related Bellman equation describes optimal policy
- Solution to Bellman equation gives optimal policy
- Linear Programming can be used to solve MDPs and SSPs in time polynomial in size of state space and actions