Robust Markov Decision Process

Beyond (and back to) Rectangularity

Vineet Goyal, Julien Grand-Clement

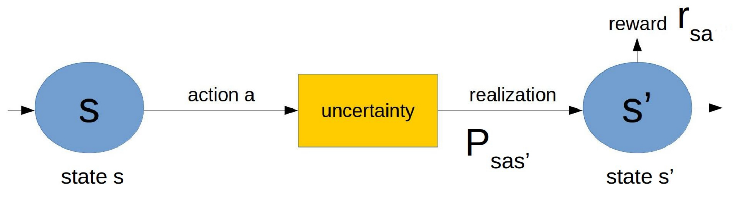
IEOR Department, Columbia University

What is this talk about?

Uncertainty in transition kernels of Markov Chains.

Geometry of the uncertainty set and tractability.

Markov Decision Process (MDP)



MDP: states $\mathbb{S}$, actions $\mathbb{A}$, reward (r_{sa}), transition probabilities $(P_{sas'})$.

Decision Maker:

Policy π : Being in state s, what action(s) should I choose?

Goal: Given transition probabilities $\boldsymbol{P} = (P_{sas'})$,

$$\text{solve } \max_{\pi \in \Pi} R(\pi, \boldsymbol{P}), \text{ where } R(\pi, \boldsymbol{P}) = E^{\pi, \boldsymbol{P}} \left[\sum_{t=0}^{\infty} \lambda^t r_{s_t a_t} \,\middle|\, \boldsymbol{s}_0 = \boldsymbol{p}_0 \right].$$

Bellman (1957), Howard (1960); Puterman (1994).

Main Results for MDP

At optimality: Stationary, Markovian, deterministic optimal policy

$$\pi^* : \mathbb{S} \to \mathbb{A}.$$

Efficiently solved:

- LP formulation, Policy Iteration (strongly polynomial: Ye (2011)).

Main Results for MDP

At optimality: Stationary, Markovian, deterministic optimal policy

$$\pi^* : \mathbb{S} \to \mathbb{A}.$$

Efficiently solved:

- LP formulation, Policy Iteration (strongly polynomial: Ye (2011)).
- Value Iteration:

Main Results for MDP

At optimality: Stationary, Markovian, deterministic optimal policy

$$\pi^* : \mathbb{S} \to \mathbb{A}.$$

Efficiently solved:

- LP formulation, Policy Iteration (strongly polynomial: Ye (2011)).
- Value Iteration:

  Define Value V_s = expected reward starting from state s.

Main Results for MDP

At optimality: Stationary, Markovian, deterministic optimal policy

$$\pi^* : \mathbb{S} \to \mathbb{A}.$$

Efficiently solved:

- LP formulation, Policy Iteration (strongly polynomial: Ye (2011)).
- Value Iteration:

  Define Value V_s = expected reward starting from state s.

  Bellman equation:

  $$V_s^* = F_s(\boldsymbol{V^*}, \boldsymbol{P}) = \max_a \{ r_{sa} + \lambda \cdot \boldsymbol{P_{sa}^T V^*} \}, \ \forall\, s.$$

Main Results for MDP

At optimality: Stationary, Markovian, deterministic optimal policy

$$\pi^* : \mathbb{S} \to \mathbb{A}.$$

Efficiently solved:

- LP formulation, Policy Iteration (strongly polynomial: Ye (2011)).
- Value Iteration:

  Define Value V_s = expected reward starting from state s.

  Bellman equation:

  $$V_s^* = F_s(\boldsymbol{V^*}, \boldsymbol{P}) = \max_a \{ \; r_{sa} + \lambda \cdot \boldsymbol{P_{sa}^T V^*} \; \}, \; \forall \; s.$$

  Fixed point of a contraction: $\lambda < 1$.

Main Results for MDP

At optimality: Stationary, Markovian, deterministic optimal policy

$$\pi^* : \mathbb{S} \to \mathbb{A}.$$

Efficiently solved:

- LP formulation, Policy Iteration (strongly polynomial: Ye (2011)).
- Value Iteration:

  Define Value V_s = expected reward starting from state s.

  Bellman equation:

  $$V_s^* = F_s(\boldsymbol{V^*}, \boldsymbol{P}) = \max_a \{ \, r_{sa} + \lambda \cdot \boldsymbol{P_{sa}^T} \boldsymbol{V^*} \, \}, \; \forall \, s.$$

  Fixed point of a contraction: $\lambda < 1$.

  $\to$ Iterate:

  $$V_s^{k+1} = F_s(\boldsymbol{V^k}, \boldsymbol{P}), \; \forall \, s; \; \boldsymbol{V^0} \in \mathbb{R}_+^S.$$

Uncertainty in MDP

- Intrinsic uncertainty: Variance of the expected reward ($\rightarrow$ risk).

Uncertainty in MDP

- Intrinsic uncertainty: Variance of the expected reward ($\rightarrow$ risk).
- **External uncertainty:** Parameters are estimated from the data.

Uncertainty in MDP

- Intrinsic uncertainty: Variance of the expected reward ($\rightarrow$ risk).
- **External uncertainty:** Parameters are estimated from the data.
- Optimal policy π^{nom} for nominal transition matrices:

Uncertainty in MDP

- Intrinsic uncertainty: Variance of the expected reward ($\rightarrow$ risk).
- **External uncertainty:** Parameters are estimated from the data.
- Optimal policy π^{nom} for nominal transition matrices:

  may severely deteriorate for different but close matrices.

Uncertainty in MDP

- Intrinsic uncertainty: Variance of the expected reward ($\rightarrow$ risk).
- **External uncertainty:** Parameters are estimated from the data.
- Optimal policy π^{nom} for nominal transition matrices:

  may severely deteriorate for different but close matrices.
- Robust optimization: $\boldsymbol{P} \in \mathbb{P}$.

Uncertainty in MDP

- Intrinsic uncertainty: Variance of the expected reward ($\rightarrow$ risk).
- **External uncertainty:** Parameters are estimated from the data.
- Optimal policy π^{nom} for nominal transition matrices:

  may severely deteriorate for different but close matrices.
- Robust optimization: $\boldsymbol{P} \in \mathbb{P}$.

  Decision maker chooses policy π,

Uncertainty in MDP

- Intrinsic uncertainty: Variance of the expected reward ($\to$ risk).
- **External uncertainty:** Parameters are estimated from the data.
- Optimal policy π^{nom} for nominal transition matrices:

  may severely deteriorate for different but close matrices.
- Robust optimization: $\boldsymbol{P} \in \mathbb{P}$.

  Decision maker chooses policy π,

  Nature chooses $\boldsymbol{P} \in \mathbb{P}$ adversarially,

Uncertainty in MDP

- Intrinsic uncertainty: Variance of the expected reward ($\to$ risk).
- **External uncertainty:** Parameters are estimated from the data.
- Optimal policy π^{nom} for nominal transition matrices:

  may severely deteriorate for different but close matrices.
- Robust optimization: $\boldsymbol{P} \in \mathbb{P}$.

  Decision maker chooses policy π,

  Nature chooses $\boldsymbol{P} \in \mathbb{P}$ adversarially,

  Decision maker obtains $R(\pi, \boldsymbol{P})$.

Uncertainty in MDP

- Intrinsic uncertainty: Variance of the expected reward ($\to$ risk).
- **External uncertainty:** Parameters are estimated from the data.
- Optimal policy π^{nom} for nominal transition matrices:

  may severely deteriorate for different but close matrices.
- Robust optimization: $\boldsymbol{P} \in \mathbb{P}$.

  Decision maker chooses policy π,

  Nature chooses $\boldsymbol{P} \in \mathbb{P}$ adversarially,

  Decision maker obtains $R(\pi, \boldsymbol{P})$.

Our goal

Solve $\max_{\pi \in \Pi} \min_{\boldsymbol{P} \in \mathbb{P}} R(\pi, \boldsymbol{P})$.

Related Works

- For general uncertainty set $\mathbb{P}$:

  "$\min_{P \in \mathbb{P}} R(\pi, P) \geq \gamma$?" is strongly NP-Hard (Wiesemann et al. (2012)).

Related Works

- For general uncertainty set $\mathbb{P}$:

  "$\min_{P \in \mathbb{P}} R(\pi, P) \geq \gamma$?" is strongly NP-Hard (Wiesemann et al. (2012)).

- Tractable for some special cases.
  Iyengar (2005), El Ghaoui, Nilim (2005):

  $$\text{(s,a)-rectangularity: } \mathbb{P} = \bigtimes_{s,a} \mathbb{P}_{sa},\ (P_{sas'})_{s'} \in \mathbb{R}_+^S.$$

  Wiesemann et al. (2012):

  $$\text{s-rectangularity: } \mathbb{P} = \bigtimes_{s} \mathbb{P}_{s},\ (P_{sas'})_{as'} \in \mathbb{R}_+^{S \times A}.$$

Related Works

- **For general uncertainty set $\mathbb{P}$:**

  "$\min_{P \in \mathbb{P}} R(\pi, P) \geq \gamma$?" is strongly NP-Hard (Wiesemann et al. (2012)).

- **Tractable** for some special cases.
  Iyengar (2005), El Ghaoui, Nilim (2005):

  $$\text{(s,a)-rectangularity: } \mathbb{P} = \times_{s,a} \mathbb{P}_{sa},\ (P_{sas'})_{s'} \in \mathbb{R}_+^S.$$

  Wiesemann et al. (2012):

  $$\text{s-rectangularity: } \mathbb{P} = \times_{s} \mathbb{P}_{s},\ (P_{sas'})_{as'} \in \mathbb{R}_+^{S \times A}.$$

- Cartesian product $X \times Y$: no constraints across components.

What does Rectangularity mean?

What does Rectangularity mean?

- In MDP: transition probabilities P_{sa} are **unrelated.**

What does Rectangularity mean?

- In MDP: transition probabilities P_{sa} are **unrelated.**
- *This is **too conversative** in some applications.*

What does Rectangularity mean?

- In MDP: transition probabilities P_{sa} are **unrelated.**
- *This is **too conversative** in some applications.*
- **Healthcare:** Drug effect related to the dose.

What does Rectangularity mean?

- In MDP: transition probabilities P_{sa} are **unrelated.**
- *This is **too conversative** in some applications.*
- **Healthcare:** Drug effect related to the dose.

  different health states, same treatment $\rightarrow$ related effect, ie,

$$P_{s_1 a} \text{ and } P_{s_2 a} \text{ might be related.}$$

Factor matrix (FM) uncertainty set

Idea: Transitions depend of common underlying vectors.

$$P_{sa} = \sum_{i=1}^{r} u_{sa}^{i} w^{i},$$

$$\sum_{i=1}^{r} u_{sa}^{i} = 1, \forall\, (s, a),$$

$$(w_1, ..., w_r) = W \in \mathbb{W}.$$

Factor matrix (FM) uncertainty set

Generality of FM uncertainty set

For r $= S \times A$, this captures **any** uncertainty set $\mathbb{P}$.

"$\min_{P \in \mathbb{P}} R(\pi, P) \geq \gamma$?" is strongly NP-Hard (Wiesemann et al. (2012)).

We assume r-rectangularity:

$$\mathbb{W} = \bigtimes_{i=1,..,r} \mathbb{W}^i.$$

Idea: Coupled $\mathbb{P}$ = simple transformation of uncoupled $\mathbb{W}$.

Think about SVD / Nonnegative Matrix Factorization.

Sum-up so far

In a nutshell:

> Rectangularity of $\mathbb{P}$ leads to tractability.

Our theoretical results:

Sum-up so far

In a nutshell:

- Rectangularity of $\mathbb{P}$ leads to tractability.
- But too conservative for our applications (healthcare).

Our theoretical results:

Sum-up so far

In a nutshell:

- Rectangularity of $\mathbb{P}$ leads to tractability.
- But too conservative for our applications (healthcare).
- FM uncertainty set: captures relations across transitions.

Our theoretical results:

Sum-up so far

In a nutshell:

- Rectangularity of $\mathbb{P}$ leads to tractability.
- But too conservative for our applications (healthcare).
- FM uncertainty set: captures relations across transitions.

Our theoretical results:

- Assuming r-rectangular $\mathbb{W}$:

Sum-up so far

In a nutshell:

Rectangularity of $\mathbb{P}$ leads to tractability.

But too conservative for our applications (healthcare).

FM uncertainty set: captures relations across transitions.

Our theoretical results:

Assuming r-rectangular $\mathbb{W}$:

- Given π, we solve $\min_{W \in \mathbb{W}} R(\pi, W)$ (also in Goh et al. (2014)).

Sum-up so far

In a nutshell:

Rectangularity of $\mathbb{P}$ leads to tractability.

But too conservative for our applications (healthcare).

FM uncertainty set: captures relations across transitions.

Our theoretical results:

Assuming r-rectangular $\mathbb{W}$:

- Given π, we solve $\min_{W \in \mathbb{W}} R(\pi, W)$ (also in Goh et al. (2014)).
- We solve $\max_{\pi \in \Pi} \min_{W \in \mathbb{W}} R(\pi, W)$.

Sum-up so far

In a nutshell:

Rectangularity of $\mathbb{P}$ leads to tractability.

But too conservative for our applications (healthcare).

FM uncertainty set: captures relations across transitions.

Our theoretical results:

Assuming r-rectangular $\mathbb{W}$:

- Given π, we solve $\min_{W \in \mathbb{W}} R(\pi, W)$ (also in Goh et al. (2014)).
- We solve $\max_{\pi \in \Pi} \min_{W \in \mathbb{W}} R(\pi, W)$.
- We prove a structural result:

$$\max_{\pi \in \Pi} \min_{W \in \mathbb{W}} R(\pi, W) = \min_{W \in \mathbb{W}} \max_{\pi \in \Pi} R(\pi, W).$$

Sum-up so far

In a nutshell:

Rectangularity of $\mathbb{P}$ leads to tractability.

But too conservative for our applications (healthcare).

FM uncertainty set: captures relations across transitions.

Our theoretical results:

Assuming r-rectangular $\mathbb{W}$:

- Given π, we solve $\min_{W \in \mathbb{W}} R(\pi, W)$ (also in Goh et al. (2014)).
- We solve $\max_{\pi \in \Pi} \min_{W \in \mathbb{W}} R(\pi, W)$.
- We prove a structural result:

$$\max_{\pi \in \Pi} \min_{W \in \mathbb{W}} R(\pi, W) = \min_{W \in \mathbb{W}} \max_{\pi \in \Pi} R(\pi, W).$$

- We extend classical properties of MDPs (Maximum principle, Blackwell optimality).

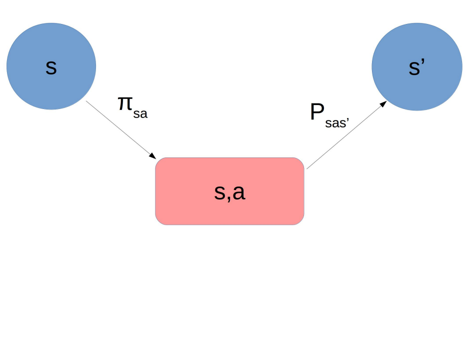
Factor matrix (FM) uncertainty set

Theorem 1: Policy Evaluation for rectangular $\mathbb{W}$

Consider a fixed policy π and $\mathbb{W}$ r-rectangular.

$\min_{W \in \mathbb{W}} R(\pi, W)$ can be casted as an MDP over $\{1, \ldots, r\}$.

Why is that the case?



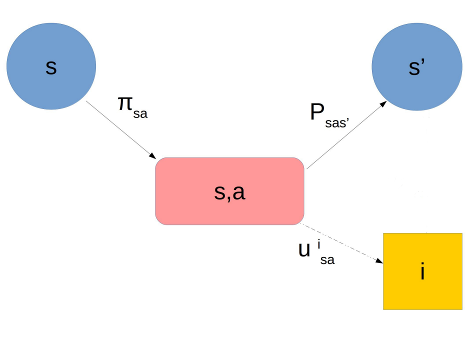
Factor matrix (FM) uncertainty set

Theorem 1: Policy Evaluation for rectangular $\mathbb{W}$

Consider a fixed policy π and $\mathbb{W}$ r-rectangular.

$\min_{W \in \mathbb{W}} R(\pi, W)$ can be casted as an MDP over $\{1, \ldots, r\}$.

Why is that the case?



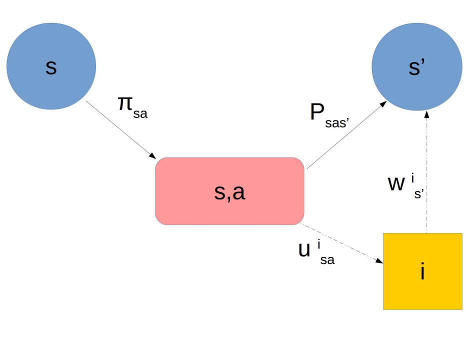
Factor matrix (FM) uncertainty set

Theorem 1: Policy Evaluation for rectangular $\mathbb{W}$

Consider a fixed policy π and $\mathbb{W}$ r-rectangular.

$\min_{W \in \mathbb{W}} R(\pi, W)$ can be casted as an MDP over $\{1, \ldots, r\}$.

Why is that the case?



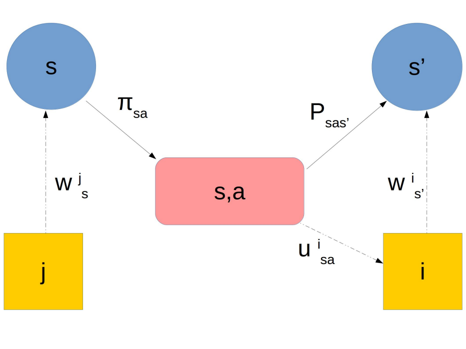
Factor matrix (FM) uncertainty set

Theorem 1: Policy Evaluation for rectangular $\mathbb{W}$

Consider a fixed policy π and $\mathbb{W}$ r-rectangular.

$\min_{W \in \mathbb{W}} R(\pi, W)$ can be casted as an MDP over $\{1, \ldots, r\}$.

Why is that the case?



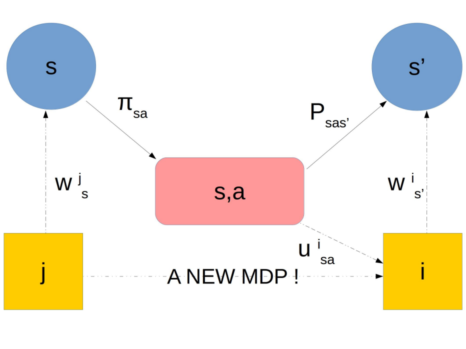
Factor matrix (FM) uncertainty set

Theorem 1: Policy Evaluation for rectangular $\mathbb{W}$

Consider a fixed policy π and $\mathbb{W}$ r-rectangular.

$\min_{W \in \mathbb{W}} R(\pi, W)$ can be casted as an MDP over $\{1, \ldots, r\}$.

Why is that the case?



Sketch: solve $\min_{W \in \mathbb{W}} R(\pi, W)$

- Define an underlying MDP on $i \in \{1, ..., r\}$.

Sketch: solve $\min_{W \in \mathbb{W}} R(\pi, W)$

- Define an underlying MDP on $i \in \{1, ..., r\}$.
- Write value vector and Bellman Recursion for this MDP:

Sketch: solve $\min_{W \in \mathbb{W}} R(\pi, W)$

- Define an underlying MDP on $i \in \{1, ..., r\}$.
- Write value vector and Bellman Recursion for this MDP:

  β_i = expected reward starting from a 'state' $i \in \{1, ..., r\}$.

Sketch: solve $\min_{W \in \mathbb{W}} R(\pi, W)$

- Define an underlying MDP on $i \in \{1, ..., r\}$.
- Write value vector and Bellman Recursion for this MDP:

  β_i = expected reward starting from a 'state' $i \in \{1, ..., r\}$.

$$\beta_i^* = H_i(\boldsymbol{\beta^*}, \pi) = \min_{\boldsymbol{w_i} \in \mathbb{W}^i} \{\boldsymbol{w_i}^T(\boldsymbol{r_\pi} + \lambda \cdot \boldsymbol{T_\pi} \boldsymbol{\beta^*})\}, \ \forall\, i.$$

Sketch: solve $\min_{W \in \mathbb{W}} R(\pi, W)$

- Define an underlying MDP on $i \in \{1, ..., r\}$.
- Write value vector and Bellman Recursion for this MDP:

  β_i = expected reward starting from a 'state' $i \in \{1, ..., r\}$.

$$\beta_i^* = H_i(\boldsymbol{\beta^*}, \pi) = \min_{\boldsymbol{w_i} \in \mathbb{W}^i} \{\boldsymbol{w_i}^T(\boldsymbol{r_\pi} + \lambda \cdot \boldsymbol{T_\pi} \boldsymbol{\beta^*})\}, \ \forall\ i.$$

Rectangularity $\Rightarrow$ feasibility of $(\boldsymbol{w_1^*}, ..., \boldsymbol{w_r^*}) \in \mathbb{W}$.

Sketch: solve $\min_{W \in \mathbb{W}} R(\pi, W)$

- Define an underlying MDP on $i \in \{1, ..., r\}$.
- Write value vector and Bellman Recursion for this MDP:

  β_i = expected reward starting from a 'state' $i \in \{1, ..., r\}$.

$$\beta_i^* = H_i(\boldsymbol{\beta^*}, \pi) = \min_{\boldsymbol{w_i} \in \mathbb{W}^i} \{\boldsymbol{w_i}^T(\boldsymbol{r_\pi} + \lambda \cdot \boldsymbol{T_\pi} \boldsymbol{\beta^*})\}, \ \forall\ i.$$

Rectangularity $\Rightarrow$ feasibility of $(\boldsymbol{w_1^*}, ..., \boldsymbol{w_r^*}) \in \mathbb{W}$.

Fixed point of a contraction ($\lambda < 1$), iterate:

$$\beta_i^{k+1} = H_i(\boldsymbol{\beta^k}, \pi), \ \forall\ i.$$

Sketch: solve $\min_{W \in \mathbb{W}} R(\pi, W)$

- Define an underlying MDP on $i \in \{1, ..., r\}$.
- Write value vector and Bellman Recursion for this MDP:

  β_i = expected reward starting from a 'state' $i \in \{1, ..., r\}$.

  $$\beta_i^* = H_i(\boldsymbol{\beta^*}, \pi) = \min_{\boldsymbol{w_i} \in \mathbb{W}^i} \{\boldsymbol{w_i}^T(\boldsymbol{r_\pi} + \lambda \cdot \boldsymbol{T_\pi \beta^*})\}, \ \forall\ i.$$

  Rectangularity $\Rightarrow$ feasibility of $(\boldsymbol{w_1^*}, ..., \boldsymbol{w_r^*}) \in \mathbb{W}$.

  Fixed point of a contraction ($\lambda < 1$), iterate:

  $$\beta_i^{k+1} = H_i(\boldsymbol{\beta^k}, \pi), \ \forall\ i.$$

- Also LP formulation, Policy Iteration.

Our results

Theorem 1: Max-min duality for rectangular $\mathbb{W}$

$$\max_{\pi \in \Pi} \min_{W \in \mathbb{W}} R(\pi, W) = \min_{W \in \mathbb{W}} \max_{\pi \in \Pi} R(\pi, W).$$

Our results

Theorem 1: Max-min duality for rectangular $\mathbb{W}$

$$\max_{\pi \in \Pi} \min_{W \in \mathbb{W}} R(\pi, W) = \min_{W \in \mathbb{W}} \max_{\pi \in \Pi} R(\pi, W).$$

- Formulation of $\min_{W \in \mathbb{W}} R(\pi, W)$ as an MDP over $\{1, \ldots, r\}$.

Our results

Theorem 1: Max-min duality for rectangular $\mathbb{W}$

$$\max_{\pi \in \Pi} \min_{W \in \mathbb{W}} R(\pi, W) = \min_{W \in \mathbb{W}} \max_{\pi \in \Pi} R(\pi, W).$$

- Formulation of $\min_{W \in \mathbb{W}} R(\pi, W)$ as an MDP over $\{1, \ldots, r\}$.

Lemma

There exists a deterministic robust optimal policy π^*.

Our results

Theorem 1: Max-min duality for rectangular $\mathbb{W}$

$$\max_{\pi \in \Pi} \min_{W \in \mathbb{W}} R(\pi, W) = \min_{W \in \mathbb{W}} \max_{\pi \in \Pi} R(\pi, W).$$

- Formulation of $\min_{W \in \mathbb{W}} R(\pi, W)$ as an MDP over $\{1, \ldots, r\}$.

Lemma

There exists a deterministic robust optimal policy π^*.

- Interpretability, implementability.

Our results

Theorem 1: Max-min duality for rectangular $\mathbb{W}$

$$\max_{\pi \in \Pi} \min_{W \in \mathbb{W}} R(\pi, W) = \min_{W \in \mathbb{W}} \max_{\pi \in \Pi} R(\pi, W).$$

- Formulation of $\min_{W \in \mathbb{W}} R(\pi, W)$ as an MDP over $\{1, \ldots, r\}$.

Lemma

There exists a deterministic robust optimal policy π^*.

- Interpretability, implementability.

Key Lemma

$$W^* \in \arg\min_{W \in \mathbb{W}} R(\pi^*, W),$$

$$\pi^* \in \arg\max_{\pi \in \Pi} R(\pi, W^*).$$

Our results

Assume r-Rectangularity:

$$(w_1, ..., w_r) \in \bigtimes_{i=1,..,r} \mathbb{W}^i.$$

Our results

Assume r-Rectangularity:

$$(w_1, \ldots, w_r) \in \bigtimes_{i=1,..,r} \mathbb{W}^i.$$

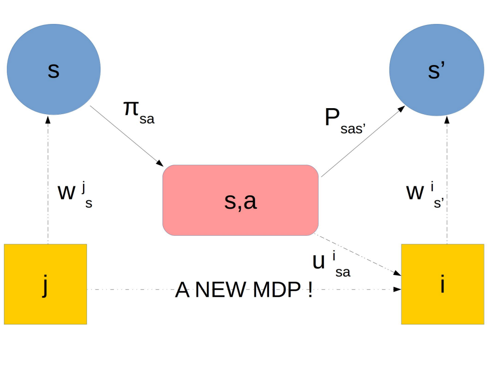
Theorem 2: Robust value iteration

Robust Value Iteration obtains (π^*, W^*) in polynomial time with

$$W^* \in \arg\min_{W \in \mathbb{W}} R(\pi^*, W),$$

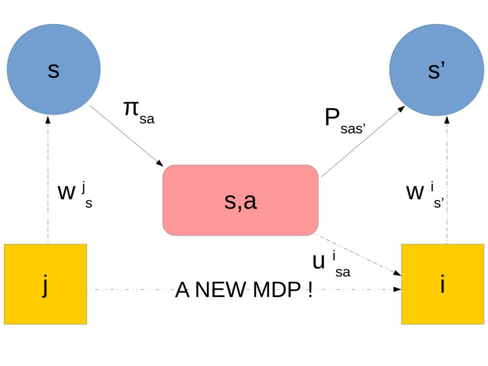
$$\pi^* \in \arg\max_{\pi \in \Pi} R(\pi, W^*).$$

Sketch: solve $\max_{\pi \in \Pi} \min_{W \in \mathbb{W}} R(\pi, W)$



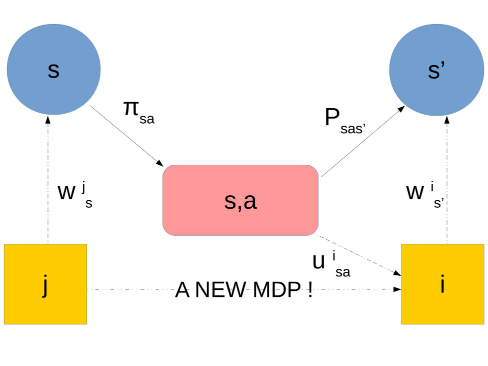
- Two competing MDP, with rectangular sets Π and $\mathbb{W}$.

Sketch: solve $\max_{\pi \in \Pi} \min_{W \in \mathbb{W}} R(\pi, W)$



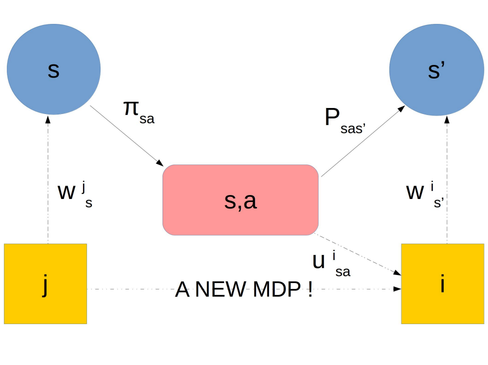
- Two competing MDP, with rectangular sets Π and $\mathbb{W}$.
- Compete for the same reward $R(\pi, W)$.

Sketch: solve $\max_{\pi\in\Pi} \min_{W\in\mathbb{W}} R(\pi, W)$



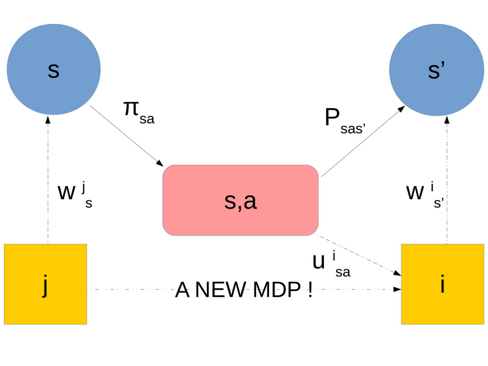
- Two competing MDP, with rectangular sets Π and $\mathbb{W}$.
- Compete for the same reward $R(\pi, W)$.
- Max-min Duality gives two interwoven Bellman Equations.

Sketch: solve $\max_{\pi \in \Pi} \min_{W \in \mathbb{W}} R(\pi, W)$



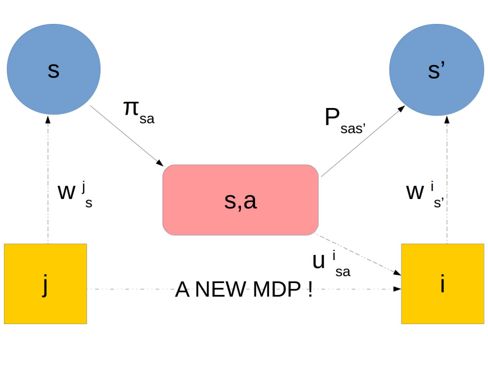
- Two competing MDP, with rectangular sets Π and $\mathbb{W}$.
- Compete for the same reward $R(\pi, W)$.
- Max-min Duality gives two interwoven Bellman Equations.
- Write two contractions mapping.

Sketch: solve $\max_{\pi \in \Pi} \min_{W \in \mathbb{W}} R(\pi, W)$



- Two competing MDP, with rectangular sets Π and $\mathbb{W}$.
- Compete for the same reward $R(\pi, W)$.
- Max-min Duality gives two interwoven Bellman Equations.
- Write two contractions mapping.
- Their fixed points are the solutions to our max-min problem.

Sketch: solve $\max_{\pi \in \Pi} \min_{W \in \mathbb{W}} R(\pi, W)$



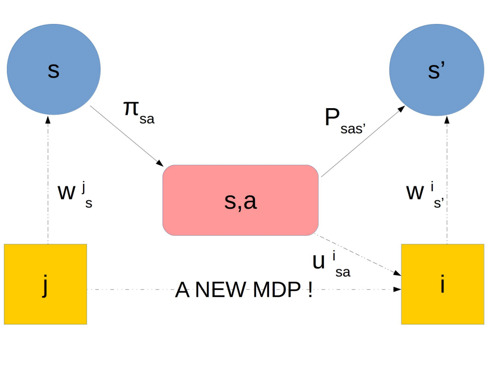
Decision Maker:

$$\pi^* \in \arg\max_{\pi \in \Pi} R(\pi, W^*).$$

Adversary:

$$W^* \in \arg\min_{W \in \mathbb{W}} R(\pi^*, W).$$

Sketch: solve $\max_{\pi \in \Pi} \min_{W \in \mathbb{W}} R(\pi, W)$



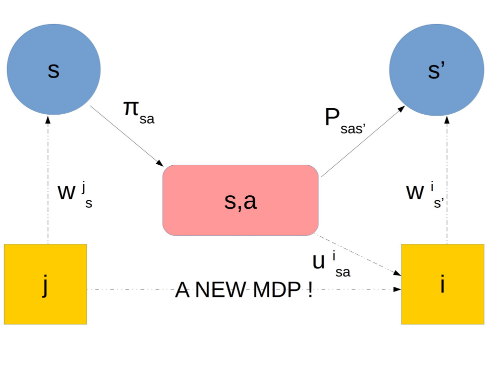
Decision Maker:

$$V_s^* = F_s(V^*, W^*), \ \forall \, s.$$

Adversary:

$$W^* \in \arg\min_{W \in \mathbb{W}} R(\pi^*, W).$$

Sketch: solve $\max_{\pi \in \Pi} \min_{W \in \mathbb{W}} R(\pi, W)$



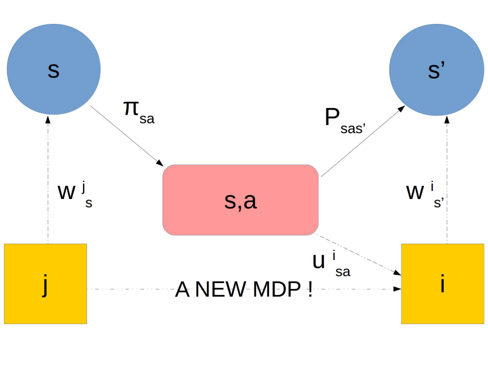
Decision Maker:

$$V_s^* = F_s(\boldsymbol{V}^*, \boldsymbol{W}^*),\ \forall\, s.$$

Adversary:

$$\beta_i^* = H_i(\boldsymbol{\beta}^*, \pi^*),\ \forall\, i.$$

Sketch: solve $\max_{\pi \in \Pi} \min_{W \in \mathbb{W}} R(\pi, W)$



Using $\boldsymbol{W}^{*T}\boldsymbol{V}^* = \boldsymbol{\beta}^*$,

$$V_s^* = f_s(\boldsymbol{V}^*), \; \forall \, s,$$
$$\beta_i^* = h_i(\boldsymbol{\beta}^*), \; \forall \, i,$$

and $f : \mathbb{R}^S \to \mathbb{R}^S, h : \mathbb{R}^r \to \mathbb{R}^r$ are contractions.

Some useful properties

Maximum principle: in classical MDP, $V^{\pi^*} \geq V^{\pi}$, for all policy π.

Robust Maximum principle

$V^{\pi^*, W^*} \geq V^{\pi, W}$, for all policy π, for all $W \in \arg\min_{W \in \mathbb{W}} R(\pi, W)$.

Blackwell optimality: the optimal nominal policy remains optimal when $\lambda \to 1$.

Robust Blackwell optimality

There exists (π^*, W^*) and $\lambda_0 \in (0, 1)$, such that for all λ in $(\lambda_0, 1)$,

$$(\pi^*, W^*) \in \arg\max_{\pi \in \Pi} \min_{W \in \mathbb{W}} R(\pi, W, \lambda).$$

Application: how to use FM uncertainty set?

- use of Non Negative Matrix Factorization (NMF).
- use of Robust maximum principle.

Use of Non Negative Matrix Factorization (NMF)

NMF $\iff$ non-negative SVD:

$$\text{given } A \geq \mathbf{0}, \text{ minimize } \|A - BC\| \text{ with } B \geq \mathbf{0}, C \geq \mathbf{0}.$$

Suppose we have estimated $\hat{\boldsymbol{P}} = (\hat{P}_{sas'})$ with confidence $\tau > 0$.

$$\text{we want } \hat{\boldsymbol{P}}_{sa} = \sum_{i=1}^{r} u_{sa}^{i} \boldsymbol{w}_i.$$

Therefore, for $\hat{\boldsymbol{P}} = (\hat{\boldsymbol{P}}_{s_1 a_1}, ..., \hat{\boldsymbol{P}}_{SA})$, we solve the NMF program

$$\min_{W \in H_1, \boldsymbol{u} \in H_2} \|\hat{\boldsymbol{P}} - W\boldsymbol{u}\|.$$

r-rectangularity: obtaining $(\hat{W}, \hat{\boldsymbol{u}})$, budget of uncertainty sets:

$$\mathbb{W}^i = \{\boldsymbol{w}_i = \hat{\boldsymbol{w}}_i + \boldsymbol{\delta} | \delta \in \mathbb{R}^S, \|\boldsymbol{\delta}\|_1 \leq \sqrt{S} \cdot \tau, \|\boldsymbol{\delta}\|_\infty \leq \tau, \boldsymbol{e}_S^\top \boldsymbol{w}_i = 1, \boldsymbol{w}_i \geq \mathbf{0}\}, \forall\, i.$$

Example: hospital management (joint with C. Chan and V. Goyal)

$300,000$ 'patients histories' at Kaiser Permanente hospitals:

Patients arrive in hospital ward with risk score $s \in \{1, ..., n\}$.
Risk score updated every six hours: $s \to s'$, Markov chain $\boldsymbol{T}$.

Question: proactive transfer of patients to the emergency room?

- Pros: better outcome when treated in emergency room.

Example: hospital management (joint with C. Chan and V. Goyal)

$300,000$ 'patients histories' at Kaiser Permanente hospitals:

Patients arrive in hospital ward with risk score $s \in \{1, ..., n\}$.
Risk score updated every six hours: $s \to s'$, Markov chain $\boldsymbol{T}$.

Question: proactive transfer of patients to the emergency room?

- Pros: better outcome when treated in emergency room.
- Cons: crowded emergency room, patients get 'bumped'.

Example: hospital management (joint with C. Chan and V. Goyal)

$300,000$ 'patients histories' at Kaiser Permanente hospitals:

Patients arrive in hospital ward with risk score $s \in \{1, ..., n\}$.
Risk score updated every six hours: $s \to s'$, Markov chain $\boldsymbol{T}$.

Question: proactive transfer of patients to the emergency room?

- Pros: better outcome when treated in emergency room.
- Cons: crowded emergency room, patients get 'bumped'.
- Build an MDP (action: send or not the patient at risk score s).

  Using robust maximum principle, we prove

Proposition: structure of optimal policies.

The optimal nominal policy π^{nom} is threshold.

The optimal robust policy π^{rob} is threshold, and $thr(\pi^{\text{rob}}) \leq thr(\pi^{\text{nom}})$.

Example: hospital management (joint with C. Chan and V. Goyal)

300,000 ‘patients histories’ at Kaiser Permanente hospitals:

Patients arrive in hospital ward with risk score $s \in \{1, ..., n\}$.
Risk score updated every six hours: $s \to s'$, Markov chain $\boldsymbol{T}$.

Question: proactive transfer of patients to the emergency room?

- Pros: better outcome when treated in emergency room.
- Cons: crowded emergency room, patients get ‘bumped’.
- Build an MDP (action: send or not the patient at risk score s).

Using robust maximum principle, we prove

Proposition: structure of optimal policies.

The optimal nominal policy π^{nom} is threshold.

The optimal robust policy π^{rob} is threshold, and
$thr(\pi^{\text{rob}}) \leq thr(\pi^{\text{nom}})$.

Example: π^{nom} admits top 8% patients, π^{rob} admits top 27.1%.

Conclusion and future work

Our contributions:

Assuming $\mathbb{W}$ is rectangular:

- Efficient algorithm for $\min_{W \in \mathbb{W}} R(\pi, W)$.
- **Robust Value Iteration for** $\max_{\pi \in \Pi} \min_{W \in \mathbb{W}} R(\pi, W)$.
- Equilibrium result:

$$\exists\ (\pi^*, W^*),\ W^* \in \min_{W \in \mathbb{W}} R(\pi^*, W),\ \pi^* \in \max_{\pi \in \Pi} R(\pi, W^*).$$

$\Rightarrow$ strong duality + deterministic optimal robust policy.

- Robust maximum principle, robust Blackwell optimality.

Future work:

- Robust MDP with long-run average reward.

Conclusion and future work

Our contributions:

Assuming $\mathbb{W}$ is rectangular:

- Efficient algorithm for $\min_{W \in \mathbb{W}} R(\pi, W)$.
- **Robust Value Iteration for $\max_{\pi \in \Pi} \min_{W \in \mathbb{W}} R(\pi, W)$.**
- Equilibrium result:

$$\exists\ (\pi^*, W^*),\ W^* \in \min_{W \in \mathbb{W}} R(\pi^*, W),\ \pi^* \in \max_{\pi \in \Pi} R(\pi, W^*).$$

$\Rightarrow$ strong duality + deterministic optimal robust policy.

- Robust maximum principle, robust Blackwell optimality.

Future work:

- Robust MDP with long-run average reward.
- Fixed parameter r + non-rectangular $\mathbb{W}$.

References

J. Goh, B. Mohsen, S.A. Zenios, S. Sundeep, and D. Moore. Data uncertainty in markov chains: Application to cost-effectiveness analyses of medical innovations. Operations Research (forthcoming), 2014.

G. Iyengar. Robust dynamic programming. Mathematics of Operations Research, 30(2):257–280, 2005.

A. Nilim and L. El Ghaoui. Robust control of markov decision processes with uncertain transition probabilities. Operations Research, 53(5):780–798, 2005.

W. Wiesemann, D. Kuhn, and B. Rustem. Robust markov decision processes. Operations Research, 38(1):153–183, 2013.

Thank you! Any questions?

jg3728@columbia.edu