On the interplay between average and discount optimality in robust MDPs

Julien Grand-Clément (HEC Paris)
Marek Petrik (University of New Hampshire)
Nicolas Vieille (HEC Paris)

ICCOPT 2025

This talk in one slide

Main objective:

Solve (robust) MDPs with average return

Why it's interesting?

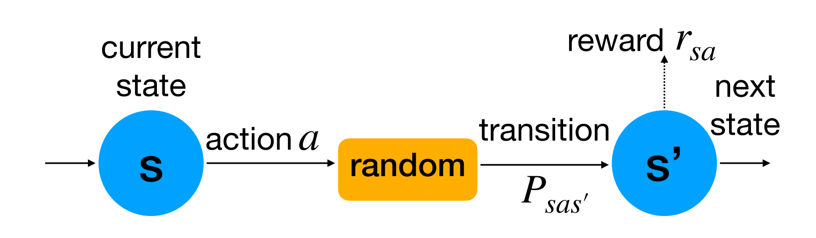
Well-studied for MDPs and stochastic games...

... largely understudied for robust MDPs

Main results:

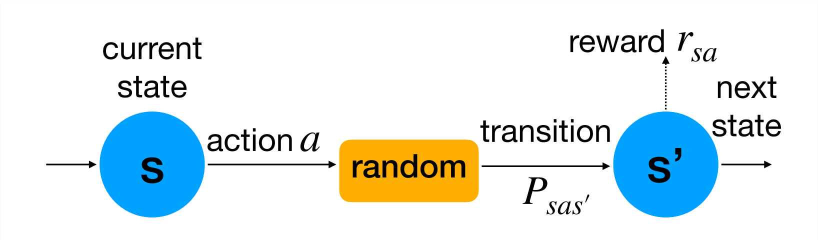
1. Properties of average optimal policies for robust MDPs
2. Computing average opt. policies by solving discounted problems

Setup for robust Markov decision process



- Finite set of states and actions
- History-dependent policy $\pi \in \Pi_{\mathrm{H}}$: maps finite histories to actions
- Transition probabilities $\boldsymbol{P} = (P_{sas'})$, unknown: $\boldsymbol{P} \in \mathcal{U}$

Setup for robust Markov decision process



- Finite set of states and actions
- History-dependent policy $\pi \in \Pi_{\mathrm{H}}$: maps finite histories to actions
- Transition probabilities $\boldsymbol{P} = (P_{sas'})$, unknown: $\boldsymbol{P} \in \mathcal{U}$
  This talk: $\mathcal{U}$ convex compact, sa-rectangular:

$$\mathcal{U} = \times_{(s,a)\in\mathcal{S}\times\mathcal{A}} \mathcal{U}_{sa}, \quad \mathcal{U}_{sa} \subset \Delta(\mathcal{S})$$

Discounted and average returns

Given a policy $\pi \in \Pi_S$ and some transitions $\boldsymbol{P} \in \mathcal{U}$:

Discounted return: for a *discount factor* $\gamma \in [0, 1)$,

$$R_\gamma(\pi, \boldsymbol{P}) \quad = \quad (1-\gamma) \sum_{t=0}^{\infty} \gamma^t \cdot \mathbb{E}^{\pi, \boldsymbol{P}} \left[r_{s_t a_t} \right]$$

Discounted and average returns

Given a policy $\pi \in \Pi_S$ and some transitions $\boldsymbol{P} \in \mathcal{U}$:

Discounted return: for a *discount factor* $\gamma \in [0, 1)$,

$$R_\gamma(\pi, \boldsymbol{P}) \quad = \quad (1-\gamma) \sum_{t=0}^{\infty} \gamma^t \cdot \mathbb{E}^{\pi, \boldsymbol{P}} \left[r_{s_t a_t} \right]$$

Average return:

$$R_{\mathsf{AVG}}(\pi, \boldsymbol{P}) \quad = \quad \lim_{T \to +\infty} \frac{1}{T+1} \sum_{t=0}^{T} \mathbb{E}^{\pi, \boldsymbol{P}} \left[r_{s_t a_t} \right]$$

Discounted and average returns

Given a policy $\pi \in \Pi_S$ and some transitions $\boldsymbol{P} \in \mathcal{U}$:

Discounted return: for a *discount factor* $\gamma \in [0, 1)$,

$$R_\gamma(\pi, \boldsymbol{P}) \quad = \quad (1-\gamma) \sum_{t=0}^{\infty} \gamma^t \cdot \mathbb{E}^{\pi, \boldsymbol{P}} \left[r_{s_t a_t} \right]$$

Average return:

$$R_{\mathsf{AVG}}(\pi, \boldsymbol{P}) \quad = \quad \lim_{T \to +\infty} \frac{1}{T+1} \sum_{t=0}^{T} \mathbb{E}^{\pi, \boldsymbol{P}} \left[r_{s_t a_t} \right]$$

Hardy-Littlewood: $\lim_{\gamma \to 1} R_\gamma(\pi, \boldsymbol{P}) = R_{\mathsf{AVG}}(\pi, \boldsymbol{P})$

Discounted and average returns

Given a policy $\pi \in \Pi_S$ and some transitions $\boldsymbol{P} \in \mathcal{U}$:

Discounted return: for a *discount factor* $\gamma \in [0, 1)$,

$$R_\gamma(\pi, \boldsymbol{P}) \quad = \quad (1-\gamma) \sum_{t=0}^{\infty} \gamma^t \cdot \mathbb{E}^{\pi, \boldsymbol{P}} \left[r_{s_t a_t} \right]$$

Average return:

$$R_{\mathsf{AVG}}(\pi, \boldsymbol{P}) \quad = \quad \lim_{T \to +\infty} \frac{1}{T+1} \sum_{t=0}^{T} \mathbb{E}^{\pi, \boldsymbol{P}} \left[r_{s_t a_t} \right]$$

Hardy-Littlewood: $\lim_{\gamma \to 1} R_\gamma(\pi, \boldsymbol{P}) = R_{\mathsf{AVG}}(\pi, \boldsymbol{P})$

Blackwell [Bla62]: for $\gamma \to 1$, discount opt. policies are average opt.

Main objective in this talk: Find a policy π solving

$$\sup_{\pi \in \Pi_{\mathsf{H}}} \inf_{\boldsymbol{P} \in \mathcal{U}} R_{\mathsf{AVG}}(\pi, \boldsymbol{P}) \tag{1}$$

Main objective in this talk: Find a policy π solving

$$\sup_{\pi \in \Pi_{\mathsf{H}}} \inf_{\boldsymbol{P} \in \mathcal{U}} R_{\mathsf{AVG}}(\pi, \boldsymbol{P}) \tag{1}$$

Main difficulties: R_{AVG} is discontinuous, lim/sup/inf may not exist, Bellman operator not a contraction, no ergodic/unichain assumption...

Main objective in this talk: Find a policy π solving

$$\sup_{\pi \in \Pi_{\mathrm{H}}} \inf_{\boldsymbol{P} \in \mathcal{U}} R_{\mathrm{AVG}}(\pi, \boldsymbol{P}) \tag{1}$$

Main difficulties: R_{AVG} is discontinuous, lim/sup/inf may not exist, Bellman operator not a contraction, no ergodic/unichain assumption...

[GCPV23]: stationary deterministic optimal policies exist for (1).

How to compute average optimal policies?

Main objective in this talk: Find a policy π solving

$$\sup_{\pi \in \Pi_{\mathrm{H}}} \inf_{\boldsymbol{P} \in \mathcal{U}} R_{\mathrm{AVG}}(\pi, \boldsymbol{P}) \tag{1}$$

Main difficulties: R_{AVG} is discontinuous, lim/sup/inf may not exist, Bellman operator not a contraction, no ergodic/unichain assumption...

[GCPV23]: stationary deterministic optimal policies exist for (1).

How to compute average optimal policies?

Sketch of our approach:

- "Optimal discounted policies are average optimal for γ large enough"
- $\Rightarrow$ "let's just solve discounted models for γ large enough"

Definition: The Blackwell discount factor

The *Blackwell discount factor* $\gamma_{\text{bw}} \in [0, 1)$ is the smallest discount factor such that the set of stationary discount optimal policies does not change for all γ in $(\gamma_{\text{bw}}, 1)$.

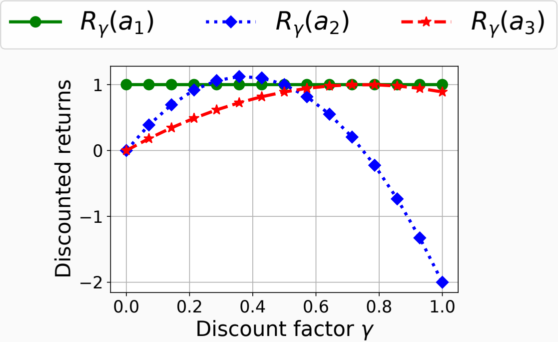


Figure 1: Example with three policies a_1, a_2, a_3

Definition: The Blackwell discount factor

The *Blackwell discount factor* $\gamma_{\text{bw}} \in [0, 1)$ is the smallest discount factor such that the set of stationary discount optimal policies does not change for all γ in $(\gamma_{\text{bw}}, 1)$.

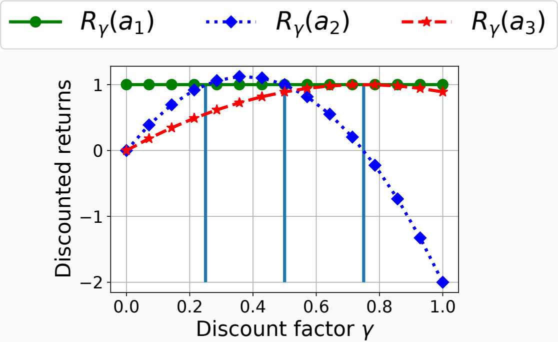


Figure 2: Example with three policies a_1, a_2, a_3

Definition: The Blackwell discount factor

The *Blackwell discount factor* $\gamma_{\mathsf{bw}} \in [0, 1)$ is the smallest discount factor such that the set of stationary discount optimal policies does not change for all γ in $(\gamma_{\mathsf{bw}}, 1)$.

Theorem 5 in [Bla62] [1]:

The Blackwell discount factor exists for finite MDPs.

Extension by [Sma66] for finite MDPs:

The interval $[0, 1)$ can be partitioned into finitely many subintervals, inside which the set of stationary discount optimal policies is constant.

[1] Blackwell proved something slightly weaker

Definition: The Blackwell discount factor

The *Blackwell discount factor* $\gamma_{\mathrm{bw}} \in [0,1)$ is the smallest discount factor such that the set of stationary discount optimal policies does not change for all γ in $(\gamma_{\mathrm{bw}}, 1)$.

Theorem 5 in [Bla62] [1]:

The Blackwell discount factor exists for finite MDPs.

Extension by [Sma66] for finite MDPs:

The interval $[0,1)$ can be partitioned into finitely many subintervals, inside which the set of stationary discount optimal policies is constant.

1. Upper bound on γ_{bw} for MDPs?
2. Existence and upper bounds for robust MDPs?

[1] Blackwell proved something slightly weaker

Bound on the Blackwell discount factor γ_{bw} for MDPs

$\gamma \mapsto R_\gamma(\pi)$ is a *rational function* of $\gamma \in [0, 1)$:

$$R_\gamma(\pi) = \frac{\text{poly}_1(\gamma)}{\text{poly}_2(\gamma)}$$

Bound on the Blackwell discount factor γ_{bw} for MDPs

$\gamma \mapsto R_\gamma(\pi)$ is a *rational function* of $\gamma \in [0,1)$:

$$R_\gamma(\pi) = \frac{\mathrm{poly}_1(\gamma)}{\mathrm{poly}_2(\gamma)}$$

$R_\gamma(\pi) = R_\gamma(\pi')$ is a polynomial equation in γ: $Q(\gamma) = 0$, and $Q(1) = 0$

Bound on the Blackwell discount factor γ_{bw} for MDPs

$\gamma \mapsto R_\gamma(\pi)$ is a *rational function* of $\gamma \in [0,1)$:

$$R_\gamma(\pi) = \frac{\text{poly}_1(\gamma)}{\text{poly}_2(\gamma)}$$

$R_\gamma(\pi) = R_\gamma(\pi')$ is a polynomial equation in γ: $Q(\gamma) = 0$, and $Q(1) = 0$

Root separation: [Lag69],[Had93],[Mah62],[Rum79]...

Bound on the Blackwell discount factor γ_{bw} for MDPs

$\gamma \mapsto R_\gamma(\pi)$ is a *rational function* of $\gamma \in [0,1)$:

$$R_\gamma(\pi) = \frac{\text{poly}_1(\gamma)}{\text{poly}_2(\gamma)}$$

$R_\gamma(\pi) = R_\gamma(\pi')$ is a polynomial equation in γ: $Q(\gamma) = 0$, and $Q(1) = 0$

Root separation: [Lag69],[Had93],[Mah62],[Rum79]...

Let $Q = \sum_{i=0}^{n} a_i X^i$ with $a_i \in \mathbb{Z}, \max_i |a_i| \leq H$ and $Q(1) = 0$.

$\exists$ $\text{SEP}(n, H) > 0$ such that $Q(x) \neq 0$ for $1 - \text{SEP}(n, H) < x < 1$.

Bound on the Blackwell discount factor γ_{bw} for MDPs

$\gamma \mapsto R_\gamma(\pi)$ is a *rational function* of $\gamma \in [0,1)$:

$$R_\gamma(\pi) = \frac{\text{poly}_1(\gamma)}{\text{poly}_2(\gamma)}$$

$R_\gamma(\pi) = R_\gamma(\pi')$ is a polynomial equation in γ: $Q(\gamma) = 0$, and $Q(1) = 0$

Root separation: [Lag69],[Had93],[Mah62],[Rum79]...

Let $Q = \sum_{i=0}^{n} a_i X^i$ with $a_i \in \mathbb{Z}, \max_i |a_i| \leq H$ and $Q(1) = 0$.

$\exists$ $\text{SEP}(n,H) > 0$ such that $Q(x) \neq 0$ for $1 - \text{SEP}(n,H) < x < 1$.

Key property for our application

For $\gamma > 1 - \text{SEP}(n,H)$, "discounted returns can't intersect!"

Bound on the Blackwell discount factor γ_{bw} for MDPs

$\gamma \mapsto R_\gamma(\pi)$ is a *rational function* of $\gamma \in [0,1)$:

$$R_\gamma(\pi) = \frac{\text{poly}_1(\gamma)}{\text{poly}_2(\gamma)}$$

$R_\gamma(\pi) = R_\gamma(\pi')$ is a polynomial equation in γ: $Q(\gamma) = 0$, and $Q(1) = 0$

Root separation: [Lag69],[Had93],[Mah62],[Rum79]...

Let $Q = \sum_{i=0}^{n} a_i X^i$ with $a_i \in \mathbb{Z}, \max_i |a_i| \leq H$ and $Q(1) = 0$.

$\exists$ $\text{SEP}(n, H) > 0$ such that $Q(x) \neq 0$ for $1 - \text{SEP}(n, H) < x < 1$.

Key property for our application

For $\gamma > 1 - \text{SEP}(n, H)$, "discounted returns can't intersect!"

$\Rightarrow \gamma_{\text{bw}} \leq 1 - \text{SEP}(n, H)$

Bound on the Blackwell discount factor γ_{bw} for MDPs

$\gamma \mapsto R_\gamma(\pi)$ is a *rational function* of $\gamma \in [0, 1)$:

$$R_\gamma(\pi) = \frac{\text{poly}_1(\gamma)}{\text{poly}_2(\gamma)}$$

$R_\gamma(\pi) = R_\gamma(\pi')$ is a polynomial equation in γ: $Q(\gamma) = 0$, and $Q(1) = 0$

Root separation: [Lag69],[Had93],[Mah62],[Rum79]...

Let $Q = \sum_{i=0}^{n} a_i X^i$ with $a_i \in \mathbb{Z}, \max_i |a_i| \leq H$ and $Q(1) = 0$.

$\exists$ $\text{SEP}(n, H) > 0$ such that $Q(x) \neq 0$ for $1 - \text{SEP}(n, H) < x < 1$.

Key property for our application

For $\gamma > 1 - \text{SEP}(n, H)$, "discounted returns can't intersect!"

$\Rightarrow \gamma_{bw} \leq 1 - \text{SEP}(n, H)$

Remains to bound degree/height of Q $\rightarrow$ use "closed-form" for $R_\gamma(\pi)$

Main bound on the Blackwell discount factor γ_{bw} for MDPs

Theorem [GCP24]

Consider a finite MDP instance with:

- M = maximum rewards and common denominator for transitions
- S = number of states

Then

$$1 - \gamma_{\mathrm{bw}} \geq \Omega\left(\frac{1}{(2M)^{S^2}}\right).$$

Main bound on the Blackwell discount factor γ_{bw} for MDPs

Theorem [GCP24]

Consider a finite MDP instance with:

- M = maximum rewards and common denominator for transitions
- S = number of states

Then

$$1 - \gamma_{bw} \geq \Omega\left(\frac{1}{(2M)^{S^2}}\right).$$

Note 1: no assumption on MDP instance (unichain, mixing time, etc.)!

Main bound on the Blackwell discount factor γ_{bw} for MDPs

Theorem [GCP24]

Consider a finite MDP instance with:

- M = maximum rewards and common denominator for transitions
- S = number of states

Then

$$1 - \gamma_{\mathrm{bw}} \geq \Omega\left(\frac{1}{(2M)^{S^2}}\right).$$

Note 1: no assumption on MDP instance (unichain, mixing time, etc.)!

Note 2: MDPs can be solved in $\tilde{O}\left(|\log(1-\gamma)|\right)$ [Ye05]

$\Rightarrow$ weakly-polytime algorithms for computing average optimal policies.

The case of robust MDPs

What about robust MDPs?

The Blackwell discount factor exists γ_{bw} for $\mathcal{U}$ sa-rec. AND:

- [TB07]: based on ℓ_∞-ball
- [GGC22]: $\mathcal{U}$ polytope
- [WVA$^+$23]: unichain assumption + average optimal policy unique.

Q: Existence of γ_{bw} for general sa-rectangular, compact convex $\mathcal{U}$?

Counterexample to existence the Blackwell discount factor γ_{bw}

Theorem

The Blackwell discount factor **may not exist**, even for sa-rectangular convex compact uncertainty set $\mathcal{U}$.

Long story short: worst-case discounted returns oscillate as $\gamma \to 1$

Counterexample to existence the Blackwell discount factor γ_{bw}

Theorem

The Blackwell discount factor **may not exist**, even for sa-rectangular convex compact uncertainty set $\mathcal{U}$.

Long story short: worst-case discounted returns oscillate as $\gamma \to 1$

Jérôme Bolte, ICCOPT, Monday July 2021 2025:

"Oscillations are always hidden behind monsters"

Counterexample to existence the Blackwell discount factor γ_{bw}

Theorem

The Blackwell discount factor **may not exist**, even for sa-rectangular convex compact uncertainty set $\mathcal{U}$.

Long story short: worst-case discounted returns oscillate as $\gamma \to 1$

Jérôme Bolte, ICCOPT, Monday July 2021 2025:

"Oscillations are always hidden behind monsters"

We construct an instance with one state s and two actions a_1, a_2 s.t.:

- Action a_1 is optimal for $\gamma = 1 - \frac{1}{2k}$
- Action a_2 is optimal for $\gamma = 1 - \frac{1}{2k+1}$

Counterexample to existence the Blackwell discount factor γ_{bw}

Theorem

The Blackwell discount factor **may not exist**, even for sa-rectangular convex compact uncertainty set $\mathcal{U}$.

Long story short: worst-case discounted returns oscillate as $\gamma \to 1$

Jérôme Bolte, ICCOPT, Monday July 2021 2025:

"Oscillations are always hidden behind monsters"

We construct an instance with one state s and two actions a_1, a_2 s.t.:

- Action a_1 is optimal for $\gamma = 1 - \frac{1}{2k}$
- Action a_2 is optimal for $\gamma = 1 - \frac{1}{2k+1}$

Intuition: the next two functions oscillate and intersect as $\gamma \to 1$:

$$\gamma \mapsto \min_{\boldsymbol{P} \in \mathcal{U}_{sa_1}} R_\gamma(a_1, \boldsymbol{P})$$

$$\gamma \mapsto \min_{\boldsymbol{P} \in \mathcal{U}_{sa_2}} R_\gamma(a_2, \boldsymbol{P})$$

Preventing oscillations with definability

Definable functions [Cos00] (definition and o-minimality: see (2)):

- "Building blocks": multinomials and exp
- Stable under several operations:
  If f, g are definable, then so are $f + g, f \circ g, f \times g, f/g, -f, f^{-1}$
- Stable by max and min:
  Pointwise max and min of definable functions are definable
- Definable sets = graph of definable functions

Preventing oscillations with definability

Definable functions [Cos00] (definition and o-minimality: see (2)):

- "Building blocks": multinomials and exp
- Stable under several operations:
  If f, g are definable, then so are $f + g, f \circ g, f \times g, f/g, -f, f^{-1}$
- Stable by max and min:
  Pointwise max and min of definable functions are definable
- Definable sets = graph of definable functions

Examples: KL divergences, ℓ_p norms, Wasserstein distance

Preventing oscillations with definability

Definable functions [Cos00] (definition and o-minimality: see (2)):

- "Building blocks": multinomials and exp
- Stable under several operations:
  If f, g are definable, then so are $f + g, f \circ g, f \times g, f/g, -f, f^{-1}$
- Stable by max and min:
  Pointwise max and min of definable functions are definable
- Definable sets = graph of definable functions

Examples: KL divergences, ℓ_p norms, Wasserstein distance

Example: $\mathcal{U}_{sa} = \{\boldsymbol{p} \in \Delta(\mathcal{S}) \mid f(\boldsymbol{p}, \hat{\boldsymbol{p}}) \leq \alpha\}$ is definable if f is definable

Why do we care?

Existence the Blackwell discount factor γ_{bw}

Definability prevents oscillations:

Monotonicity Lemma

If $f:(a,b)\rightarrow\mathbb{R}$ is definable, we can partition (a,b) into *finitely* many subintervals, in which f is either constant or strictly monotone.

Discounted returns can not oscillate when $\mathcal{U}$ is definable:

Lemma

If $\mathcal{U}$ is definable, then $\gamma \mapsto \min_{\boldsymbol{P}\in\mathcal{U}} R_\gamma(\pi,\boldsymbol{P})$ is definable.

Existence the Blackwell discount factor γ_{bw}

Definability prevents oscillations:

Monotonicity Lemma

If $f : (a, b) \to \mathbb{R}$ is definable, we can partition (a, b) into *finitely* many subintervals, in which f is either constant or strictly monotone.

Discounted returns can not oscillate when $\mathcal{U}$ is definable:

Lemma

If $\mathcal{U}$ is definable, then $\gamma \mapsto \min_{\boldsymbol{P} \in \mathcal{U}} R_\gamma(\pi, \boldsymbol{P})$ is definable.

Putting everything together:

Theorem

Let $\mathcal{U}$ be an sa-rectangular convex compact uncertainty set.

If $\mathcal{U}$ is definable, then the Blackwell discount factor γ_{bw} exists.

Existence the Blackwell discount factor γ_{bw}

Definability prevents oscillations:

Monotonicity Lemma

If $f : (a, b) \to \mathbb{R}$ is definable, we can partition (a, b) into *finitely* many subintervals, in which f is either constant or strictly monotone.

Discounted returns can not oscillate when $\mathcal{U}$ is definable:

Lemma

If $\mathcal{U}$ is definable, then $\gamma \mapsto \min_{\boldsymbol{P} \in \mathcal{U}} R_\gamma(\pi, \boldsymbol{P})$ is definable.

Putting everything together:

Theorem

Let $\mathcal{U}$ be an sa-rectangular convex compact uncertainty set.

If $\mathcal{U}$ is definable, then the Blackwell discount factor γ_{bw} exists.

Next question: how to bound γ_{bw} away from 1?

Bound on the Blackwell discount factor γ_{bw} for robust MDPs

Theorem [GCP24]

Consider a finite MDP instance with $\mathcal{U}$ sa-rectangular and:

- $M =$ maximum rewards and common denominator for transitions
- $S =$ number of states
- $\mathcal{U}_{sa} = \ell_1$ or ℓ_∞ balls around nominal transition probabilities.

Then

$$1 - \gamma_{\text{bw}} \geq \Omega\left(\frac{1}{(4M)^{S^2}}\right).$$

Bound on the Blackwell discount factor γ_{bw} for robust MDPs

Theorem [GCP24]

Consider a finite MDP instance with $\mathcal{U}$ sa-rectangular and:

- $M =$ maximum rewards and common denominator for transitions
- $S =$ number of states
- $\mathcal{U}_{sa} = \ell_1$ or ℓ_∞ balls around nominal transition probabilities.

Then

$$1 - \gamma_{\text{bw}} \geq \Omega\left(\frac{1}{(4M)^{S^2}}\right).$$

RMDPs can be solved in $\tilde{O}\left((1-\gamma)^{-1}\right)$...

... So we *don't* obtain a polytime algorithm!

Open Questions and Future Work

More in the papers:

Bounding γ_{bw} for robust MDPs [GCP24]

A complete treatment of average optimality for sa-rec. RMDPs [GCPV23]

The case of s-rec. RMDPs [GCPV23, GCV25]

A more refined analysis of γ_{bw} for stochastic games [GGCK25]

Next steps:

sa-rec. RMDPS: computing average optimal policies?

The case of ϵ-optimal policies?

Unichain/irreducible, weakly-communicating, absorbing, etc.

Open Questions and Future Work

More in the papers:

Bounding γ_{bw} for robust MDPs [GCP24]

A complete treatment of average optimality for sa-rec. RMDPs [GCPV23]

The case of s-rec. RMDPs [GCPV23, GCV25]

A more refined analysis of γ_{bw} for stochastic games [GGCK25]

Next steps:

sa-rec. RMDPS: computing average optimal policies?

The case of ϵ-optimal policies?

Unichain/irreducible, weakly-communicating, absorbing, etc.

Thank you!

David Blackwell.
Discrete dynamic programming.
The Annals of Mathematical Statistics, pages 719–726, 1962.

Michel Coste.
An introduction to o-minimal geometry.
Istituti editoriali e poligrafici internazionali Pisa, 2000.

Julien Grand-Clément and Marek Petrik.
Reducing blackwell and average optimality to discounted mdps via the blackwell discount factor.
Advances in Neural Information Processing Systems, 36, 2024.

Julien Grand-Clement, Marek Petrik, and Nicolas Vieille.
Beyond discounted returns: Robust markov decision processes with average and blackwell optimality.
arXiv preprint arXiv:2312.03618, 2023.

Julien Grand-Clément and Nicolas Vieille.
Playing against a stationary opponent.
arXiv preprint arXiv:2503.15346, 2025.

Vineet Goyal and Julien Grand-Clément.
Robust Markov decision processes: Beyond rectangularity.
Mathematics of Operations Research, 2022.

Stéphane Gaubert, Julien Grand-Clément, and Ricardo D Katz.
Thresholds for sensitive optimality and blackwell optimality in stochastic games.
arXiv preprint arXiv:2506.18545, 2025.

J. Hadamard.
Étude sur les propriétés des fonctions entières et en particulier d'une fonction considéré par Riemann.
Journal de Mathématiques Pures et Appliquées, 58:171–215, 1893.

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

J. L. Lagrange.
Sur la résolution des équations numériques.
Mémoires de l'Académie royale des Sciences et Belles-Lettres de Berlin, XXIII, 1769.

K. Mahler.
On some inequalities for polynomials in several variables.
J. London Math. Soc, 37(1):341–344, 1962.

M. Mignotte and M. Waldschmidt.
On algebraic numbers of small height: linear forms in one logarithm.
Journal of Number Theory, 47(1):43–62, 1994.

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

Siegfried M Rump.
Polynomial minimum root separation.
Mathematics of Computation, 33(145):327–336, 1979.

Richard D Smallwood.
Optimum policy regions for markov processes with discounting.
Operations Research, 14(4):658–669, 1966.

Ambuj Tewari and Peter L Bartlett.
Bounded parameter Markov decision processes with average reward criterion.
In *International Conference on Computational Learning Theory*, pages 263–277. Springer, 2007.

Yue Wang, Alvaro Velasquez, George Atia, Ashley Prater-Bennette, and Shaofeng Zou.
Robust average-reward markov decision processes.
In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 15215–15223, 2023.

C. K. Yap.
Fundamental problems of algorithmic algebra, volume 49.
Oxford University Press Oxford, 2000.

Y. Ye.
A new complexity result on solving the Markov decision problem.
Mathematics of Operations Research, 30(3):733–749, 2005.

Rigorous definition of definability

Definition: definable set and definable function [Cos00]

A subset of $\mathbb{R}^n$ is *definable* if it is the image, under a canonical projection $\mathbb{R}^{n+k} \to \mathbb{R}^n$ that eliminates any set of k variables, of a set of the form

$$\{\boldsymbol{x} \in \mathbb{R}^{n+k} \mid \mathrm{Poly}(x_1, ..., x_{n+k}, \exp(x_1), ..., \exp(x_{n+k})) = 0\} \qquad (2)$$

A function is definable if its graph is definable.

Back to main body: slide 14

Main results in [GCPV23]

Uncertainty set $\mathcal{U}$	Discount optimality	Average optimality	Blackwell optimality
Singleton (MDPs)	stationary, deterministic	stationary, deterministic	stationary, deterministic
sa-rectangular, compact	stationary, deterministic	**stationary, deterministic**	• **may not exist** • $\exists\ \pi$ **stationary deterministic,** π ϵ-**Blackwell optimal,** $\forall \epsilon > 0$ • π **also average optimal**
sa-rectangular, compact, definable	stationary, deterministic	**stationary, deterministic**	• **stationary, deterministic** • π **also average optimal**
s-rectangular, compact convex	stationary, randomized	• **may not exist** • **may be history-dependent, randomized**	**may not exist**

Main results in [GCPV23]

Our main results for the *average return*:

$$\sup_{\pi \in \Pi_{\mathsf{H}}} \inf_{\boldsymbol{P} \in \mathcal{U}} \quad \mathbb{E}^{\pi,\boldsymbol{P}} \left[\limsup_{T \to +\infty} \frac{1}{T+1} \sum_{t=0}^{T} r_{s_t a_t s_{t+1}} \right].$$

1. For sa-rectangular RMDPs:
   - Optimality of stationary deterministic policies
   - Strong duality (existence of a value)
   - “All” optimality criteria ($\liminf, \limsup$) are equivalent
   - Optimal average value $= \lim_{\gamma \to 1} \mathrm{VAL}_\gamma$

Main results in [GCPV23]

Our main results for the *average return*:

$$\sup_{\pi \in \Pi_{\mathrm{H}}} \inf_{\boldsymbol{P} \in \mathcal{U}} \quad \mathbb{E}^{\pi, \boldsymbol{P}} \left[\limsup_{T \to +\infty} \frac{1}{T+1} \sum_{t=0}^{T} r_{s_t a_t s_{t+1}} \right].$$

1. For sa-rectangular RMDPs:
   - Optimality of stationary deterministic policies
   - Strong duality (existence of a value)
   - "All" optimality criteria ($\liminf, \limsup$) are equivalent
   - Optimal average value $= \lim_{\gamma \to 1} \mathrm{VAL}_\gamma$
2. For s-rectangular RMDPs:
   - Non-existence of optimal policies in general
   - The Big Match: Markovian policies are optimal
   - Optimality criteria are not equivalent

Main results in [GCPV23]

Our main results for *Blackwell optimality*:

For sa-rectangular RMDPs:

- Blackwell optimal policies may not exist in general
- ϵ-Blackwell optimal stationary policies always exist
- Non-Lipschitzness of the discounted value functions as $\gamma \to 1$
- *Definable* uncertainty sets $\Rightarrow$ existence of stationary Blackwell optimal policies

For s-rectangular RMDPs:

- Blackwell optimal policies may not exist in general