The Value Function Polytope In Reinforcement Learning

Robert Dadashi
AI Resident @ Google Brain Montreal

Collaborators

Adrien Ali Taiga

Nicolas Le Roux

Dale Schuurmans

Marc G. Bellemare

Problematic

Question:

What is the geometry of the space of possible value functions for a given Markov decision process ?

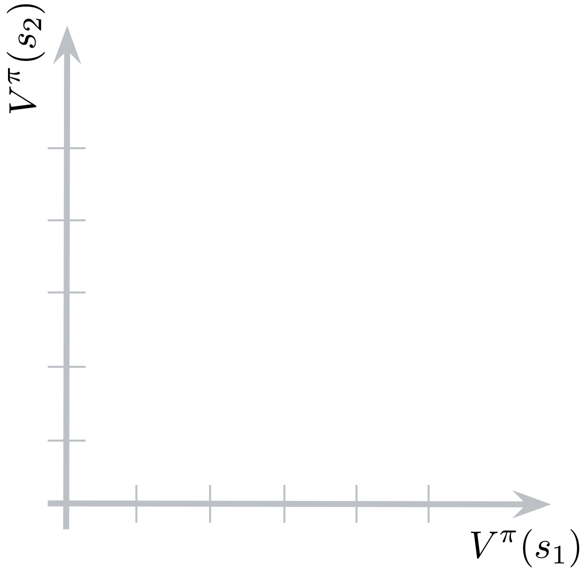
Motivation:

- Relationship between policy space and value function space
- Better understanding of the dynamics of existing algorithms
- New formalism of representation learning in RL

Illustration

Consider a Markov decision process with 2 states: R, P, γ

$$\pi_1 \sim \mathcal{P}(\mathcal{A}) \times \mathcal{P}(\mathcal{A}) \to V^{\pi_1} = \begin{bmatrix} 3 \\ 5 \end{bmatrix}$$

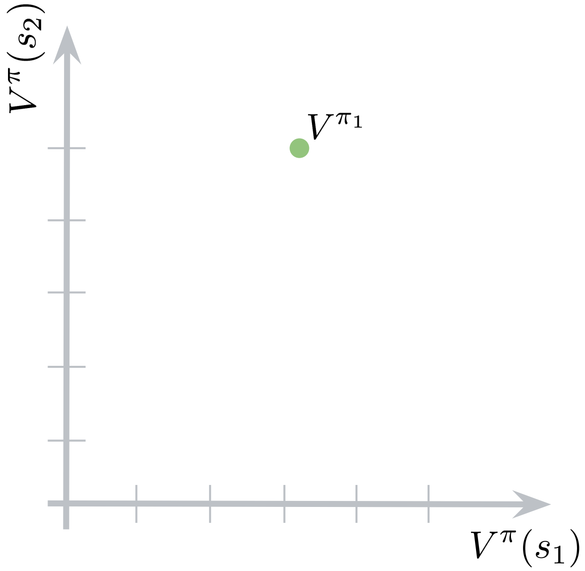


What is the geometry of the space of value functions for a given MDP ?

Illustration

Consider a Markov decision process with 2 states: R, P, γ

$$\pi_1 \sim \mathcal{P}(\mathcal{A}) \times \mathcal{P}(\mathcal{A}) \to V^{\pi_1} = \begin{bmatrix} 3 \\ 5 \end{bmatrix}$$



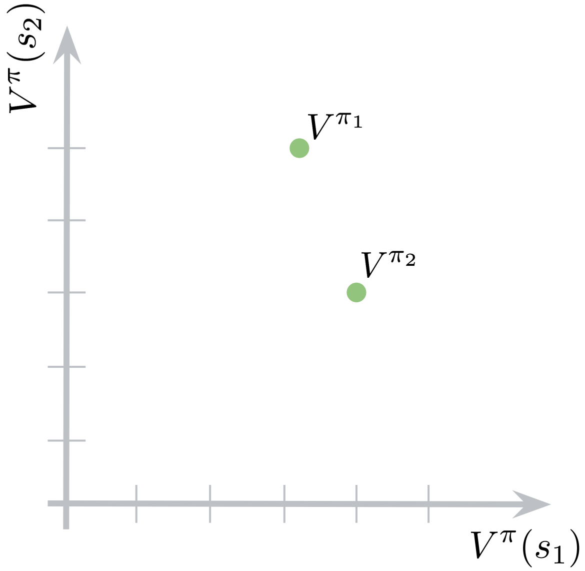
What is the geometry of the space of value functions for a given MDP ?

Illustration

Consider a Markov decision process with 2 states: R, P, γ

$$\pi_1 \sim \mathcal{P}(\mathcal{A}) \times \mathcal{P}(\mathcal{A}) \rightarrow V^{\pi_1} = \begin{bmatrix} 3 \\ 5 \end{bmatrix}$$

$$\pi_2 \sim \mathcal{P}(\mathcal{A}) \times \mathcal{P}(\mathcal{A}) \rightarrow V^{\pi_2} = \begin{bmatrix} 4 \\ 3 \end{bmatrix}$$



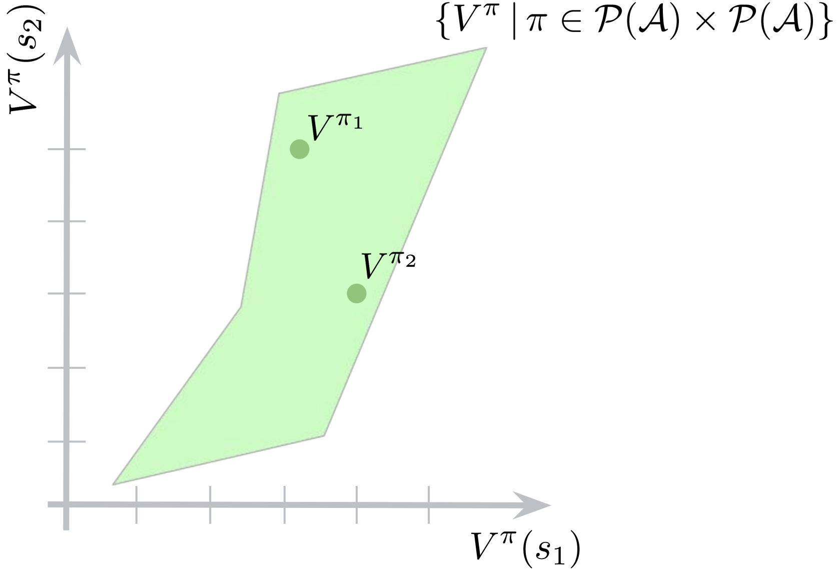
What is the geometry of the space of value functions for a given MDP ?

Illustration

Consider a Markov decision process with 2 states: R, P, γ

$$\pi_1 \sim \mathcal{P}(\mathcal{A}) \times \mathcal{P}(\mathcal{A}) \to V^{\pi_1} = \begin{bmatrix} 3 \\ 5 \end{bmatrix}$$

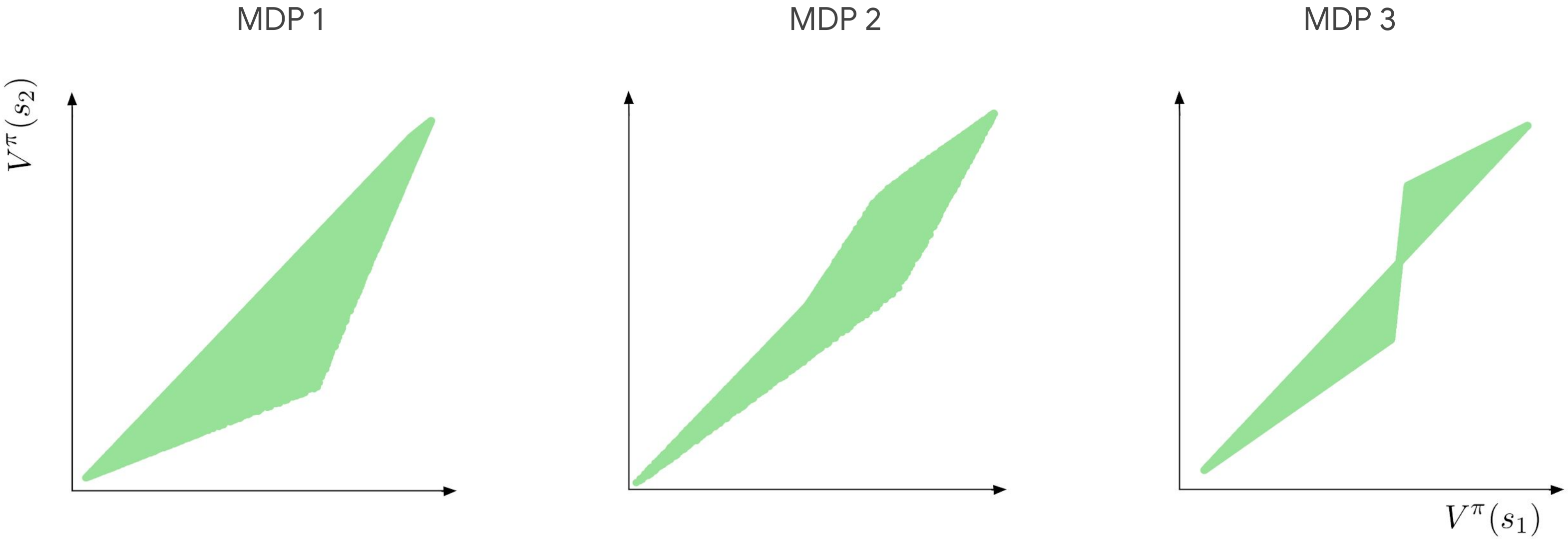
$$\pi_2 \sim \mathcal{P}(\mathcal{A}) \times \mathcal{P}(\mathcal{A}) \to V^{\pi_2} = \begin{bmatrix} 4 \\ 3 \end{bmatrix}$$


$\{V^\pi \mid \pi \in \mathcal{P}(\mathcal{A}) \times \mathcal{P}(\mathcal{A})\}$
$V^\pi(s_2)$
V^{π_1}
V^{π_2}
$V^\pi(s_1)$

What is the geometry of the space of value functions for a given MDP ?

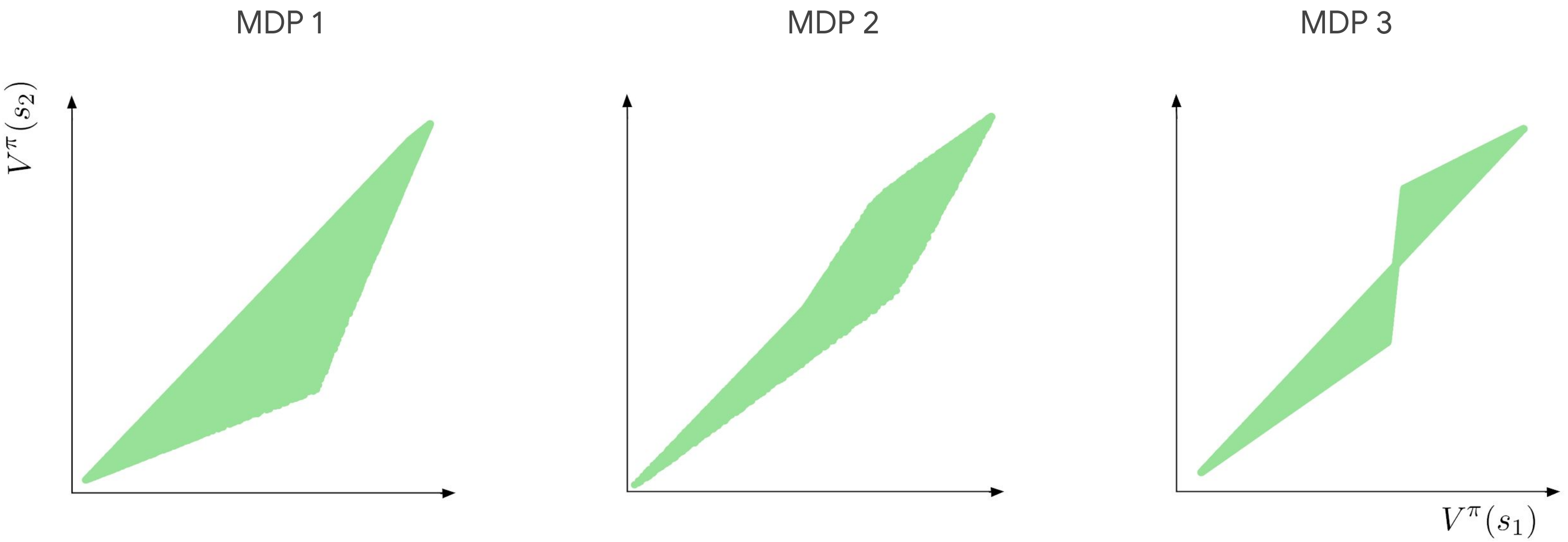
Main result

What is the geometry of the space of value functions for a given MDP ?



Main result

What is the geometry of the space of value functions for a given MDP ?



MDP 1

MDP 2

MDP 3

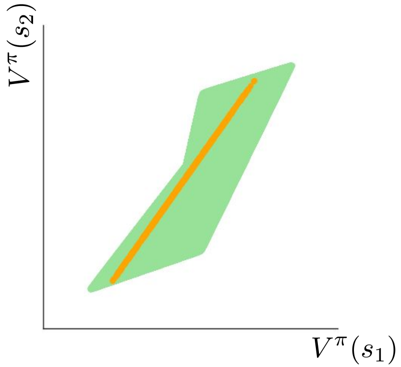
Theorem:

The ensemble of value functions is a possibly non-convex polytope.

Building blocks

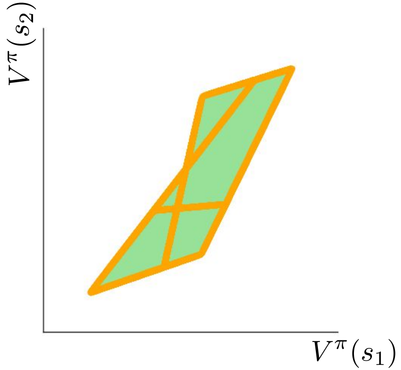
The *line* Theorem

The value functions of mixtures of two policies that differ in only one state describe a line in value function space.



The *boundary* Theorem

The boundary of the space of value functions is included in the image of the boundary of the space of policies.



Google

Algorithms in the polytope

Value Iteration

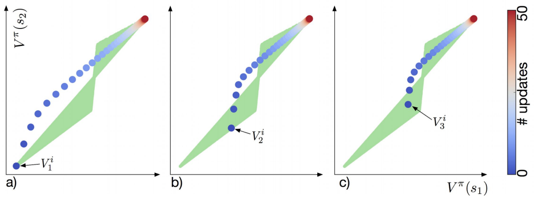


Figure 7. Value iteration dynamics for three initialization points.

Policy Iteration

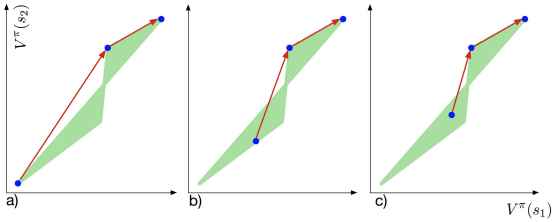


Figure 8. Policy iteration. The red arrows show the sequence of value functions (blue) generated by the algorithm.

Policy Gradient

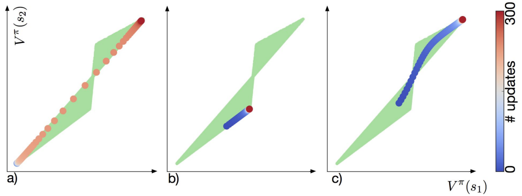


Figure 9. Value functions generated by policy gradient.

Policy Gradient + entropy

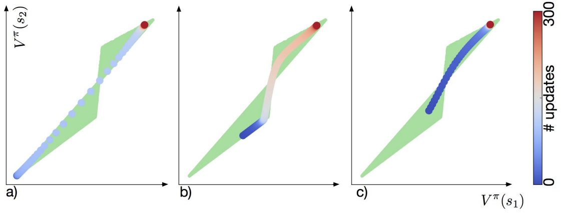


Figure 10. Value functions generated by policy gradient with entropy, for three different initialization points.

CEM + CEM-CN

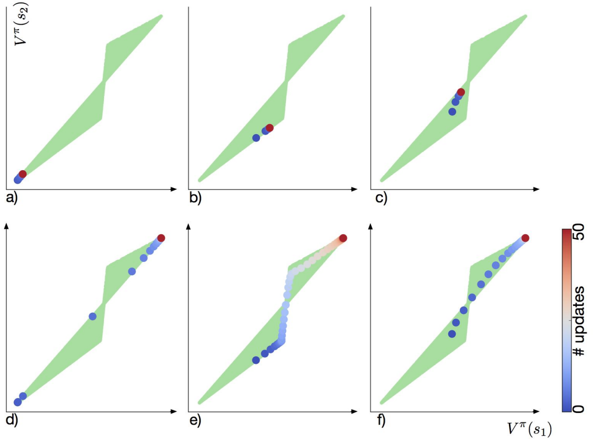


Figure 12. The cross-entropy method without noise (CEM) (a, b, c); with constant noise (CEM-CN) (d, e, f).

Ongoing work

- Representation learning in Reinforcement Learning
- New actor-critic algorithms

Thank you

Poster # 119

dadashi@google.com