# The Value Function Semi-Algebraic Set in Partially Observable Markov Decision Processes

Ryan A. Anderson[1] Guido Montúfar[1,2,3]

(1) UCLA Statistics & Data Science, (2) UCLA Mathematics, (3) MPI MiS

ICML 2026

# The Geometry of Feasible Value Functions

- A (PO)MDP is a tuple $\langle \mathcal{S}, \mathcal{A}, \mathcal{O}, \alpha, \beta, r, \gamma \rangle$. Each policy $\pi \in \Delta_{\mathcal{A}}^{\mathcal{O}}$ induces a value function $V^\pi \in \mathbb{R}^{\mathcal{S}}$ solving the Bellman equation

$$(I - \gamma P^\pi) V^\pi = r^\pi, \quad \gamma \in (0, 1)$$

- We study the **feasible set** of value functions as $\pi$ ranges over all policies $\mathcal{V} = \{V^\pi : \pi \in \Delta_{\mathcal{A}}^{\mathcal{O}}\}$
- For fully observable MDPs, [DTR+19] showed $\mathcal{V}$ is a (non-convex) union of **polytopes** (closed, bounded, cut out by halfspaces), with a *line theorem* governing single-state policy changes
- **Open question:** what is the geometry of $\mathcal{V}$ under partial observability?

# Bellman Equation as a Parametric Linear System

- We regard the Bellman equation as a **parametric linear system** $A(p)x - b(p) = 0$, affine in the policy parameters $p = (p_{o,a}) \in \Delta_{\mathcal{A}}^{\mathcal{O}}$:

$$A(p) = A^0 + \sum_{(o,a)} A^{(o,a)} p_{o,a}, \quad b(p) = \sum_{(o,a)} b^{(o,a)} p_{o,a}$$

- $\mathcal{V}$ is the **solution set** of this system over the policy polytope $\Delta_{\mathcal{A}}^{\mathcal{O}}$
- For affine parametric systems, [Hla12] gives a necessary and sufficient condition: $x$ solves $A(p)x = b(p)$ for some $p \in [p]$ iff for every $y \in \mathbb{R}^n$

$$y^\top (A(p^c)x - b(p^c)) \leq \sum_{k=1}^{K} p_k^\Delta |y^\top (A^k x - b^k)|$$

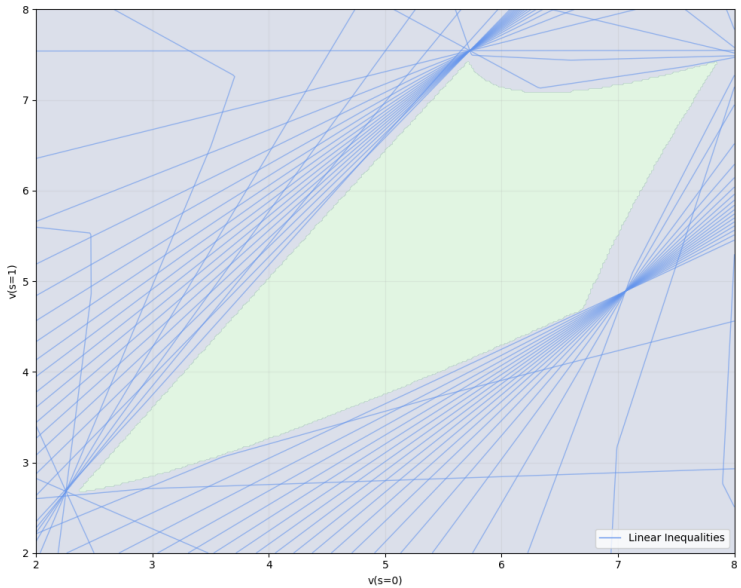
- This is an **infinite family** of piecewise linear inequalities, one per $y$
- Issue: this requires a hyperrectangle parametrization, but the policy polytope $\Delta_{\mathcal{A}}^{\mathcal{O}}$ is not a box — we write it as an **intersection** of box parametrizations $S = \bigcap_i S_i$

# Solution Set of the Bellman Equation

- Applying [Hla12] to each box parametrization $(B_i, c_i)$ yields an exact characterization: $x \in \mathcal{V}$ iff for every $y$ and every anchor $i$,

$$y^\top (B_i(v^c)x - c_i(v^c)) \leq \sum_{(o,a)} v^{\Delta}_{o,a} |y^\top (B_i^{(o,a)} x - c_i^{(o,a)})|$$

- In the fully observable case this collapses to **finitely many** piecewise linear inequalities, recovering $\min_a Q^x_a(s) \leq x_s \leq \max_a Q^x_a(s)$, i.e. $x_s \in \mathrm{conv}\{Q^x_a(s)\}$

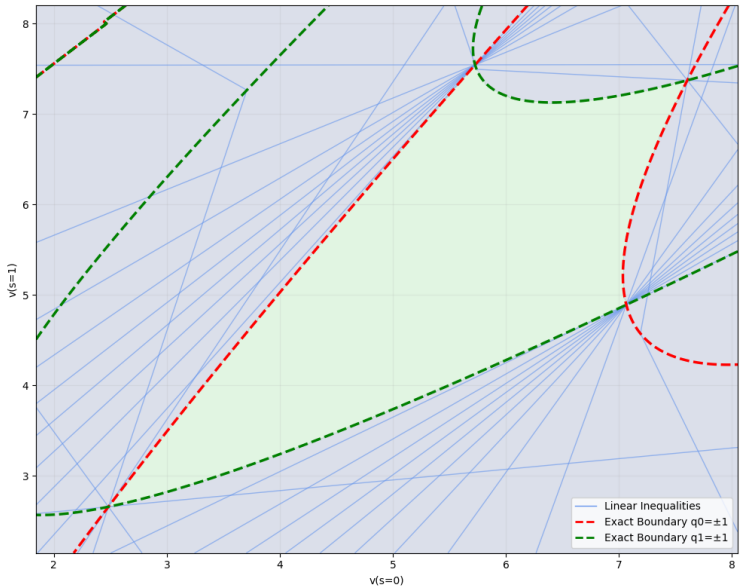


# From Infinite Inequalities to a Finite Description

- We want description in finitely many polynomial conditions ⇒ **zonotope condition** on the parametric system
- Let $D$ be a matrix with $k$-th column $D_k(x) = p_k^{\Delta}(A^k x - b^k)$, and $C(x) = A(p^c)x - b(p^c)$. Then $x \in \mathcal{V}$ iff there exist $q^{(i)} \in \mathbb{R}^{\mathcal{O}\times\mathcal{A}_i}$ with

$$D_i(x)\, q^{(i)} = C_i(x), \quad -1 \le q^{(i)} \le 1$$

- Entries of $D_i, C_i$ are **polynomial** in $x$ ⇒ by Tarski–Seidenberg, $\mathcal{V}$ is a **semi-algebraic set**



# Quantifier-Free Semi-Algebraic Description

- We eliminate the quantifier over $q^{(i)}$ by considering a new parametric system: stack the Bellman columns $u_{o,a}(x)$ into $M(x)$ with the row-sum matrix $R$ and obtain $C(x)\,P = f(x),\ P \geq 0$ with

$$C(x) = \begin{pmatrix} M(x) \\ R \end{pmatrix}, \quad f(x) = \begin{pmatrix} x \\ \mathbf{1}_{|\mathcal{O}|} \end{pmatrix}$$

- Stratifying by rank $\rho = \operatorname{rank} C(x)$, we obtain a decomposition of the solution set to the Bellman equation $S$:

$$S = \bigcup_{\rho} \bigcup_{|I|=\rho} \bigcup_{|B|=\rho} S_{\rho,I,B}$$

- Each piece $S_{\rho,I,B}$ is cut out by **determinantal** conditions: rank certificates, consistency of $[C(x)|f(x)]$, and sign polynomials

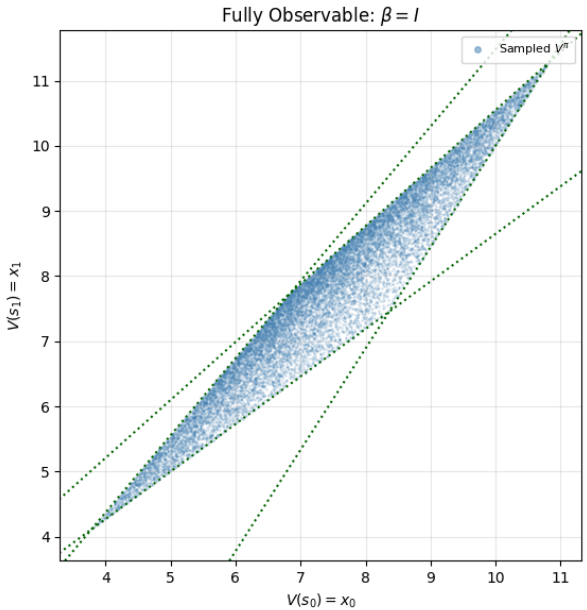
$$\det C_{I,B,t}(x) \cdot \det C_{I,B}(x) \geq 0, \quad t = 1, \ldots, \rho$$

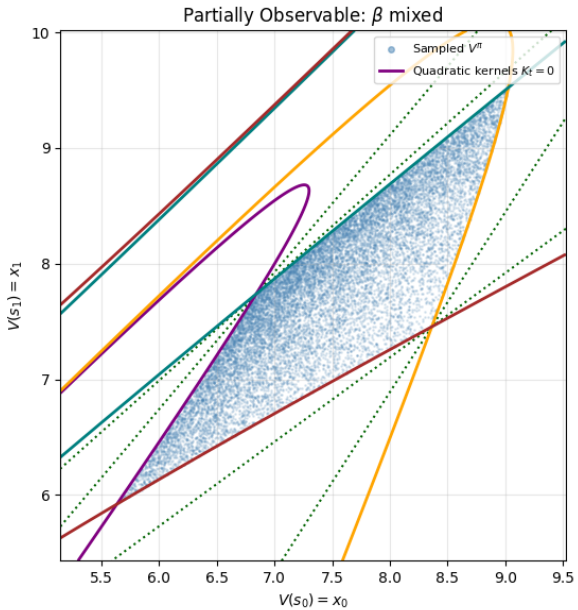
# Computing Boundary Components for $|S| = |A| = |O| = 2$

- On the minimal $2 \times 2 \times 2$ instance, the sign polynomials factor (computed in Macaulay2 [GS]) as

$$\det C_{B,t}(x) \cdot \det C_B(x) = Q_0(x)\, Q_1(x) \cdot K_t(x; \alpha, r, \gamma, \beta)$$

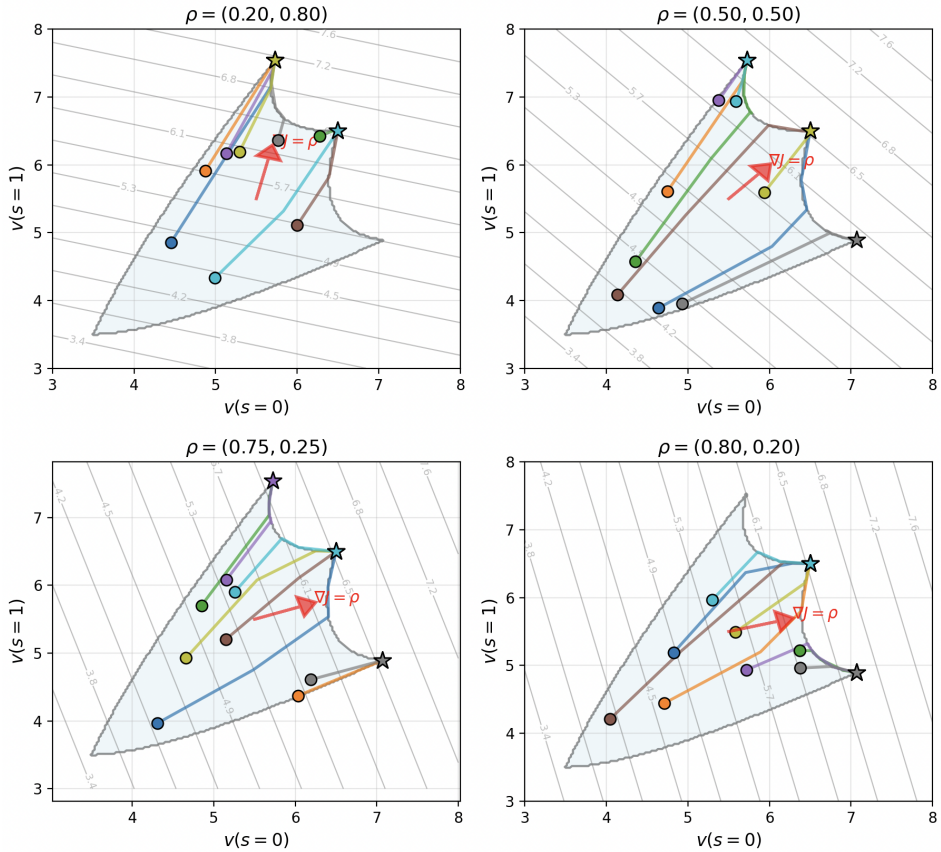
- **Fully observable** ($\beta = I$): $K_t$ splits into **linear** factors $\Rightarrow$ fully observable (potentially non-convex) polytope
- **Partially observable**: $K_t$ is an **irreducible quadratic** and $\partial\mathcal{V}$ acquires curved arcs



# Linear Programming Fails to Reach Optima in POMDPs

- In MDPs, the optimal policy maximizes $J = \sum_s \rho_s V^\pi(s)$ and is **independent** of the initial distribution $\rho$ [Put05]
- Under partial observability the curved boundary breaks this: optimal policies **depend on** $\rho$, and **isolated local maxima** of long-term reward emerge



# Empirical Confirmation

- Across random POMDPs, policy gradient from 50 restarts shows value spread **1–2 orders of magnitude** larger than the matched fully observable baseline
- Distinct local optima are the norm, and their number grows with the state space size $|\mathcal{S}|$
- Adding finite memory **mitigates but does not erase** the spread — the geometric complexity of $\mathcal{V}$ persists
- **Takeaway:** partial observability turns the MDP polytope into a semi-algebraic region with intrinsically curved boundary, explaining $\rho$-dependence and multiple optima

Thank you! `raanderson@g.ucla.edu`

# References I


[DTR+19] Robert Dadashi, Adrien Ali Taiga, Nicolas Le Roux, Dale Schuurmans, and Marc G. Bellemare. The value function polytope in reinforcement learning. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 1486–1495. PMLR, 2019.

[GS] Daniel R. Grayson and Michael E. Stillman. Macaulay2, a software system for research in algebraic geometry. Available at `http://www2.macaulay2.com`.

[Hla12] Milan Hladík. Enclosures for the solution set of parametric interval linear systems. *Int. J. Appl. Math. Comput. Sci.*, 22(3):561–574, 2012.

[Put05] Martin L. Puterman. *Markov decision processes: discrete stochastic dynamic programming*. Wiley online library. Wiley-Interscience, Hoboken, New Jersey, 2005.