Should I Stay or Should I Go?

A Guide to Solving Stochastic Games with Value Iteration

Jan Křetínský[1,2] Tobias Meggendorfer[1,3] Maximilian Weininger[1,3]

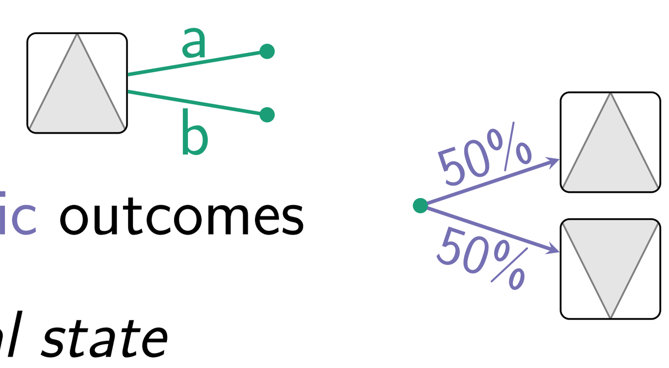
[1]Technical University of Munich [2]Masaryk University [3]Institute of Science and Technology Austria

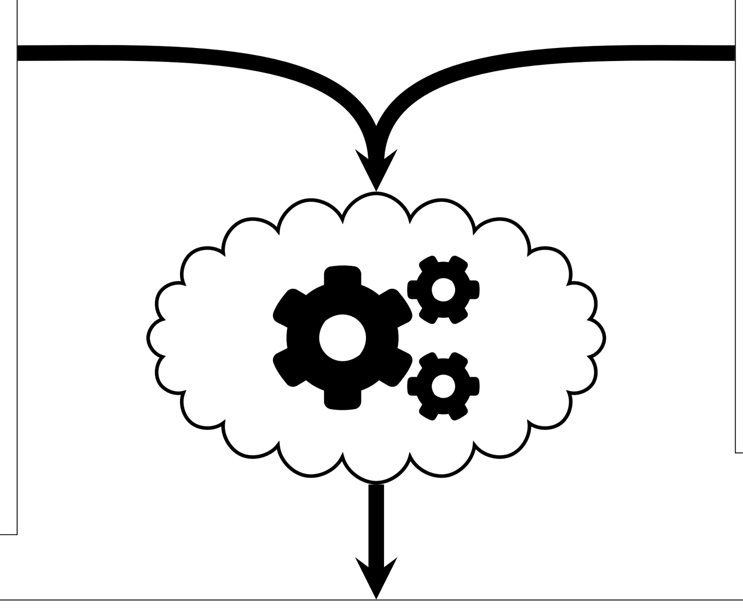
Spoiler: You should stay!

The Setting

Stochastic Games

- Graph game between two players
  s Maximizer and t Minimizer
- Each state has actions
- Actions have probabilistic outcomes
- Initially: Token on *initial state*
- Repeatedly: Player owning vertex with token chooses action, token moves according to probability





Objectives

Assign a number to each *play*

Examples:

- Reachability: 1 if reached state, 0 otherwise
- Safety: 1 if avoided state, 0 otherwise
- Total reward: Actions yield money, get total income
- Mean payoff: Actions yield money, get average income per step

Value of the Game

If both players play optimally, what will they get in expectation?

A Classical Solution Approach: Value Iteration

Step 1

Initialize lower and upper bounds for each state with surely correct values

$L(s) \leftarrow 0$, $U(s) \leftarrow 1$

Step 2

Update lower and upper estimates by playing optimally for one step

$L(s) \leftarrow \max_{a \in \mathsf{Act}(s)} \sum_{s'} p(s,a)(s') \cdot L(s')$

Step 3

Repeat Step 2 until lower and upper bounds are close enough

$U(s) - L(s) < \varepsilon$

The Issue: Step 3 might never finish because of spurious fixpoints

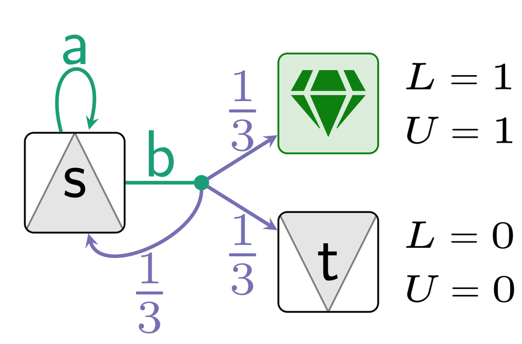
This is *known but unsolved* in generality for nearly a decade!

A Modern Idea: Deflating and Inflating

In one sentence: VI only “sees” finite horizon

⇒ Provide “infinite horizon” information: What happens if players stay *forever*?

Reachability

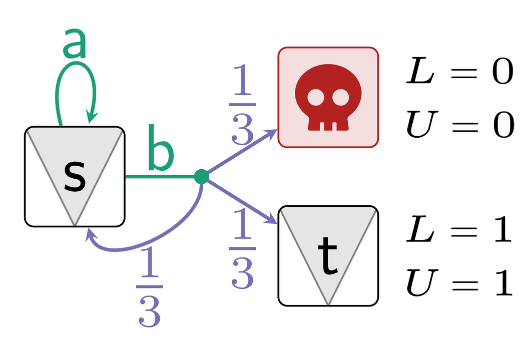


$L(s)$	0	$\frac{1}{3}$	$\frac{4}{9}$	$\frac{13}{27}$	…	Converges (action: b)
$U(s)$	1	1	1	1	…	Stuck! (action: a)

- Staying ⇒ 0 (not reaching goal)
- Using best *exit* b ⇒ $\frac{2}{3}$, then $\frac{5}{9}$, …

Solution: Deflate $U(s)$ to $\max\{0, \frac{2}{3}\}$

Safety

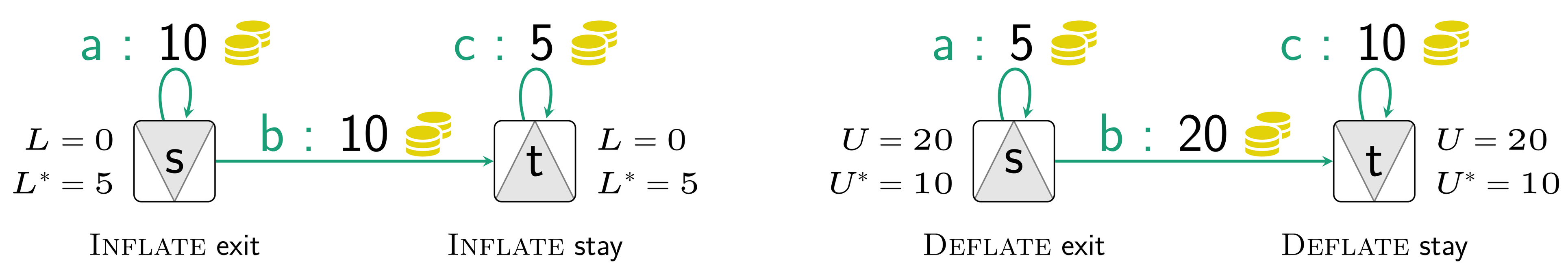


$L(s)$	0	0	0	0	…	Stuck! (action: a)
$U(s)$	1	$\frac{2}{3}$	$\frac{5}{9}$	$\frac{14}{27}$	…	Converges (action: b)

- Staying ⇒ 1 (avoiding sink)
- Using best *exit* b ⇒ $\frac{1}{3}$, then $\frac{4}{9}$, …

Solution: Inflate $L(s)$ to $\min\{1, \frac{1}{3}\}$

Mean Payoff



- Staying value non-trivial
- Both Inflate and Deflate

Key Takeaway

SOUND

Updates are correct

When you are asked “Should I stay or should I go?” answer with staying values and best exits!

COMPLETE

Updates ensures convergence

Unifies and extends previous works [Brá+14; HM14; Bai+17; Ash+17] (MDP) [Kel+18; Pha+20] (SG)

UNIVERSAL

Reasoning is objective independent

Your best exit? Save the webpage and read the paper!

Jan Křetínský, Tobias Meggendorfer, and Maximilian Weininger. “Stopping Criteria for Value Iteration on Stochastic Games with Quantitative Objectives.” In: *LICS*. 2023

[Ash+17] Pranav Ashok et al. “Value Iteration for Long-Run Average Reward in Markov Decision Processes.” In: *CAV (1)*. 2017.

[Bai+17] Christel Baier et al. “Ensuring the Reliability of Your Model Checker: Interval Iteration for Markov Decision Processes.” In: *CAV (1)*. 2017.

[Brá+14] Tomás Brázdil et al. “Verification of Markov Decision Processes Using Learning Algorithms.” In: *ATVA*. 2014.

[HM14] Serge Haddad and Benjamin Monmege. “Reachability in MDPs: Refining Convergence of Value Iteration.” In: *RP*. 2014.

[Kel+18] Edon Kelmendi et al. “Value Iteration for Simple Stochastic Games: Stopping Criterion and Learning Algorithm.” In: *CAV (1)*. 2018.

[Pha+20] Kittiphon Phalakarn et al. “Widest Paths and Global Propagation in Bounded Value Iteration for Stochastic Games.” In: *CAV (2)*. 2020.