Scalable Solutions to Zero-Sum Partially Observable Stochastic Games through Belief Aggregation with Approximation Guarantees

40th AAAI Conference on Artificial Intelligence
Singapore, January 20-27, 2026

Dr. Kim Hammar and Prof. Tansu Alpcan
{kim.hammar, tansu.alpcan}@unimelb.edu.au

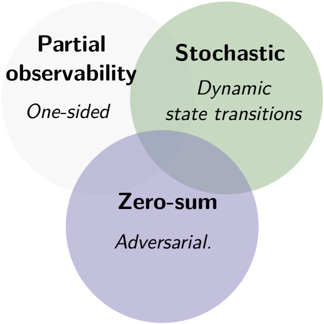
POSTERA CRESCAM LAUDE
THE UNIVERSITY OF MELBOURNE

ZS-OS-Partially Observed Stochastic Games

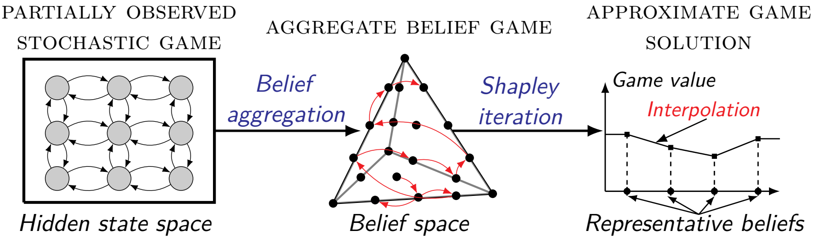
- **Applications**
  - Cybersecurity.
  - Robotics.
  - Learning theory.

- **State-of-the-art**
  - Heuristic search value iteration (HSVI).
  - Deep reinforcement learning (e.g., NFSP).

- **Limitations of Current Methods**
  - HSVI offers theoretical guarantees but is not scalable.
  - Deep reinforcement learning is scalable but offers no theoretical guarantees.



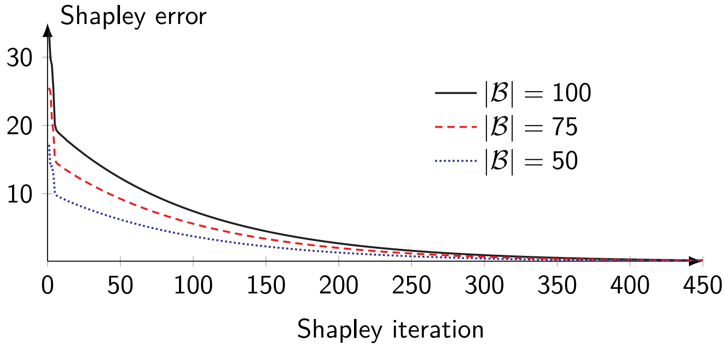
Our Method: **S**hapley Iteration with **A**ggregated **B**eliefs



- We convert the POSG into an **aggregate belief game**.
  - A (fully observed) stochastic game with finite belief space.
- We solve the aggregate belief game through Shapley iteration.
- We use the solution to the aggregate game to approximate an optimal solution to the original game via interpolation.

Proposition 1 (**Convergence (Informal)**)

SAB converges for any POSG.



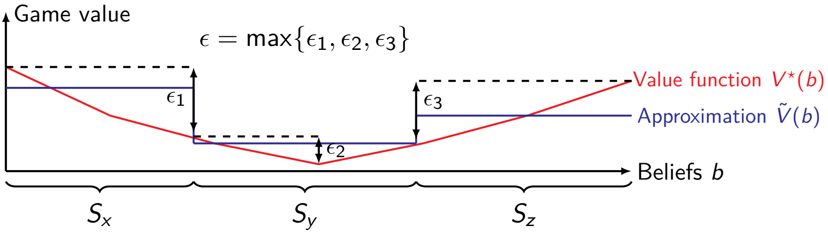
Proposition 2 (**Approximation error bound (Informal)**)

The difference between the estimated value function $\tilde{V}$ and the true value function $V^\star$ is bounded as

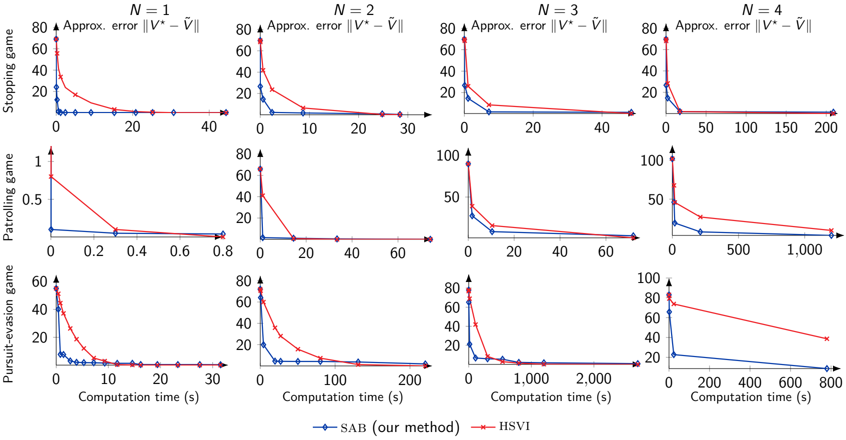
$$|\tilde{V}(b) - V^\star(b)| \leq \frac{\epsilon}{1-\gamma}, \qquad \textit{for all } b \in B,$$

where γ is the discount factor and ϵ ***is a finite constant*** *defined by*

$$\epsilon = \max_{x \in \mathcal{B}} \sup_{b,b' \in S_x} |V^\star(b) - V^\star(b')|, \quad S_x = \{b \mid b \in B, \phi_{bx} = 1\}.$$

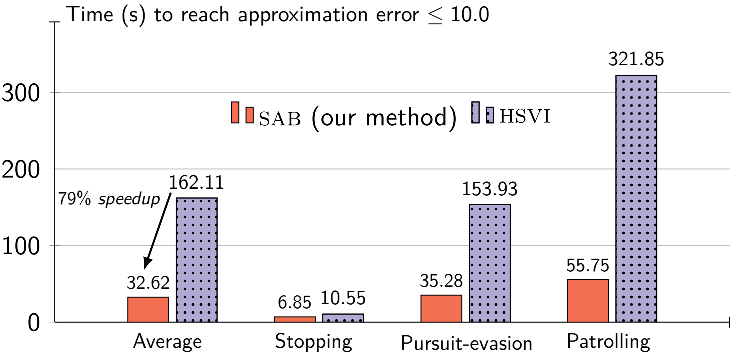


Comparison with **Heuristic Search Value Iteration**



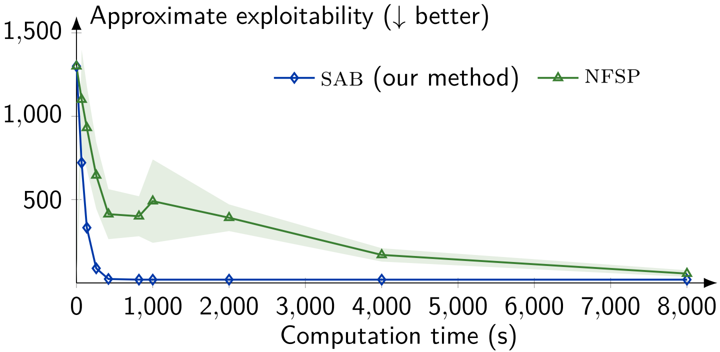
The parameter N controls the size of each game.

Computational Efficiency



Time to reach an approximation error of 10.0 or less across the evaluation games with size $N \in \{1, 2, 3, 4\}$.

Scalability



Comparison between SAB and a deep reinforcement learning method (NFSP) on a large game where HSVI is computationally intractable.

Conclusion

- We present SAB: **S**hapley iteration with **A**ggregated **B**eliefs.
  - A new method for approximately solving zero-sum POSGs.
  - Provides theoretical approximation guarantees.
  - Computationally scalable and flexible.
  - *Outperforms HSVI and NFSP across three example POSGs.*