Safe Exploration for Interactive Machine Learning

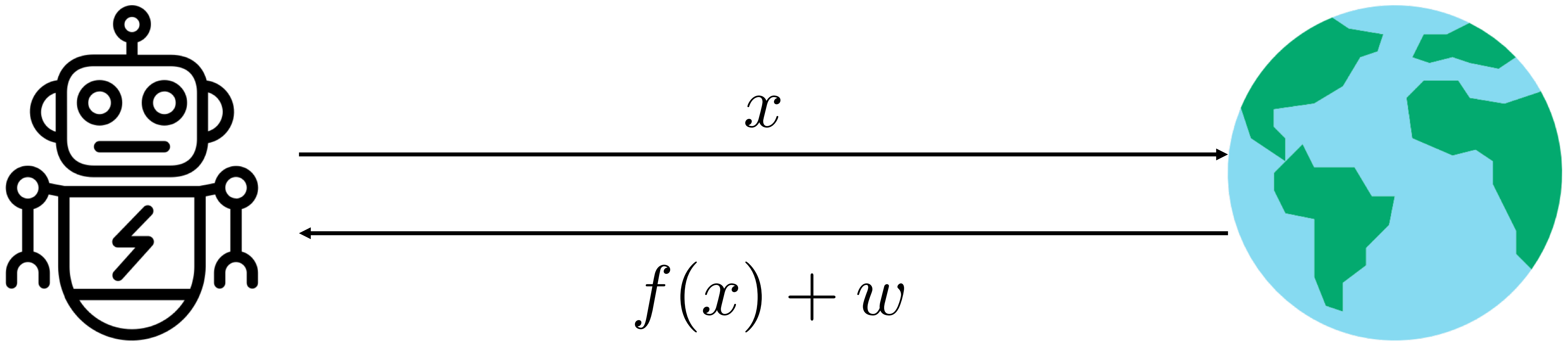
Matteo Turchetta, Felix Berkenkamp, Andreas Krause

ETH zürich

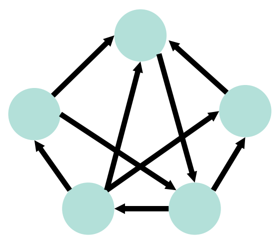
MAX-PLANCK-GESELLSCHAFT

Interactive Machine Learning

- Agent can query **noisy** values of an **unknown** function
- Use data to make **informed** queries



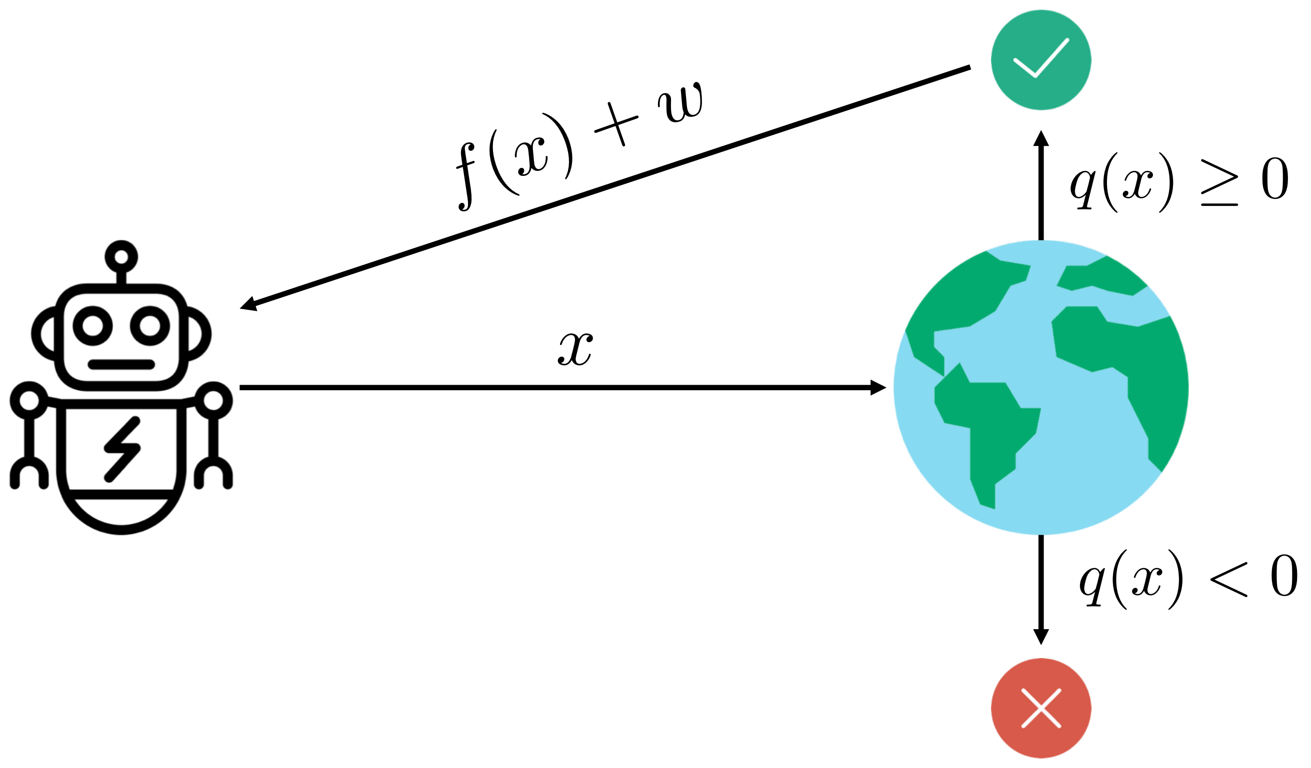
- Available queries may depend from previous ones: model dependency with directed graph



- Includes: Bayesian optimization, active learning and exploration of deterministic Markov decision processes

Safety constrained interactive machine learning

Unknown safety **constraint** q(x)>0 that must be satisfied at **all times**

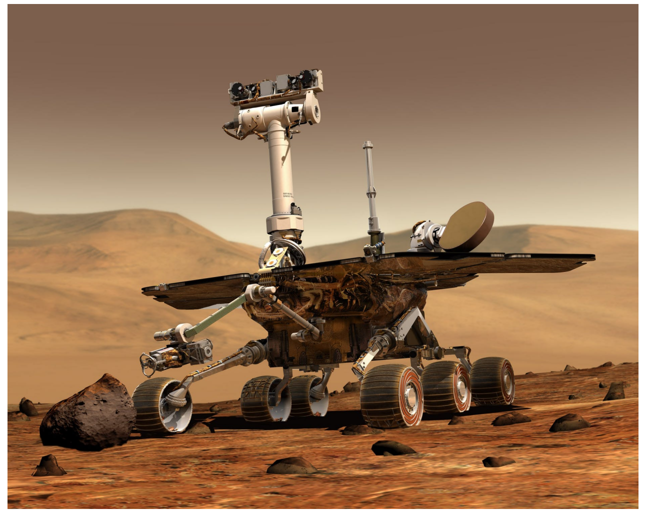


Encompasses many problems

Therapy design

[Sui et al. 2015],
[Sui et al. 2018]

Mars exploration



[Turchetta et al. 2016],
[Wachi et al. 2018]

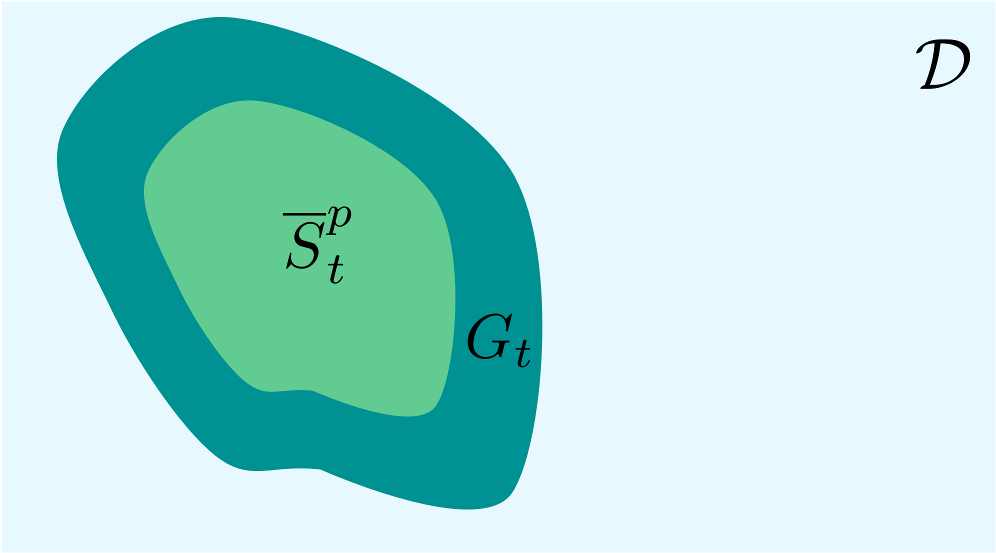
Model free RL

[Berkenkamp et al. 2016]

Existing approaches

Build a **conservative estimate** of the decisions that are safe to evaluate $\overline{S}_t^p$

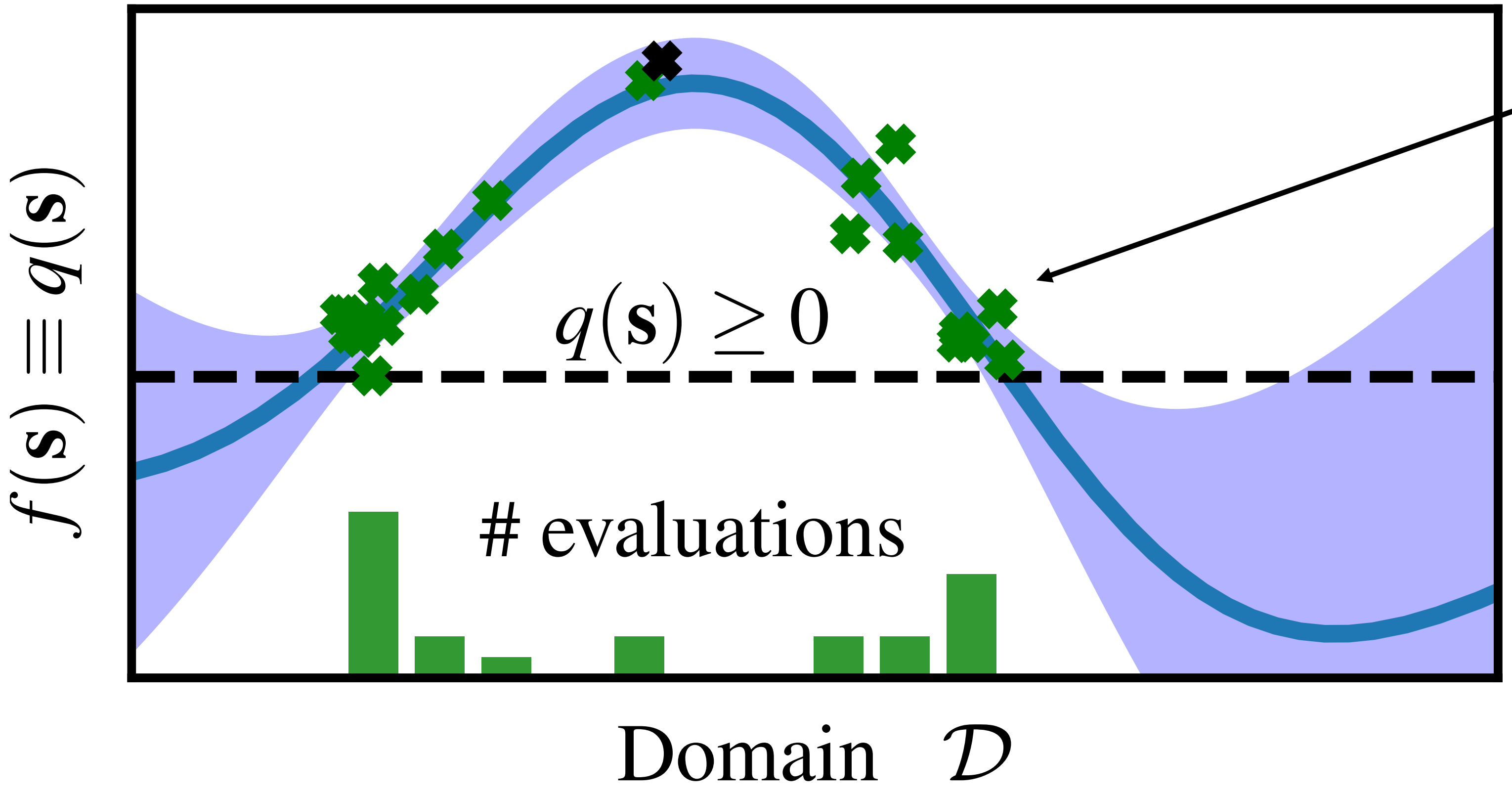
Uniformly reduce uncertainty on the boundary of this region G_t



Treating the **expansion** of the safe set as a **proxy objective** can be **wasteful**

Example: 1D optimization task

StageOPT [Sui et al. 2018]

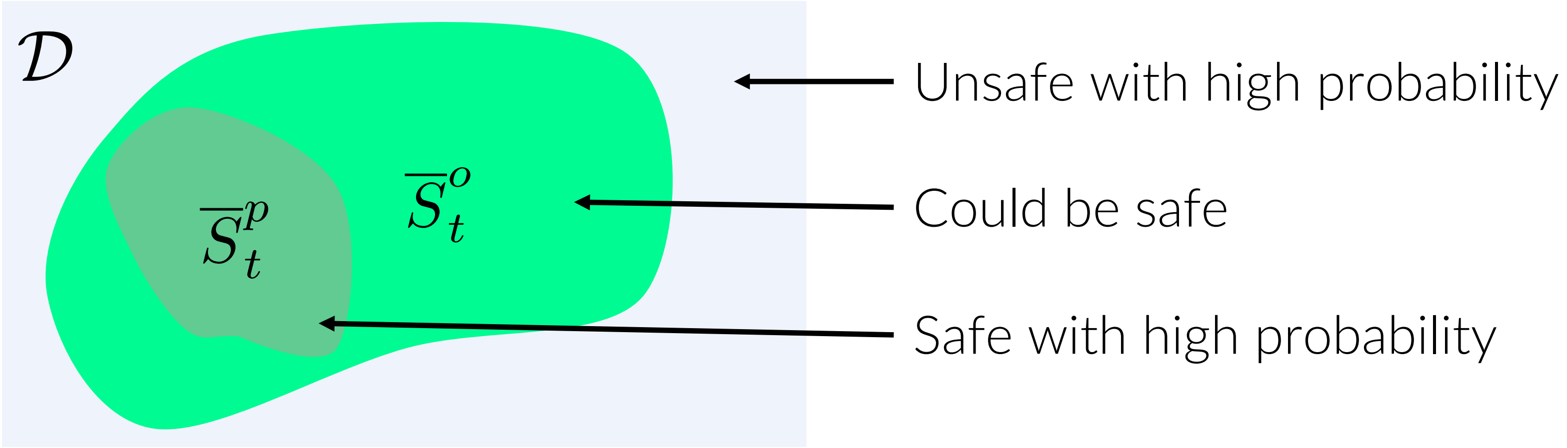


Many unnecessary samples when optimum has already been found

Goal Oriented Safe Exploration separates IML task and safety

Idea: Let existing IML algorithms solve the task and build add-on module to deal with safety

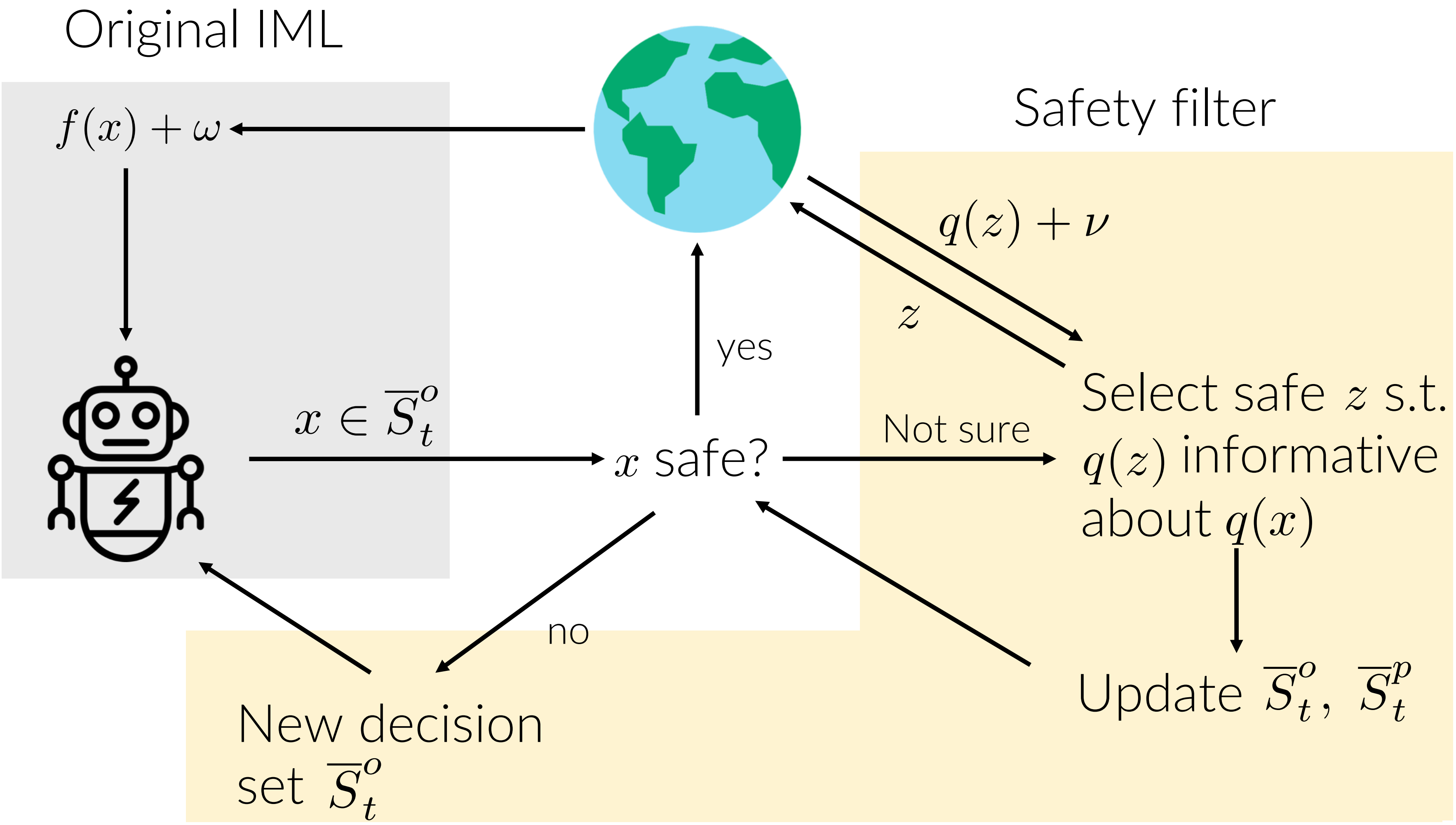
Consider the set of optimistically safe points $\overline{S}_t^o$



Exploit existing IML algorithms

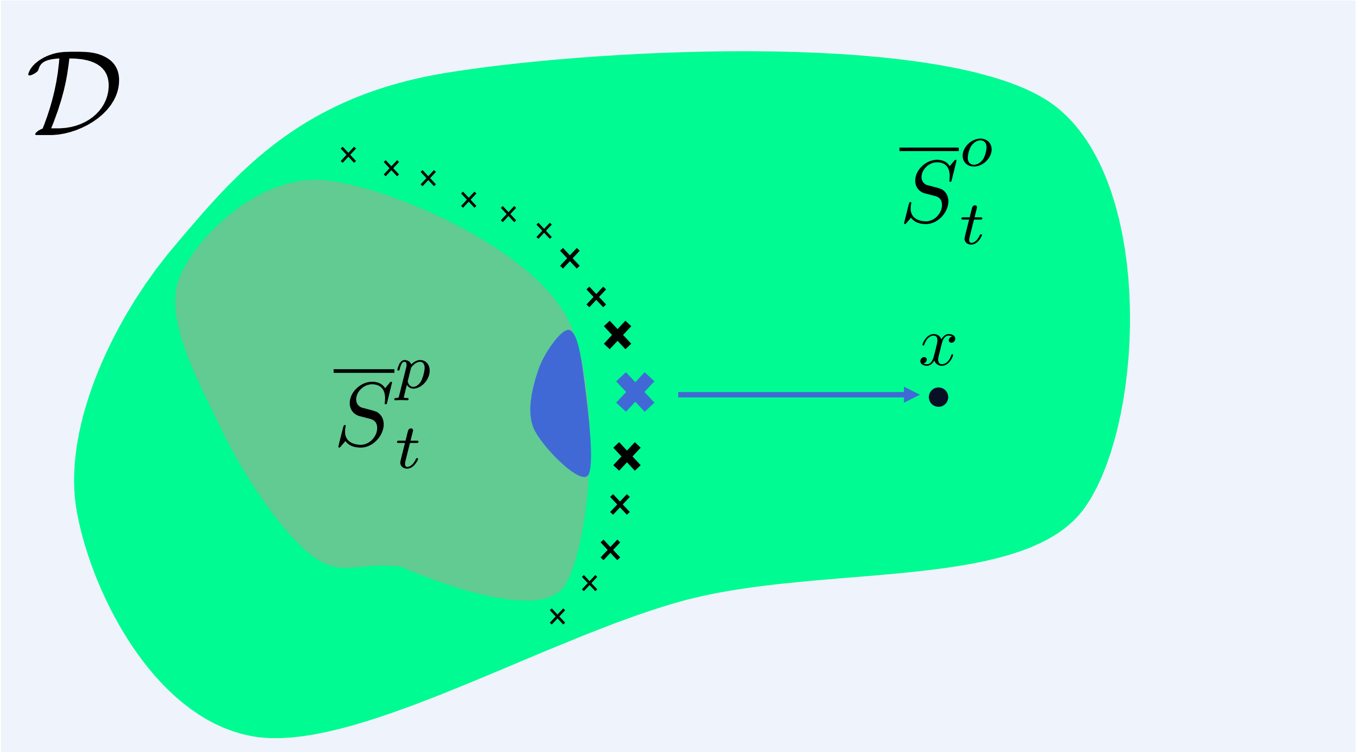
Learn about safety only when necessary

IML algorithm considers only plausibly safe decisions

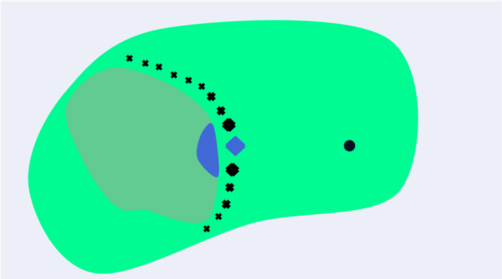


Heuristic-based expansion of the safe set

- Define a heuristic $h_t : \mathcal{D} \rightarrow \mathbb{R}$ to measure how informative $q(z)$ is about $q(x)$
- Order uncertain points by heuristic value (cross size)
- Find the point with highest heuristic, ✕
- Explore the safe points that could add ✕ to the safe set (blue shaded region)



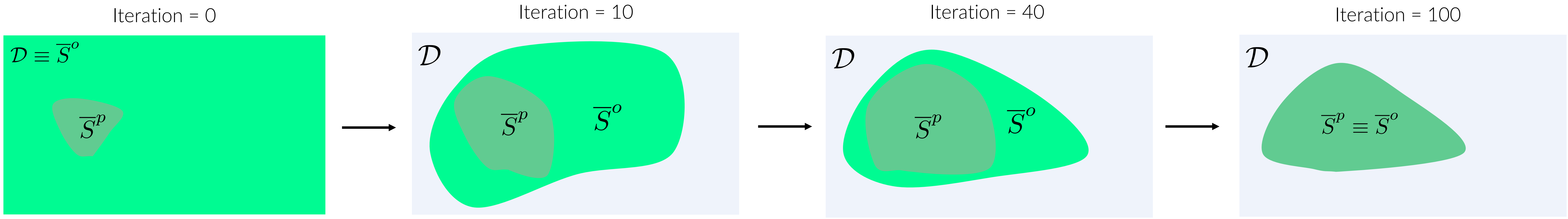
Previous methods	GoOSE
• Breadth-first search like • Reason about uncertainty inside the safe set	• A* like • Reason about uncertainty outside the safe set



Guarantees

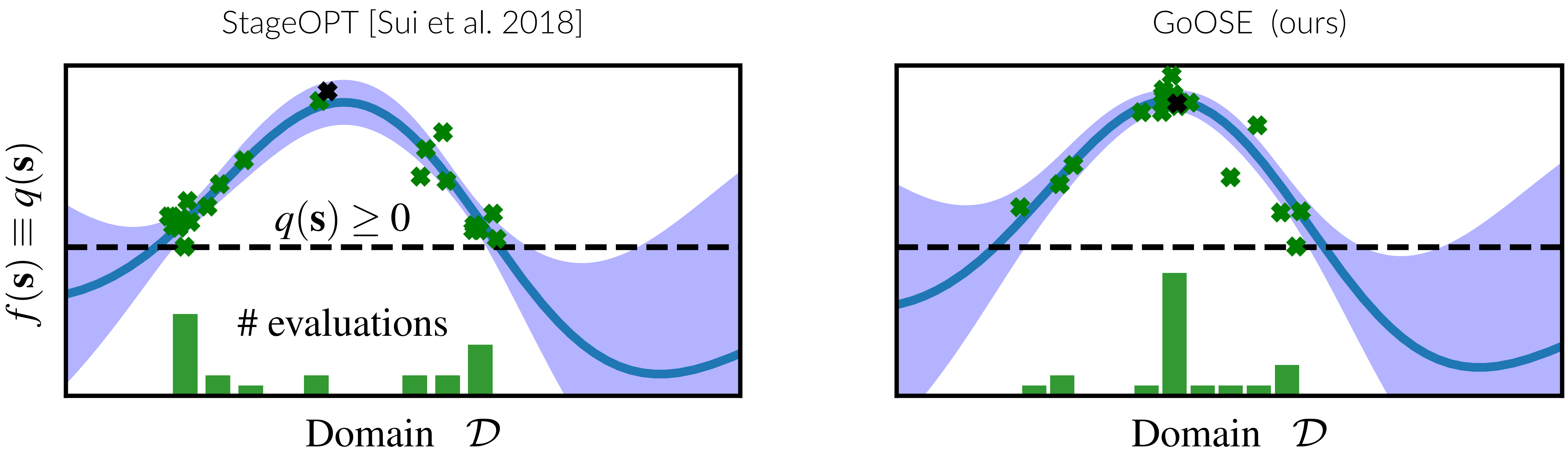
- Sampling inside $\overline{S}_t^p$ guarantees safety with high probability

- If necessary for the IML algorithm, the optimistic and pessimistic estimates of the safe set converge to a natural notion of largest safe reachable set up to a tolerance in a finite number of time steps



- Thus, except for a finite amount of iterations dedicated to the expansion of the safe set, the IML algorithm performs as if it had knowledge of the largest safe reachable set from the beginning (e.g. retains no-regret properties)

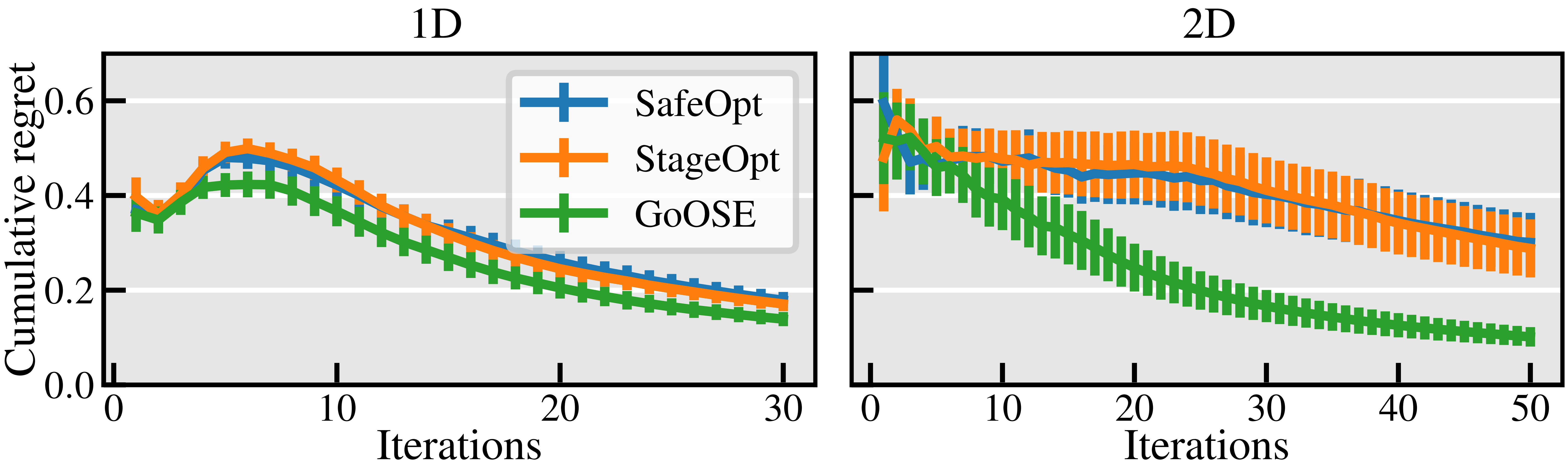
Qualitative comparison for a 1D optimization task



Quantitative comparison for optimization task

Algorithms: SafeOPT [Sui et al. 2015], StageOPT [Sui et al. 2018], GoOSE (ours)

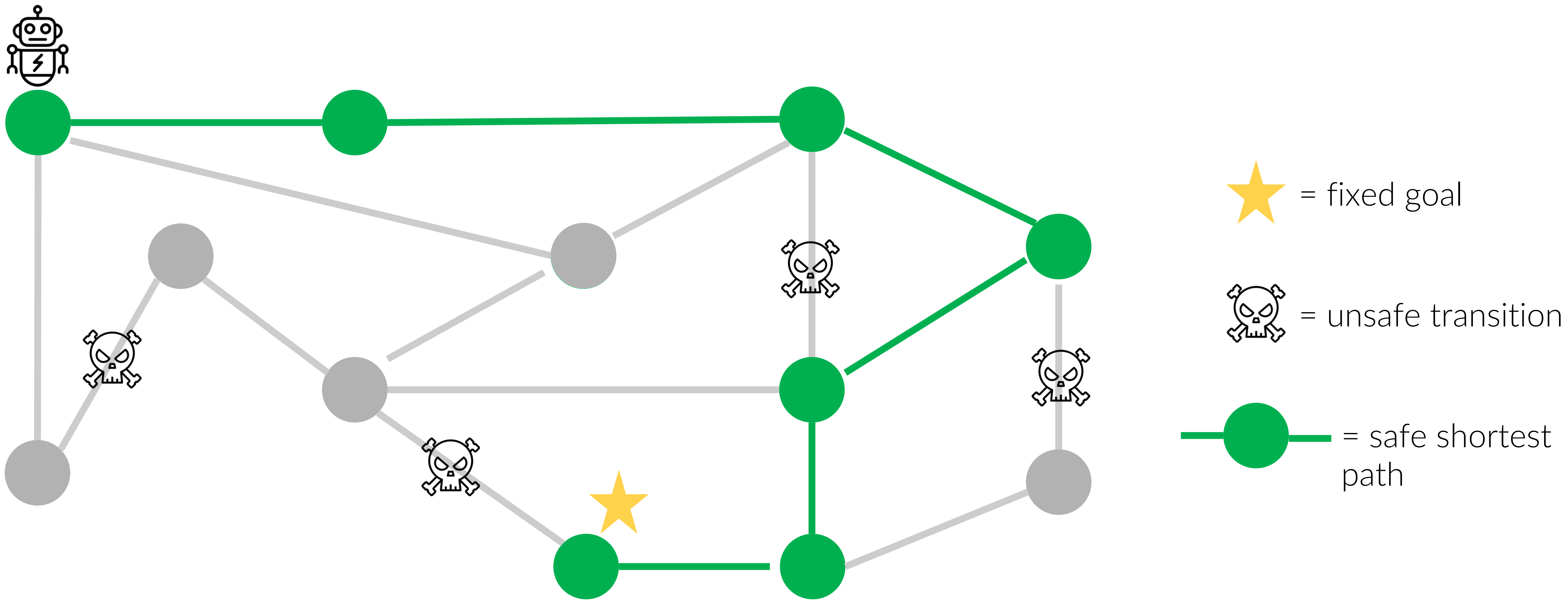
Safe average regret: $\frac{1}{T}\sum_{t}^{T} \mathrm{argmax}_{x \in A(S_0)} f(x) - f(x_t)$ where $A(S_0)$ is the largest safe set reachable from S_0



Safe shortest path in deterministic MDPs

Assumptions:

- Known, deterministic model
- Unsafe transitions unknown a priori

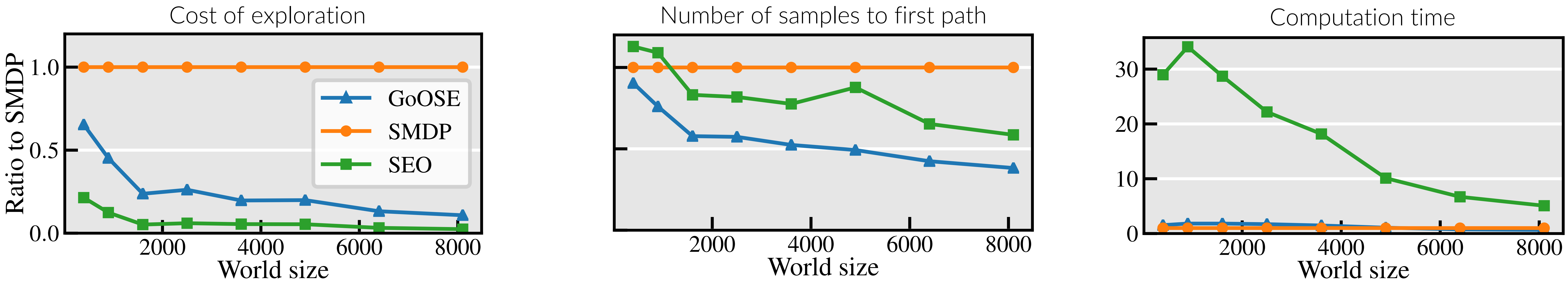


Comparison for safe shortest path in deterministic MDPs

Algorithms: SMDP [Turchetta al. 2016], SEO [Wachi et al. 2018] (optimizes exploration cost), GoOSE (ours, optimizes sample efficiency)

Setting: 100 random synthetic squared maps with size 20,30,…,90 = 800 synthetic maps

Plot: geometric mean of ratio with respect to uninformed baseline (SMDP)



Setting: 4 start-goal destination pairs on 16 maps of different areas on Mars = 64 scenarios

Table: geometric mean of ratio wrt SMDP

	GoOSE	SEO
Sample	**30.0** %	38.4 %
Cost	12.7 %	**0.7** %
Time	**37.8** %	518 %

Conclusions

We introduced GoOSE, an add-on module for general IML algorithms that:

- Provides high probability **safety guarantees**
- **Preserves properties** over the IML algorithm over the largest safe reachable set
- Is applicable to a **wide range of problems**, including safe Bayesian optimization, safe active learning and safe exploration in deterministic Markov decision processes
- Greatly **improves** the empirical **sample efficiency** over existing methods while retaining the same worst case sample complexity