Transductive Active Learning

with Application to Safe Bayesian Optimization

Jonas Hübotter

with Bhavya Sukhija, Lenart Treven, Yarden As, Andreas Krause

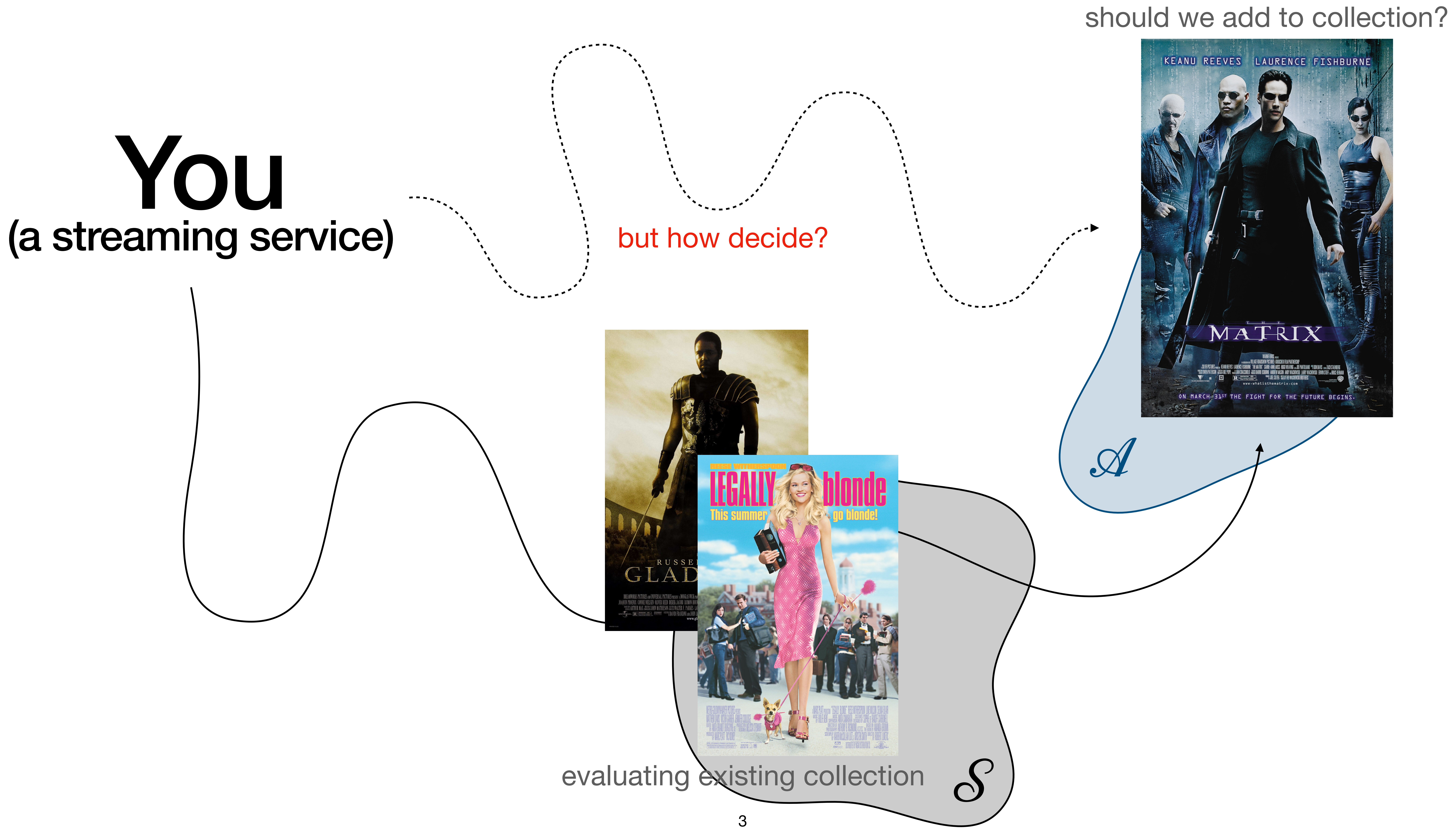



should we add to collection?

You
(a streaming service)

but how decide?

evaluating existing collection

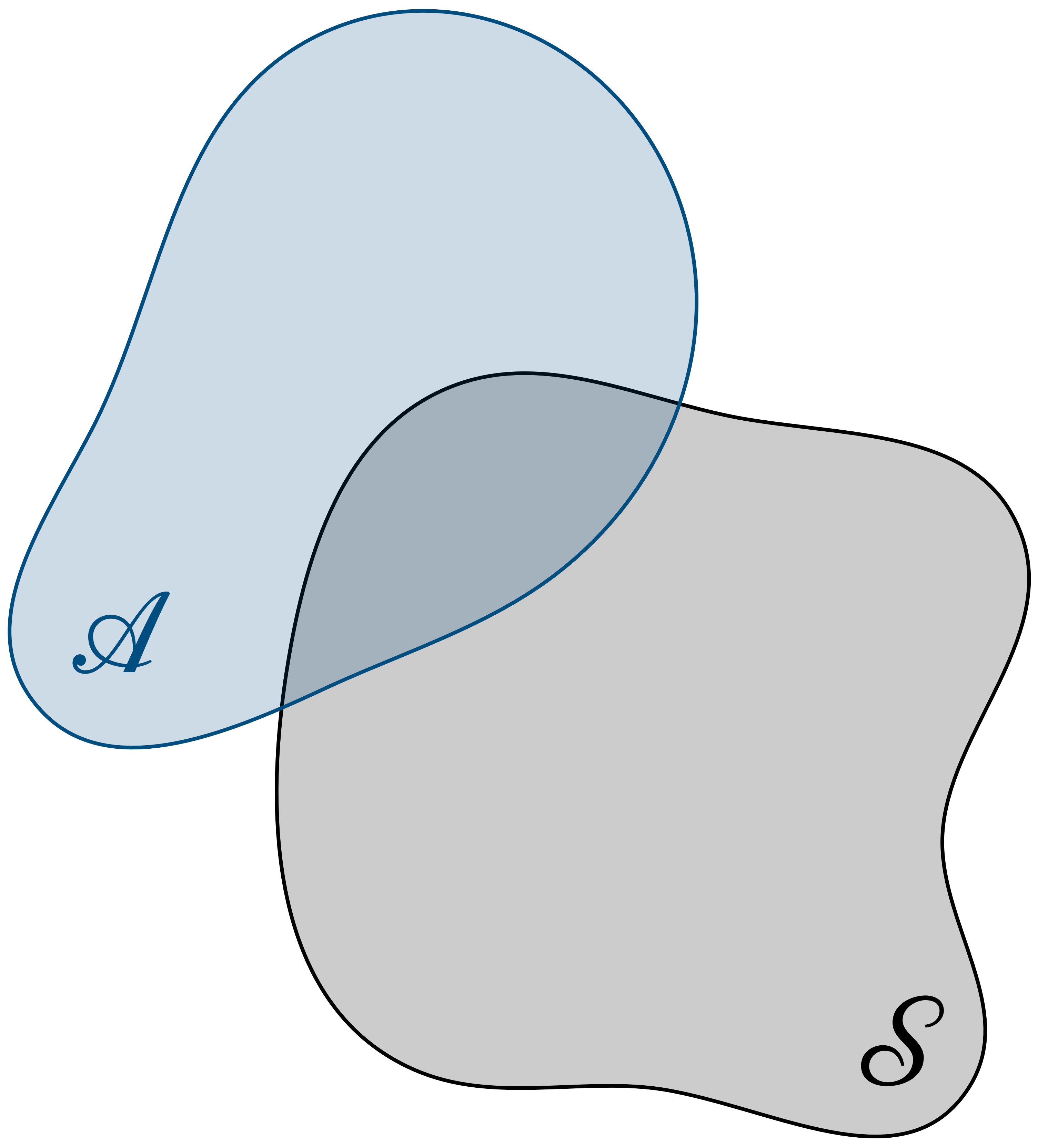

should we add to collection?
You
(a streaming service)
but how decide?
KEANU REEVES
LAURENCE FISHBURNE
THE MATRIX
ON MARCH 31ST THE FIGHT FOR THE FUTURE BEGINS.
$\mathscr{A}$
LEGALLY blonde
This summer
go blonde!
evaluating existing collection
$\mathscr{S}$

Setting

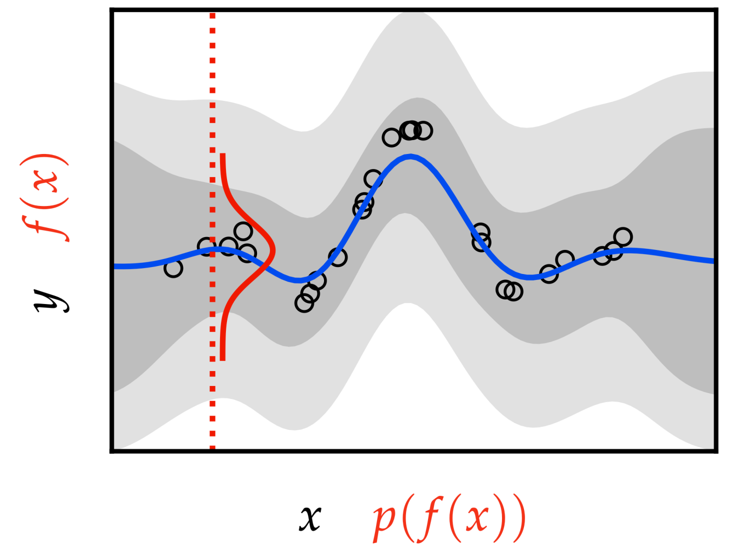
- $\mathcal{S} \subseteq \mathcal{X}$ - sample space
- $\mathcal{A} \subseteq \mathcal{X}$ - target space
- Unknown function f over $\mathcal{X}$

Goal: Learn f within $\mathcal{A}$ by sampling from $\mathcal{S}$

We call this **Transductive Active Learning**

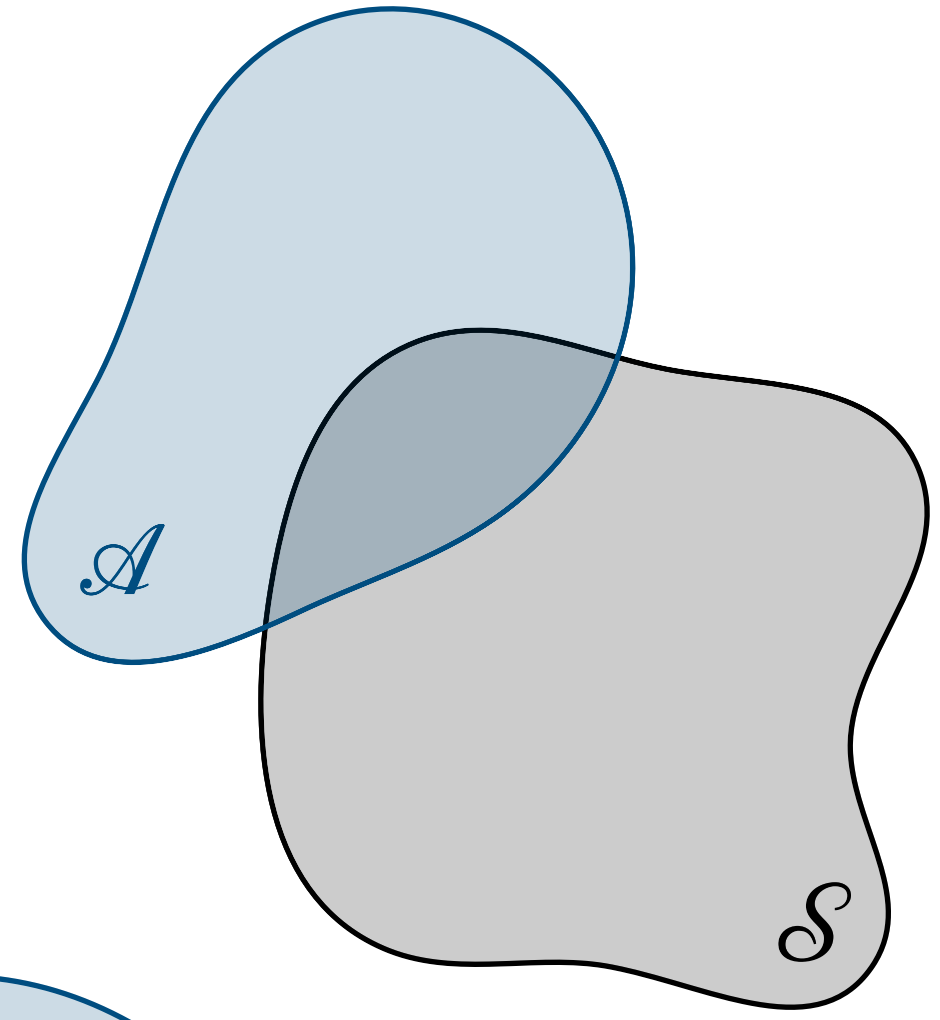


Assume for us: $\mathcal{S}$, $\mathcal{A}$ finite, and we model f by a Gaussian process



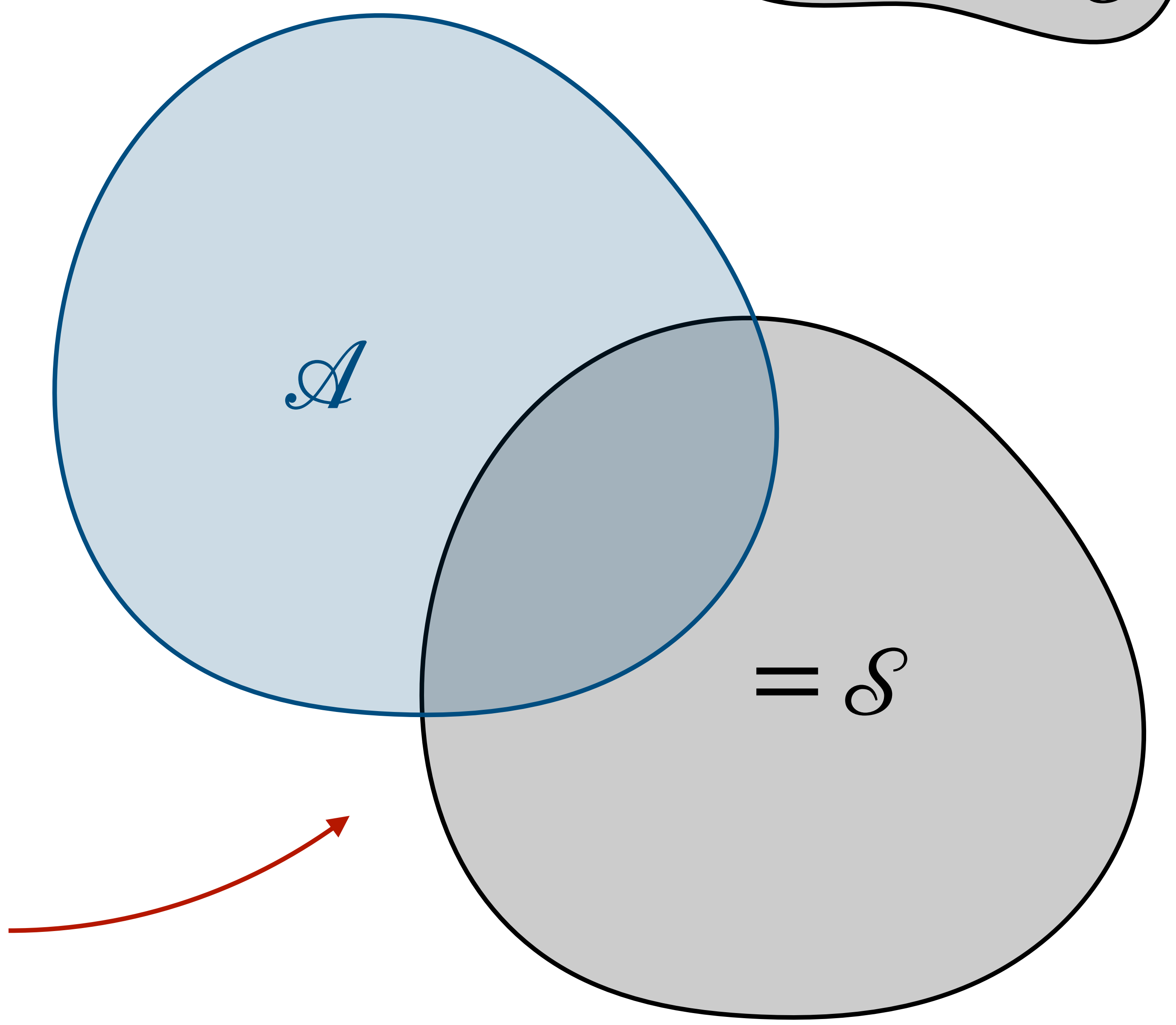
Transductive Active Learning

“only learn what is needed to solve a given task”



(Inductive) Active Learning

“learn as much as you can”



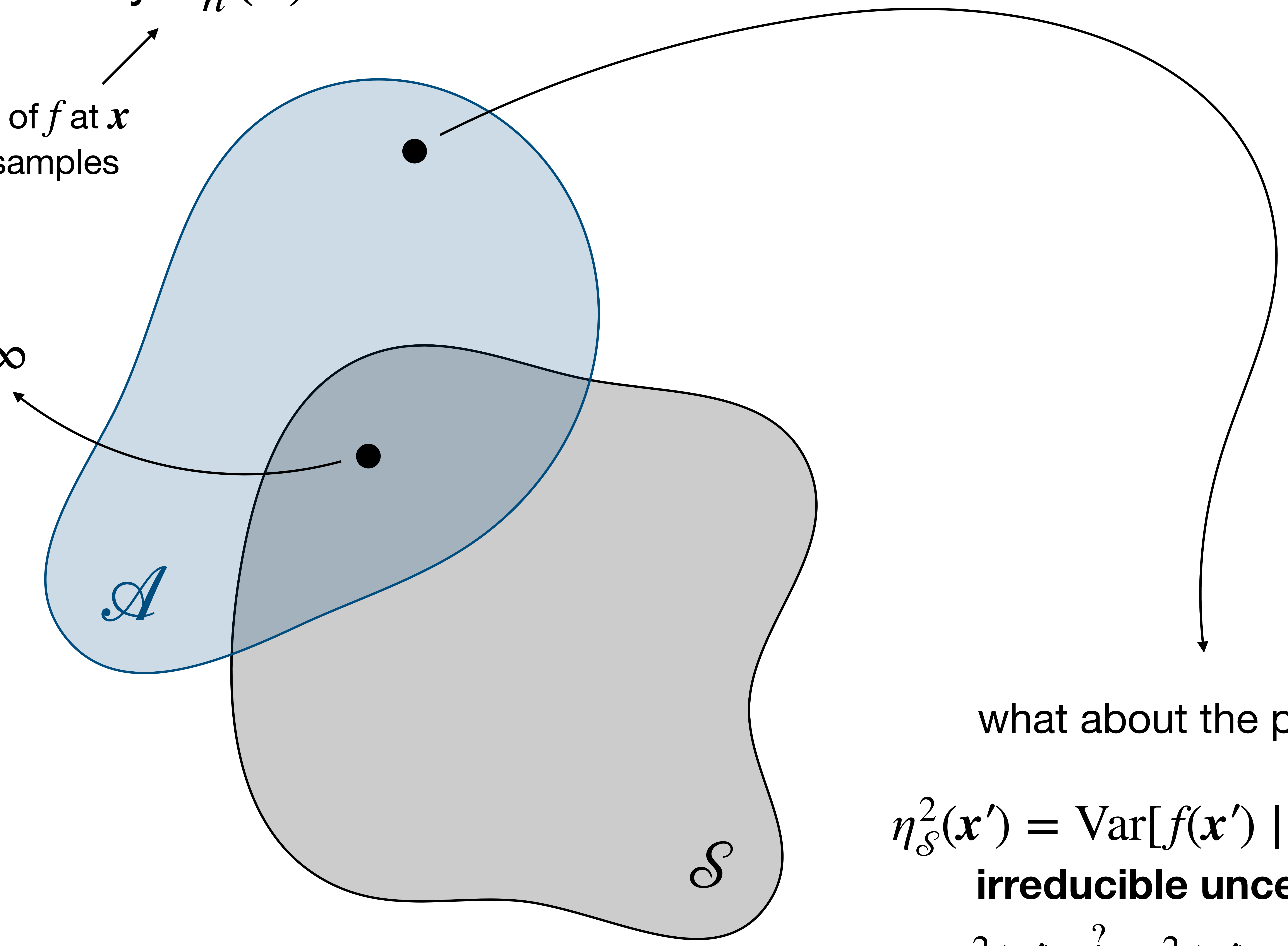
studied in most prior works

Goal: Reduce uncertainty $\sigma_n^2(\boldsymbol{x})$ at $\boldsymbol{x} \in \mathscr{A}$

variance of f at $\boldsymbol{x}$ after n samples

$\sigma_n^2(\boldsymbol{x}) \to 0$ as $n \to \infty$

e.g., by repeatedly sampling $\boldsymbol{x}$



what about the point $\boldsymbol{x}'$?

$\eta_{\mathscr{S}}^2(\boldsymbol{x}') = \mathrm{Var}[f(\boldsymbol{x}') \mid \boldsymbol{f}(\mathscr{S})]$ is the **irreducible uncertainty**:

$\sigma_n^2(\boldsymbol{x}') \stackrel{?}{\to} \eta_{\mathscr{S}}^2(\boldsymbol{x}')$ as $n \to \infty$

An Algorithmic Framework for TAL

Proposal: select the next sample to minimize *posterior* uncertainty within $\mathcal{A}$

VTL: $$\boldsymbol{x}_n = \arg\min_{\boldsymbol{x}_n \in \mathcal{S}} \sum_{\boldsymbol{x} \in \mathcal{A}} \mathrm{Var}[f(\boldsymbol{x}) \mid D_{n-1}, (\boldsymbol{x}_n, f(\boldsymbol{x}_n) + \varepsilon)]$$

MacKay, 1992; Seo et al., 2000; Yu et al., 2006

Generalization bound for VTL (informal). $\forall \boldsymbol{x}' \in \mathcal{A}$:

$$\sigma_n^2(\boldsymbol{x}') \leq \eta_{\mathcal{S}}^2(\boldsymbol{x}') + C \log n / \sqrt{n}$$

(C is a constant)

An Algorithmic Framework for TAL

Proposal: select the next sample to minimize *posterior* uncertainty within $\mathcal{A}$

VTL: $$\boldsymbol{x}_n = \arg\min_{\boldsymbol{x}_n \in \mathcal{S}} \sum_{\boldsymbol{x} \in \mathcal{A}} \mathrm{Var}[f(\boldsymbol{x}) \mid D_{n-1}, (\boldsymbol{x}_n, f(\boldsymbol{x}_n) + \varepsilon)]$$

<u>Agnostic</u> error bound for VTL (informal). If $f \in \mathcal{H}_k(\mathcal{X})$, then $\forall \boldsymbol{x}' \in \mathcal{A}$ wp $1 - \delta$:

$$|f(\boldsymbol{x}') - \underbrace{\mathbb{E}[f(\boldsymbol{x}') \mid D_n]}_{\text{prediction}}|^2 \leq \beta_n^2(\delta)\Big[\underbrace{\eta_{\mathcal{S}}^2(\boldsymbol{x}')}_{\text{irreducible}} + \underbrace{C \log n/\sqrt{n}}_{\text{reducible}}\Big]$$

(C is a constant)

Push test

Tuning legged locomotion controllers via safe bayesian optimization (Widmer et al.; 2023)

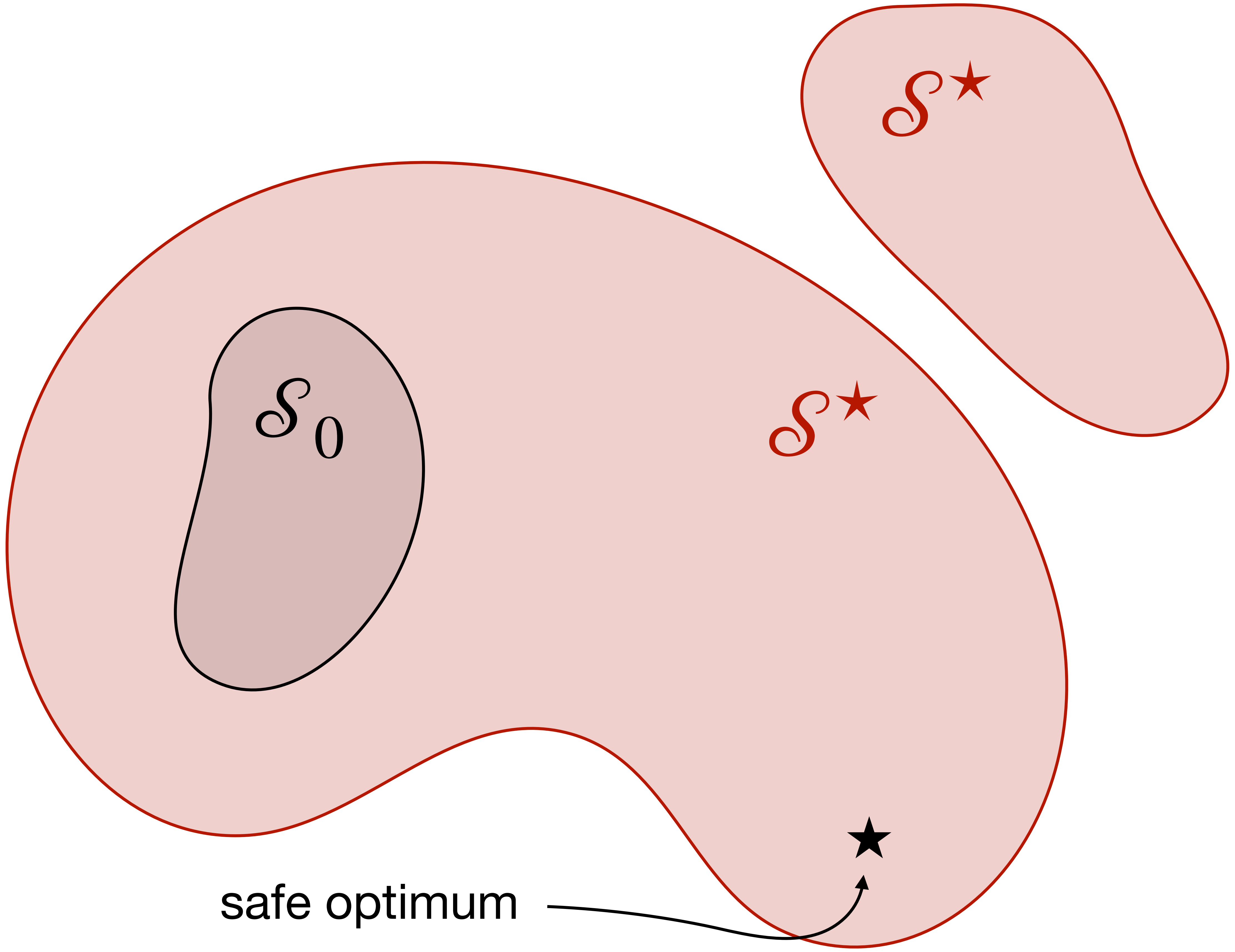
Example: Safe Bayesian Optimization

Under constraint $c^\star$ inducing *true* safe set $\mathcal{S}^\star = \{\boldsymbol{x} \mid c^\star(\boldsymbol{x}) \geq 0\}$, find $\arg\max_{\boldsymbol{x}\in\mathcal{S}^\star} f^\star(\boldsymbol{x})$.

Assumption: well-calibrated model $\boldsymbol{f}$

- $l_n^f(\boldsymbol{x}) \leq f^\star(\boldsymbol{x}) \leq u_n^f(\boldsymbol{x})$,
  $l_n^c(\boldsymbol{x}) \leq c^\star(\boldsymbol{x}) \leq u_n^c(\boldsymbol{x})$
- $\mathcal{S}_n = \{\boldsymbol{x} \mid l_n^c(\boldsymbol{x}) \geq 0\}$ pessimistic
  $\mathcal{S}_n^o = \{\boldsymbol{x} \mid u_n^c(\boldsymbol{x}) \geq 0\}$ optimistic

  $\rightarrow \mathcal{S}_n \subseteq \mathcal{S}^\star \subseteq \mathcal{S}_n^o$

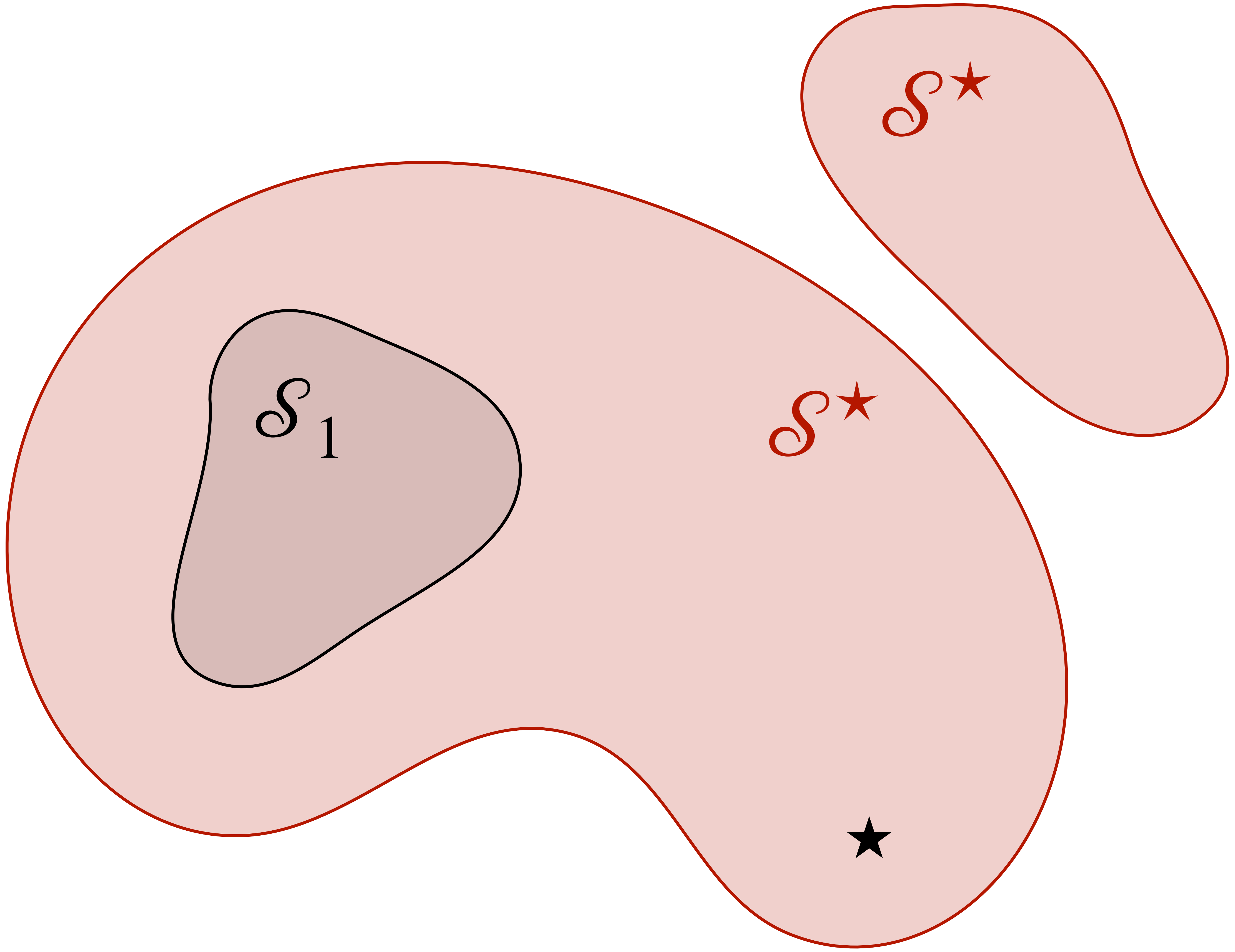


Example: Safe Bayesian Optimization

Under constraint $c^\star$ inducing *true* safe set $\mathcal{S}^\star = \{\boldsymbol{x} \mid c^\star(\boldsymbol{x}) \geq 0\}$,
find $\arg\max_{\boldsymbol{x} \in \mathcal{S}^\star} f^\star(\boldsymbol{x})$.

Assumption: well-calibrated model $\boldsymbol{f}$

- $l_n^f(\boldsymbol{x}) \leq f^\star(\boldsymbol{x}) \leq u_n^f(\boldsymbol{x})$,
  $l_n^c(\boldsymbol{x}) \leq c^\star(\boldsymbol{x}) \leq u_n^c(\boldsymbol{x})$
- $\mathcal{S}_n = \{\boldsymbol{x} \mid l_n^c(\boldsymbol{x}) \geq 0\}$ pessimistic
  $\mathcal{S}_n^o = \{\boldsymbol{x} \mid u_n^c(\boldsymbol{x}) \geq 0\}$ optimistic
  $\rightarrow \mathcal{S}_n \subseteq \mathcal{S}^\star \subseteq \mathcal{S}_n^o$

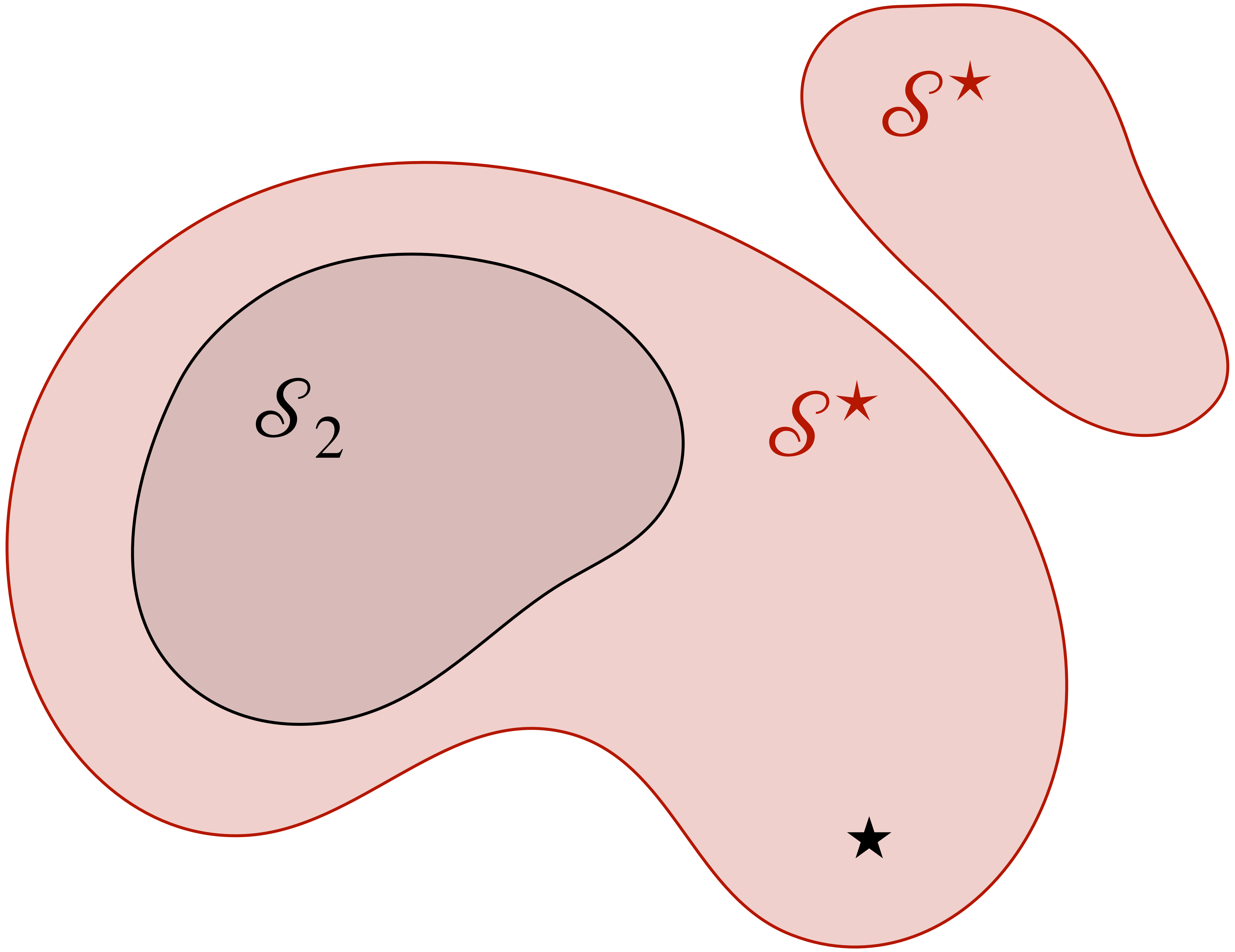


Example: Safe Bayesian Optimization

Under constraint $c^\star$ inducing *true* safe set $\mathcal{S}^\star = \{\boldsymbol{x} \mid c^\star(\boldsymbol{x}) \geq 0\}$,
find $\arg\max_{\boldsymbol{x} \in \mathcal{S}^\star} f^\star(\boldsymbol{x})$.

Assumption: well-calibrated model $\boldsymbol{f}$

- $l_n^f(\boldsymbol{x}) \leq f^\star(\boldsymbol{x}) \leq u_n^f(\boldsymbol{x})$,
  $l_n^c(\boldsymbol{x}) \leq c^\star(\boldsymbol{x}) \leq u_n^c(\boldsymbol{x})$
- $\mathcal{S}_n = \{\boldsymbol{x} \mid l_n^c(\boldsymbol{x}) \geq 0\}$ pessimistic
  $\mathcal{S}_n^o = \{\boldsymbol{x} \mid u_n^c(\boldsymbol{x}) \geq 0\}$ optimistic
  $\rightarrow \mathcal{S}_n \subseteq \mathcal{S}^\star \subseteq \mathcal{S}_n^o$

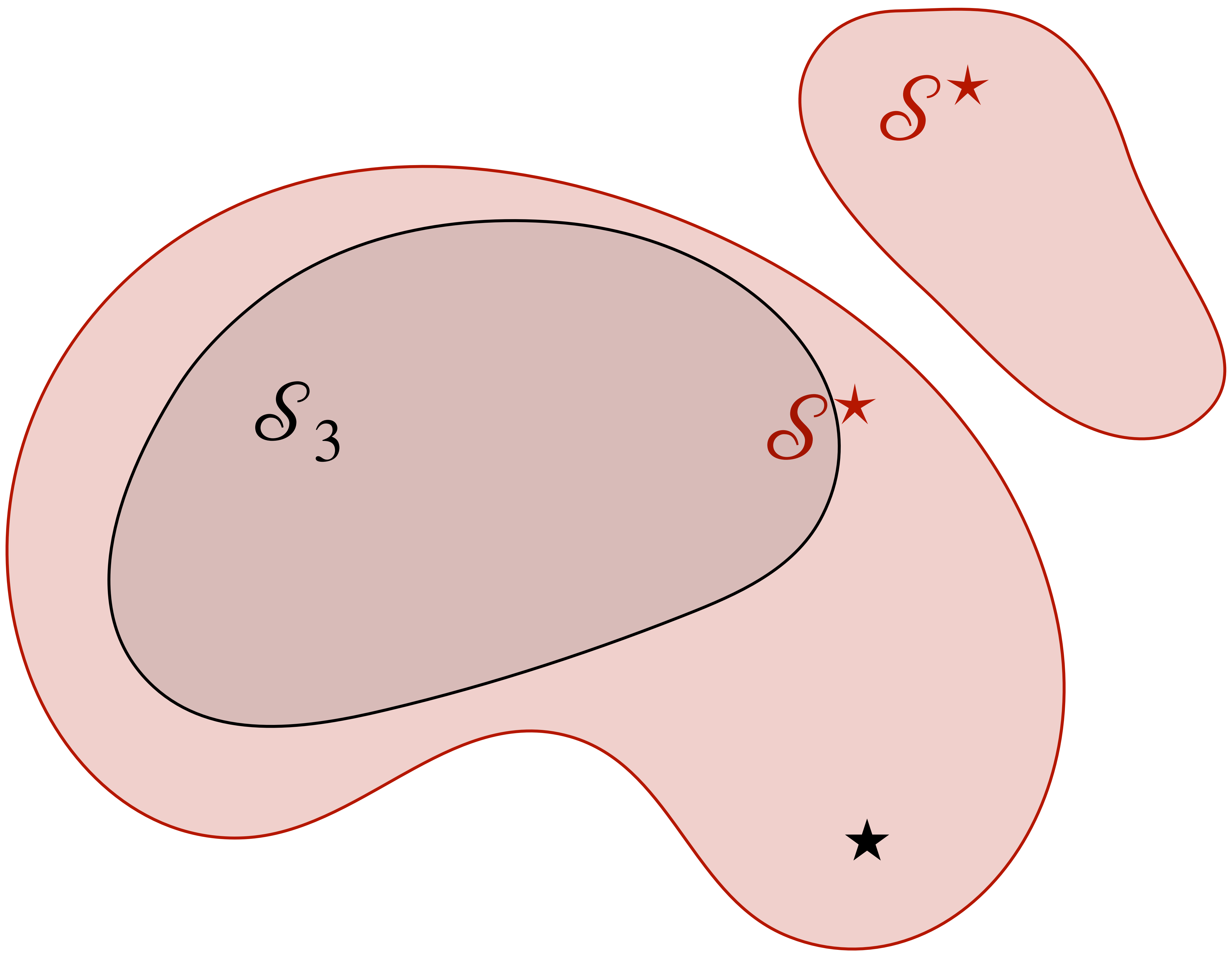


Example: Safe Bayesian Optimization

Under constraint $c^\star$ inducing *true* safe set $\mathcal{S}^\star = \{\boldsymbol{x} \mid c^\star(\boldsymbol{x}) \geq 0\}$, find $\arg\max_{\boldsymbol{x}\in\mathcal{S}^\star} f^\star(\boldsymbol{x})$.

Assumption: well-calibrated model $\boldsymbol{f}$

- $l_n^f(\boldsymbol{x}) \leq f^\star(\boldsymbol{x}) \leq u_n^f(\boldsymbol{x})$,
  $l_n^c(\boldsymbol{x}) \leq c^\star(\boldsymbol{x}) \leq u_n^c(\boldsymbol{x})$
- $\mathcal{S}_n = \{\boldsymbol{x} \mid l_n^c(\boldsymbol{x}) \geq 0\}$ pessimistic
  $\mathcal{S}_n^o = \{\boldsymbol{x} \mid u_n^c(\boldsymbol{x}) \geq 0\}$ optimistic
  $\rightarrow \mathcal{S}_n \subseteq \mathcal{S}^\star \subseteq \mathcal{S}_n^o$

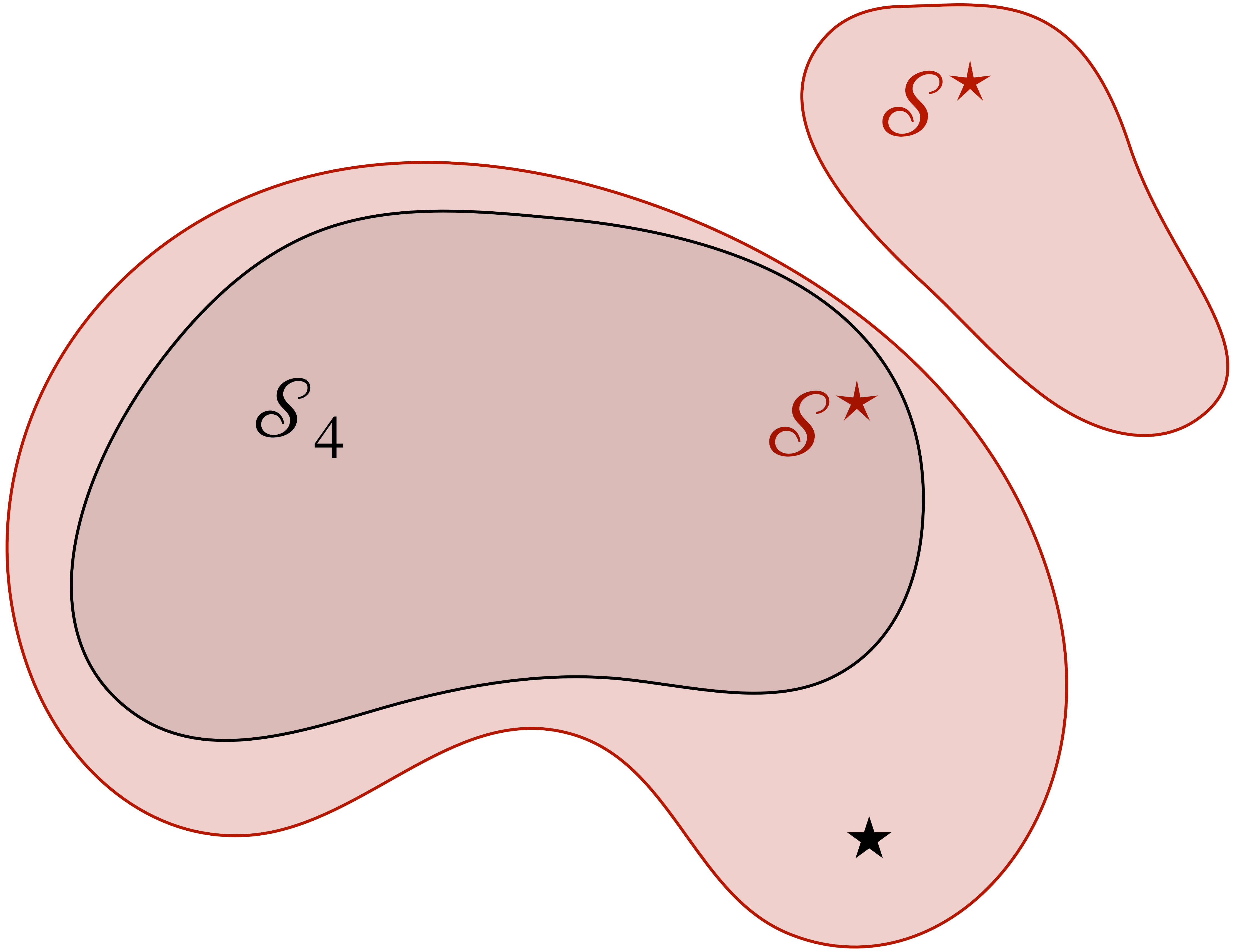


Example: Safe Bayesian Optimization

Under constraint $c^\star$ inducing *true* safe set $\mathcal{S}^\star = \{\boldsymbol{x} \mid c^\star(\boldsymbol{x}) \geq 0\}$,
find $\arg\max_{\boldsymbol{x}\in\mathcal{S}^\star} f^\star(\boldsymbol{x})$.

Assumption: well-calibrated model $\boldsymbol{f}$

- $l_n^f(\boldsymbol{x}) \leq f^\star(\boldsymbol{x}) \leq u_n^f(\boldsymbol{x})$,
  $l_n^c(\boldsymbol{x}) \leq c^\star(\boldsymbol{x}) \leq u_n^c(\boldsymbol{x})$
- $\mathcal{S}_n = \{\boldsymbol{x} \mid l_n^c(\boldsymbol{x}) \geq 0\}$ pessimistic
  $\mathcal{S}_n^o = \{\boldsymbol{x} \mid u_n^c(\boldsymbol{x}) \geq 0\}$ optimistic
  $\rightarrow \mathcal{S}_n \subseteq \mathcal{S}^\star \subseteq \mathcal{S}_n^o$



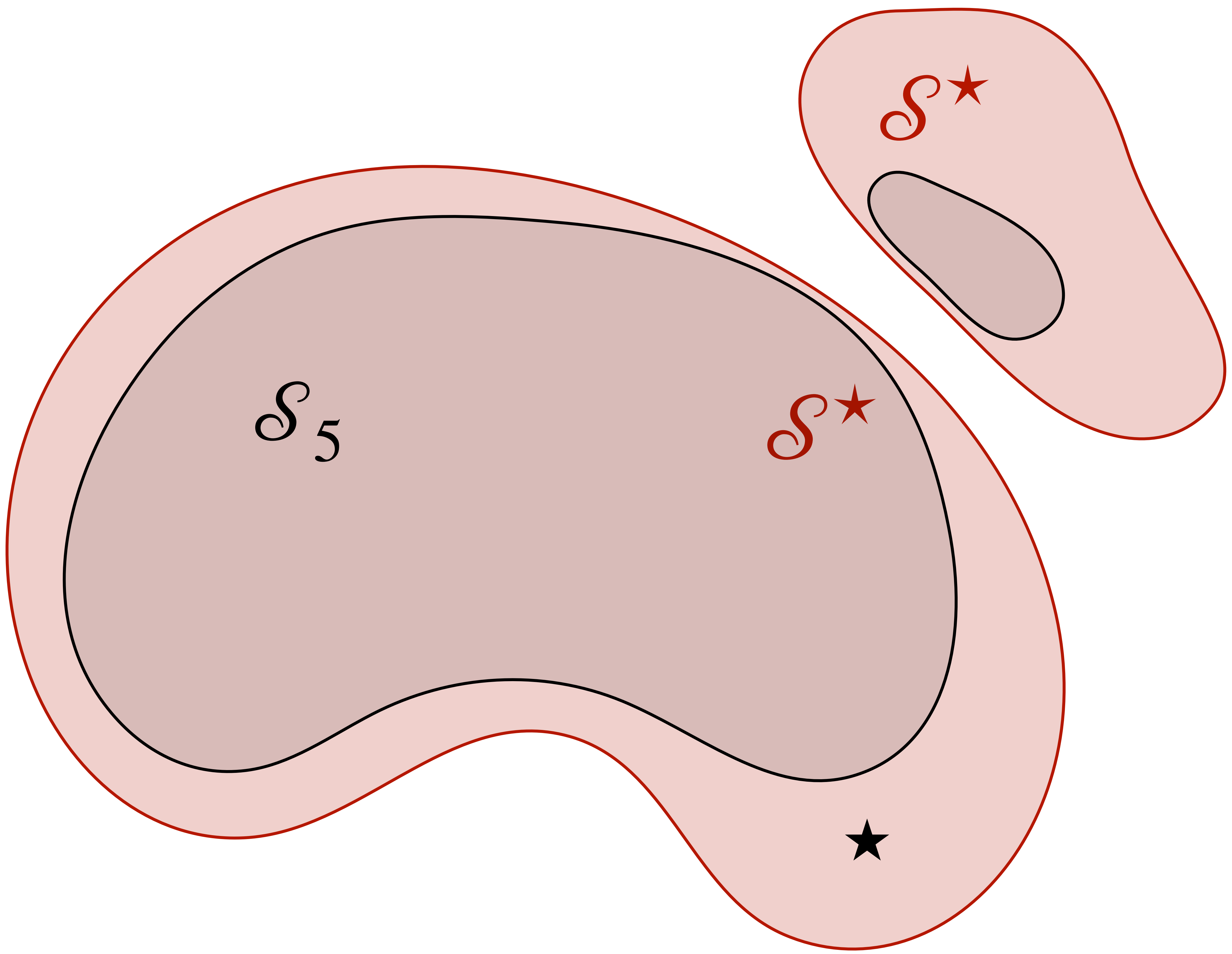
Example: Safe Bayesian Optimization

Under constraint $c^\star$ inducing *true* safe set $\mathcal{S}^\star = \{\boldsymbol{x} \mid c^\star(\boldsymbol{x}) \geq 0\}$,
find $\arg\max_{\boldsymbol{x} \in \mathcal{S}^\star} f^\star(\boldsymbol{x})$.

Assumption: well-calibrated model $\boldsymbol{f}$

- $l_n^f(\boldsymbol{x}) \leq f^\star(\boldsymbol{x}) \leq u_n^f(\boldsymbol{x})$,
  $l_n^c(\boldsymbol{x}) \leq c^\star(\boldsymbol{x}) \leq u_n^c(\boldsymbol{x})$

- $\mathcal{S}_n = \{\boldsymbol{x} \mid l_n^c(\boldsymbol{x}) \geq 0\}$ pessimistic
  $\mathcal{S}_n^o = \{\boldsymbol{x} \mid u_n^c(\boldsymbol{x}) \geq 0\}$ optimistic
  $\rightarrow \mathcal{S}_n \subseteq \mathcal{S}^\star \subseteq \mathcal{S}_n^o$



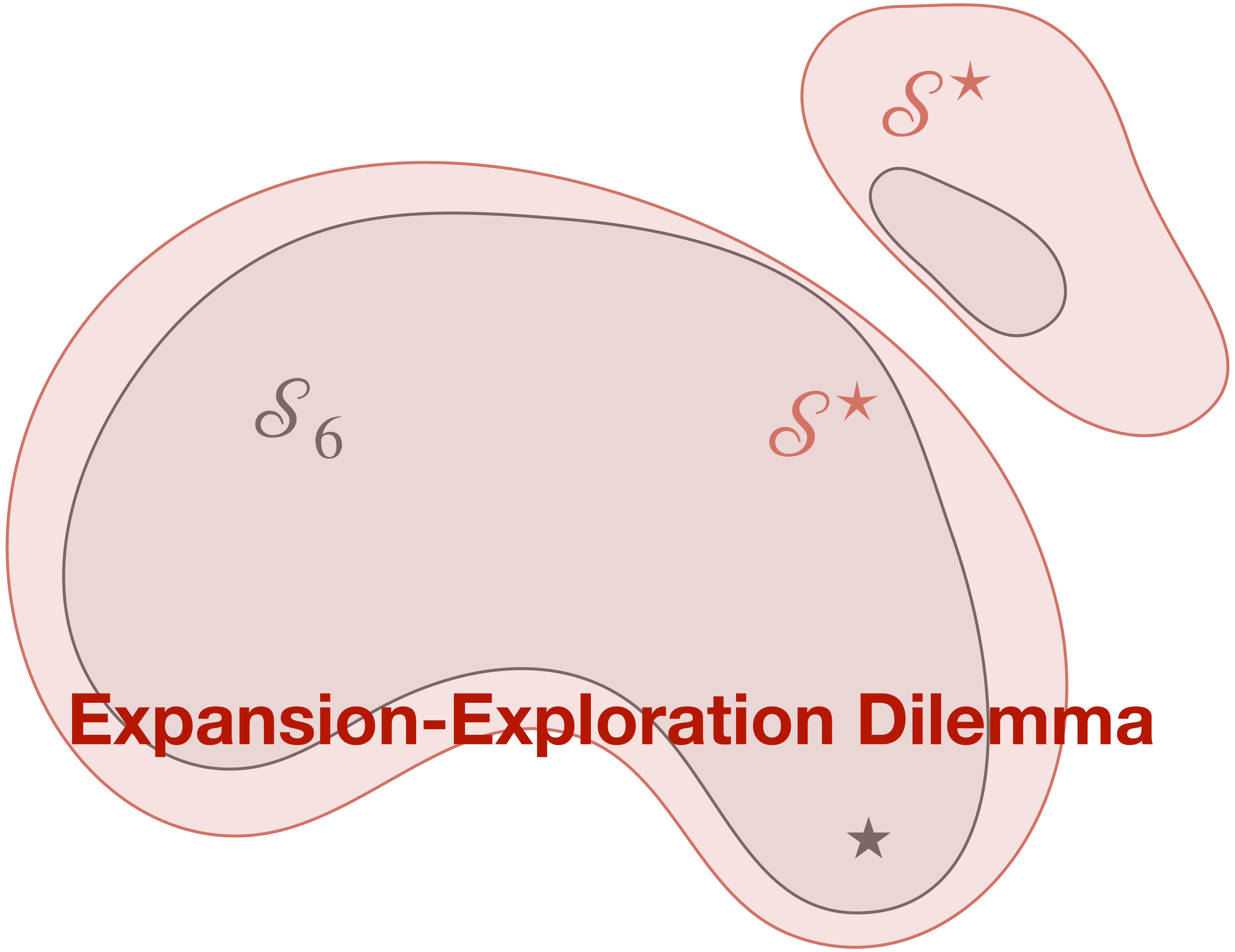
Example: Safe Bayesian Optimization

Under constraint $c^\star$ inducing *true* safe set $\mathcal{S}^\star = \{\boldsymbol{x} \mid c^\star(\boldsymbol{x}) \geq 0\}$, find $\arg\max_{\boldsymbol{x} \in \mathcal{S}^\star} f^\star(\boldsymbol{x})$.

Assumption: well-calibrated model $\boldsymbol{f}$

- $l_n^f(\boldsymbol{x}) \leq f^\star(\boldsymbol{x}) \leq u_n^f(\boldsymbol{x})$,
  $l_n^c(\boldsymbol{x}) \leq c^\star(\boldsymbol{x}) \leq u_n^c(\boldsymbol{x})$
- $\mathcal{S}_n = \{\boldsymbol{x} \mid l_n^c(\boldsymbol{x}) \geq 0\}$ pessimistic
  $\mathcal{S}_n^o = \{\boldsymbol{x} \mid u_n^c(\boldsymbol{x}) \geq 0\}$ optimistic

  $\rightarrow \mathcal{S}_n \subseteq \mathcal{S}^\star \subseteq \mathcal{S}_n^o$



Example: Safe Bayesian Optimization

Learn **potential maximizers**:

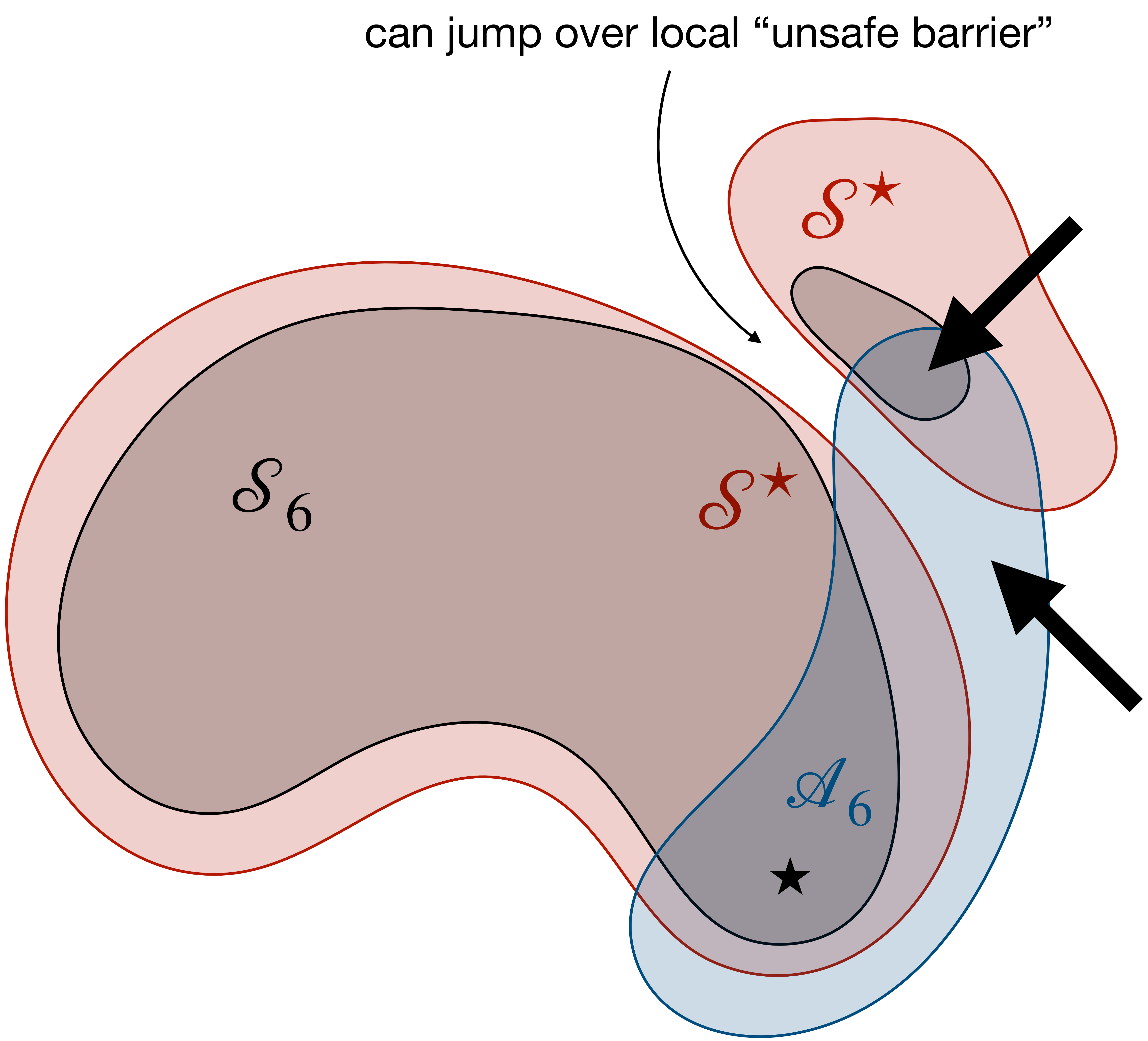
$$\mathcal{A}_n = \{\boldsymbol{x} \in \mathcal{S}_n^o \mid u_n^f(\boldsymbol{x}) \geq \max_{\boldsymbol{x}' \in \mathcal{S}_n} l_n^f(\boldsymbol{x}')\}$$

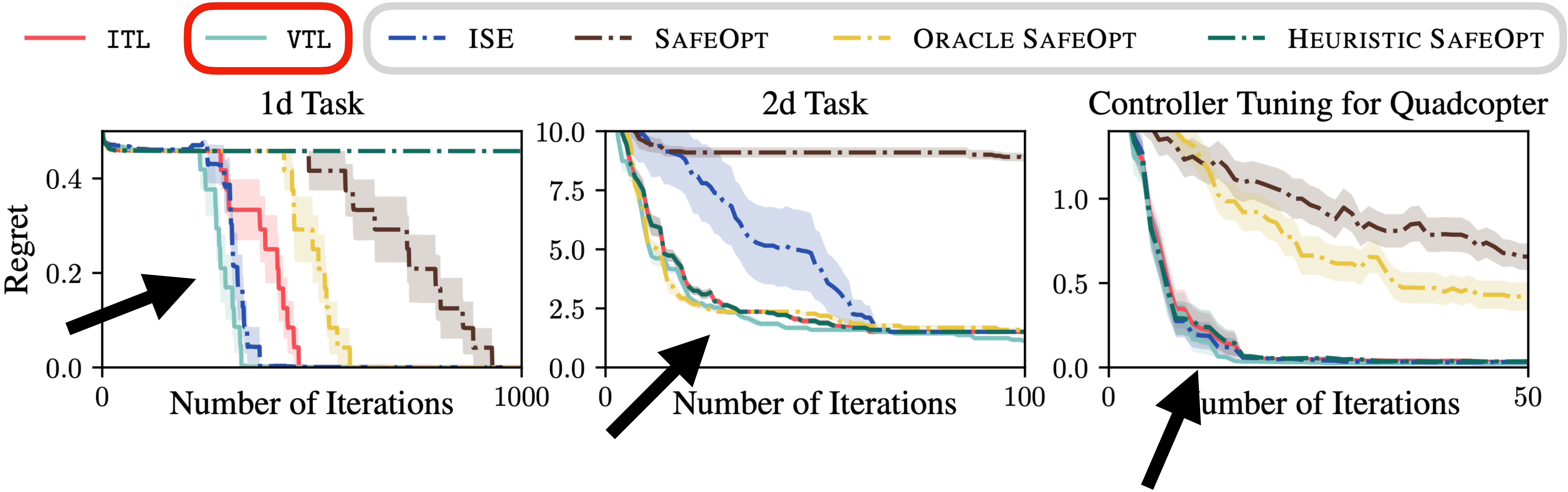
where $\mathcal{S}_n^o$ is the "optimistic" safe set

Theorem (informal):

If $f^\star, c^\star$ are sufficiently regular, VTL finds the safe reachable optimum.

Our results are *tighter* than those of prior works and *generalize* to continuous state spaces.





- VTL improves upon the sample efficiency of prior work
- **Why?** Framing Safe BO as TAL allows retrieving only the information that is needed to find the safe optimum

Summary

Transductive Active Learning ("*only learn what is needed*") is a ubiquitous problem, generalizing classical active learning ("*learn as much as you can*")

TAL has advantages over AL when

- the search space is large
- interaction time is limited
- access to parts of the search space is restricted

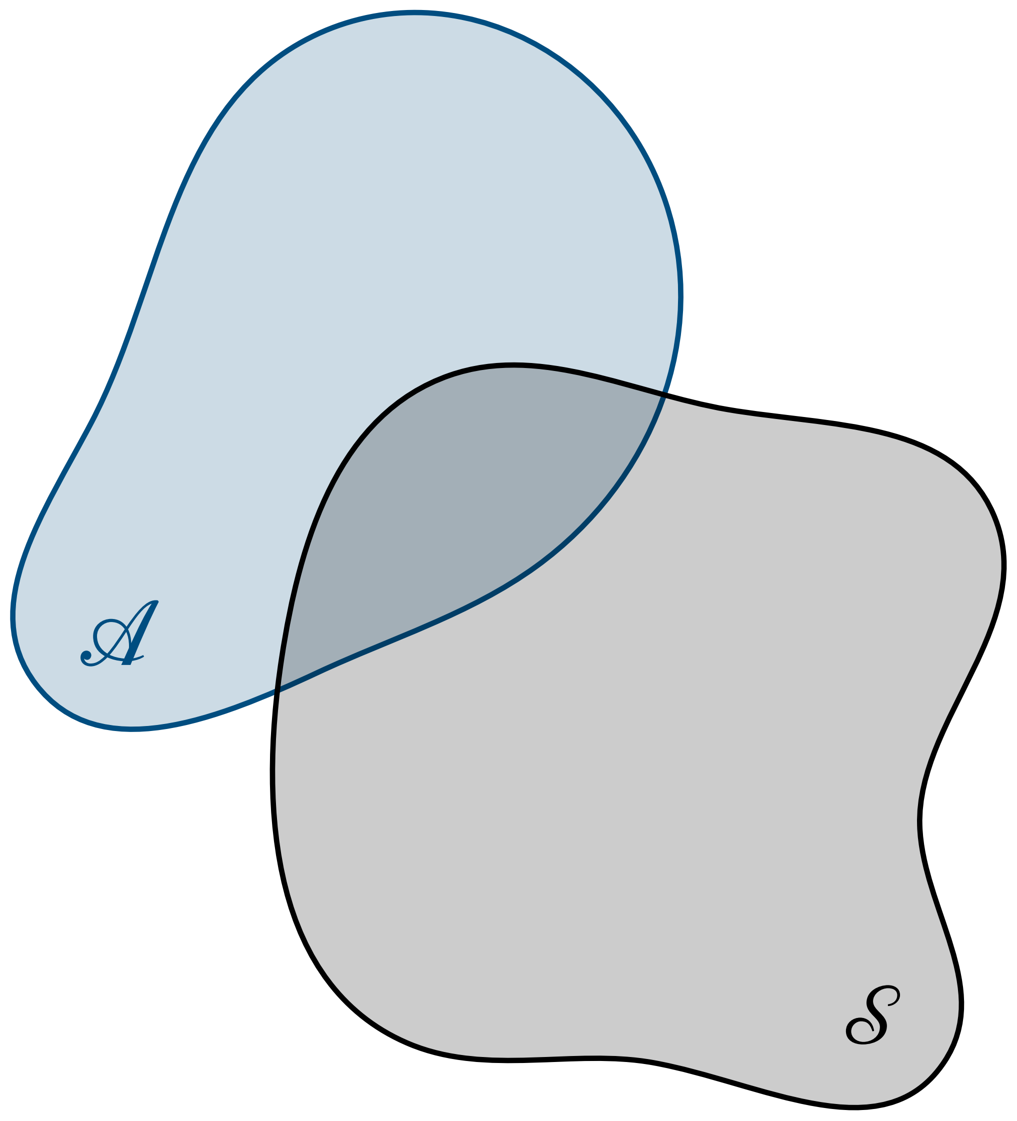
Safe Bayesian Optimization is just one such problem!

Transductive Active Learning

“only learn what is needed”

- sample space $\mathcal{S}$
- target space $\mathcal{A}$

TAL generalizes classical “inductive” active learning

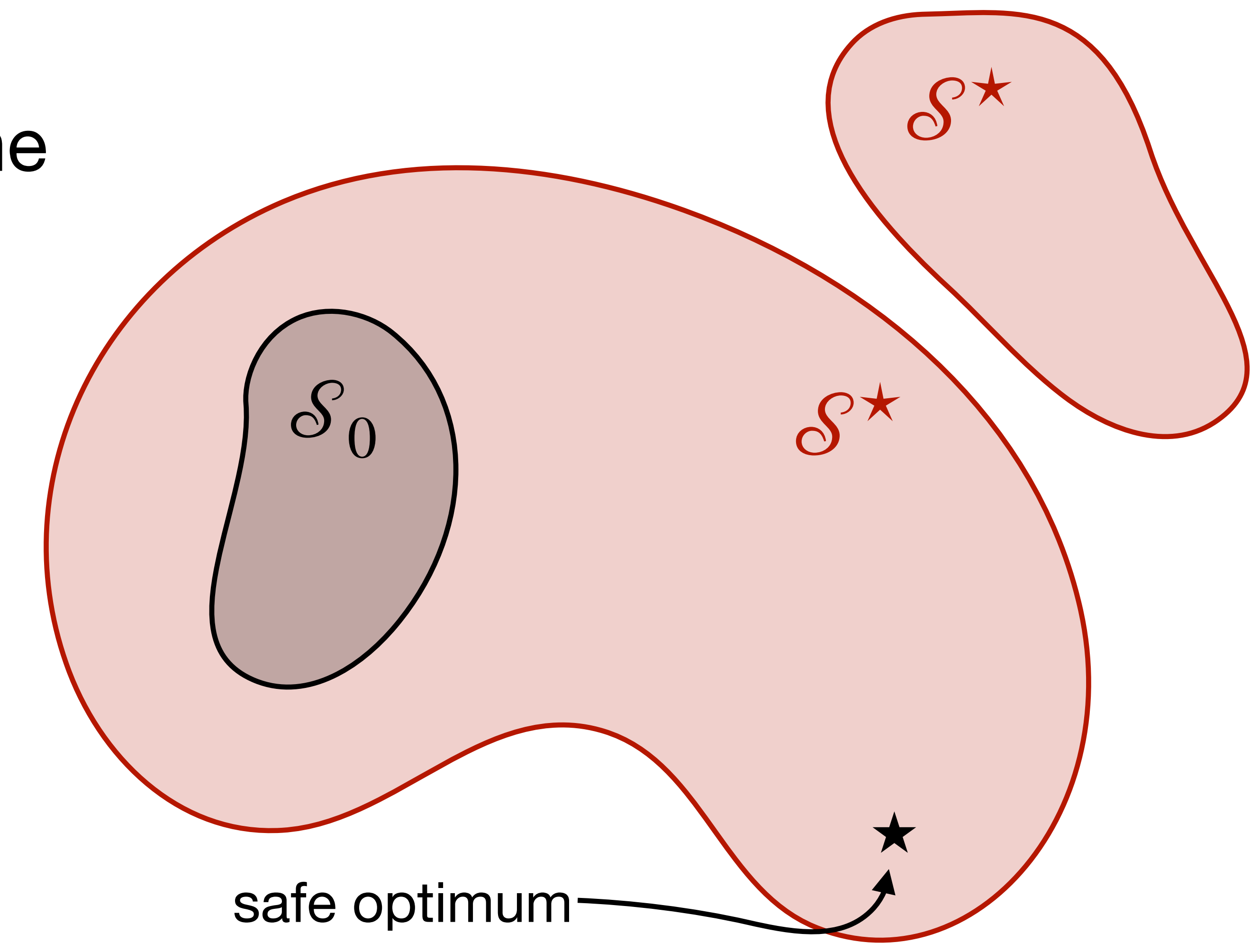


Paper:

Code:

Example: Safe BO

- target space are the *potential (safe) maximizers*



Algorithmic Framework for TAL

Proposal: minimize posterior uncertainty in $\mathcal{A}$

$$\boldsymbol{x}_n = \operatorname*{arg\,min}_{\boldsymbol{x}_n \in \mathcal{S}} \sum_{\boldsymbol{x} \in \mathcal{A}} \mathrm{Var}[f(\boldsymbol{x}) \mid D_{n-1}, (\boldsymbol{x}_n, y_n)] \quad \text{(VTL)}$$

We prove novel convergence (& generalization) guarantees

Outlook:

Active Fine-Tuning of NNs