

NeurIPS'24 & INFORMS Data
Mining Paper Competition Finalist

Cost-aware Bayesian Optimization with Adaptive Stopping via the Pandora's Box Gittins Index

On arXiv soon!

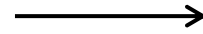
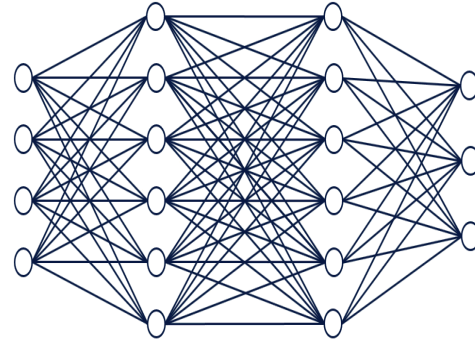
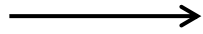
Qian Xie (Cornell ORIE)

INFORMS Applied Probability Society Conference 2025

World of Hyperparameter Optimization

Hyperparameter tuning:

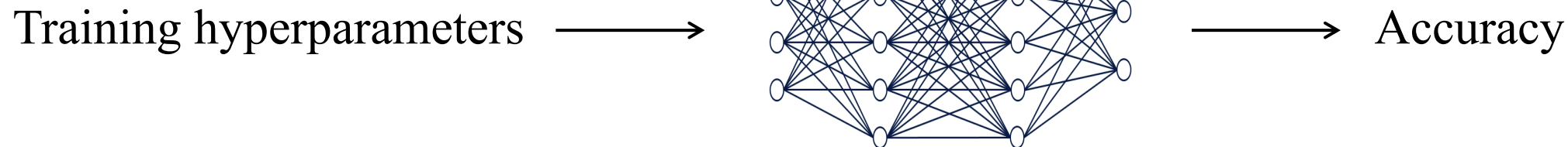
Training hyperparameters



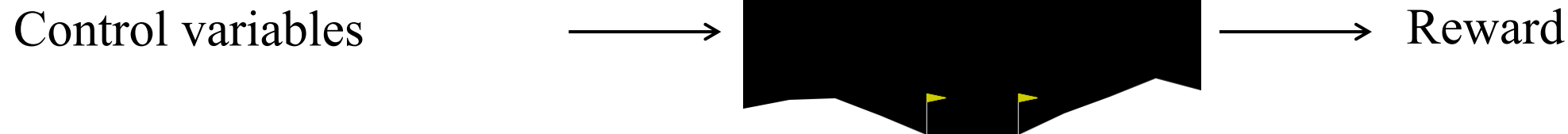
Accuracy

World of Hyperparameter Optimization

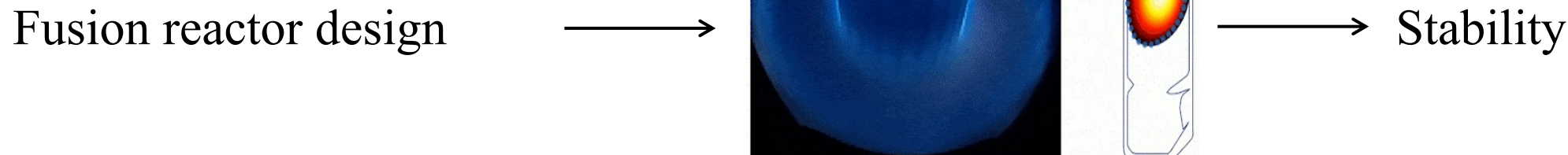
Hyperparameter tuning:



Control optimization:



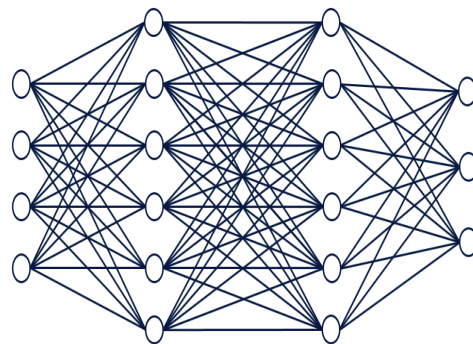
Plasma physics:



World of Hyperparameter Optimization

Hyperparameter tuning:

Training hyperparameters

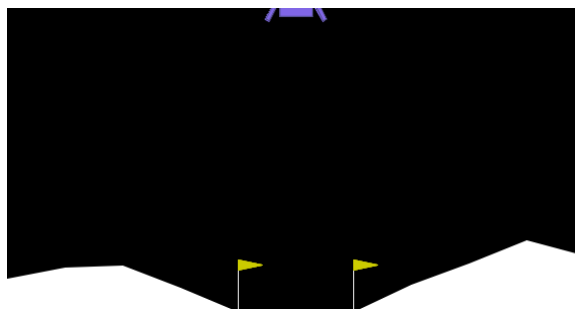


unknown &
expensive-to-evaluate

Accuracy

Control optimization:

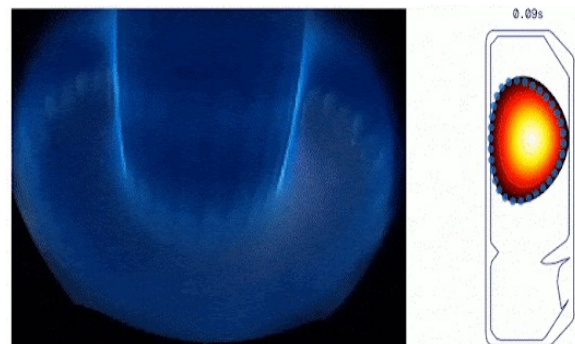
Control variables



Reward

Plasma physics:

Fusion reactor design

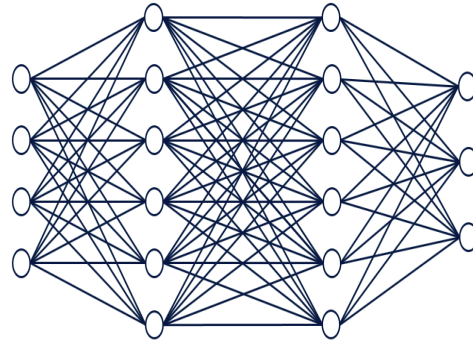


Stability

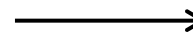
Grid Search for AutoML

Hyperparameter tuning:

Training hyperparameters



unknown &
expensive-to-evaluate



Accuracy

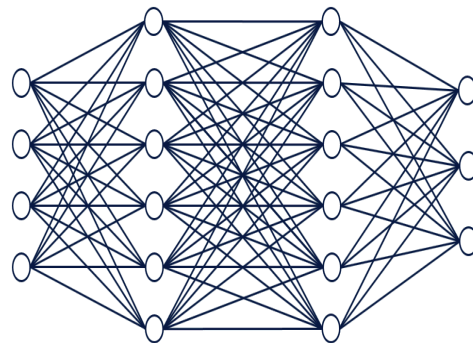
Parameter	Type	Scale	Range	Number of Options
Batch size	Integer	Log-scale	[16, 512]	10
Learning rate	Float	Log-scale	[1e-4, 1e-1]	10
Momentum	Float	Linear	[0.1, 0.99]	10
Weight decay	Float	Log-scale	[1e-5, 1e-1]	10
Number of layers	Integer	Linear	{1, 2, 3, 4}	4
Max units per layer	Integer	Log-scale	[64, 1024]	10
Dropout	Float	Linear	[0.0, 1.0]	10

40,000,000
combinations!

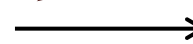
Grid Search for AutoML

Hyperparameter tuning:

Training hyperparameters



unknown &
expensive-to-evaluate



Accuracy

Time-consuming!

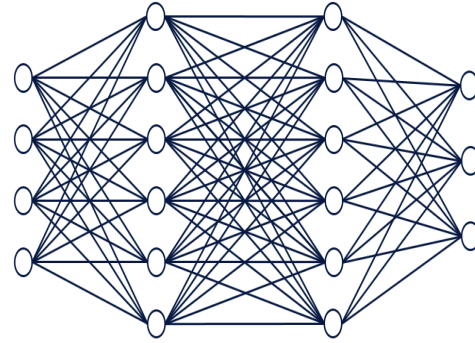
Number of Options
10
10
10
10
4
10
10

40,000,000
combinations!

Grid Search for AutoML

Hyperparameter tuning:

Training hyperparameters →



unknown &
expensive-to-evaluate

→ Accuracy

Time-consuming!

More efficient:
Bayesian optimization

Number of Options
10
10
10
10
4
10
10

40,000,000
combinations!

Bayesian Optimization

Black-box optimization:

Input hyperparameters →

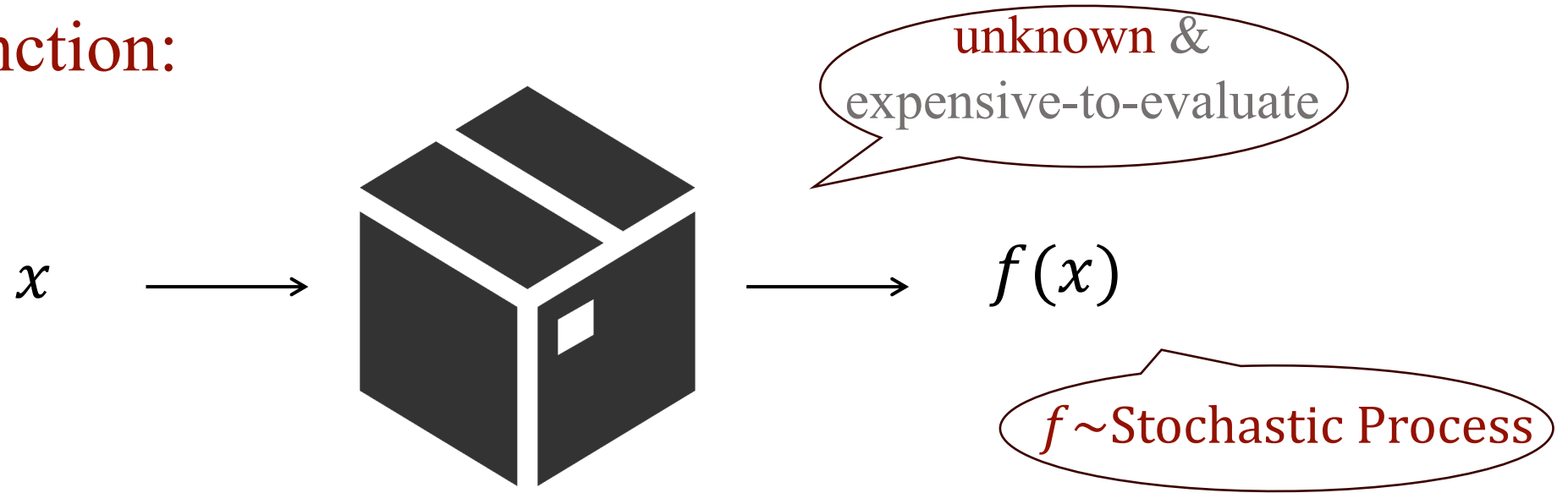


→ Performance metric

unknown &
expensive-to-evaluate

Bayesian Optimization

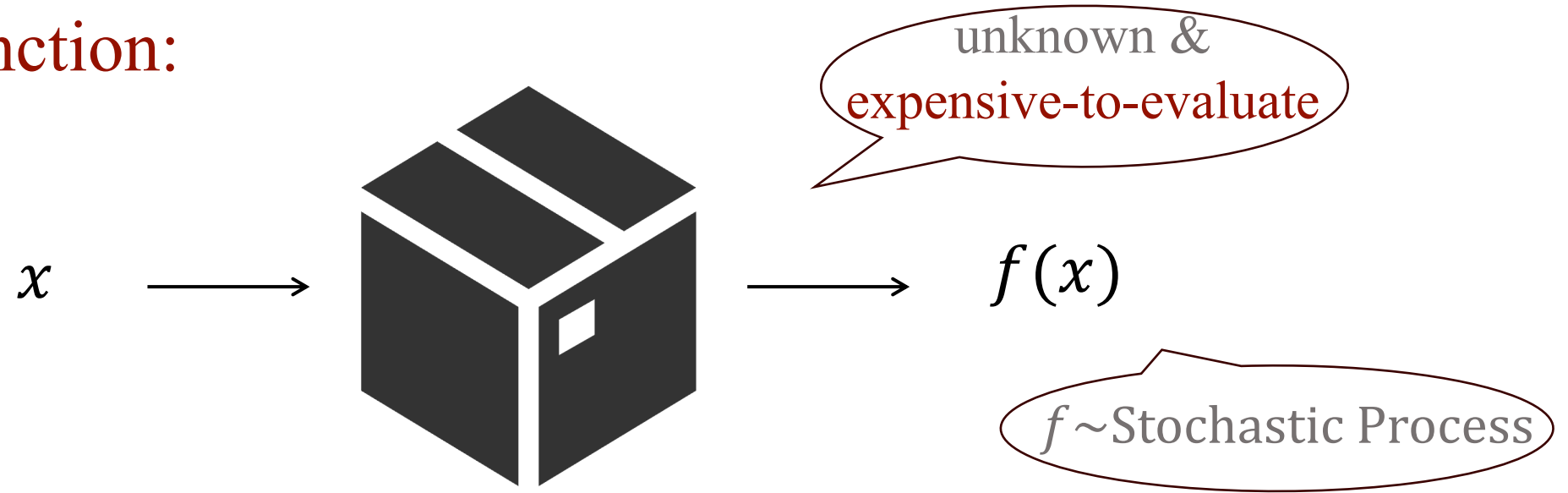
Black-box function:



Goal: $\max_{x \in \mathcal{X}} f(x)$

Bayesian Optimization

Black-box function:

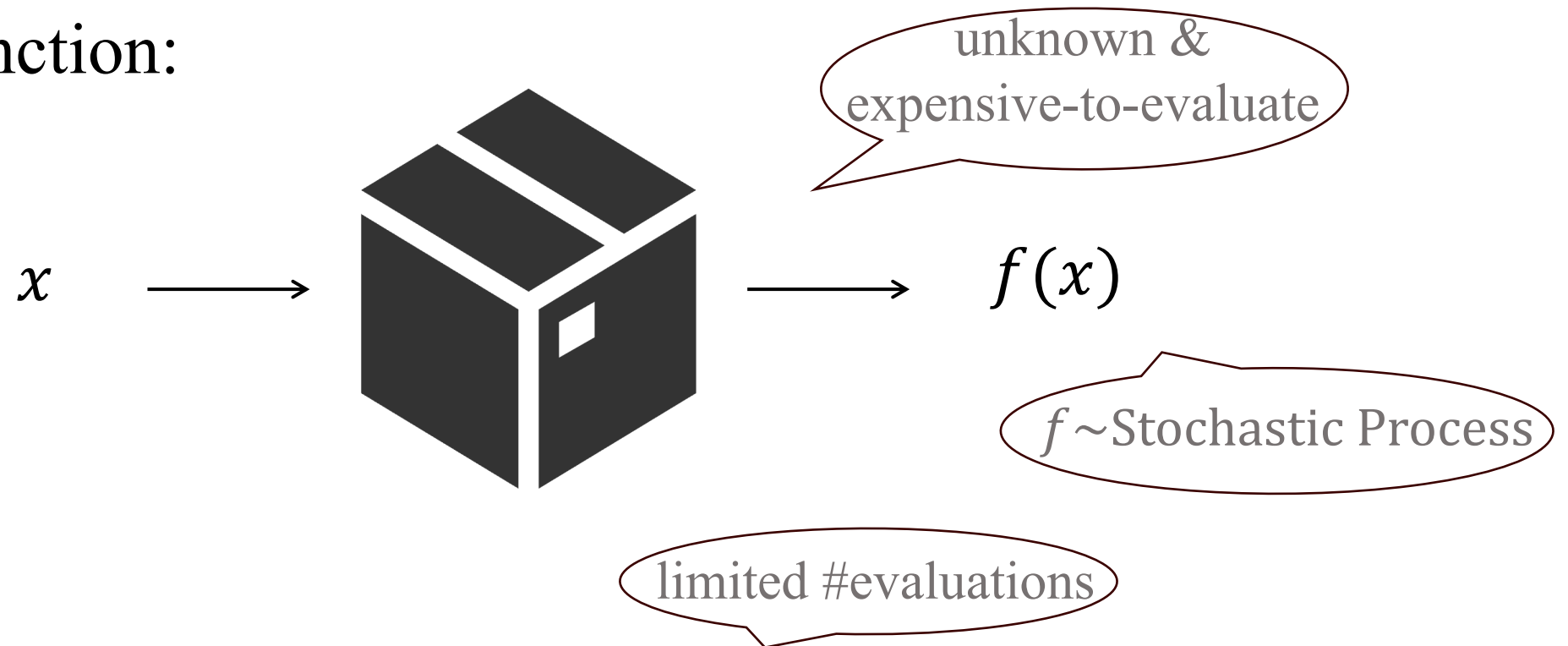


Goal: $\sup_{\text{policy}} \mathbb{E} \max_{t=1,2,\dots,T} f(x_t)$

limited #evaluations

Bayesian Optimization

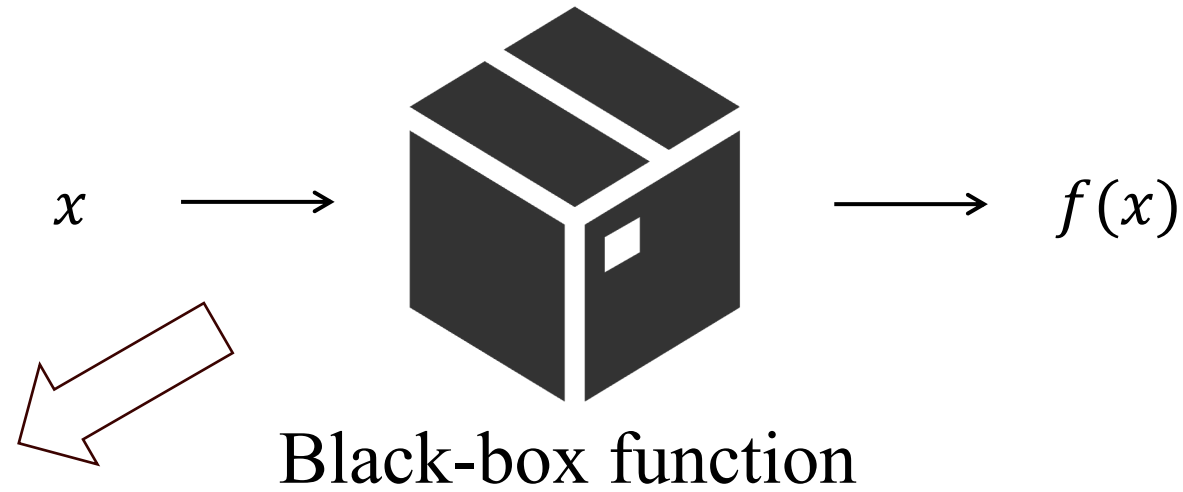
Black-box function:



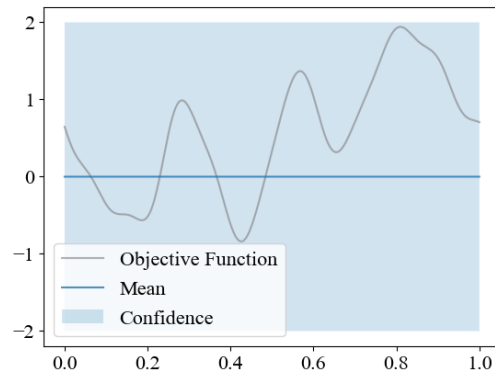
$$\text{Goal: } \sup_{\text{policy}} \mathbb{E} \max_{t=1,2,\dots,T} f(x_t)$$

Key idea: maintain probabilistic belief about f

Bayesian Optimization

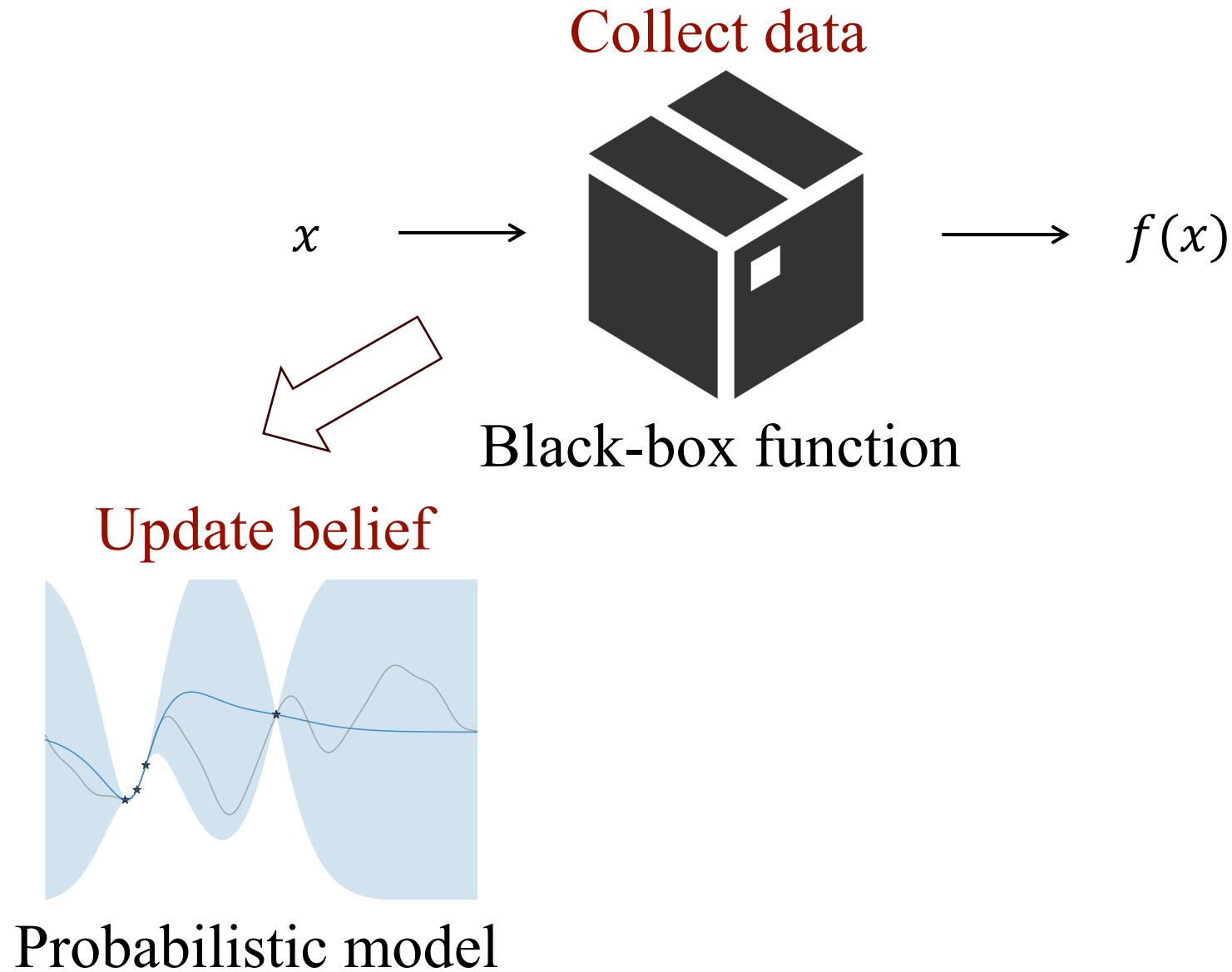


Maintain belief

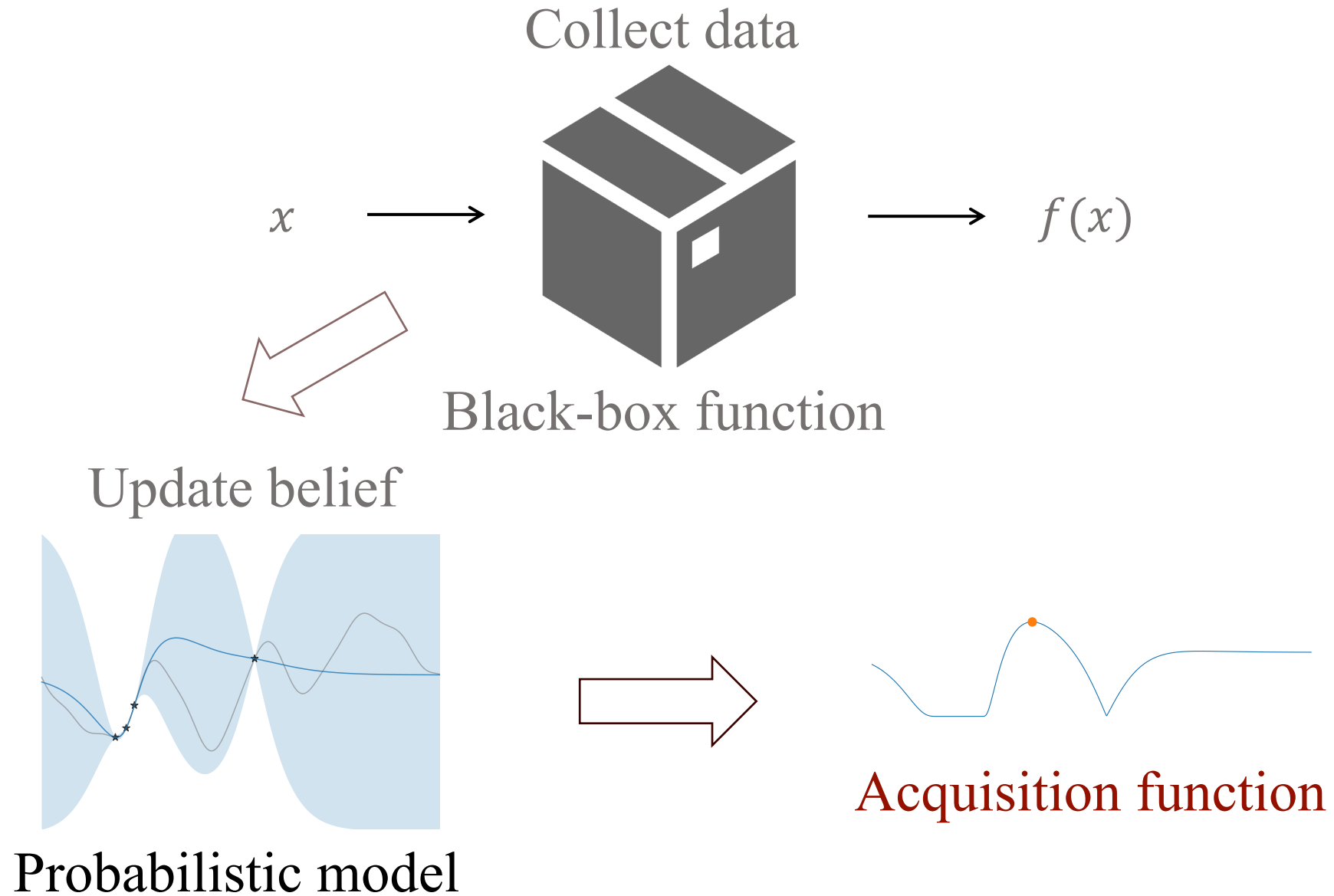


Probabilistic model

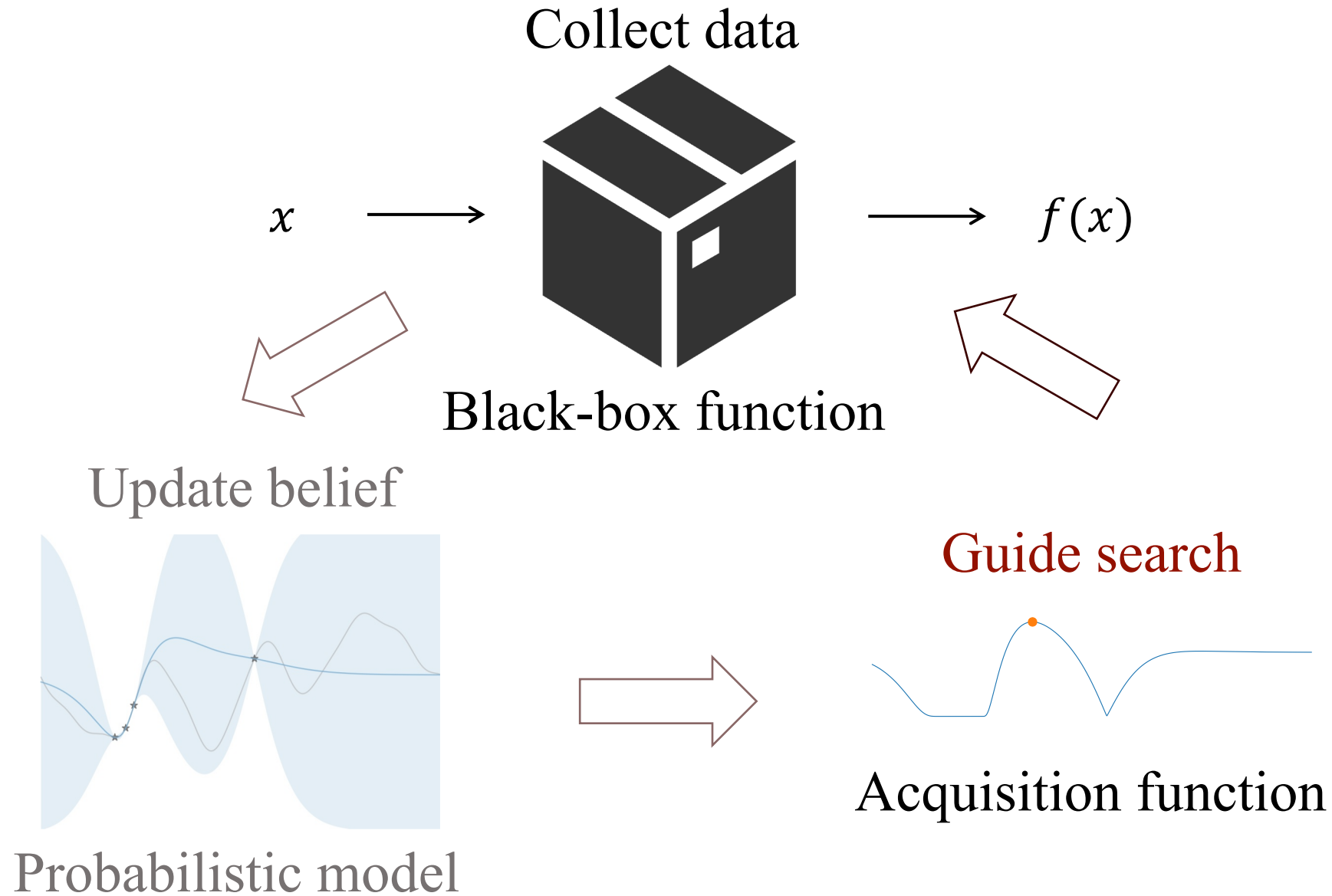
Bayesian Optimization



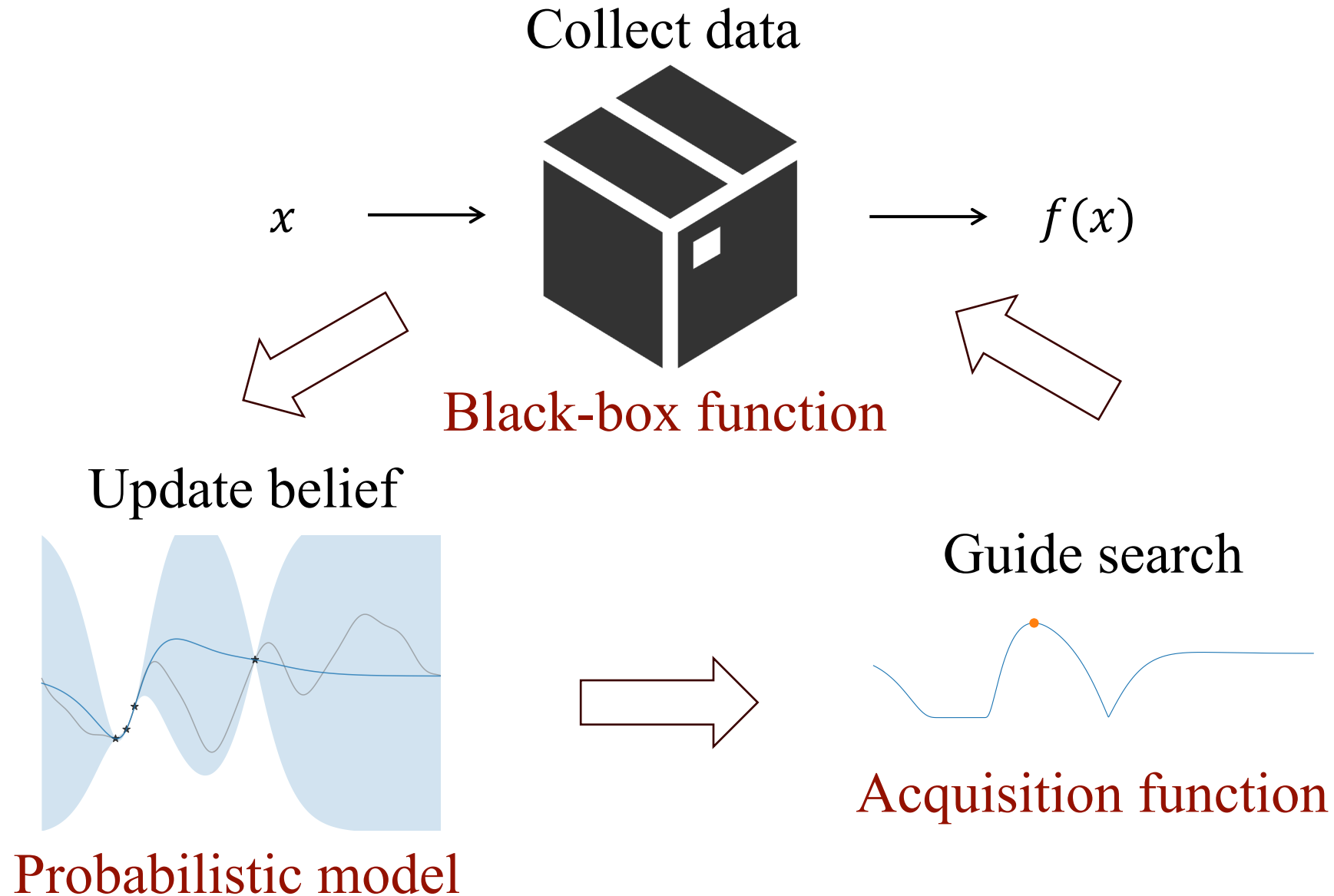
Bayesian Optimization



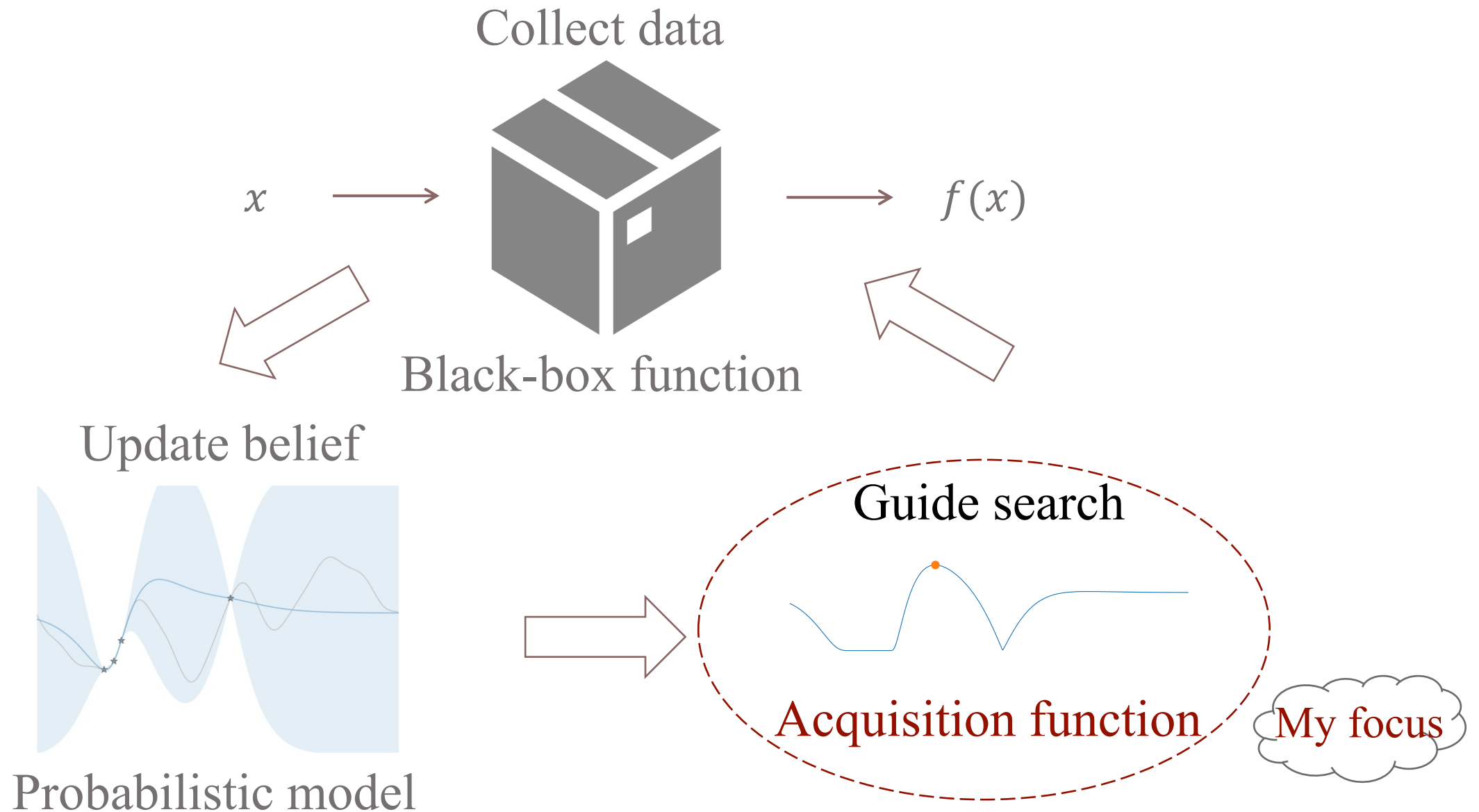
Bayesian Optimization



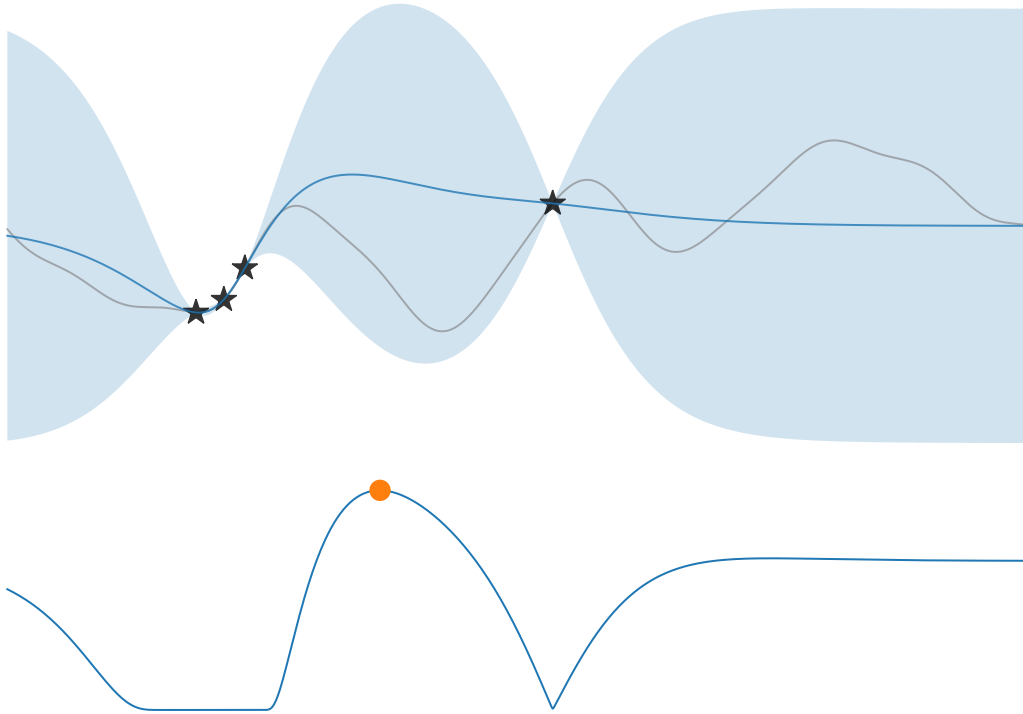
Bayesian Optimization



Bayesian Optimization

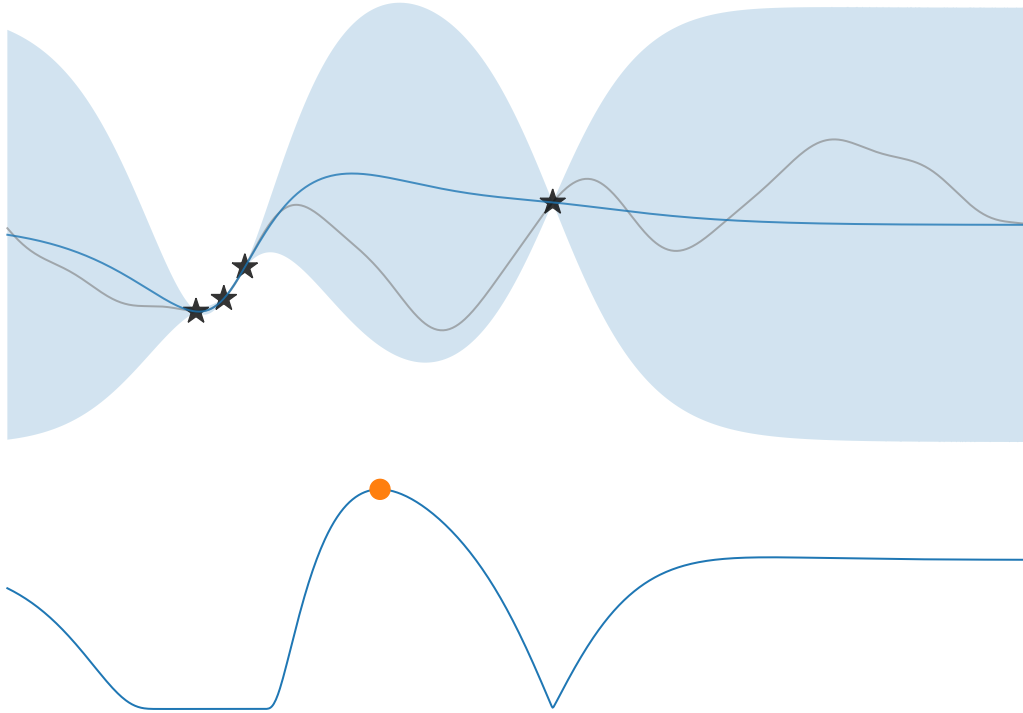


Classic Acquisition Functions



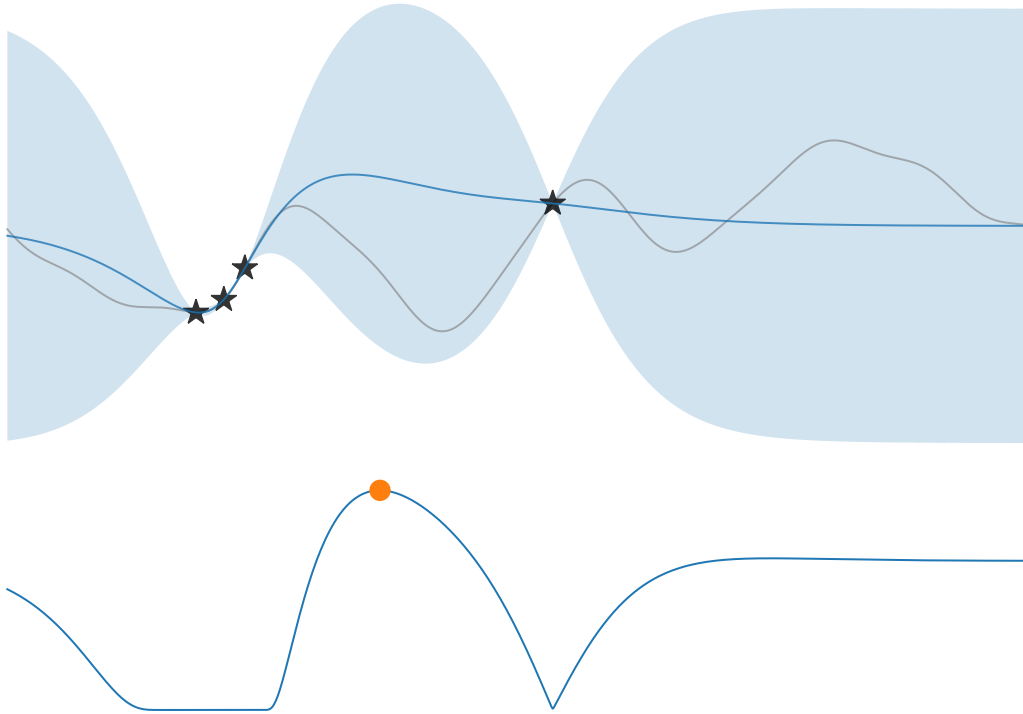
- Improvement-based
- Entropy-based
- Upper Confidence Bound
- Thompson Sampling

New Acquisition Function: Gittins Index



- Improvement-based
- Entropy-based
- Upper Confidence Bound
- Thompson Sampling
- My work: Gittins Index

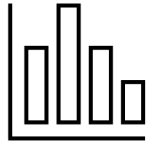
New Acquisition Function: Gittins Index



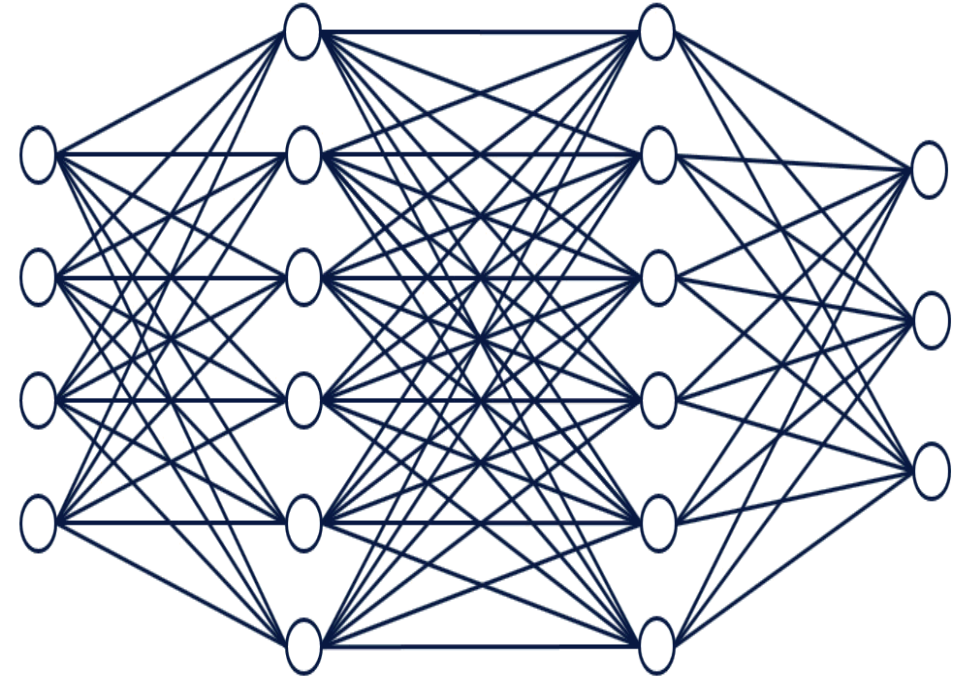
- Improvement-based
- Entropy-based
- Upper Confidence Bound
- Thompson Sampling
- My work: Gittins Index

Why another acquisition function?

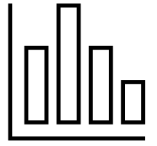
Under-explored Practical Considerations



Varying evaluation costs



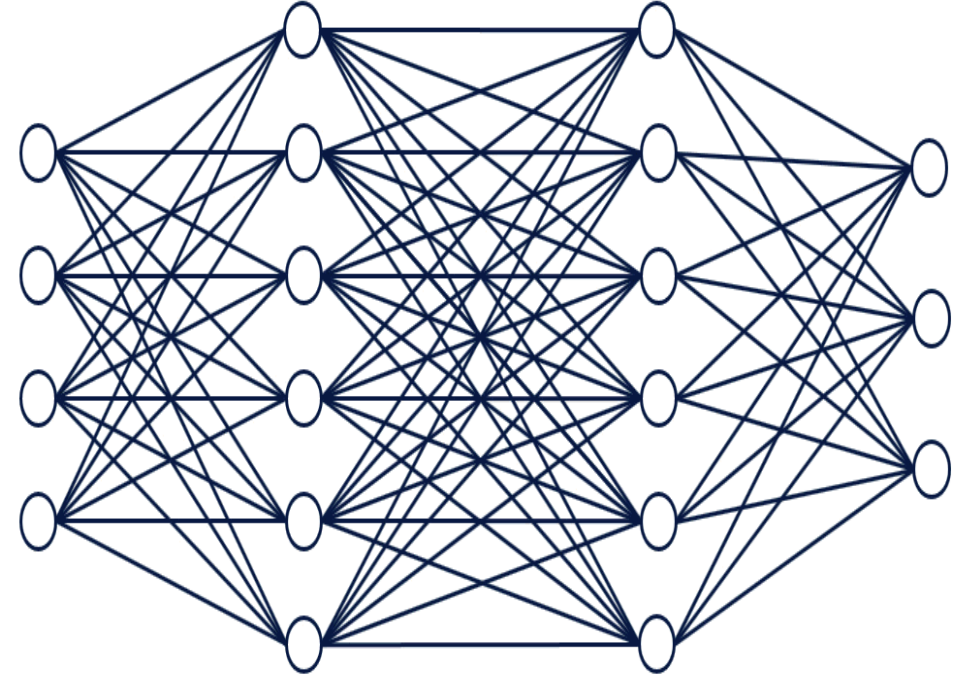
Under-explored Practical Considerations



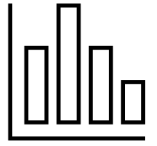
Varying evaluation costs



Smart stopping time



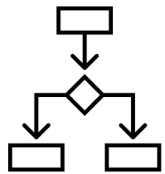
Under-explored Practical Considerations



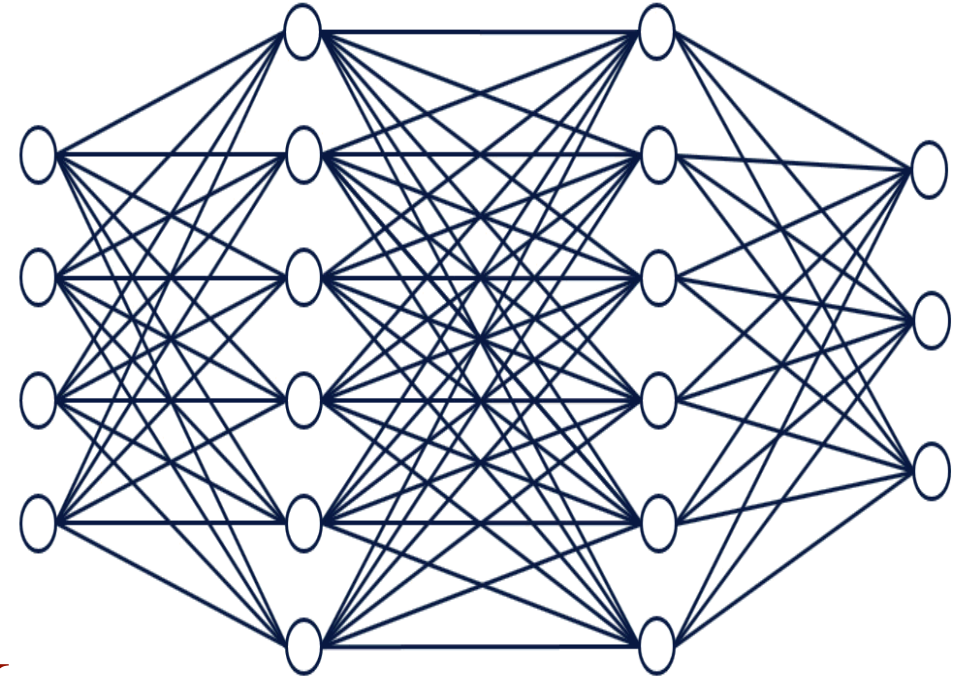
Varying evaluation costs



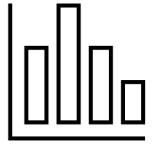
Smart stopping time



Observable multi-stage feedback



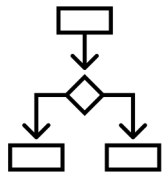
Under-explored Practical Considerations



Varying evaluation costs



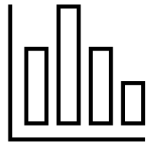
Smart stopping time



Observable multi-stage feedback

New design principle:
Gittins index

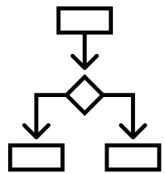
Under-explored Practical Considerations



Varying evaluation costs



Smart stopping time



Observable multi-stage feedback

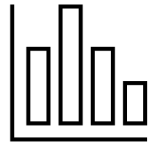
Gittins index

Cost-aware

Stopping-aware

Feedback-aware

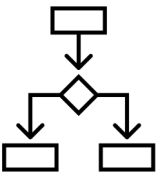
Under-explored Practical Considerations



Varying evaluation costs



Smart stopping time



Observable multi-stage feedback

Gittins index

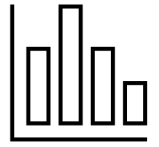
Cost-aware

Stopping-aware

Feedback-aware

Optimal in simplified problems

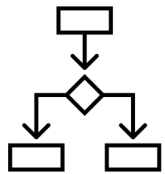
Under-explored Practical Considerations



Varying evaluation costs



Smart stopping time



Observable multi-stage feedback

In this talk

Gittins index

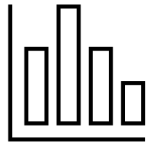
Cost-aware

Stopping-aware

Feedback-aware

Optimal in simplified problems

Coauthors



Varying evaluation costs
[NeurIPS'24]



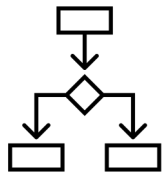
Raul Astudillo



Smart stopping time
[Under review]



Linda Cai



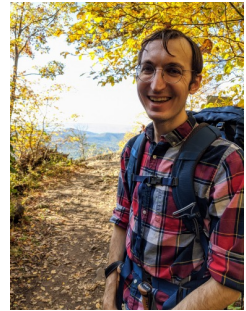
Observable multi-stage feedback
[Ongoing work]



Peter Frazier



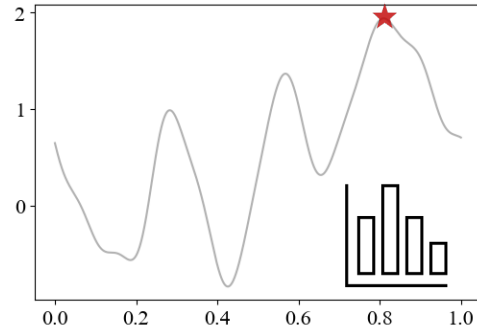
Alexander Terenin



Ziv Scully

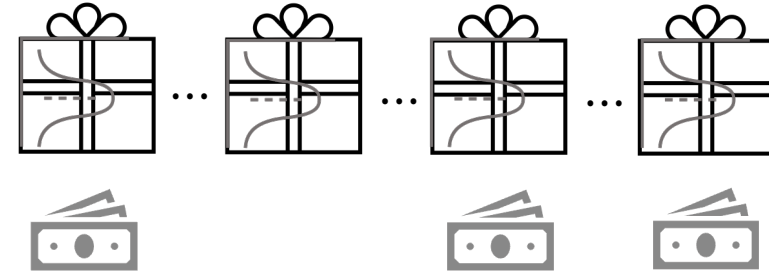
Outline

Studied Problem



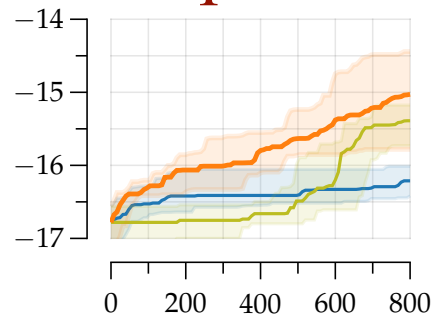
Cost-aware Bayesian optimization

Key idea



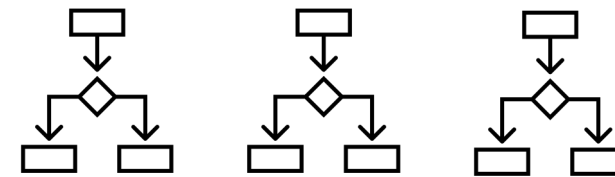
Link to simplified problem
and Gittins index theory

Impact



Competitive empirical performance

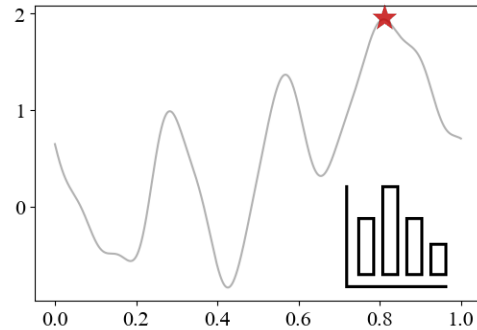
Future direction



“Exotic” Bayesian optimization

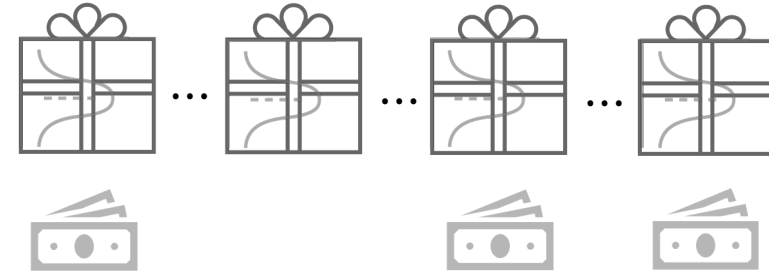
Outline

Studied Problem



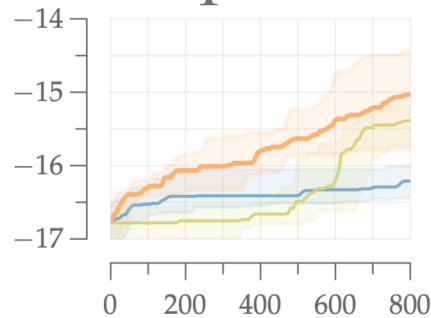
Cost-aware Bayesian optimization

Key idea



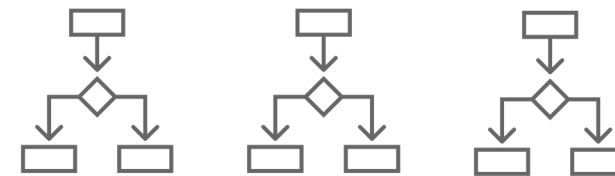
Link to simplified problem
and Gittins index theory

Impact



Competitive empirical performance

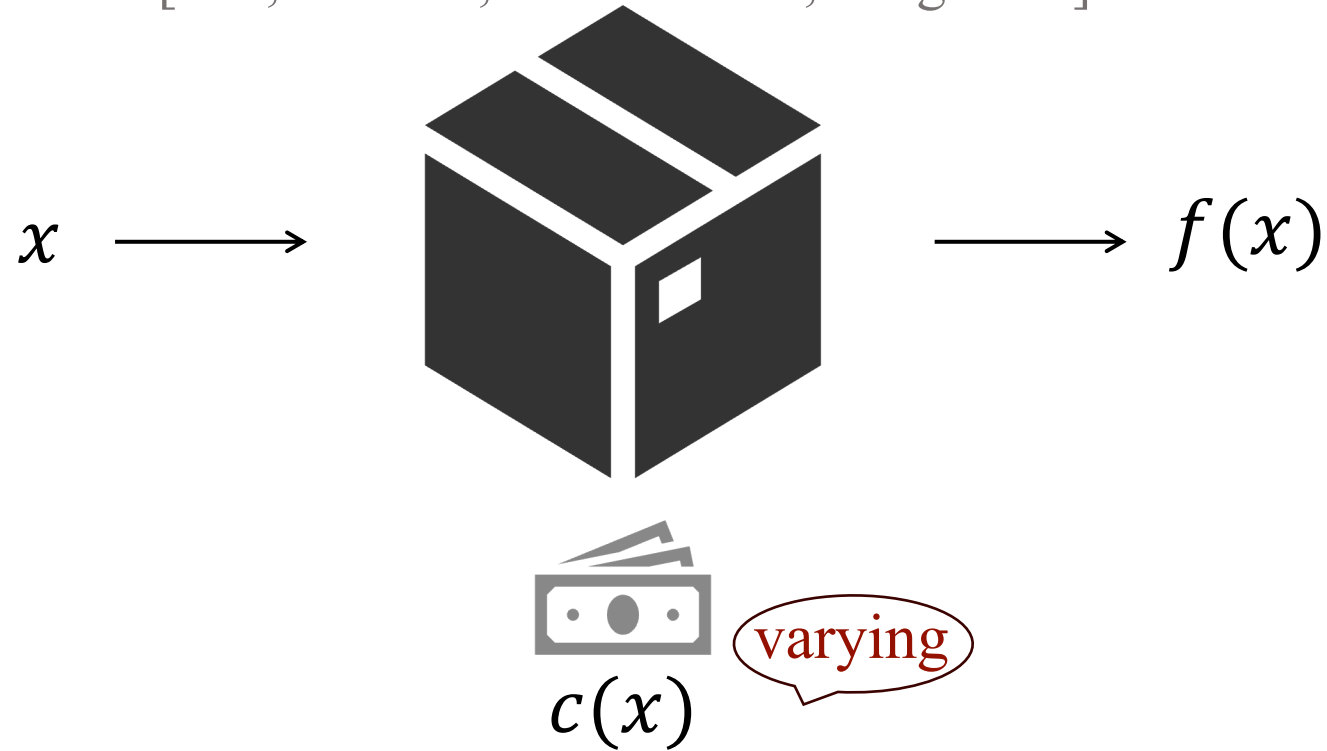
Future direction



“Exotic” Bayesian optimization

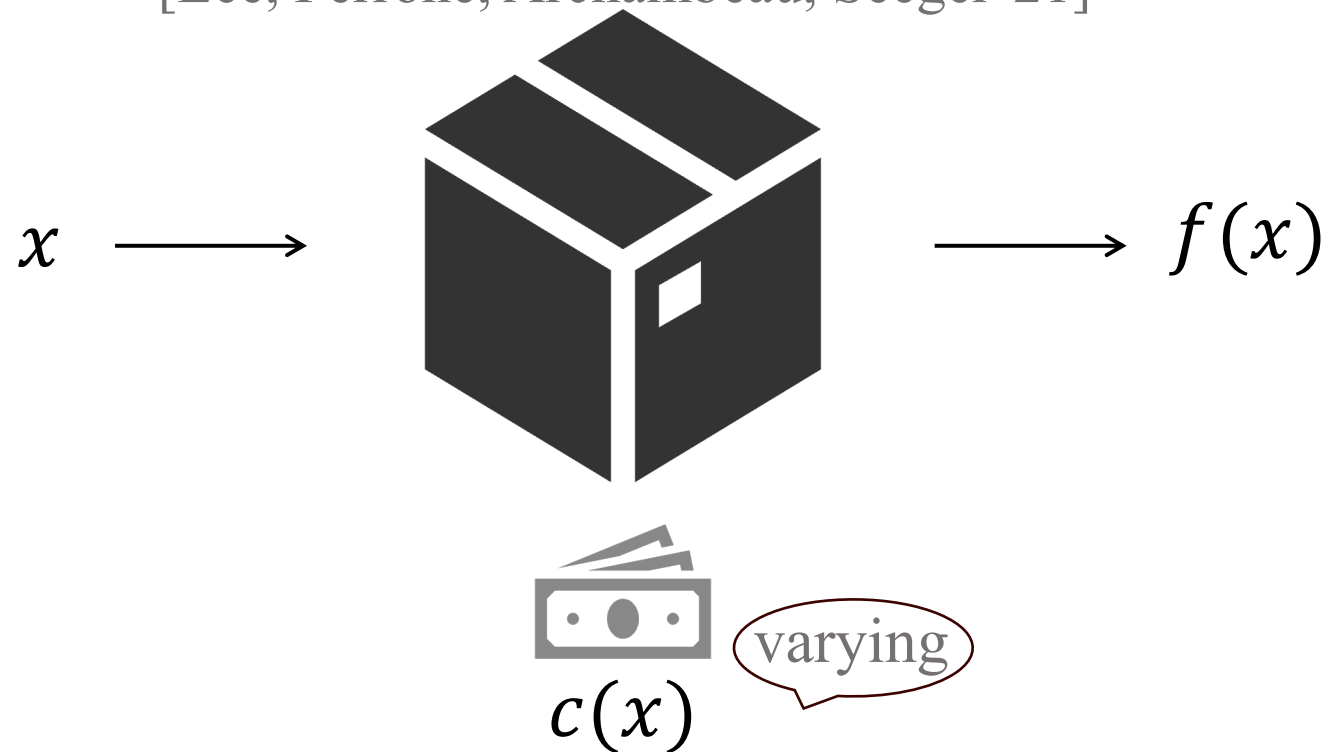
Cost-aware Bayesian Optimization

[Lee, Perrone, Archambeau, Seeger'21]



Cost-aware Bayesian Optimization

[Lee, Perrone, Archambeau, Seeger'21]

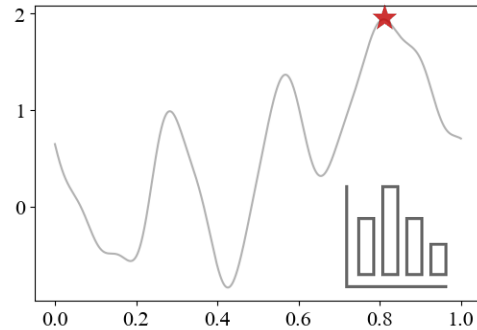


Goal: $\sup_{\text{policy}} \mathbb{E} \max_{t=1,2,\dots,T} f(x_t)$

s.t. $\sum_{t=1}^T c(x_t) \leq B$ Budget constraint

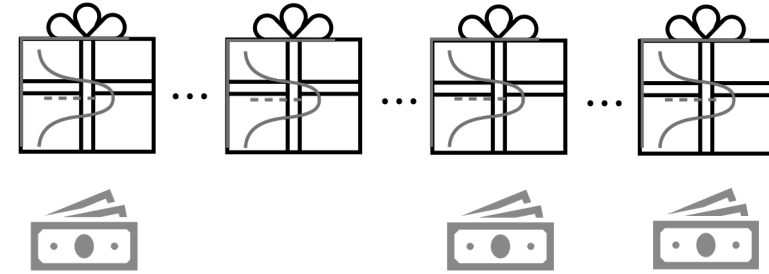
Outline

Studied Problem



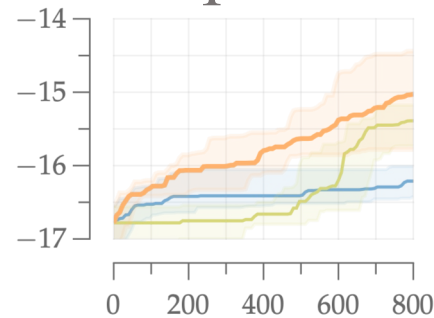
Cost-aware Bayesian optimization

Key idea



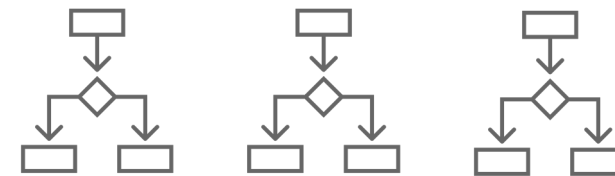
Link to simplified problem
and Gittins index theory

Impact



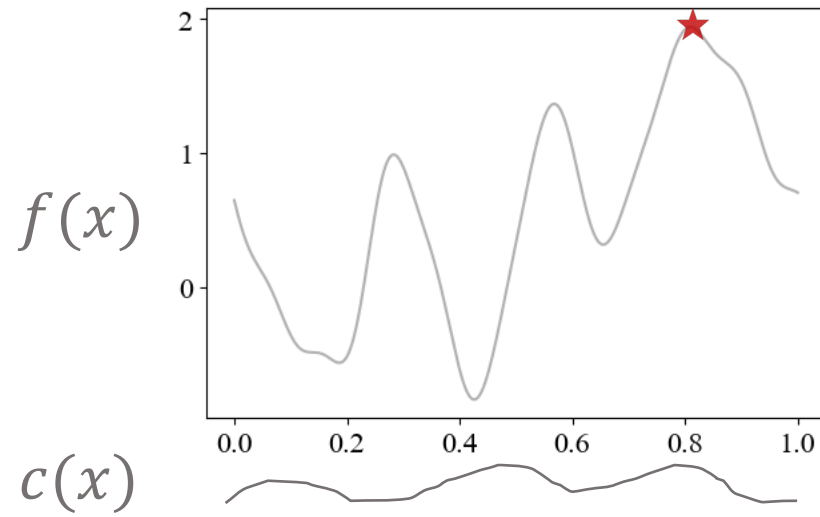
Competitive empirical performance

Future direction

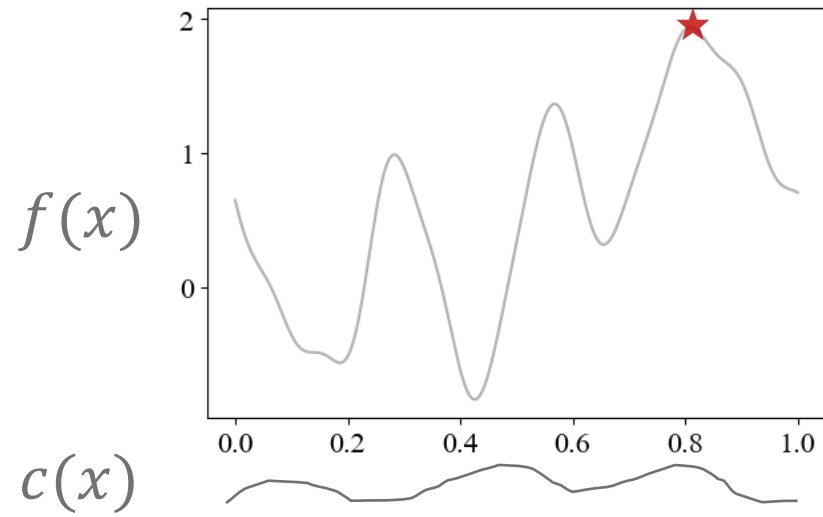


“Exotic” Bayesian optimization

Cost-aware Bayesian Optimization



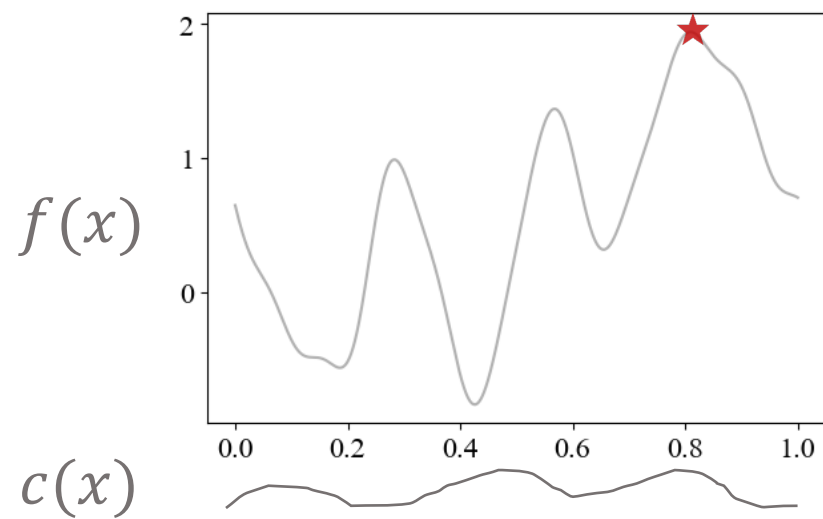
Cost-aware Bayesian Optimization



Continuous

Correlated

Cost-aware Bayesian Optimization



Continuous

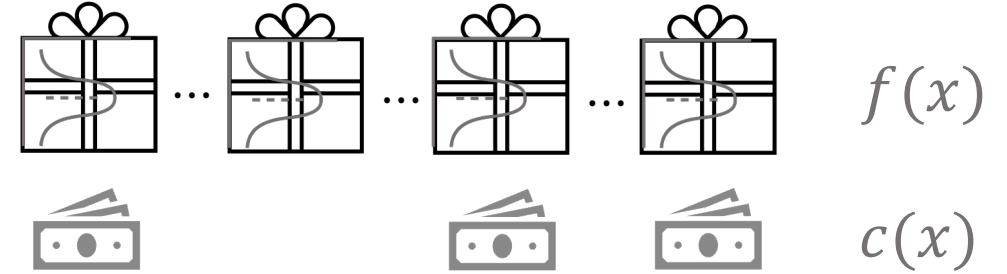
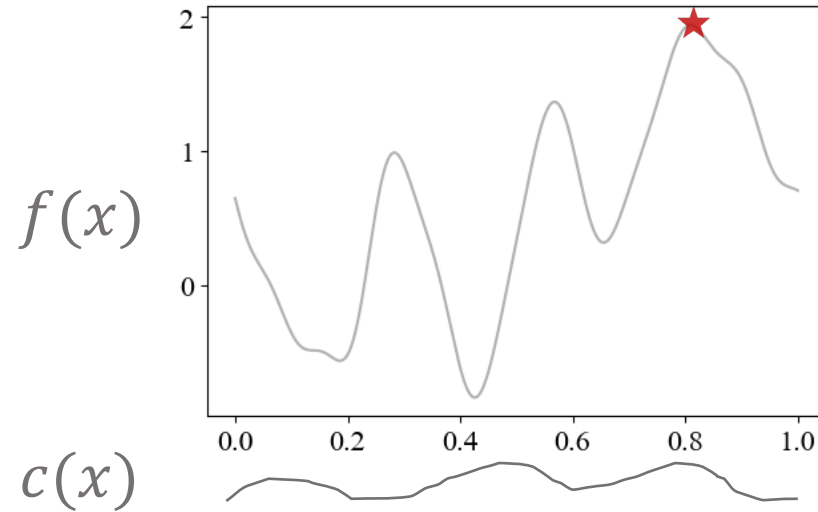
Correlated

Intractable MDP!

Cost-aware Bayesian Optimization

Pandora's Box

[Weitzman'79]



Discrete

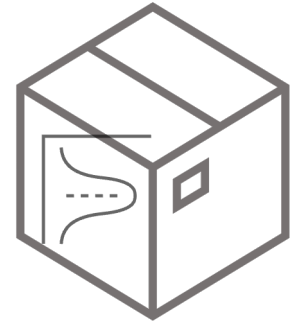
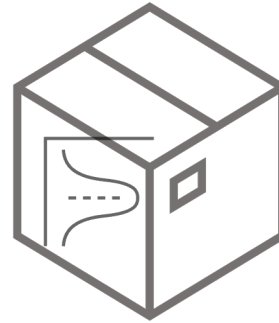
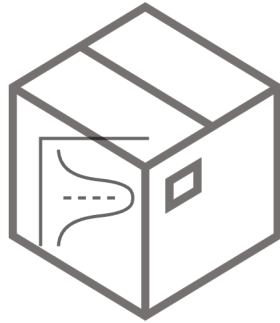
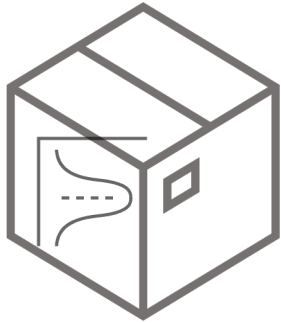


Independent

Intractable MDP!

Pandora's Box

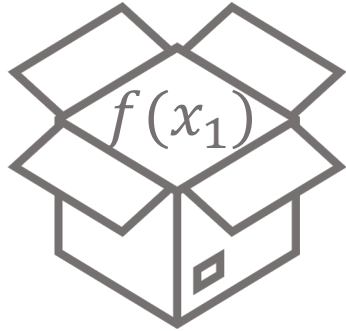
$t = 0$



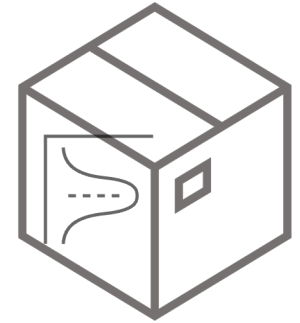
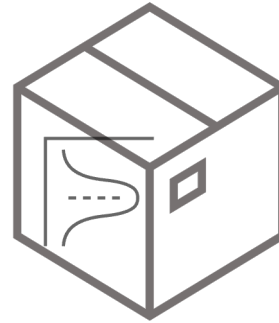
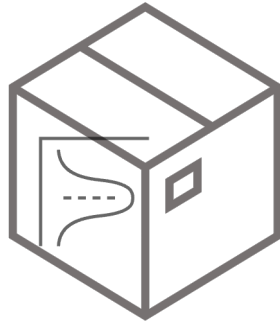
$$\sup_{\text{policy}} \mathbb{E} \left(\max_{t=1,2,\dots,T} f(x_t) - \sum_{t=1}^T c(x_t) \right)$$

Pandora's Box

$t = 1$



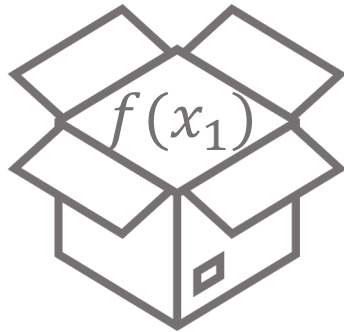
$c(x_1)$



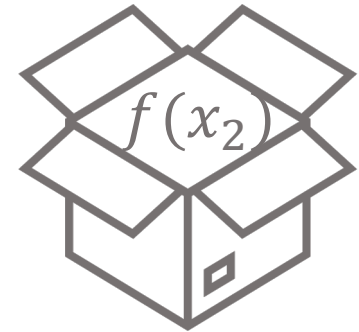
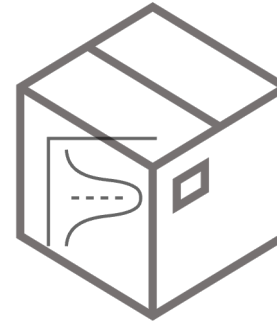
$$\sup_{\text{policy}} \mathbb{E} \left(\max_{t=1,2,\dots,T} f(x_t) - \sum_{t=1}^T c(x_t) \right)$$

Pandora's Box

$t = 2$



$c(x_1)$

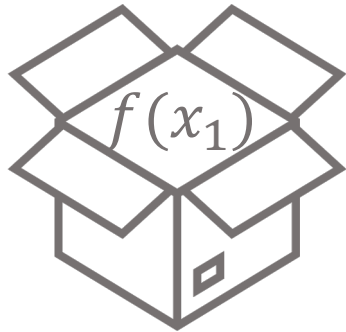


$c(x_2)$

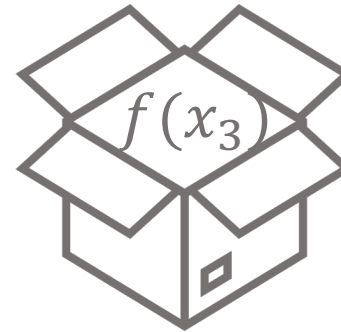
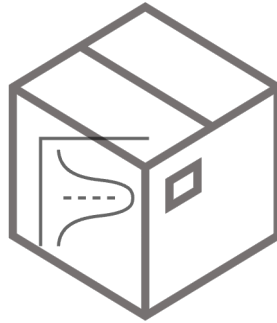
$$\sup_{\text{policy}} \mathbb{E} \left(\max_{t=1,2,\dots,T} f(x_t) - \sum_{t=1}^T c(x_t) \right)$$

Pandora's Box

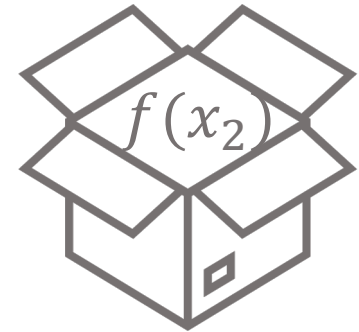
$t = 3$



$c(x_1)$



$c(x_3)$

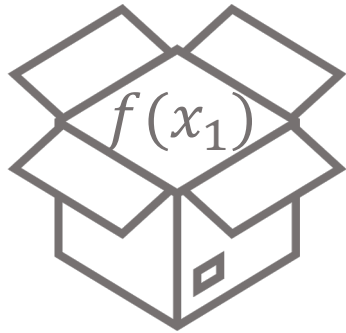


$c(x_2)$

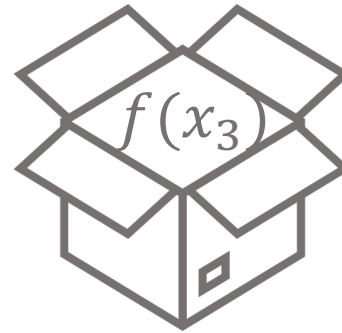
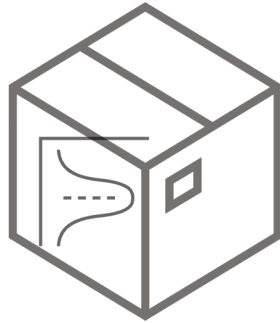
$$\sup_{\text{policy}} \mathbb{E} \left(\max_{t=1,2,\dots,T} f(x_t) - \sum_{t=1}^T c(x_t) \right)$$

Pandora's Box

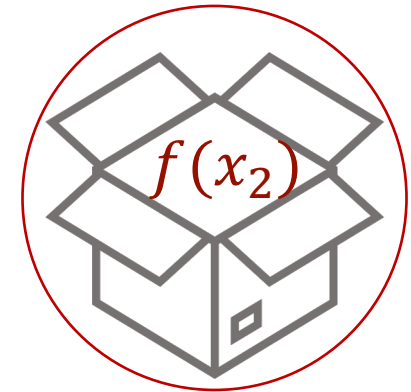
$t = T$, stop



$c(x_1)$



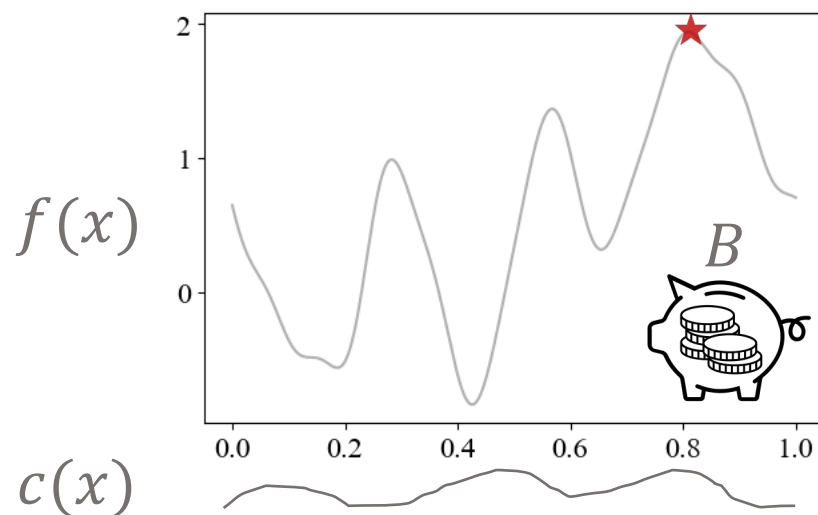
$c(x_3)$



$c(x_2)$

$$\sup_{\text{policy}} \mathbb{E} \left(\max_{t=1,2,\dots,T} f(x_t) - \sum_{t=1}^T c(x_t) \right)$$

Cost-aware Bayesian Optimization



Continuous

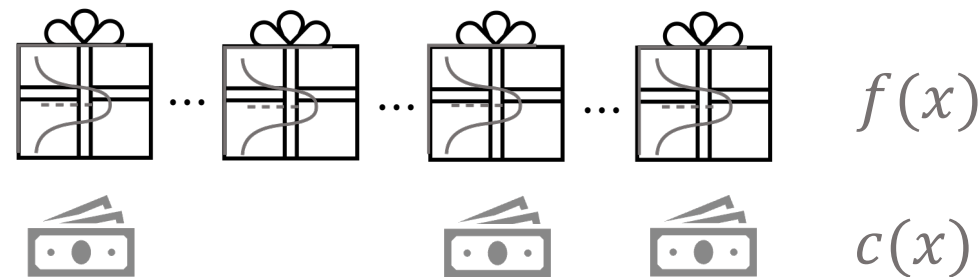
Correlated

Budget-constrained

$$\begin{aligned} \sup_{\text{policy}} \mathbb{E} \max_{t=1,2,\dots,T} f(x_t) \\ \text{s.t. } \sum_{t=1}^T c(x_t) \leq B \end{aligned}$$

Pandora's Box

[Weitzman'79]



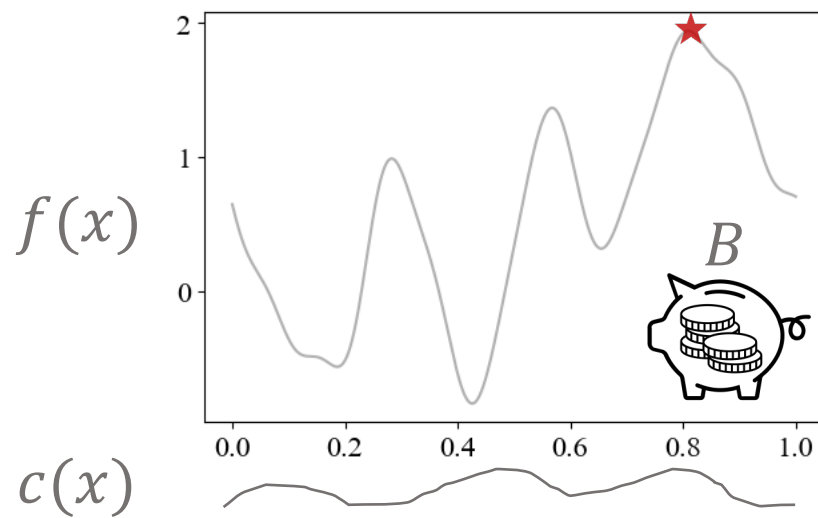
Discrete

Independent

Cost-per-sample

$$\sup_{\text{policy}} \mathbb{E} \left(\max_{t=1,2,\dots,T} f(x_t) - \sum_{t=1}^T c(x_t) \right)$$

Cost-aware Bayesian Optimization



Continuous

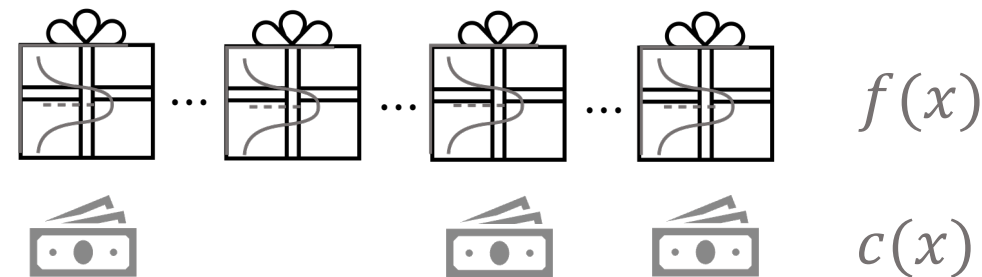
Correlated

Expected-budget-constrained

$$\begin{aligned} & \sup_{\text{policy}} \mathbb{E} \max_{t=1,2,\dots,T} f(x_t) \\ & \text{s.t. } \mathbb{E} \sum_{t=1}^T c(x_t) \leq B \end{aligned}$$

Pandora's Box

[Weitzman'79]



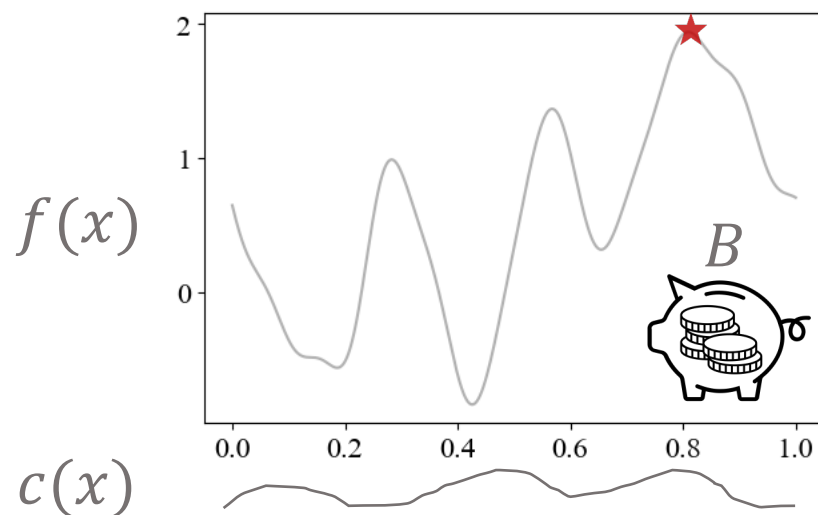
Discrete

Independent

Cost-per-sample

$$\sup_{\text{policy}} \mathbb{E} \left(\max_{t=1,2,\dots,T} f(x_t) - \sum_{t=1}^T c(x_t) \right)$$

Cost-aware Bayesian Optimization



Continuous

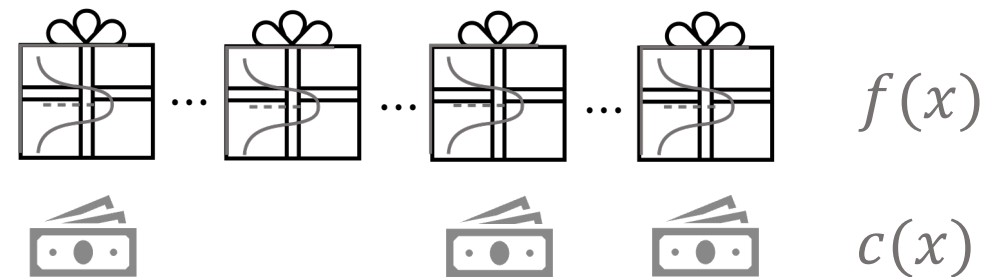
Correlated

Ebc & Cps

$$\begin{aligned} & \sup_{\text{policy}} \mathbb{E} \max_{t=1,2,\dots,T} f(x_t) \\ & \text{s.t. } \mathbb{E} \sum_{t=1}^T c(x_t) \leq B \end{aligned}$$

Pandora's Box

[Weitzman'79]



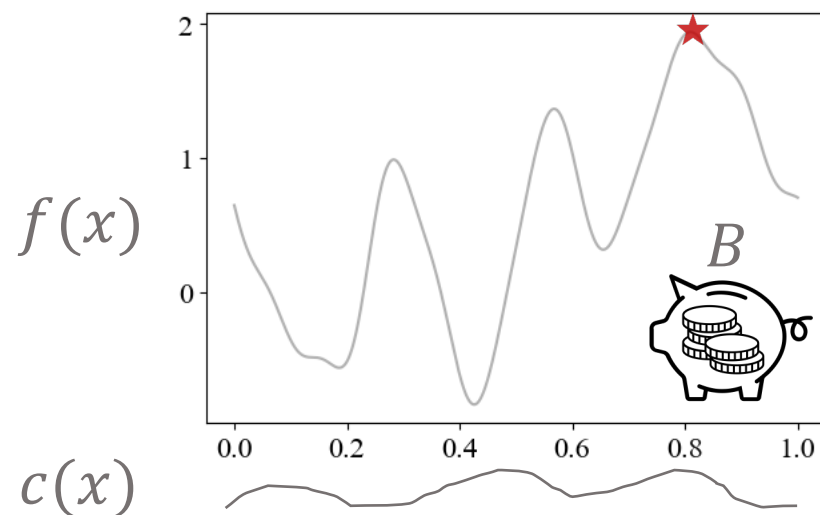
Discrete

Independent

Cost-per-sample

$$\sup_{\text{policy}} \mathbb{E} \left(\max_{t=1,2,\dots,T} f(x_t) - \sum_{t=1}^T c(x_t) \right)$$

Cost-aware Bayesian Optimization



Continuous

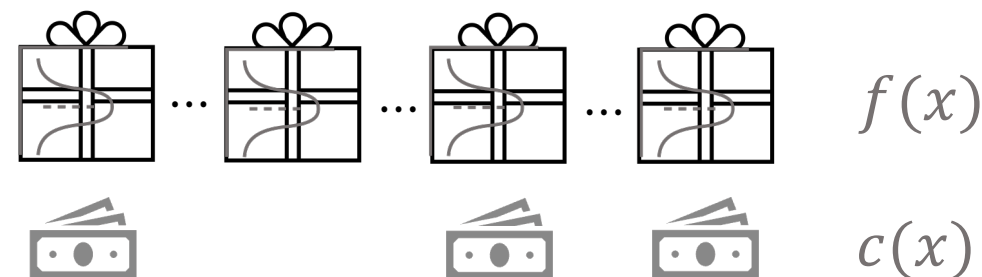
Correlated

Ebc & Cps

Intractable MDP!

Pandora's Box

[Weitzman'79]



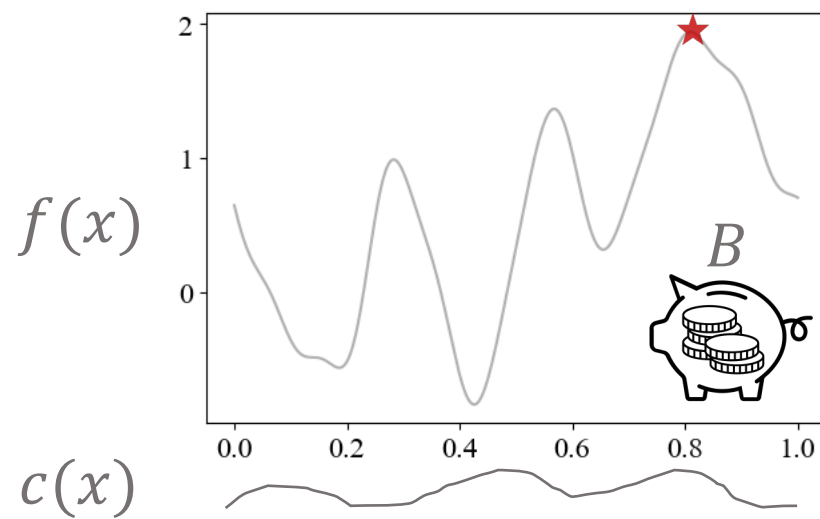
Discrete

Independent

Cost-per-sample

Optimal policy: Gittins index

Cost-aware Bayesian Optimization



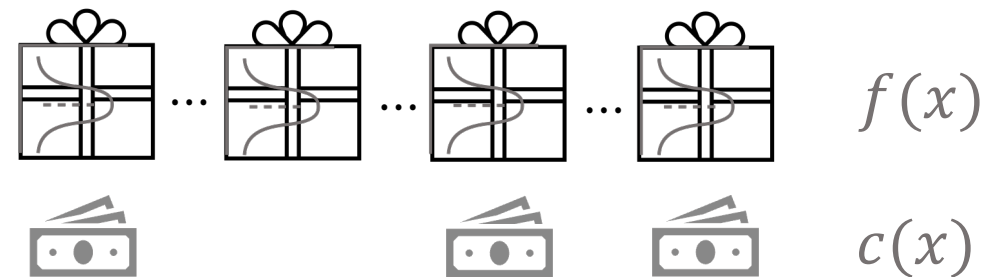
Continuous

Correlated

Ebc & Cps

Pandora's Box

[Weitzman'79]



Discrete

Independent

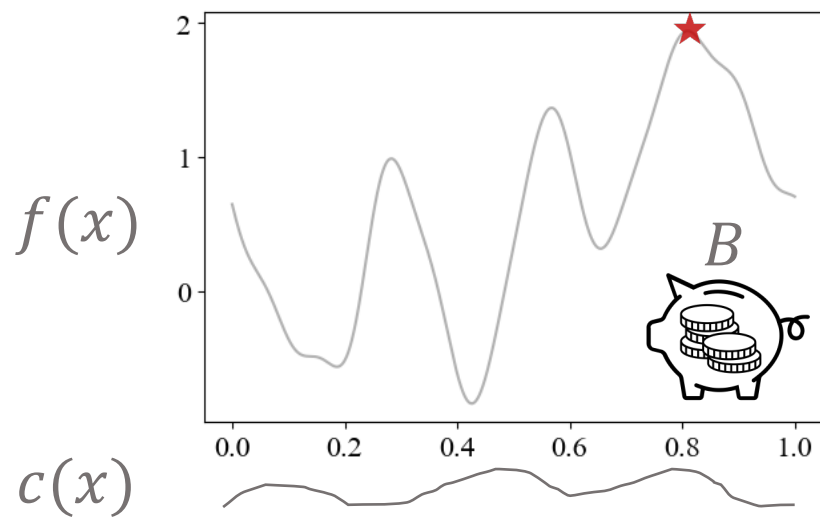
Cost-per-sample

How to translate?



Optimal policy: Gittins index

Cost-aware Bayesian Optimization



Continuous

Correlated

Ebc & Cps

Acquisition function
+ stopping rule

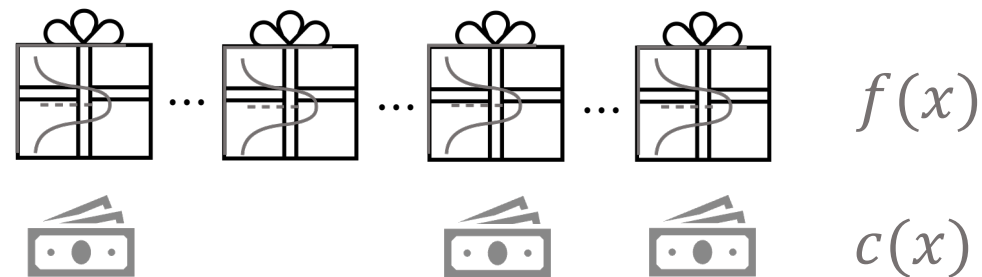
incorporate posterior



Optimal policy: Gittins index

Pandora's Box

[Weitzman'79]

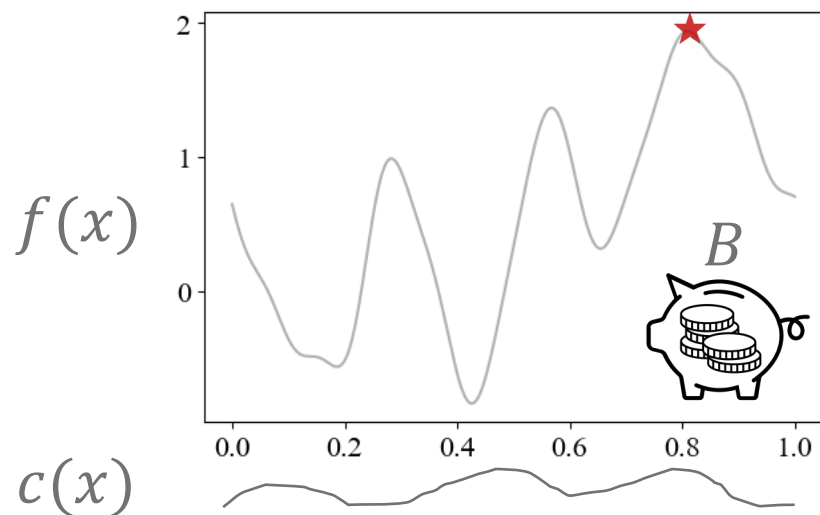


Discrete

Independent

Cost-per-sample

Cost-aware Bayesian Optimization



Continuous

Correlated

Ebc & Cps

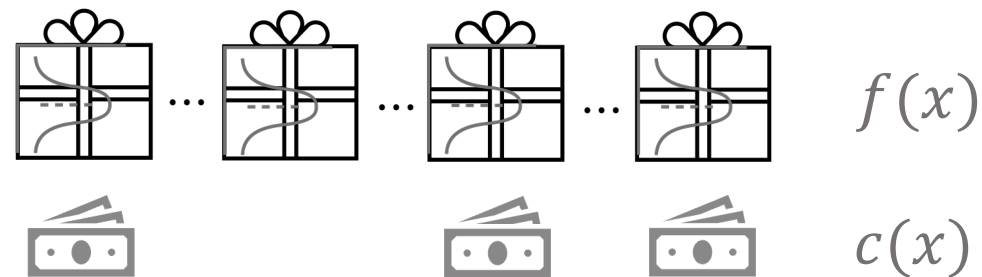
Acquisition function

+ stopping rule

Empirically good?

Pandora's Box

[Weitzman'79]



Discrete

Independent

Cost-per-sample

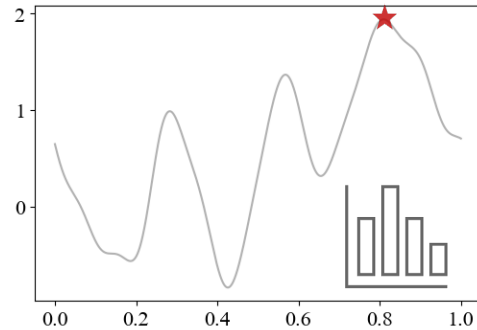
Gittins index is optimal

incorporate posterior

⇐

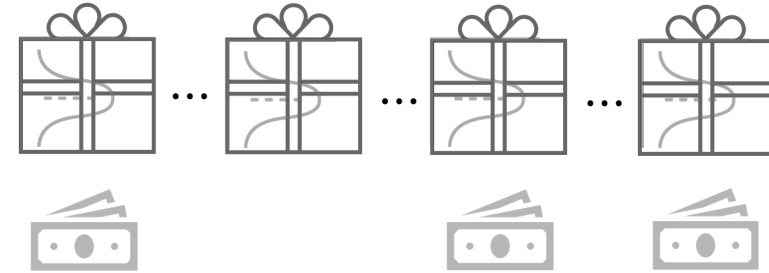
Outline

Studied Problem



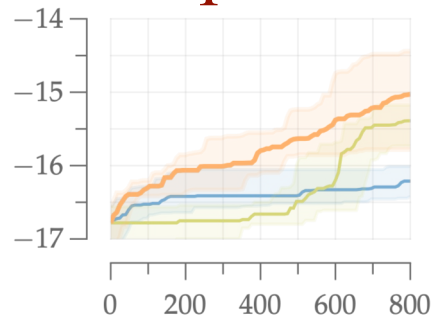
Cost-aware Bayesian optimization

Key idea



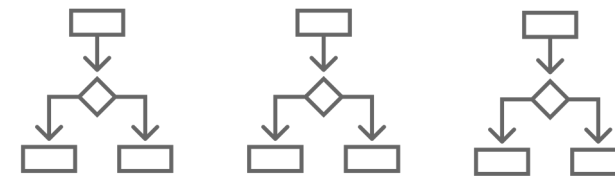
Link to Pandora's box and
Gittins index theory

Impact



Competitive empirical performance

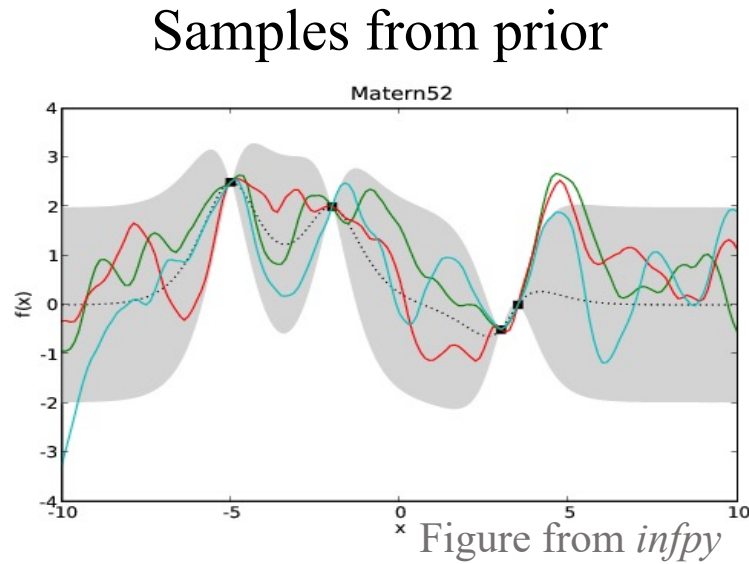
Future direction



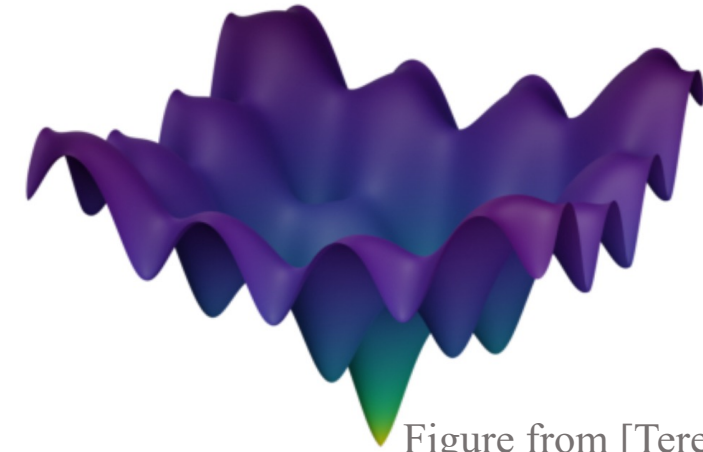
“Exotic” Bayesian optimization

Experiment Setup: Objective Functions

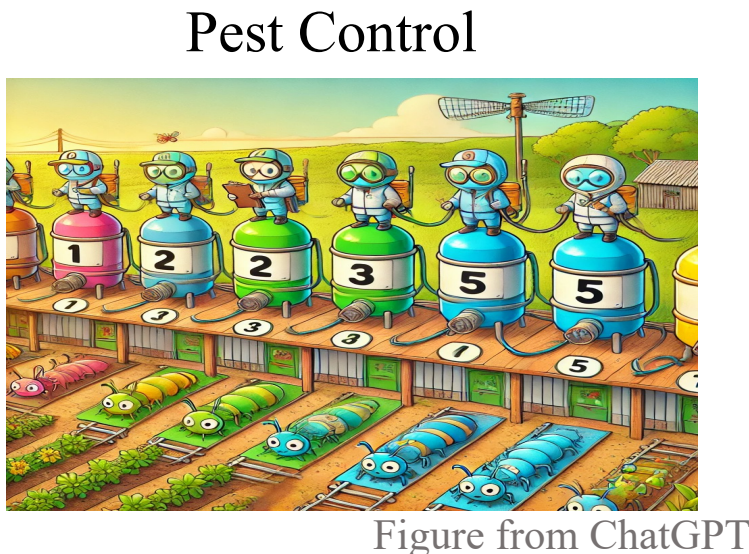
Synthetic



Ackley function



Empirical



Lunar Lander

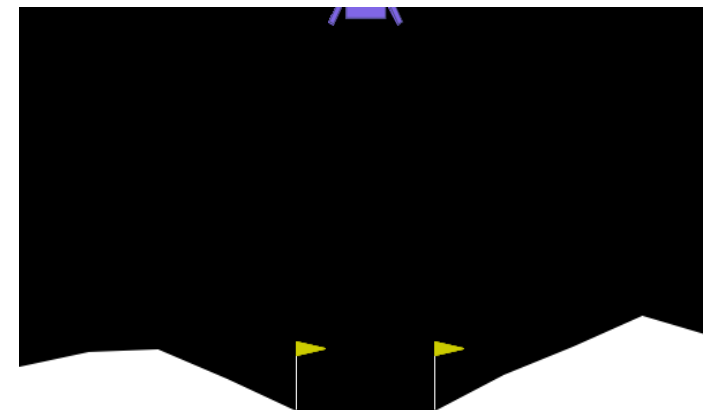
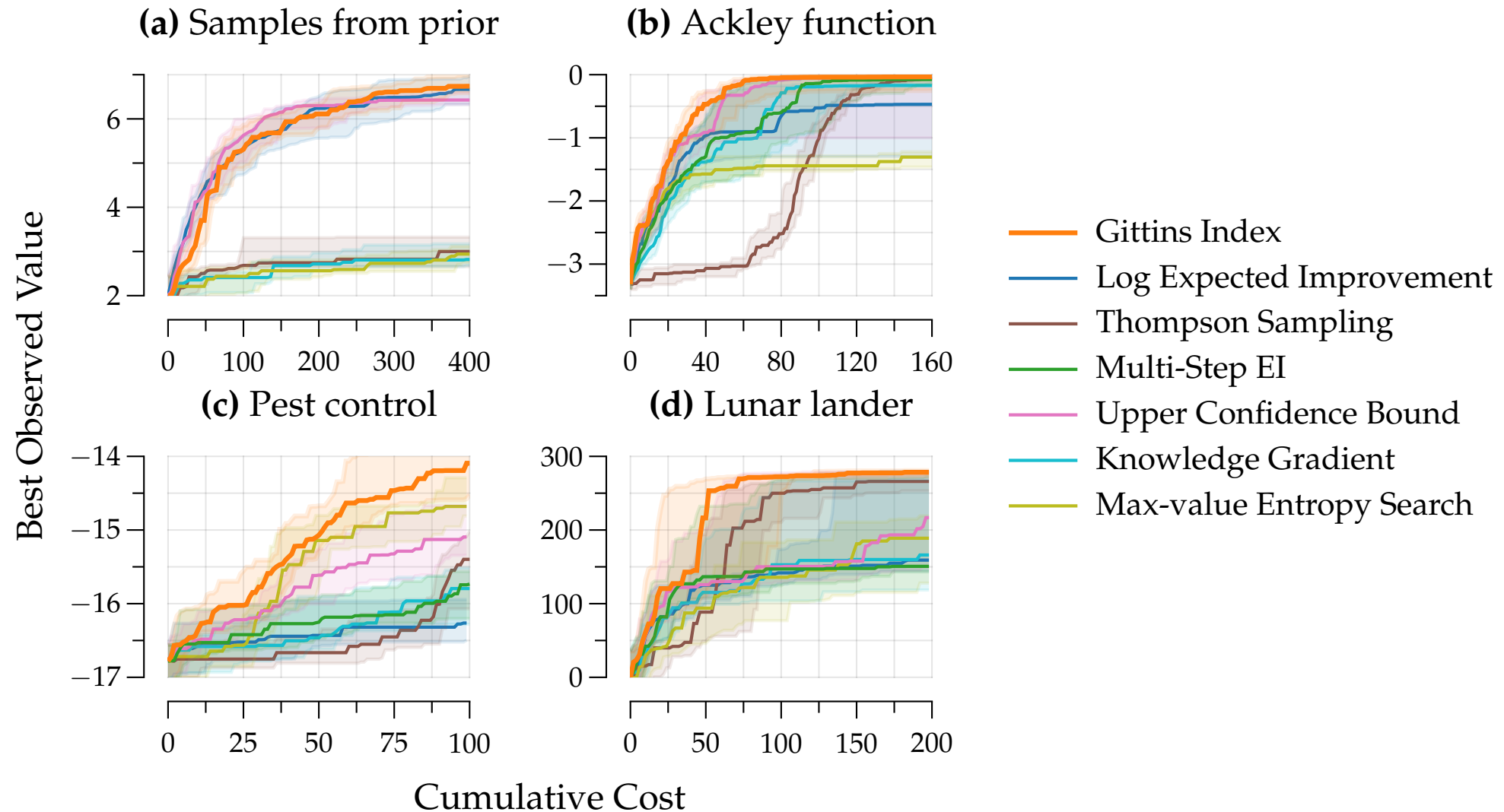
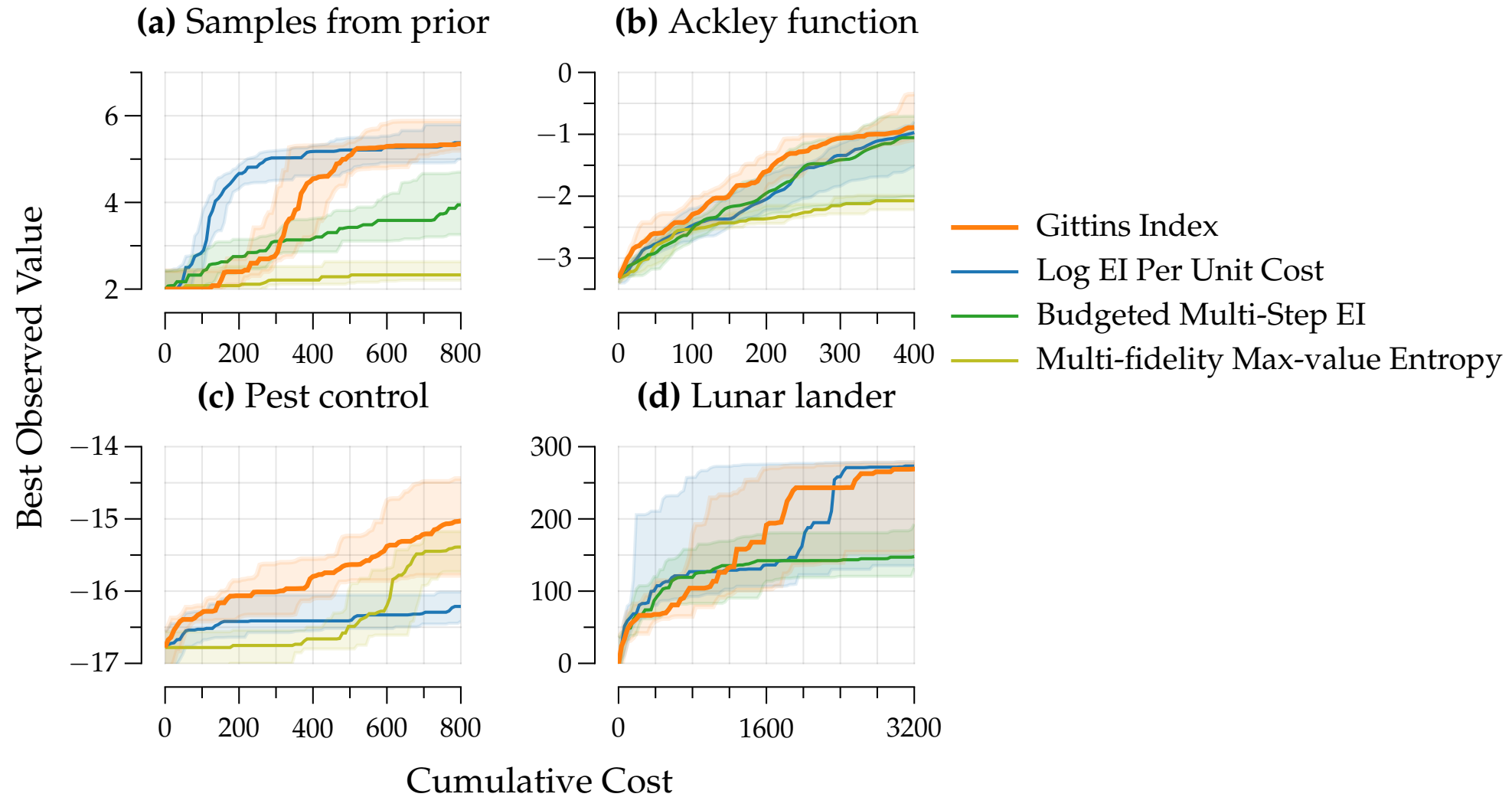


Figure from OpenAI Gym

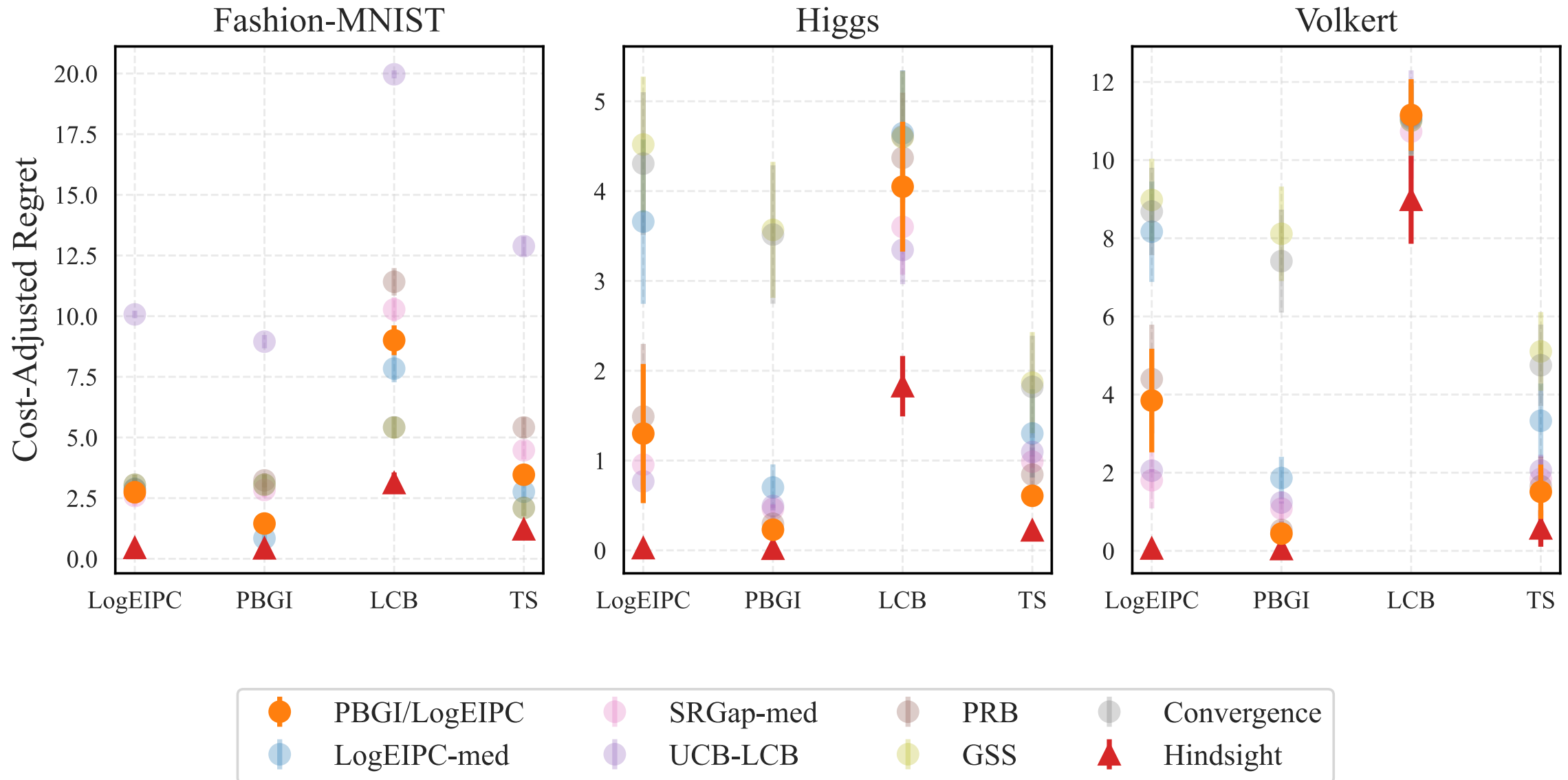
Uniform-cost: Gittins Index vs Baselines



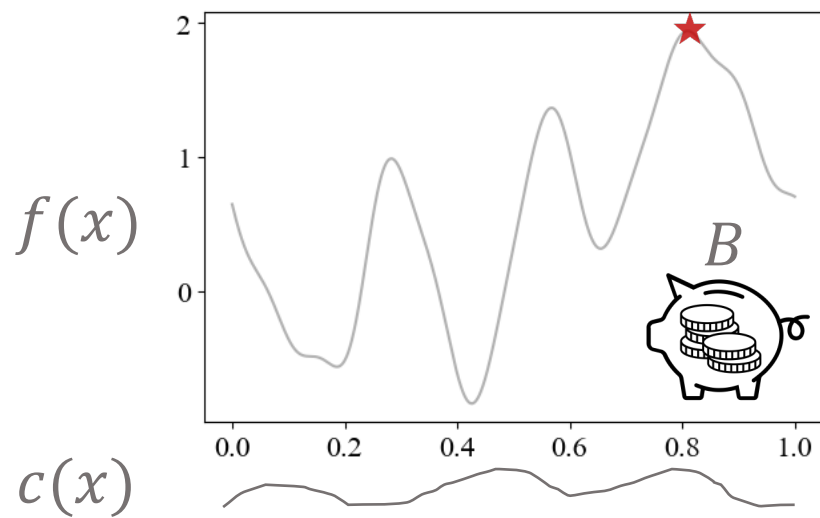
Varying-cost: Gittins Index vs Baselines



Stopping Rule: Gittins Index vs Baselines



Cost-aware Bayesian Optimization



Continuous

Correlated

Ebc & Cps

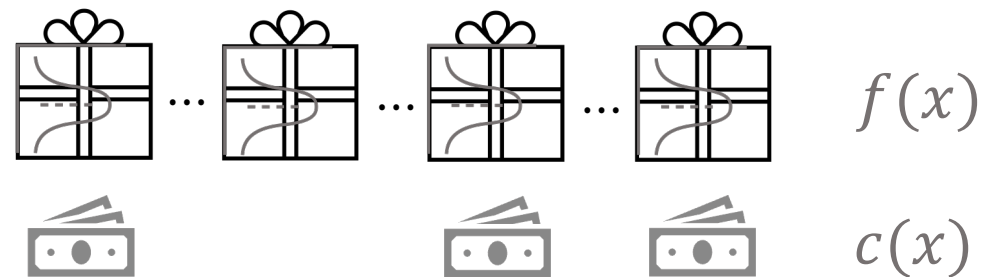
Acquisition function

+ stopping rule

Theoretical guarantee?

Pandora's Box

[Weitzman'79]



Discrete

Independent

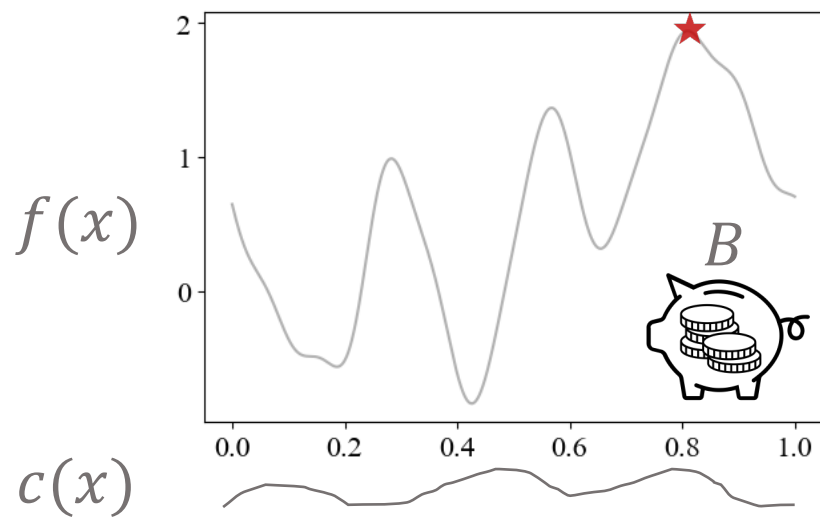
Cost-per-sample

Gittins index is optimal

incorporate posterior



Cost-aware Bayesian Optimization



Continuous

Correlated

Ebc & Cps

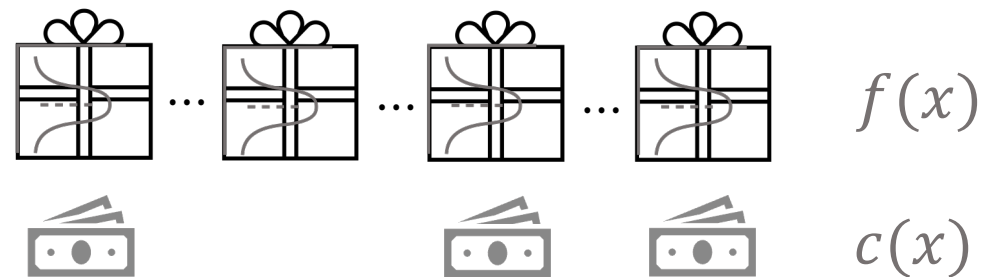
Acquisition function

+ stopping rule

Theoretical guarantee?

Pandora's Box

[Weitzman'79]



Discrete

Independent

Cost-per-sample

Gittins index is optimal

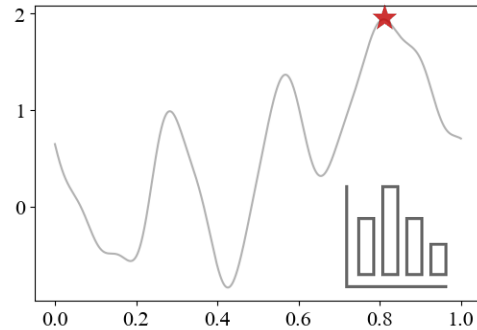
incorporate posterior

⇐

Yes! A bound on expected cost up to stopping

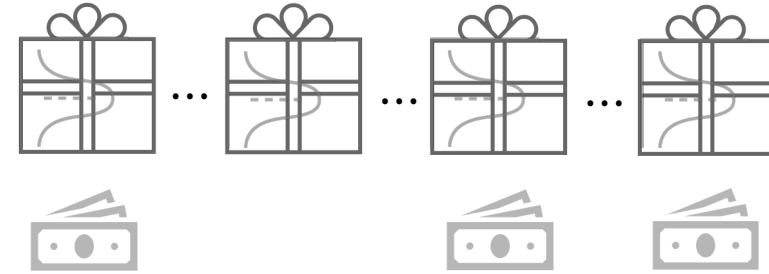
Gittins Index: A New Design Principle

Studied Problem



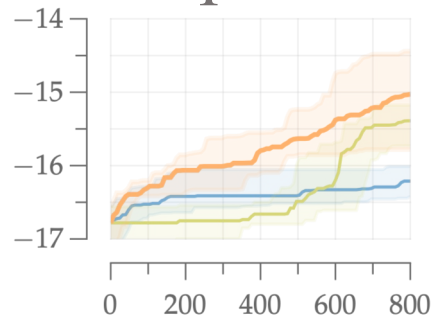
Cost-aware Bayesian optimization

Key idea



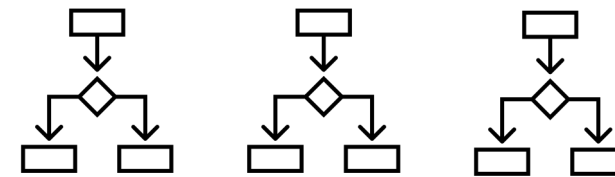
Link to Pandora's box and
Gittins index theory

Impact



Competitive empirical performance
w/ theoretical guarantee

Ongoing work



Bayesian optimization with
multi-stage feedback

Find our papers on arXiv!



"Cost-aware Bayesian Optimization
via the Pandora's Box Gittins Index."



"Cost-aware Stopping for
Bayesian Optimization."