NeurIPS'24 & INFORMS Data Mining Paper Competition Finalist

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

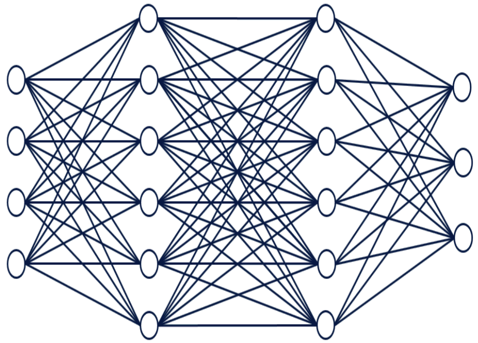
On arXiv soon!

Qian Xie (Cornell ORIE)

INFORMS Applied Probability Society Conference 2025

World of Hyperparameter Optimization

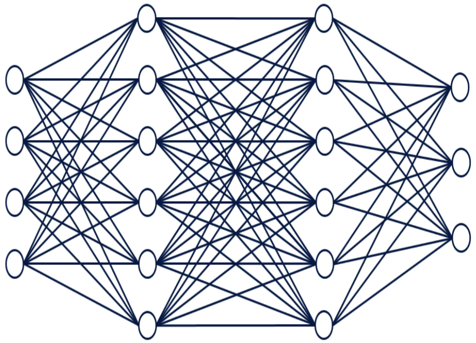
Hyperparameter tuning:

Training hyperparameters ⟶  ⟶ Accuracy

World of Hyperparameter Optimization

Hyperparameter tuning:

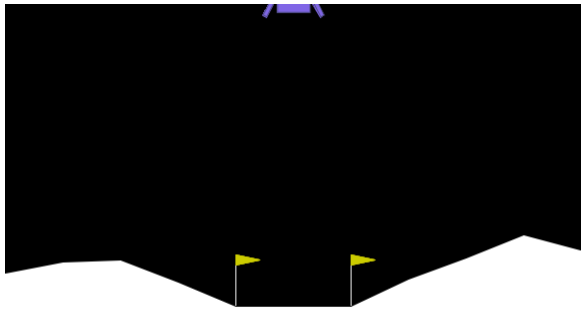
Training hyperparameters →



→ Accuracy

Control optimization:

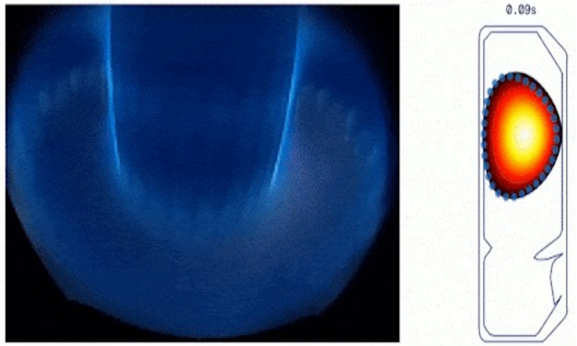
Control variables →



→ Reward

Plasma physics:

Fusion reactor design →

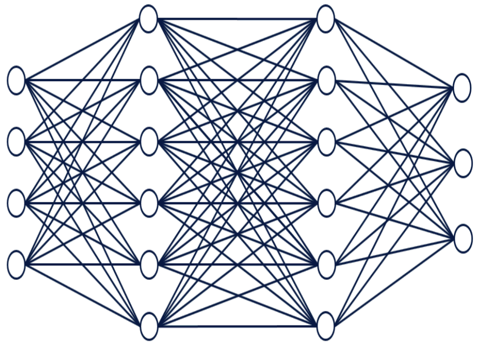


→ Stability

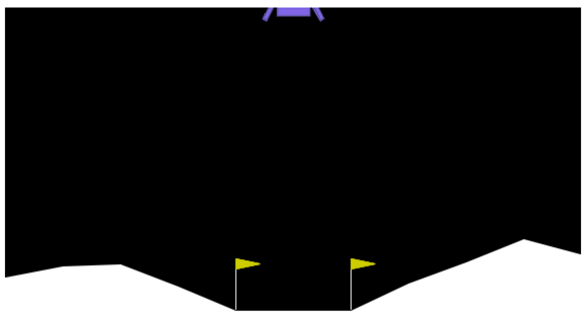
World of Hyperparameter Optimization

unknown &
expensive-to-evaluate

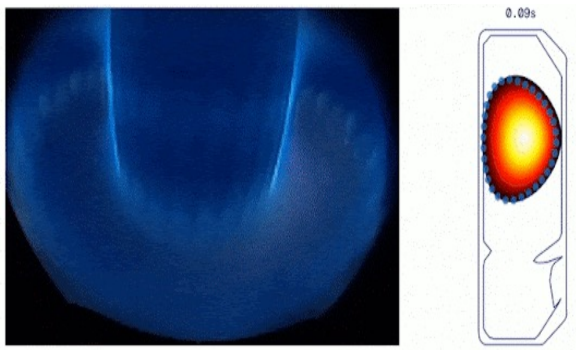
Hyperparameter tuning:

Training hyperparameters →  → Accuracy

Control optimization:

Control variables →  → Reward

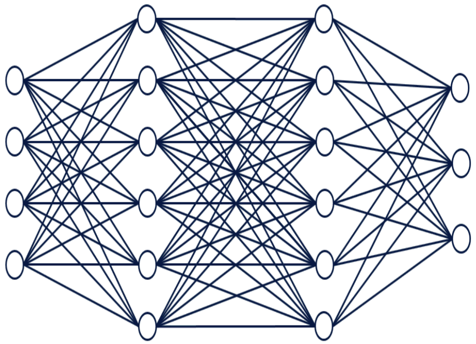
Plasma physics:

Fusion reactor design → 

→ Stability

Grid Search for AutoML

Hyperparameter tuning:

Training hyperparameters →  → Accuracy

unknown & expensive-to-evaluate

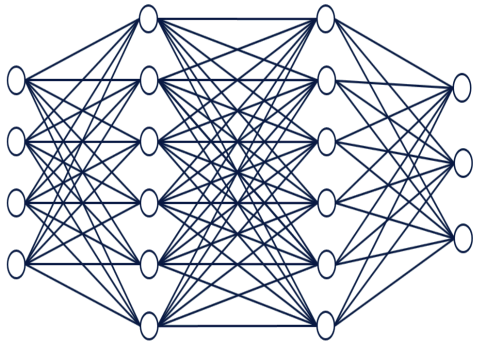
Parameter	Type	Scale	Range	Number of Options
Batch size	Integer	Log-scale	[16, 512]	**10**
Learning rate	Float	Log-scale	[1e-4, 1e-1]	**10**
Momentum	Float	Linear	[0.1, 0.99]	**10**
Weight decay	Float	Log-scale	[1e-5, 1e-1]	**10**
Number of layers	Integer	Linear	{1, 2, 3, 4}	**4**
Max units per layer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

40,000,000 combinations!

Grid Search for AutoML

Hyperparameter tuning:

unknown & expensive-to-evaluate

Training hyperparameters →  → Accuracy

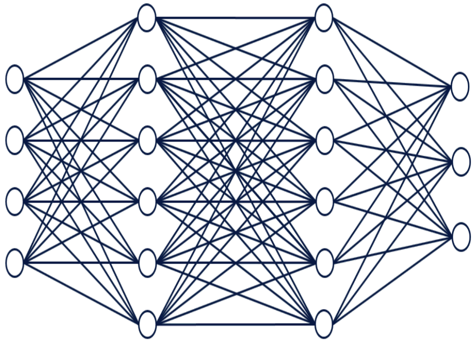
Number of Options
10
10
10
10
4
10
10

40,000,000 combinations!

Time-consuming!

Grid Search for AutoML

Hyperparameter tuning:

Training hyperparameters →  → Accuracy

unknown & expensive-to-evaluate

Number of Options
10
10
10
10
4
10
10

40,000,000 combinations!

Time-consuming!

More efficient:
Bayesian optimization

Bayesian Optimization

Black-box optimization:

unknown &
expensive-to-evaluate

Input hyperparameters ⟶  ⟶ Performance metric

Bayesian Optimization

Black-box function:

unknown &
expensive-to-evaluate

x $\longrightarrow$  $\longrightarrow$ $f(x)$

$f \sim$Stochastic Process

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

Bayesian Optimization

Black-box function:

$x \longrightarrow$  $\longrightarrow f(x)$

unknown & expensive-to-evaluate

$f\sim$Stochastic Process

limited #evaluations

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

Bayesian Optimization

Black-box function:

$x \longrightarrow$  $\longrightarrow f(x)$

unknown &
expensive-to-evaluate

$f \sim$ Stochastic Process

limited #evaluations

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

Key idea: maintain probabilistic belief about f

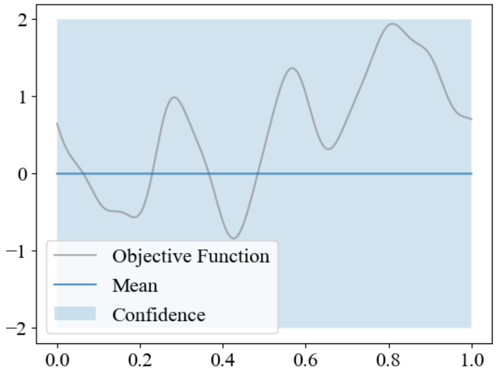
Bayesian Optimization

$x \longrightarrow$

$\longrightarrow f(x)$

Black-box function

Maintain belief



Probabilistic model

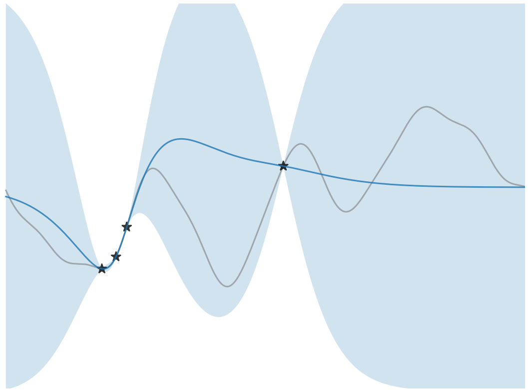
Bayesian Optimization

Collect data

$x \longrightarrow$  $\longrightarrow f(x)$

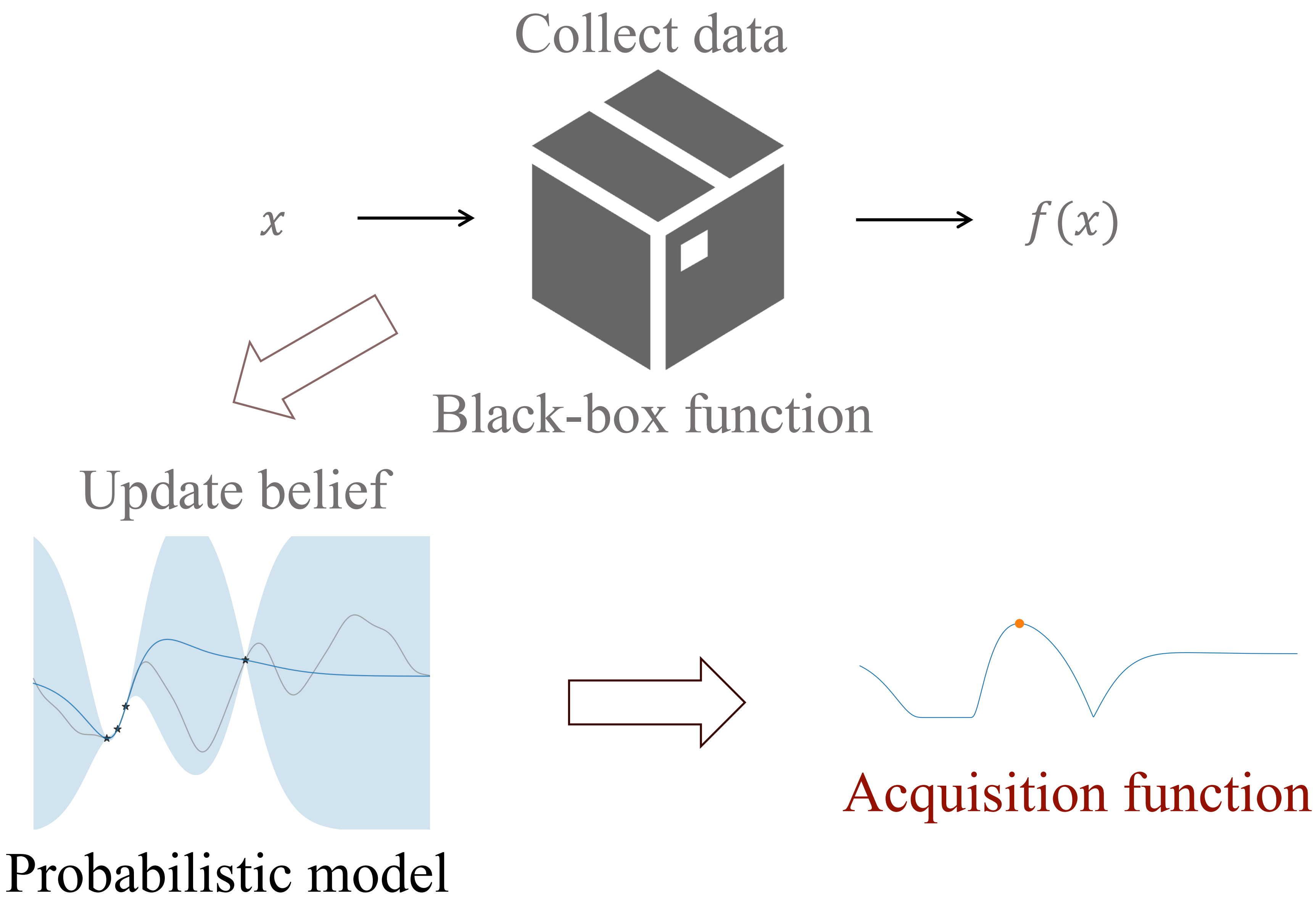
Black-box function

Update belief

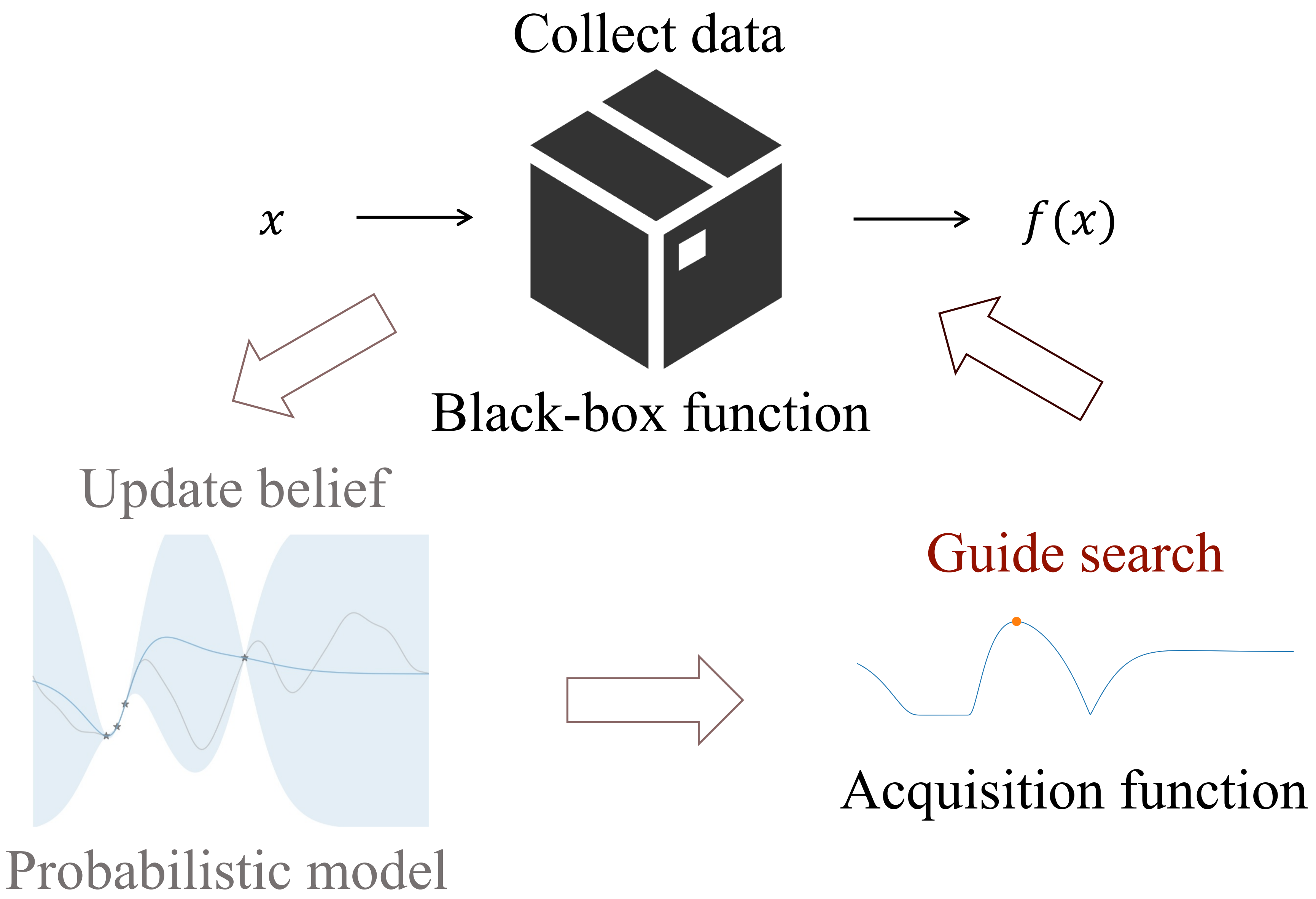


Probabilistic model

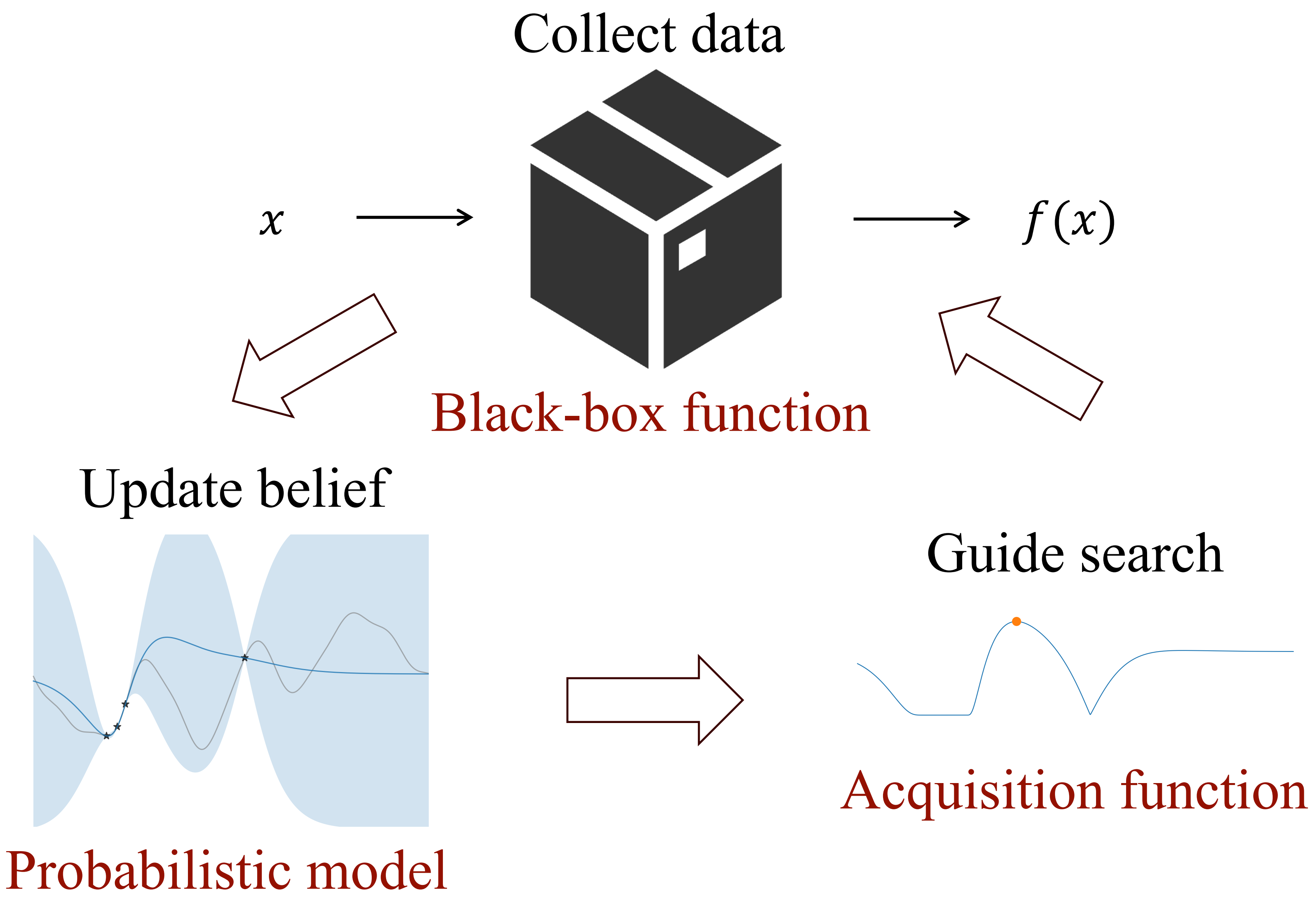
Bayesian Optimization



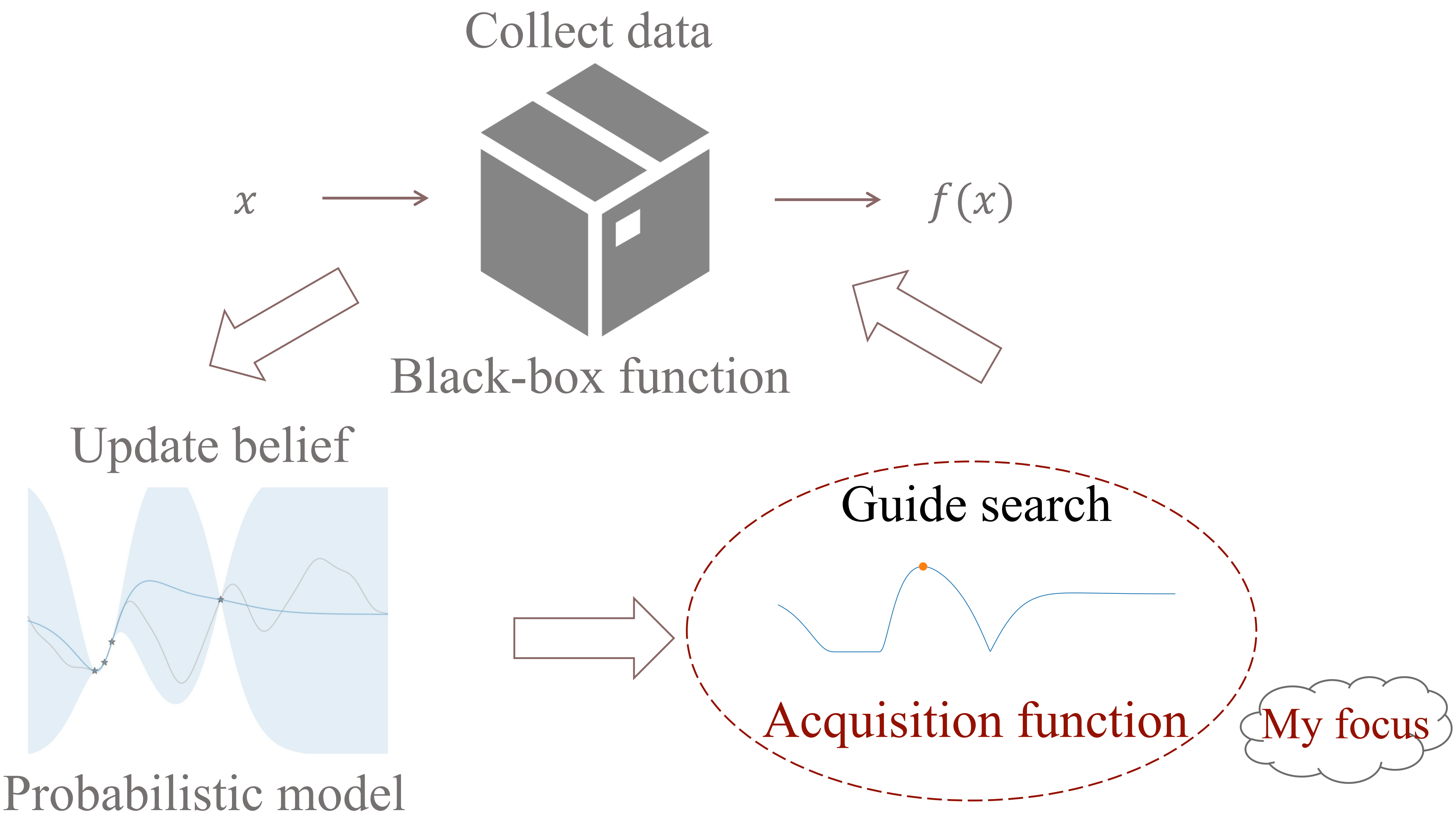
Bayesian Optimization



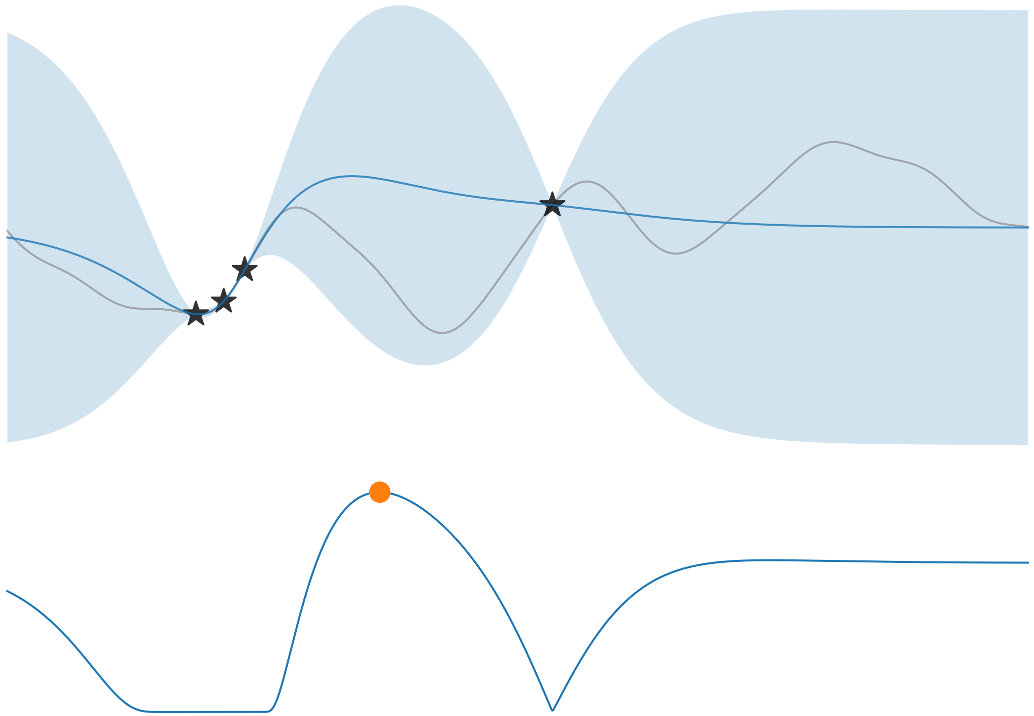
Bayesian Optimization



Bayesian Optimization

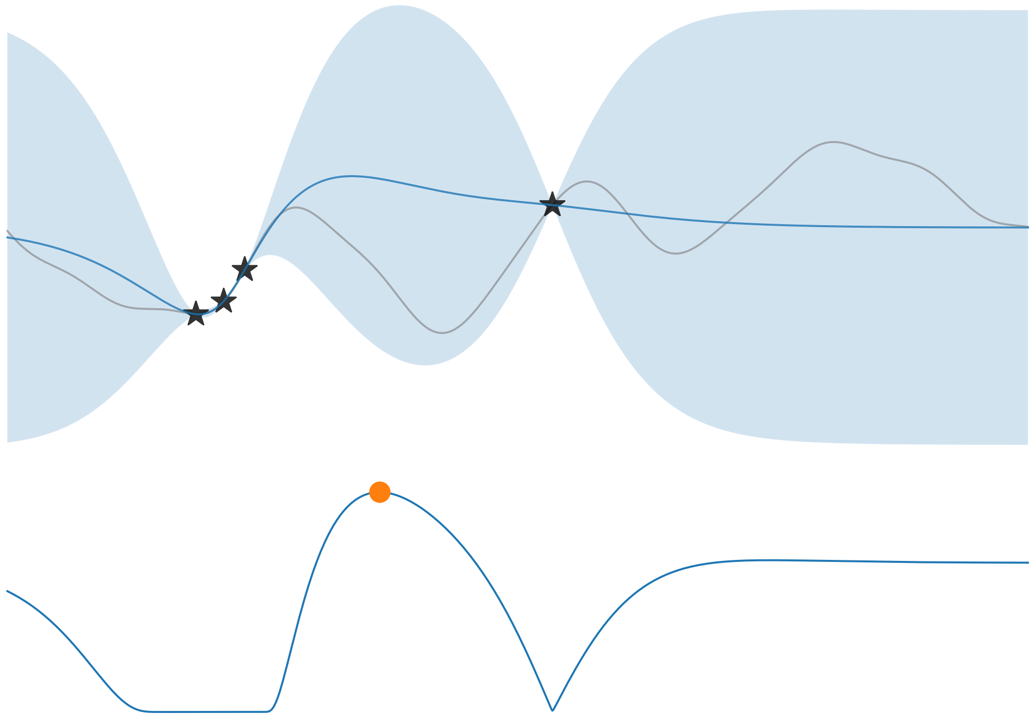


Classic Acquisition Functions



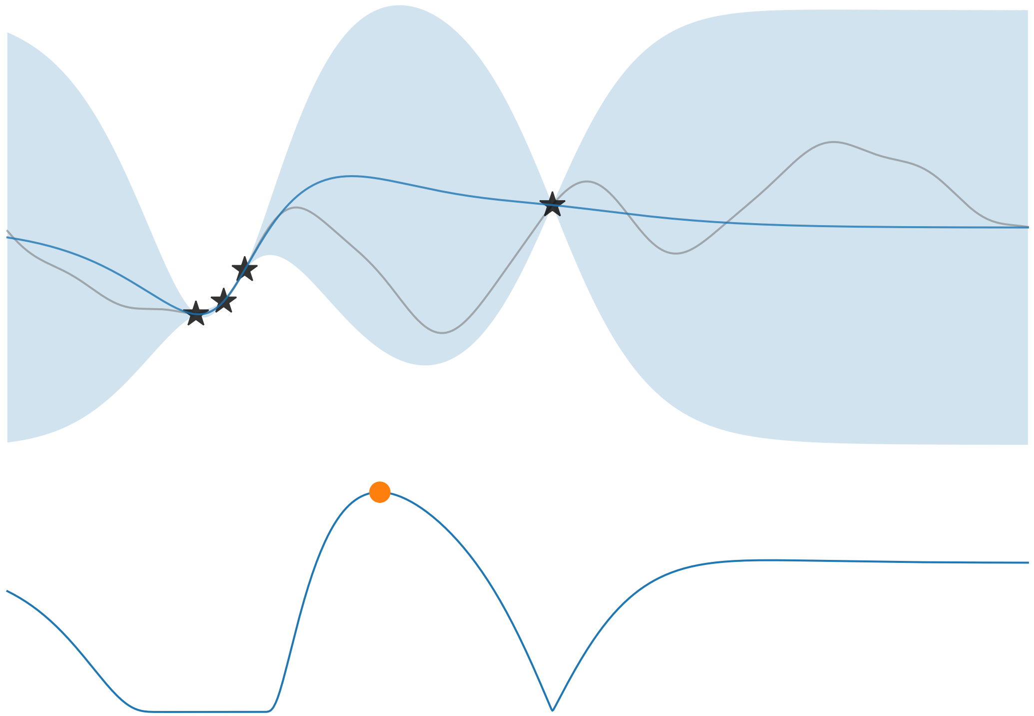
- Improvement-based
- Entropy-based
- Upper Confidence Bound
- Thompson Sampling

New Acquisition Function: Gittins Index



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

New Acquisition Function: Gittins Index

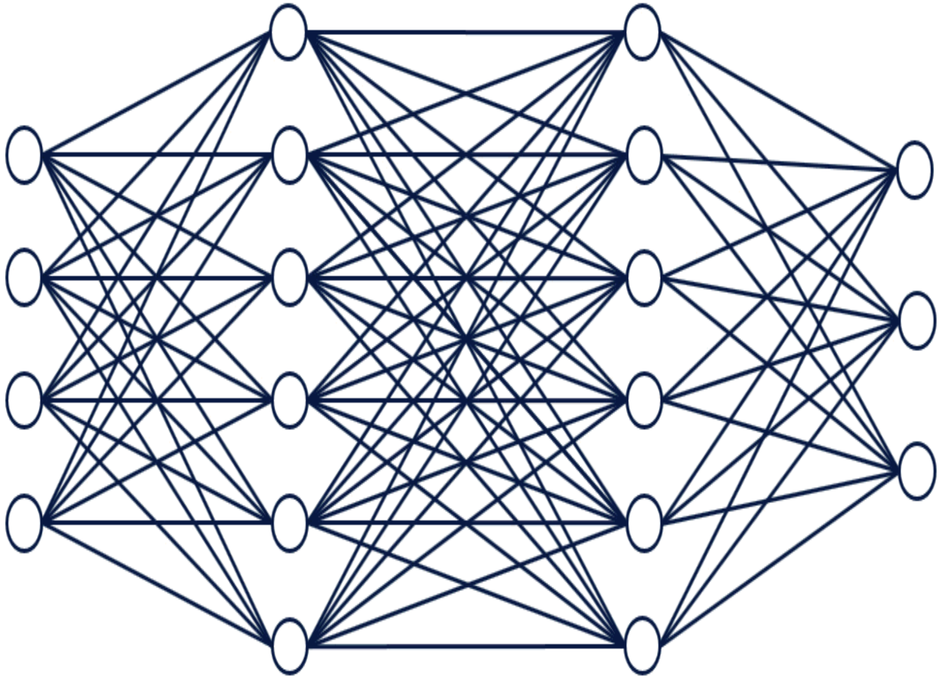


- Improvement-based
- Entropy-based
- Upper Confidence Bound
- Thompson Sampling
- My work: Gittins Index

Why another acquisition function?

Under-explored Practical Considerations

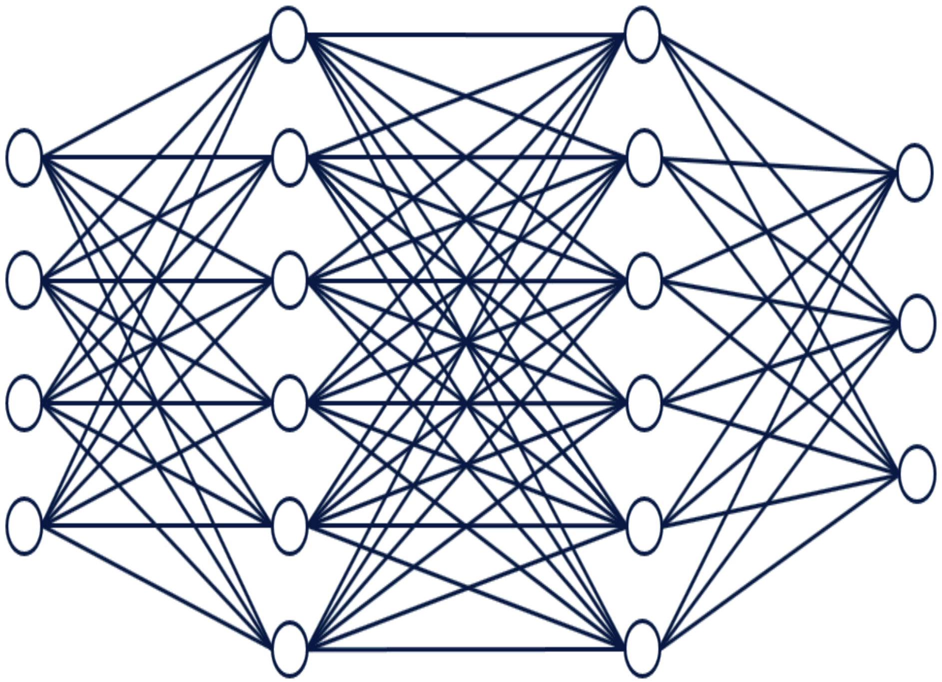
Varying evaluation costs



Under-explored Practical Considerations

Varying evaluation costs

Smart stopping time

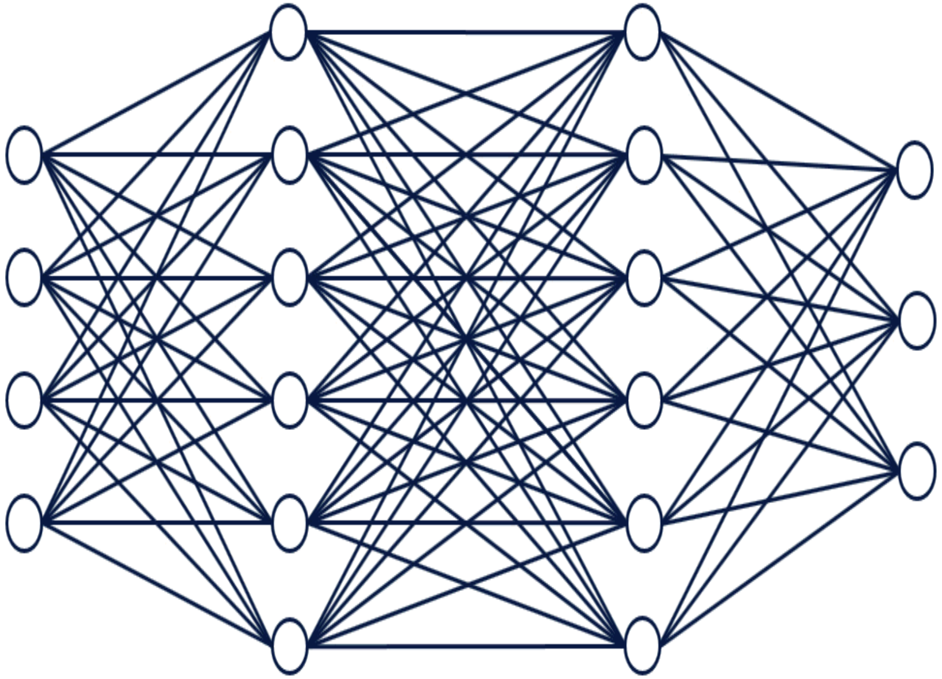


Under-explored Practical Considerations

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

Gittins index

Varying evaluation costs — Cost-aware

Smart stopping time — Stopping-aware

Observable multi-stage feedback — 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

In this talk

Smart stopping time

Observable multi-stage feedback

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


Peter Frazier

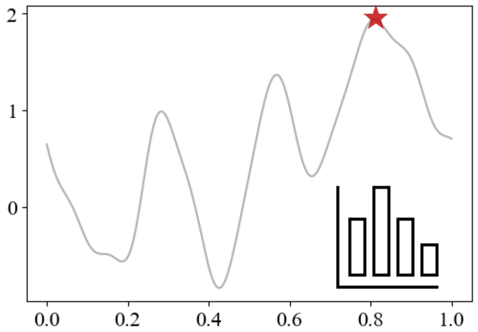

Alexander Terenin


Ziv Scully

Observable multi-stage feedback
[Ongoing work]

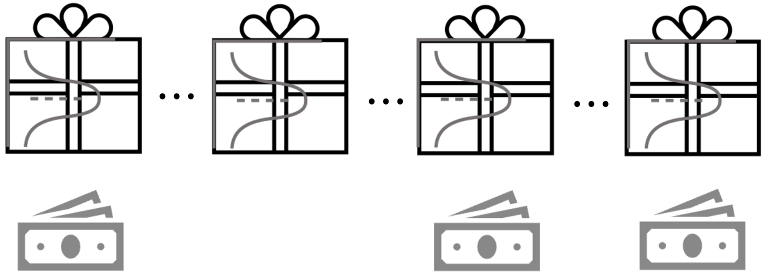
Outline

Studied Problem



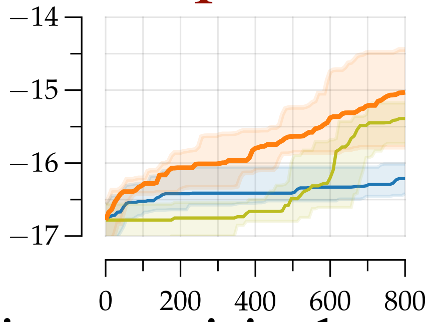
Cost-aware Bayesian optimization

Key idea



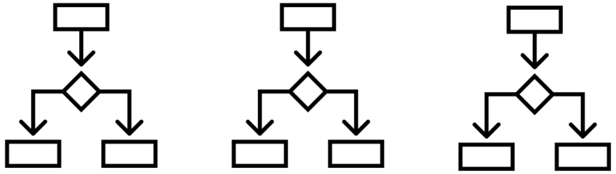
Link to simplified problem
and Gittins index theory

Impact



Competitive empirical performance

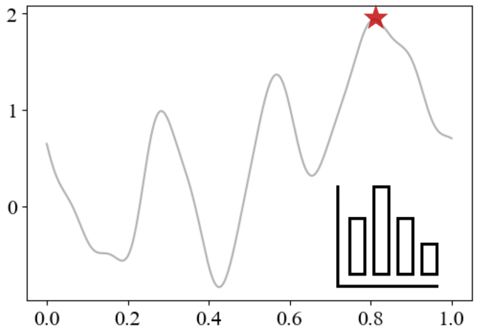
Future direction



"Exotic" Bayesian optimization

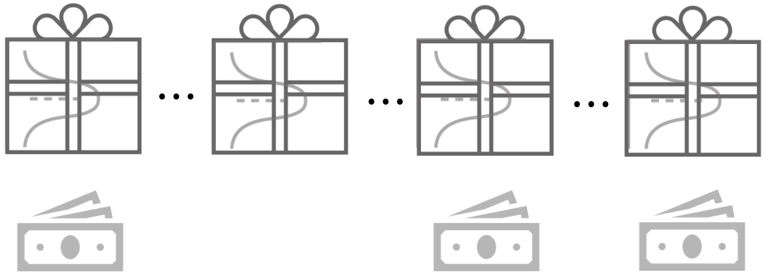
Outline

Studied Problem



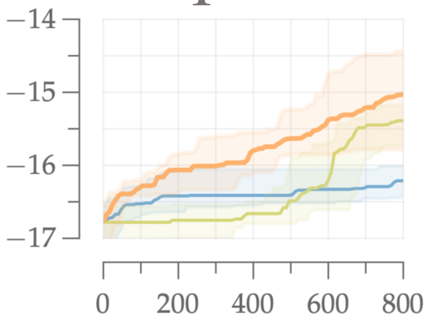
Cost-aware Bayesian optimization

Key idea



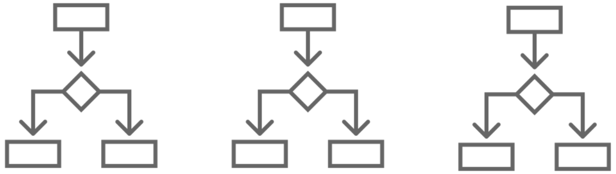
Link to simplified problem and Gittins index theory

Impact



Competitive empirical performance

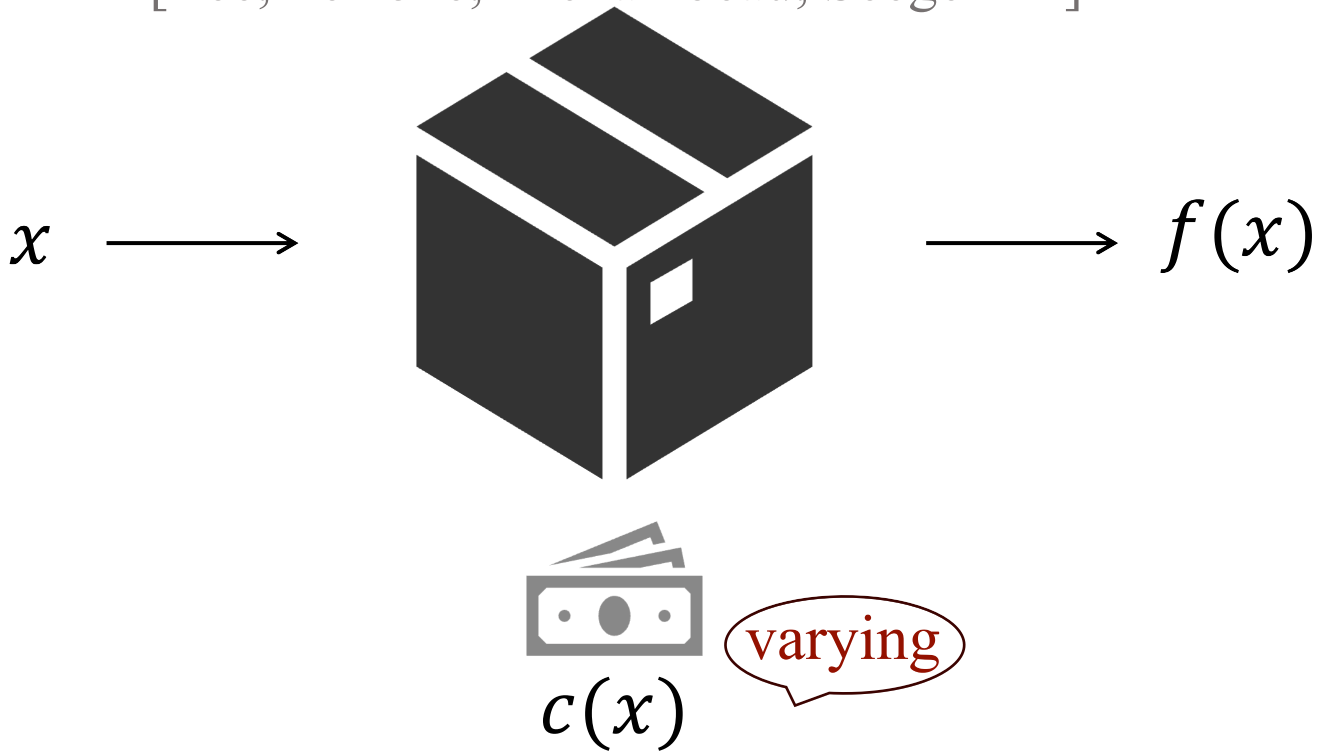
Future direction



"Exotic" Bayesian optimization

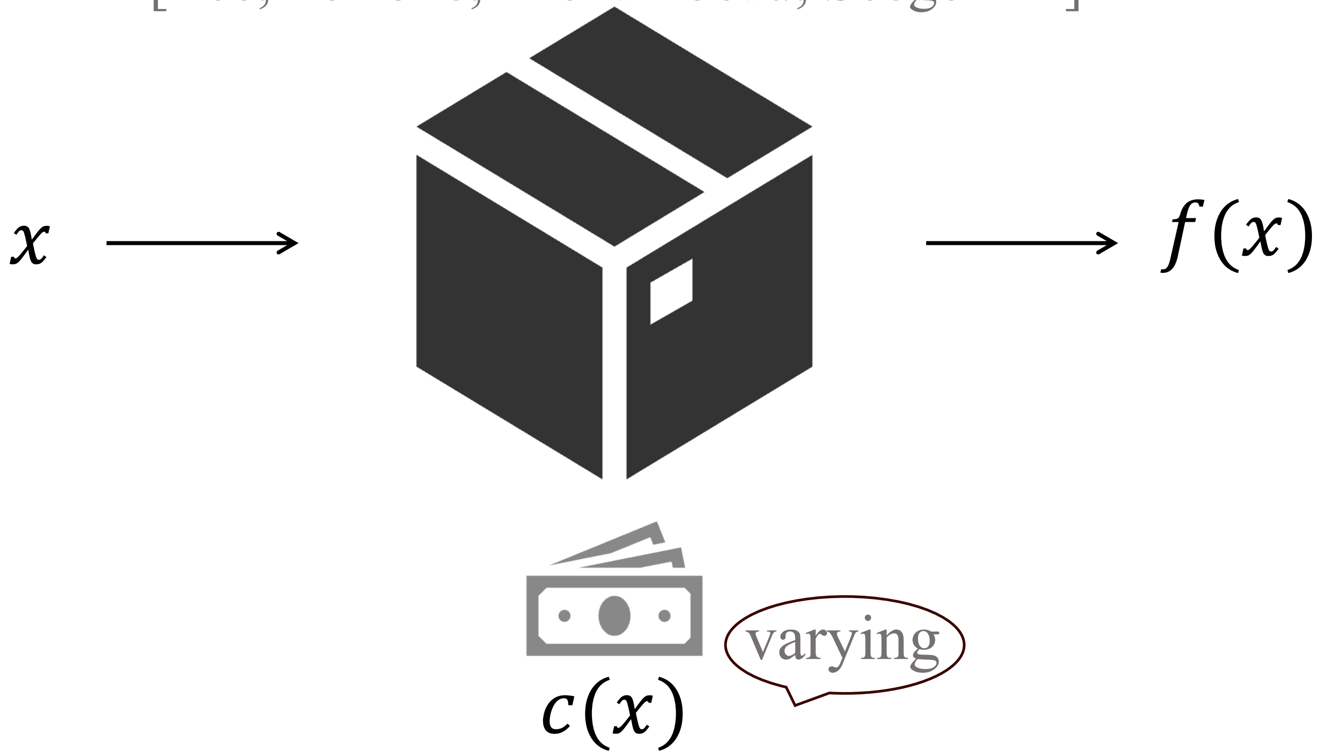
Cost-aware Bayesian Optimization

[Lee, Perrone, Archambeau, Seeger'21]



Cost-aware Bayesian Optimization

[Lee, Perrone, Archambeau, Seeger'21]

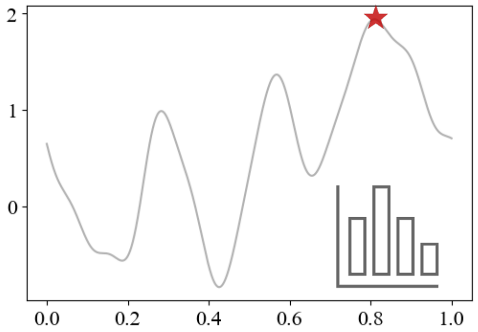


Goal: $\sup_{\text{policy}} \mathbb{E} \max_{t=1,2,\ldots,T} f(x_t)$

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

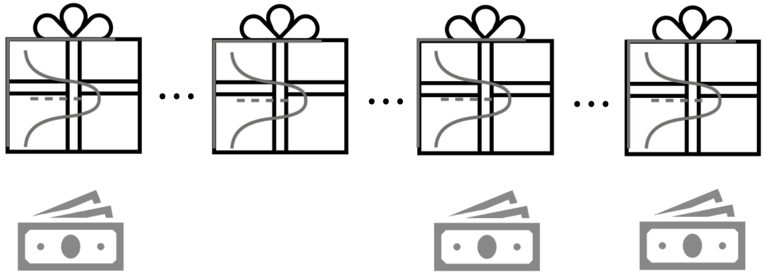
Outline

Studied Problem



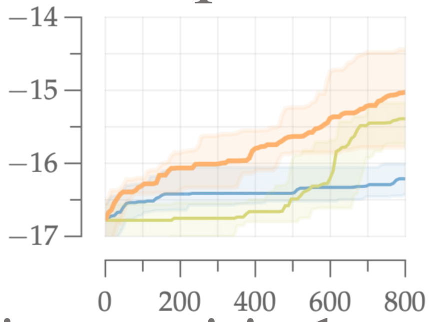
Cost-aware Bayesian optimization

Key idea



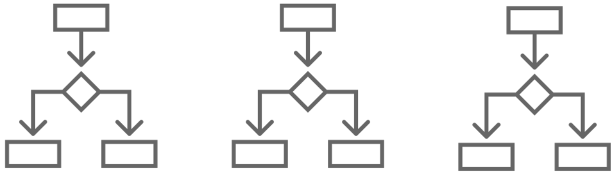
Link to simplified problem and Gittins index theory

Impact



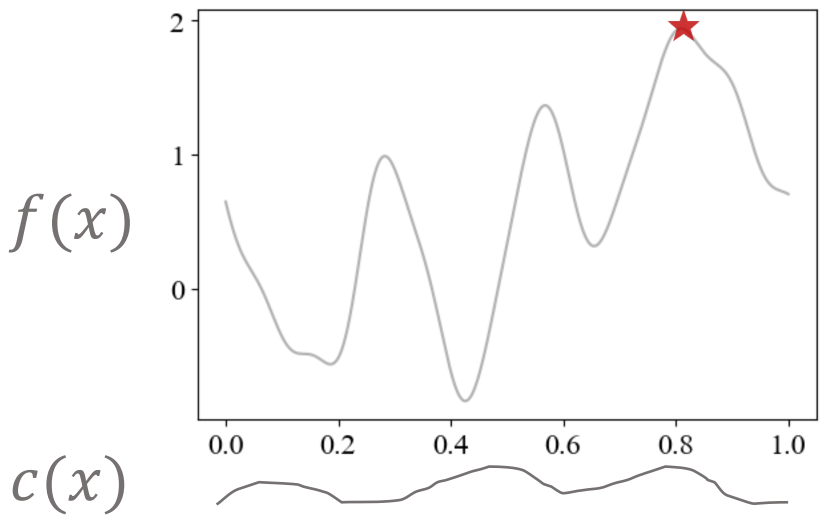
Competitive empirical performance

Future direction



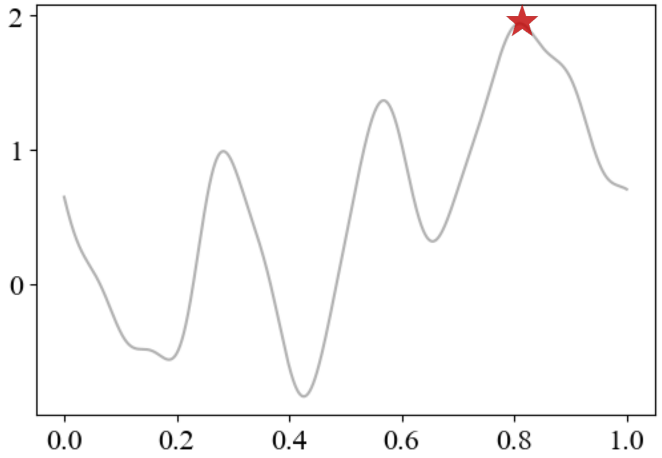
"Exotic" Bayesian optimization

Cost-aware Bayesian Optimization



Cost-aware Bayesian Optimization

$f(x)$

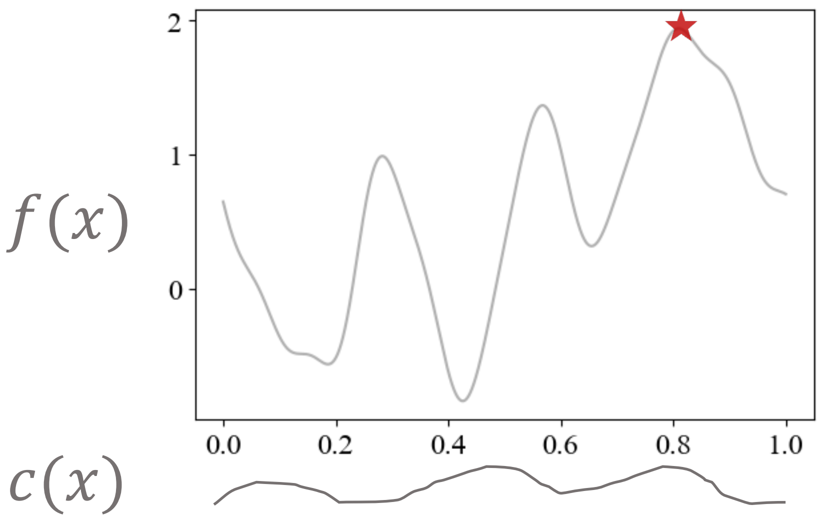


$c(x)$

Continuous

Correlated

Cost-aware Bayesian Optimization



Continuous

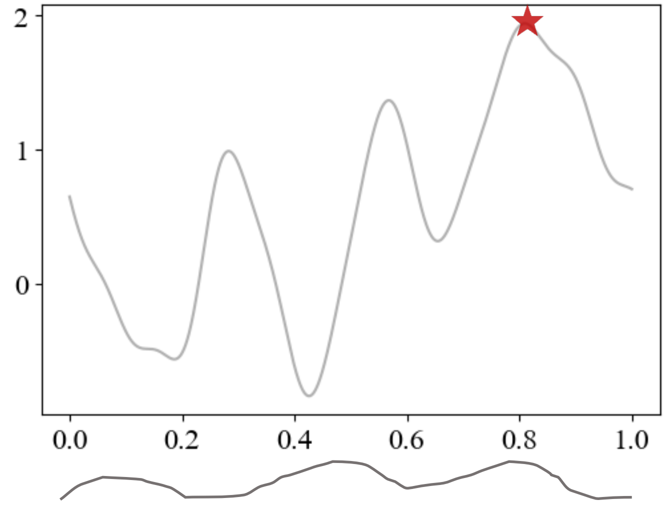
Correlated

Intractable MDP!

Cost-aware Bayesian Optimization Pandora's Box

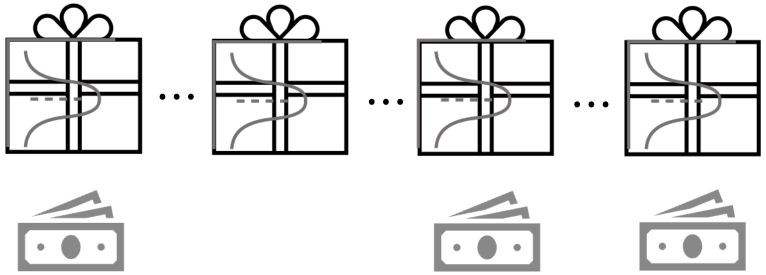
[Weitzman'79]

$f(x)$



$c(x)$

$\Longrightarrow$



$f(x)$

$c(x)$

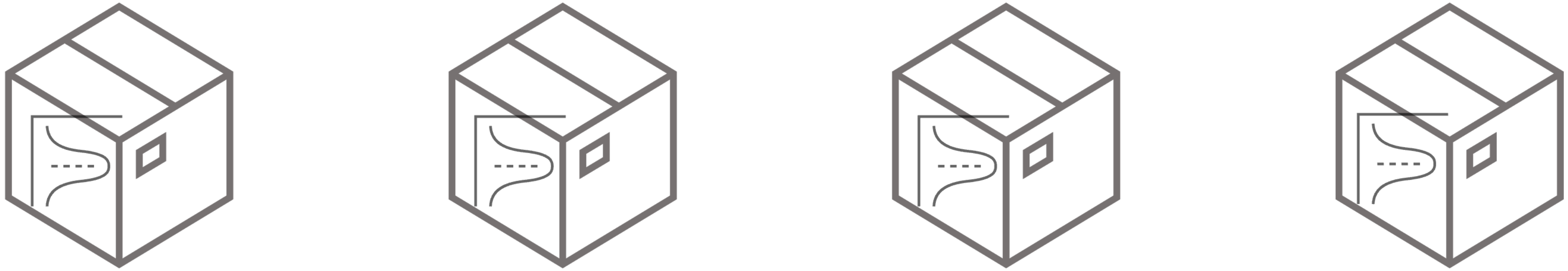
Continuous $\Longrightarrow$ Discrete

Correlated $\Longrightarrow$ Independent

Intractable MDP!

Pandora's Box

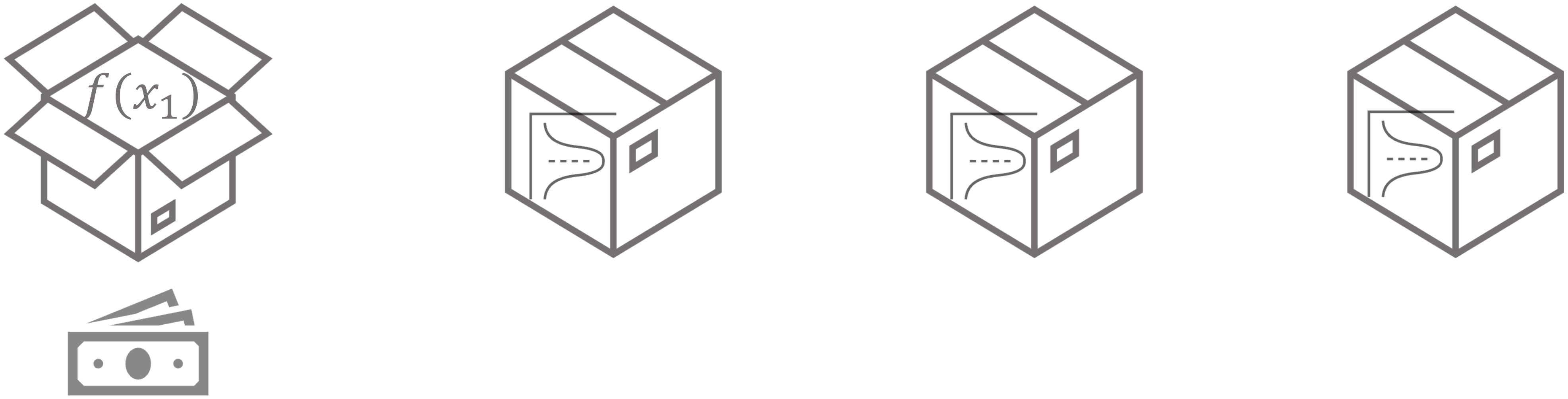
$t = 0$



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

Pandora’s Box

$t = 1$

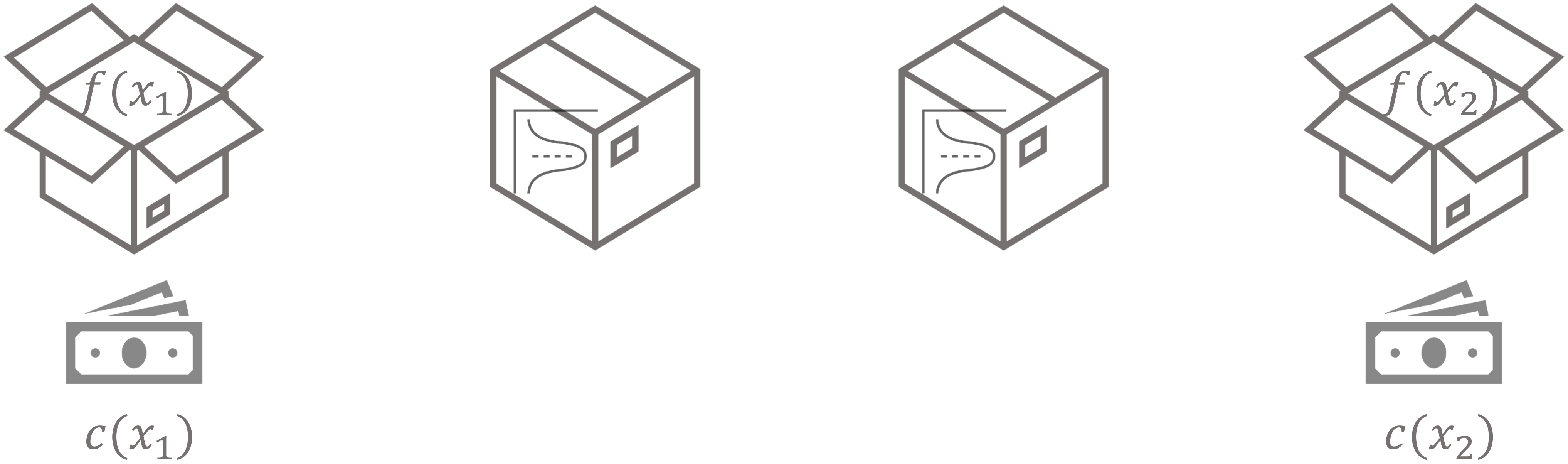


$c(x_1)$

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

Pandora's Box

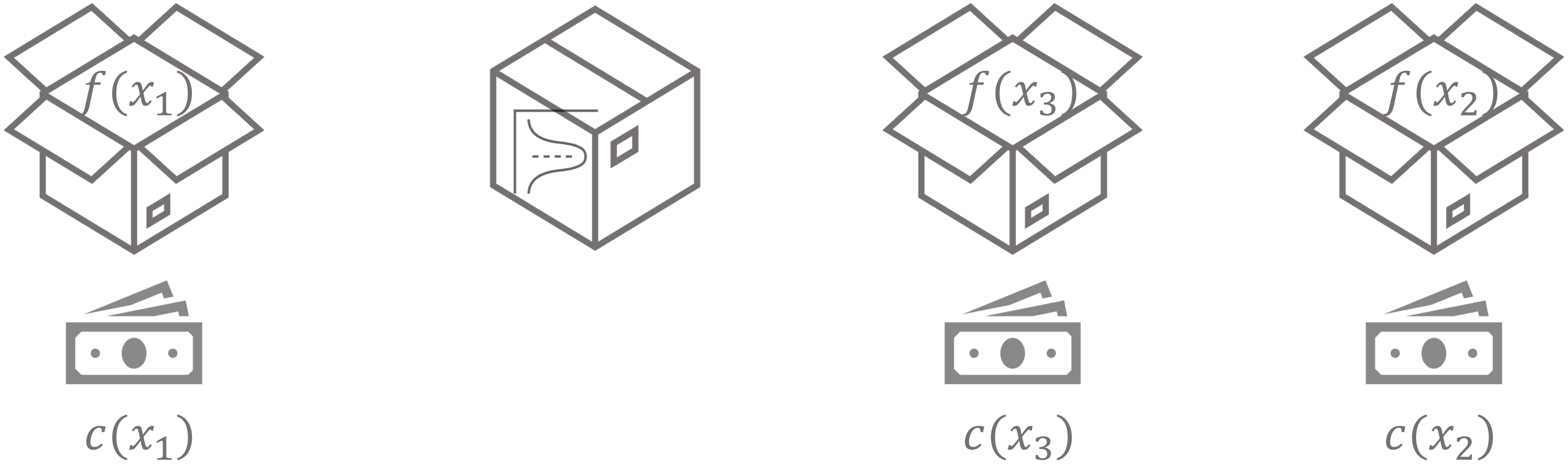
$t = 2$



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

Pandora's Box

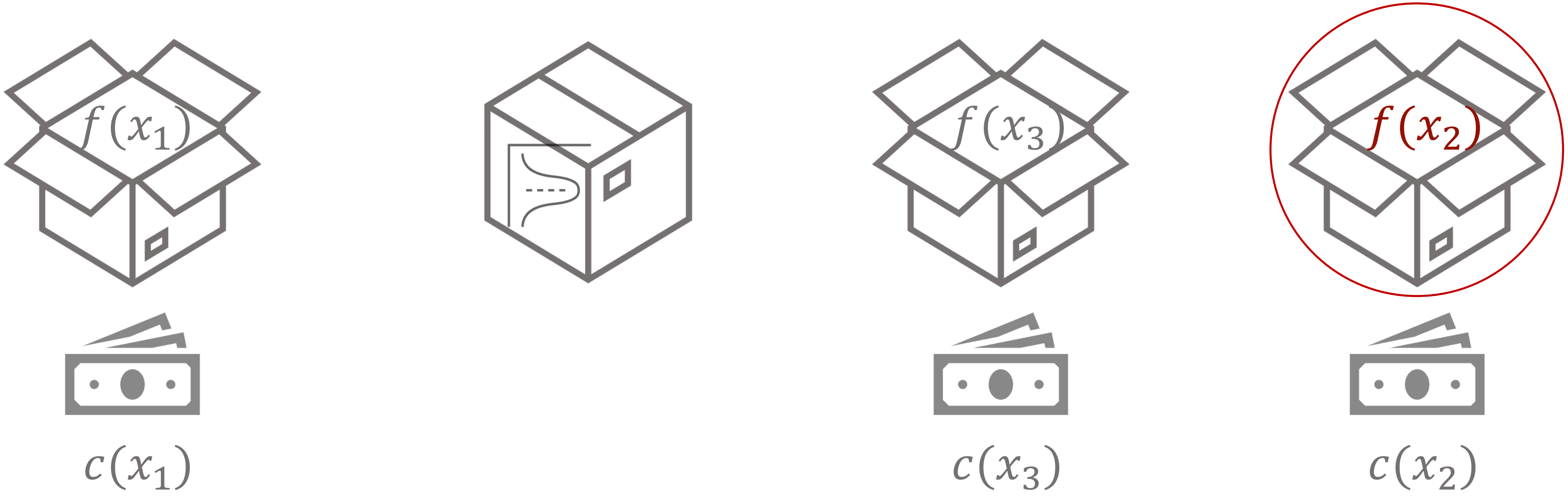
$t = 3$



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

Pandora’s Box

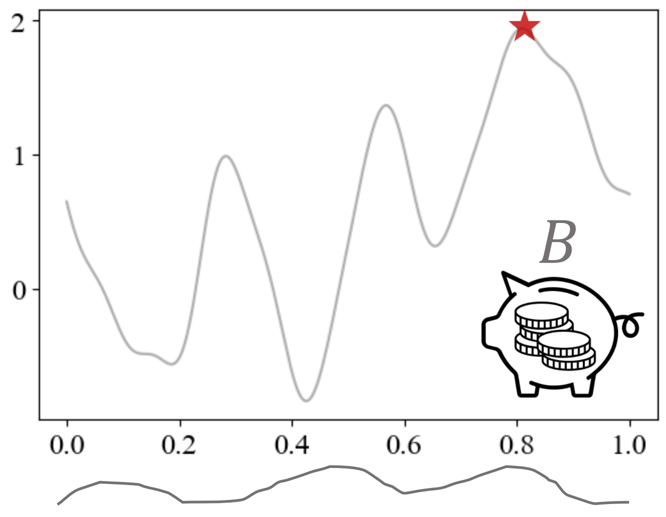
$t = T$, stop



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

Cost-aware Bayesian Optimization

$f(x)$



$c(x)$

Continuous

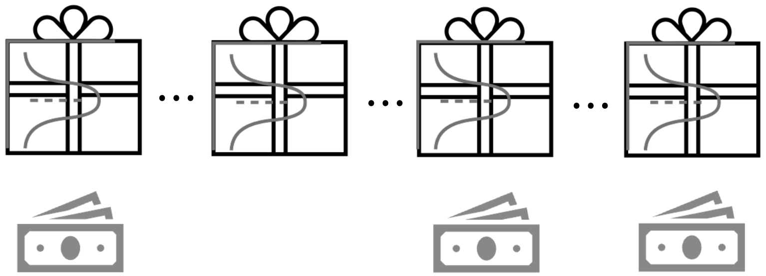
Correlated

Budget-constrained

$$\sup_{\text{policy}} \mathbb{E} \max_{t=1,2,\ldots,T} f(x_t)$$
$$\text{s.t. } \sum_{t=1}^{T} c(x_t) \le B$$

Pandora's Box

[Weitzman'79]



$f(x)$

$c(x)$

Discrete

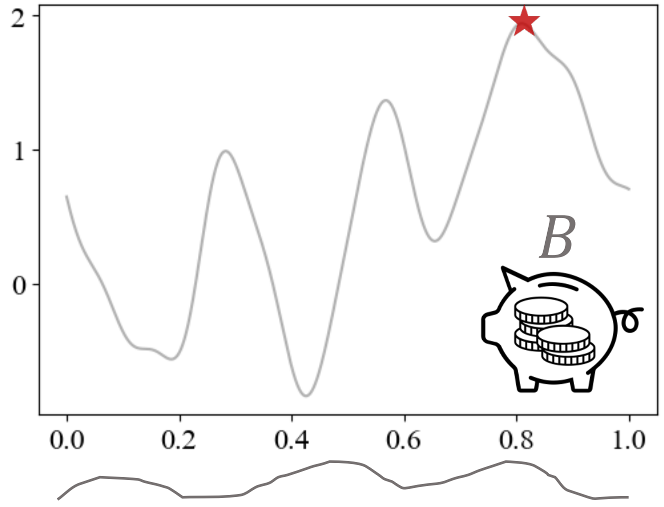
Independent

Cost-per-sample

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

Cost-aware Bayesian Optimization

$f(x)$



$c(x)$

Continuous

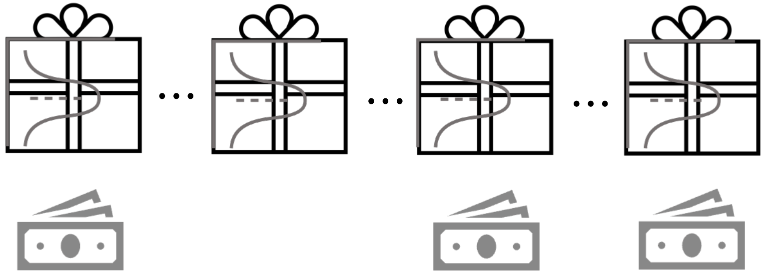
Correlated

Expected-budget-constrained

$$\sup_{\text{policy}} \mathbb{E} \max_{t=1,2,\ldots,T} f(x_t)$$
$$\text{s.t. } \mathbb{E} \sum_{t=1}^{T} c(x_t) \le B$$

Pandora's Box

[Weitzman'79]



$f(x)$

$c(x)$

Discrete

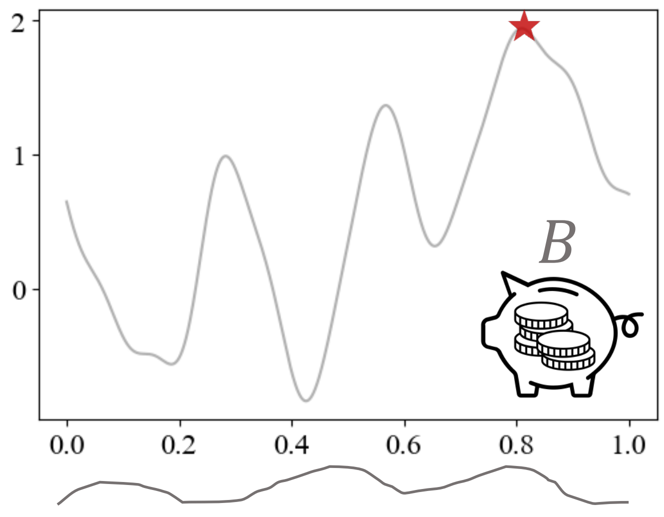
Independent

Cost-per-sample

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

Cost-aware Bayesian Optimization

$f(x)$



$c(x)$

Continuous

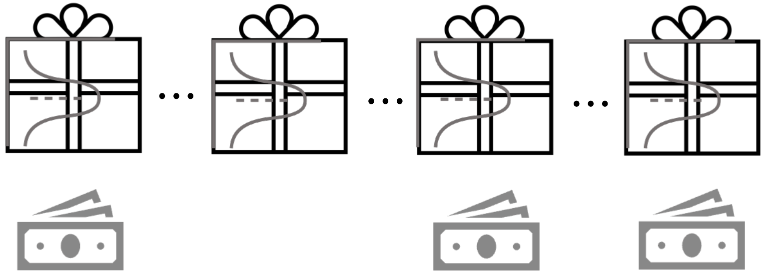
Correlated

Ebc & Cps

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

Pandora's Box

[Weitzman'79]



$f(x)$

$c(x)$

Discrete

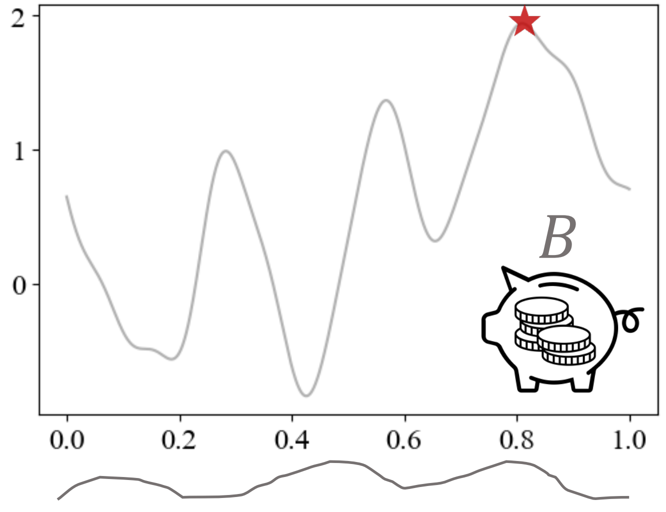
Independent

Cost-per-sample

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

Cost-aware Bayesian Optimization

$f(x)$



$c(x)$

Continuous

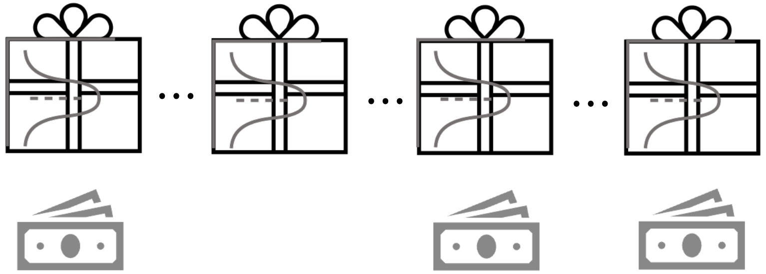
Correlated

Ebc & Cps

Intractable MDP!

Pandora's Box

[Weitzman'79]



$f(x)$

$c(x)$

Discrete

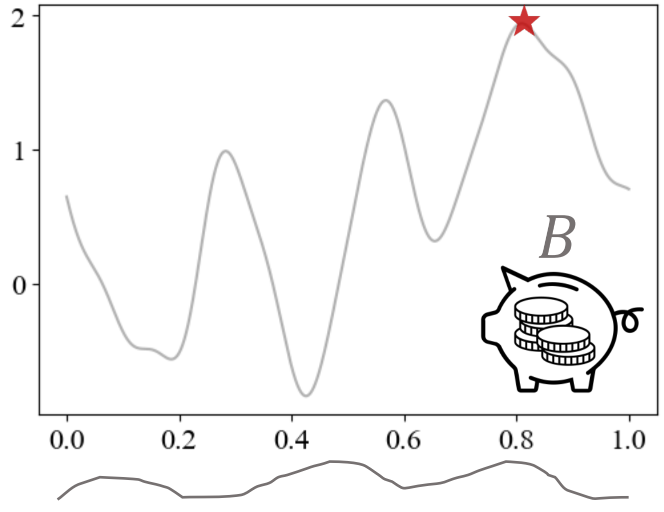
Independent

Cost-per-sample

Optimal policy: Gittins index

Cost-aware Bayesian Optimization

$f(x)$



$c(x)$

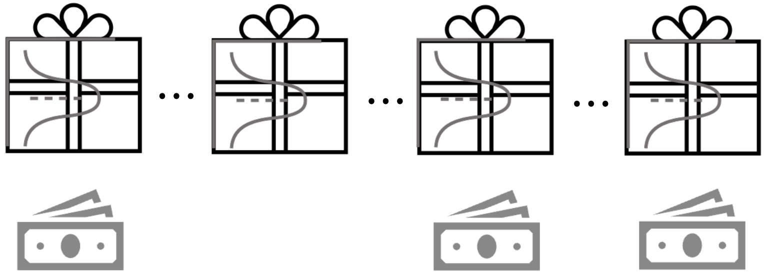
Continuous

Correlated

Ebc & Cps

Pandora's Box

[Weitzman'79]



$f(x)$

$c(x)$

Discrete

Independent

Cost-per-sample

How to translate?

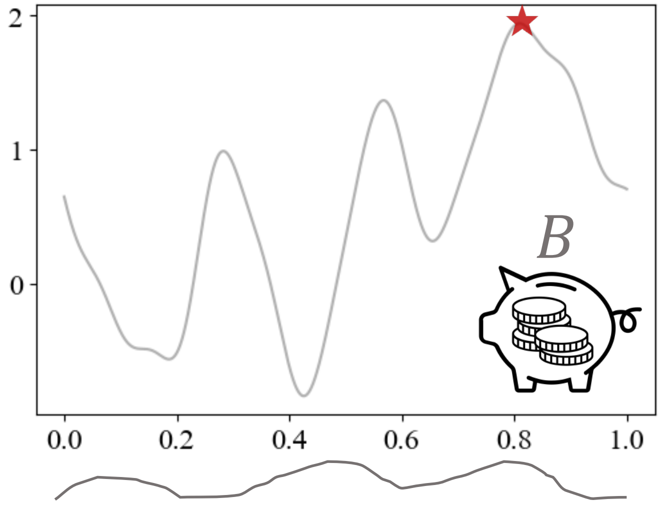
$\Longleftarrow$ Optimal policy: Gittins index

Cost-aware Bayesian Optimization

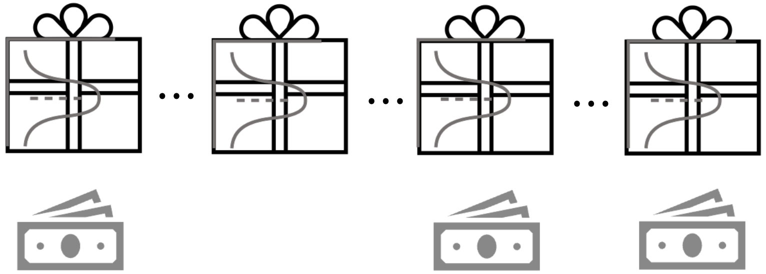
Pandora's Box

[Weitzman'79]

$f(x)$



$c(x)$



$f(x)$

$c(x)$

Continuous | Discrete

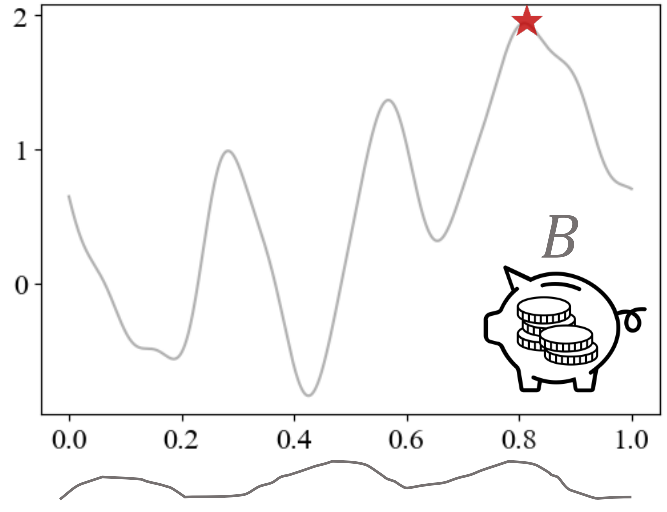
Correlated | Independent

Ebc & Cps | Cost-per-sample

incorporate posterior

Acquisition function + stopping rule $\Longleftarrow$ Optimal policy: Gittins index

Cost-aware Bayesian Optimization



Continuous

Correlated

Ebc & Cps

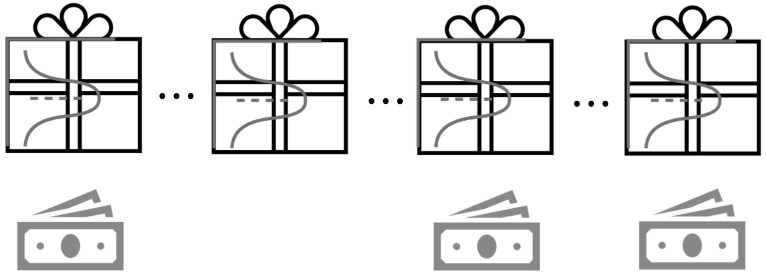
Acquisition function

+ stopping rule

Empirically good?

Pandora's Box

[Weitzman'79]



Discrete

Independent

Cost-per-sample

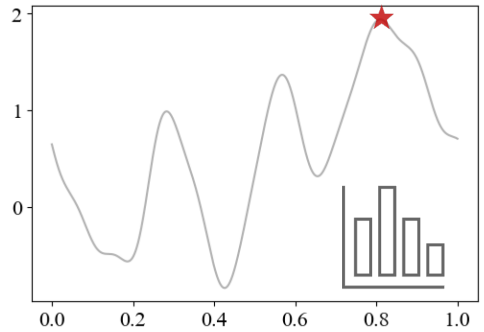
incorporate posterior

$\Longleftarrow$

Gittins index is optimal

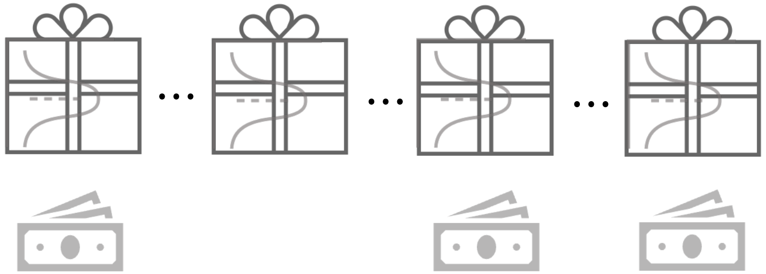
Outline

Studied Problem



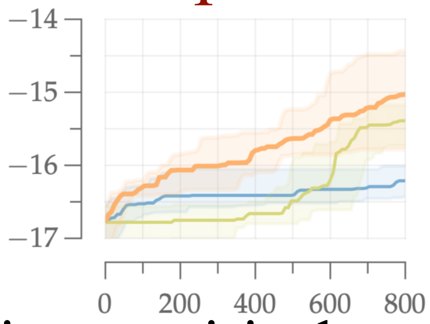
Cost-aware Bayesian optimization

Key idea



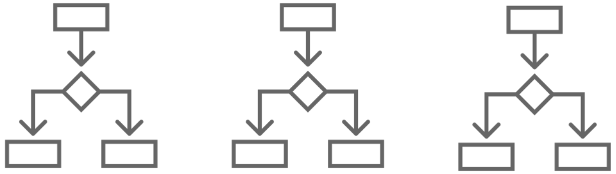
Link to Pandora's box and Gittins index theory

Impact



Competitive empirical performance

Future direction



"Exotic" Bayesian optimization

Experiment Setup: Objective Functions

Synthetic

Samples from prior

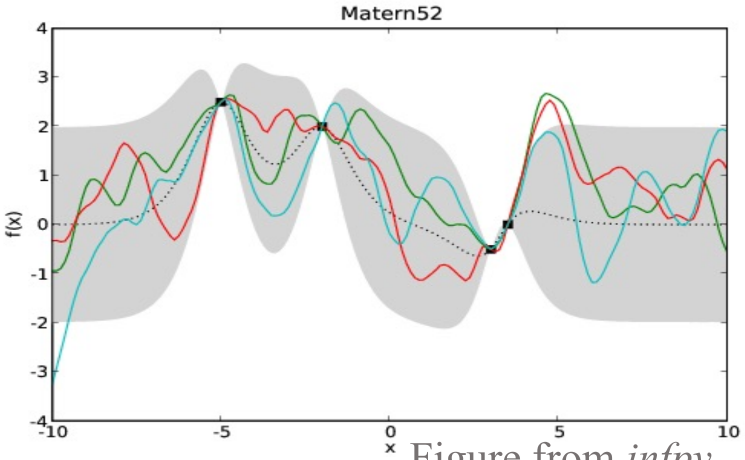


Figure from *infpy*

Ackley function

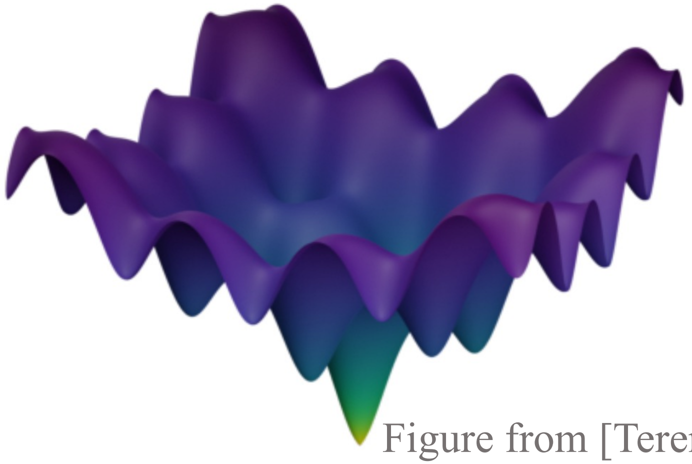


Figure from [Terenin'22]

Empirical

Pest Control

Figure from ChatGPT

Lunar Lander

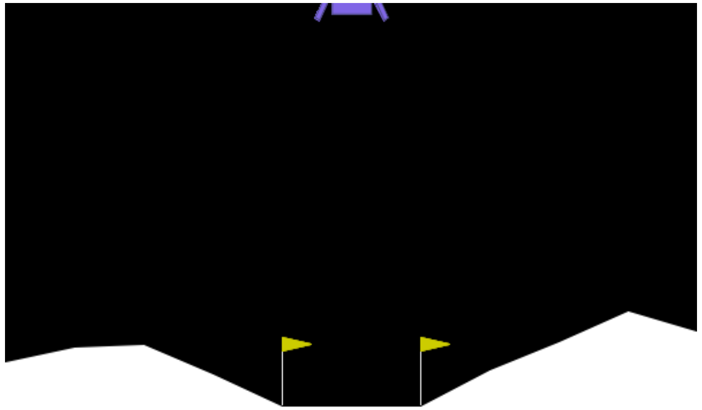
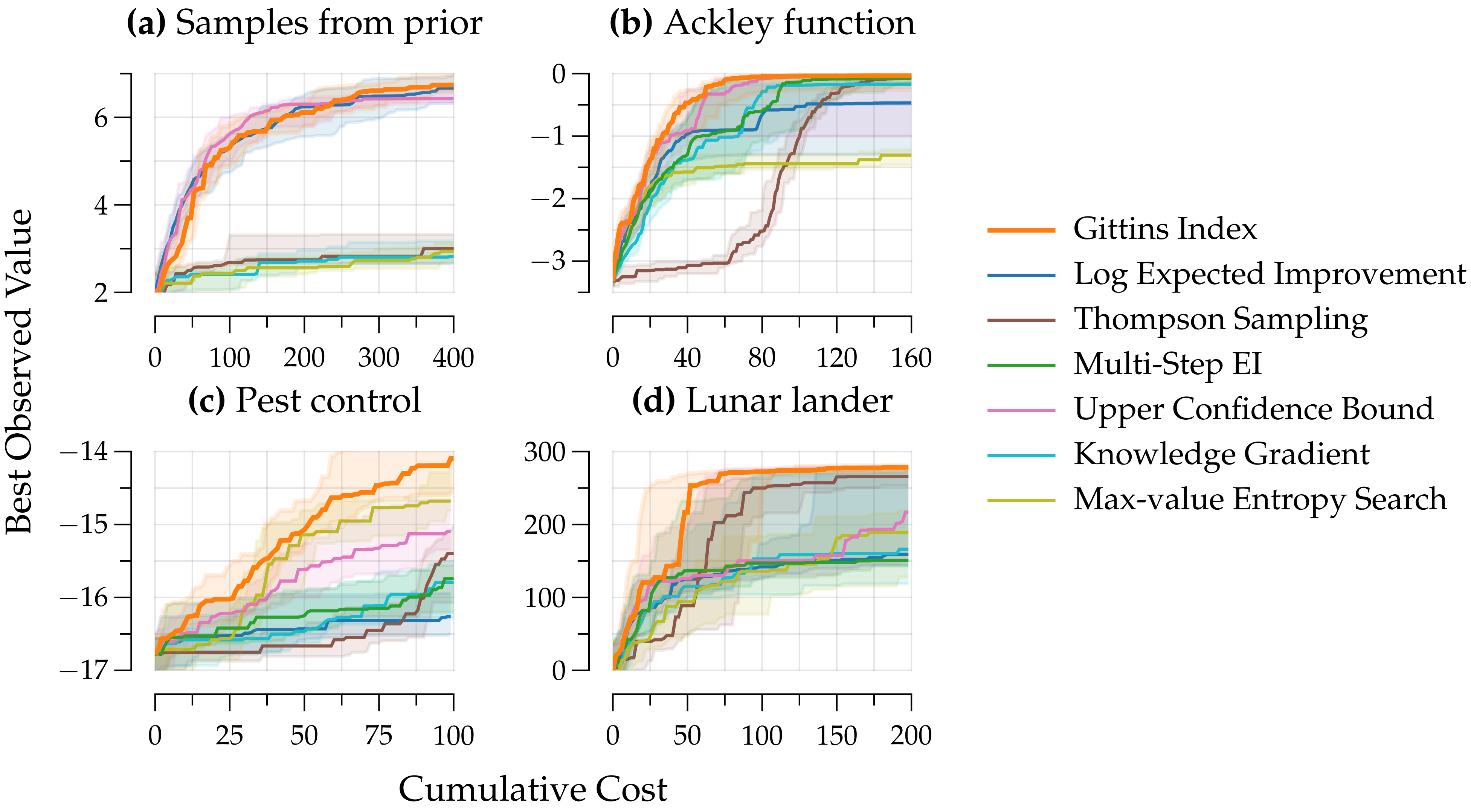
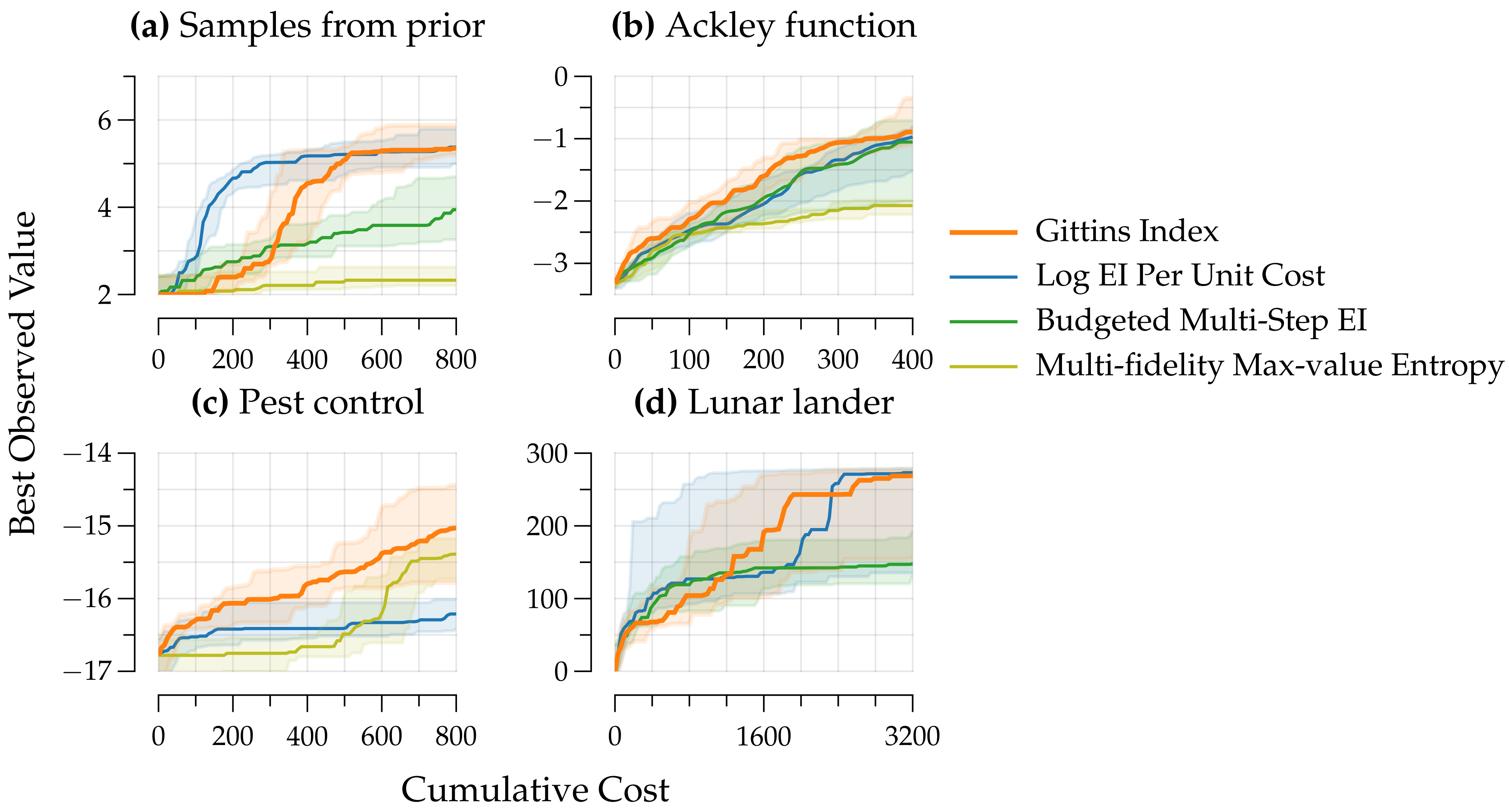


Figure from OpenAI Gym

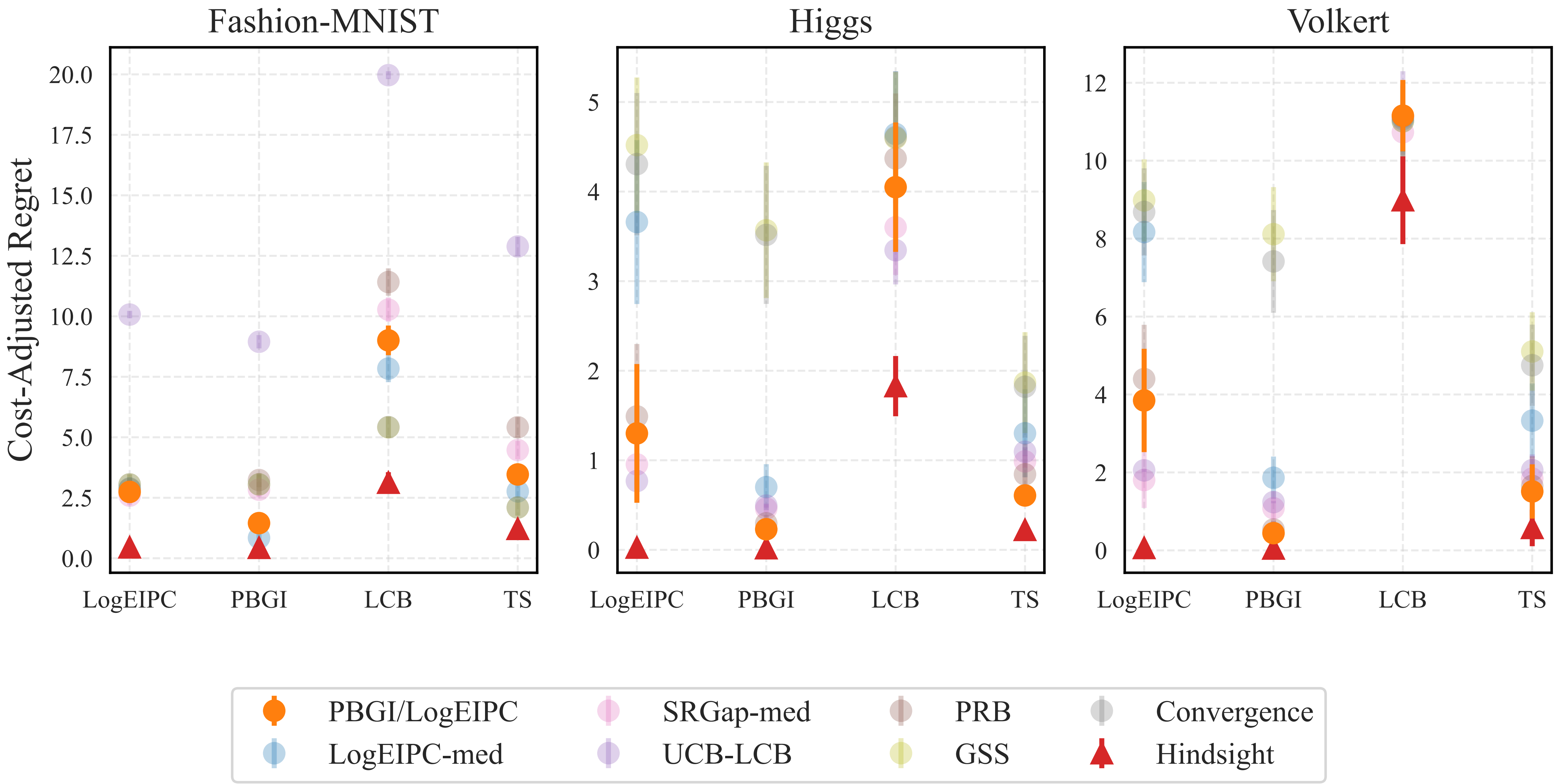
Uniform-cost: Gittins Index vs Baselines



Varying-cost: Gittins Index vs Baselines

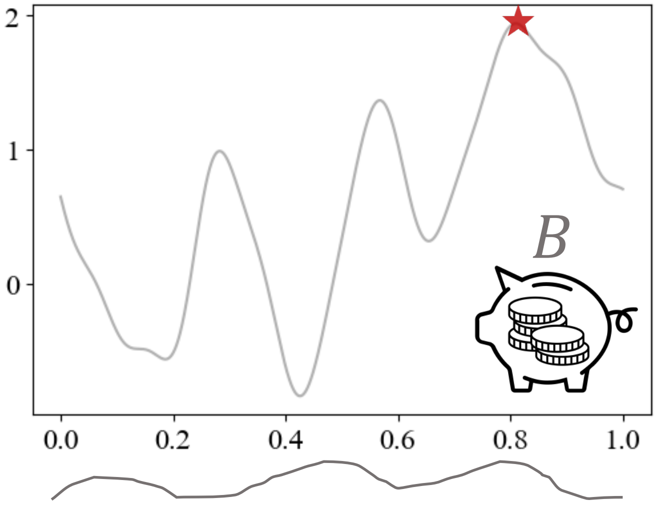


Stopping Rule: Gittins Index vs Baselines



Cost-aware Bayesian Optimization

$f(x)$



$c(x)$

Continuous

Correlated

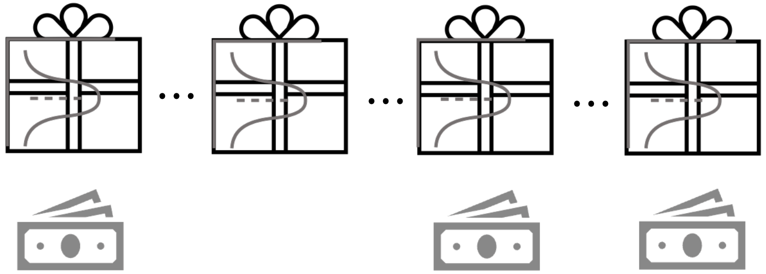
Ebc & Cps

Acquisition function
+ stopping rule

Theoretical guarantee?

Pandora's Box

[Weitzman'79]



$f(x)$

$c(x)$

Discrete

Independent

Cost-per-sample

incorporate posterior

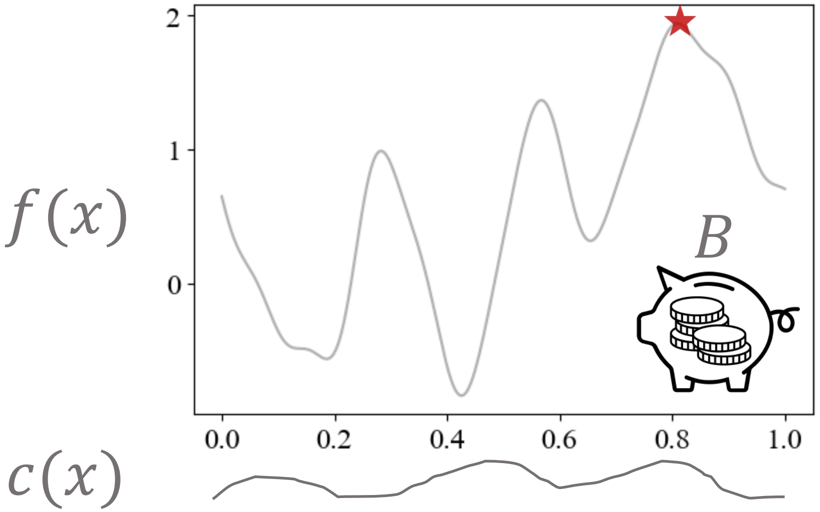
$\Longleftarrow$

Gittins index is optimal

Cost-aware Bayesian Optimization

Pandora's Box

[Weitzman'79]



Cost-aware Bayesian Optimization	Pandora's Box
Continuous	Discrete
Correlated	Independent
Ebc & Cps	Cost-per-sample

incorporate posterior

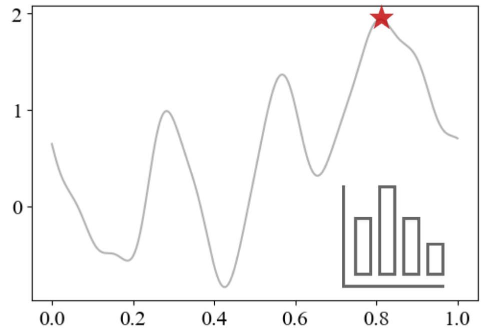
Acquisition function + stopping rule $\Longleftarrow$ Gittins index is optimal

Theoretical guarantee?

Yes! A bound on expected cost at stopping

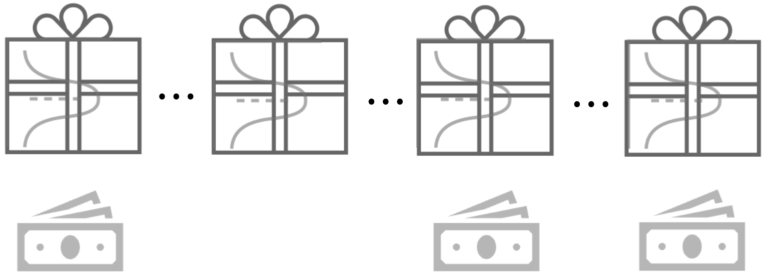
Gittins Index: A New Design Principle

Studied Problem



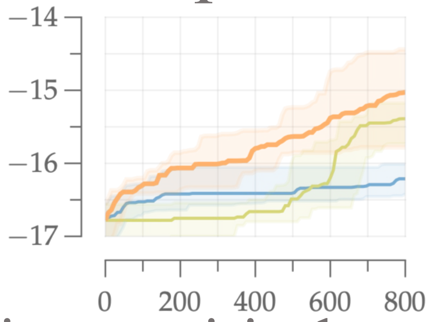
Cost-aware Bayesian optimization

Key idea



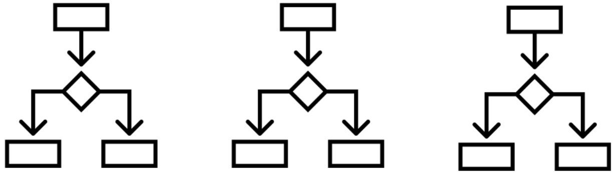
Link to Pandora’s box and Gittins index theory

Impact



Competitive empirical performance w/ theoretical guarantee

Ongoing work



Multi-stage Bayesian optimization

Find our papers on arXiv!

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

"Cost-aware Stopping for Bayesian Optimization."