

NeurIPS'24 publication

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

Qian Xie (Cornell ORIE)

Virtual Seminar Series on Bayesian Decision-making and
Uncertainty

Coauthors



Raul Astudillo



Peter Frazier



Ziv Scully

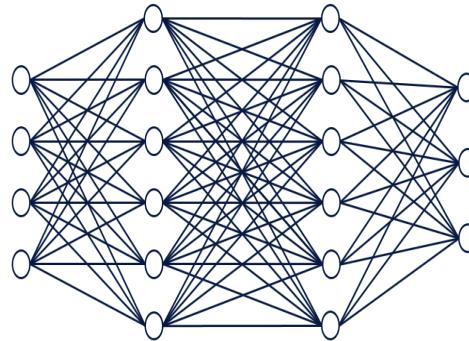


Alexander Terenin

World of Parameter Optimization

Hyperparameter tuning:

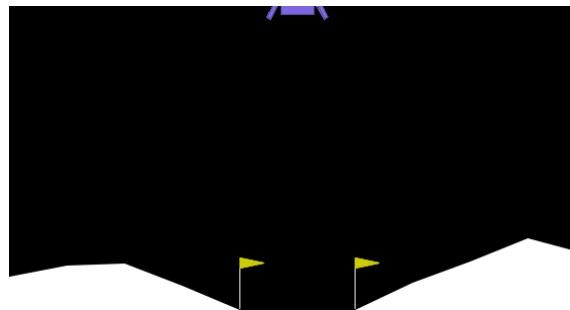
Training parameters →



Accuracy

Control optimization:

Control variables →



Reward

Adaptive experimentation:

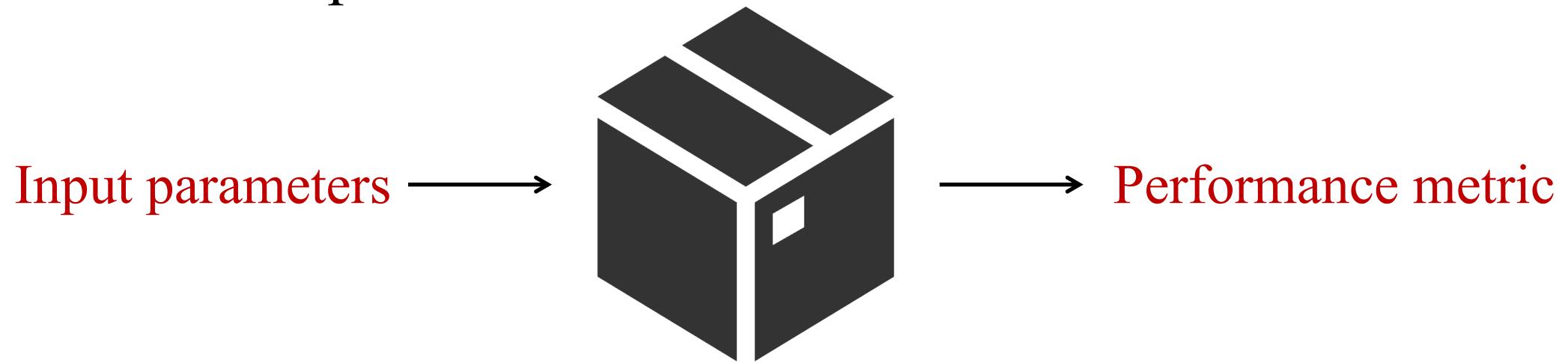
Decision variables →



Revenue

World of Parameter Optimization

Black-box optimization:

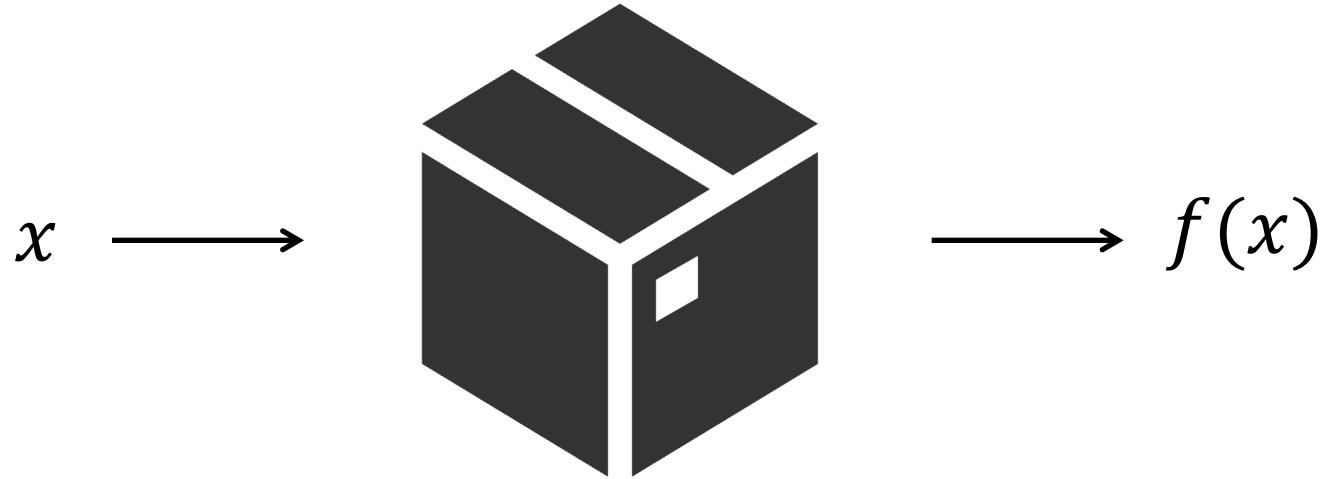


World of Parameter Optimization

Black-box function:



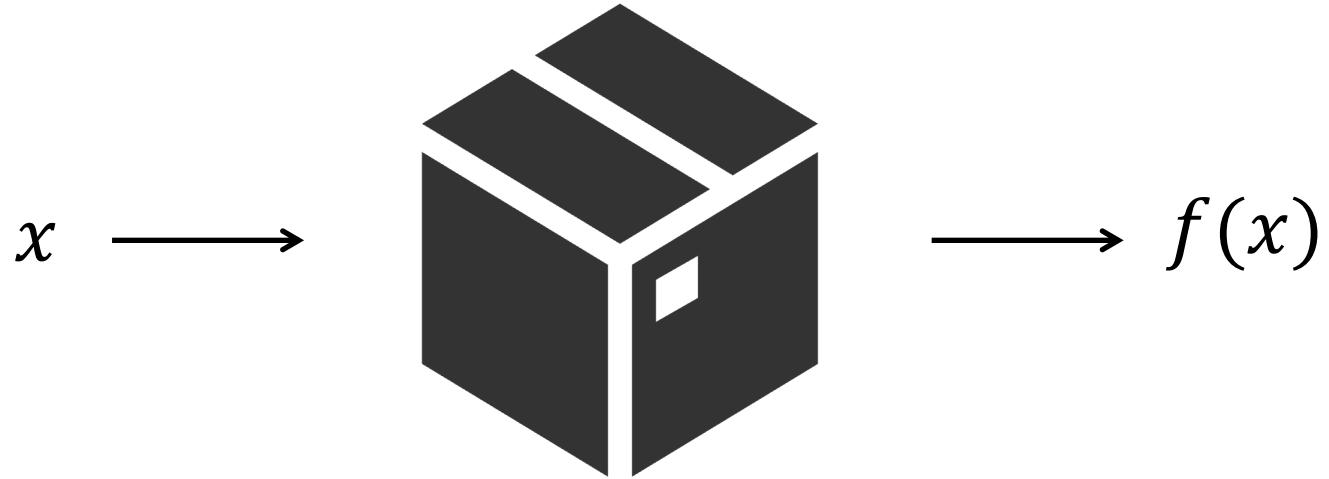
Optimizing Black-box Functions



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

$f \sim \text{Stochastic Process}$

Optimizing Black-box Functions

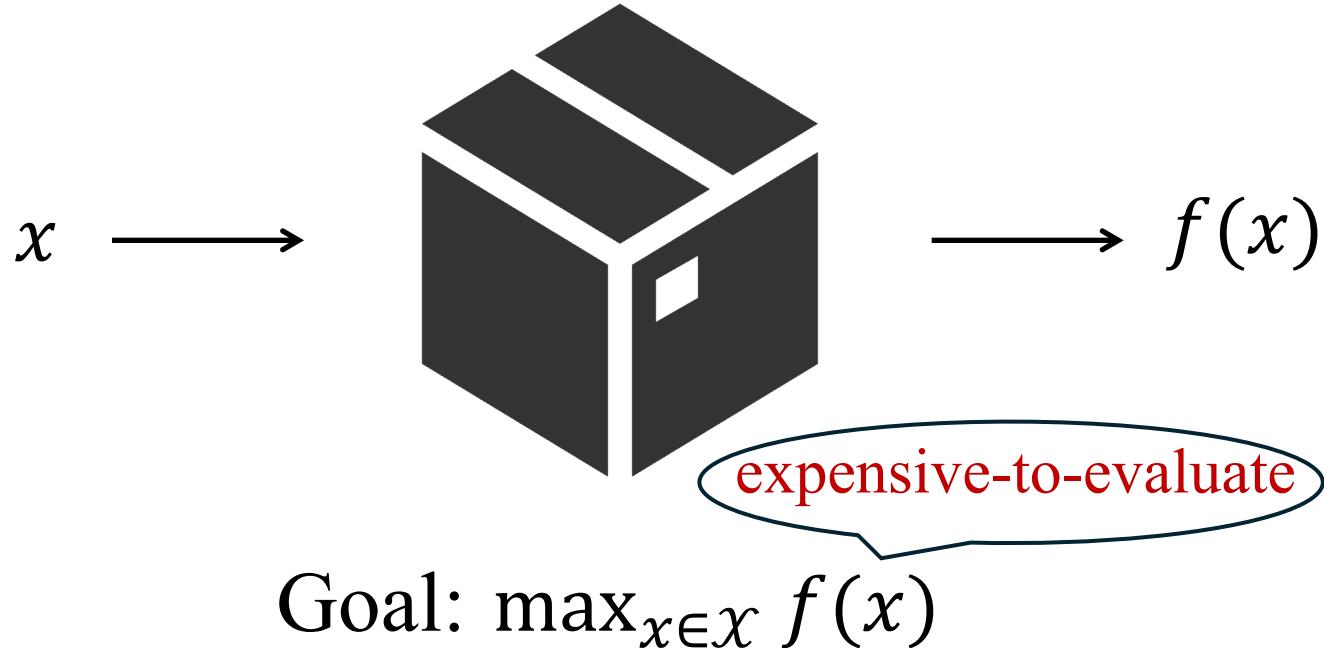


Goal: $\max_{x \in \mathcal{X}} f(x)$

$f \sim$ Stochastic Process

Grid Search? Random search?

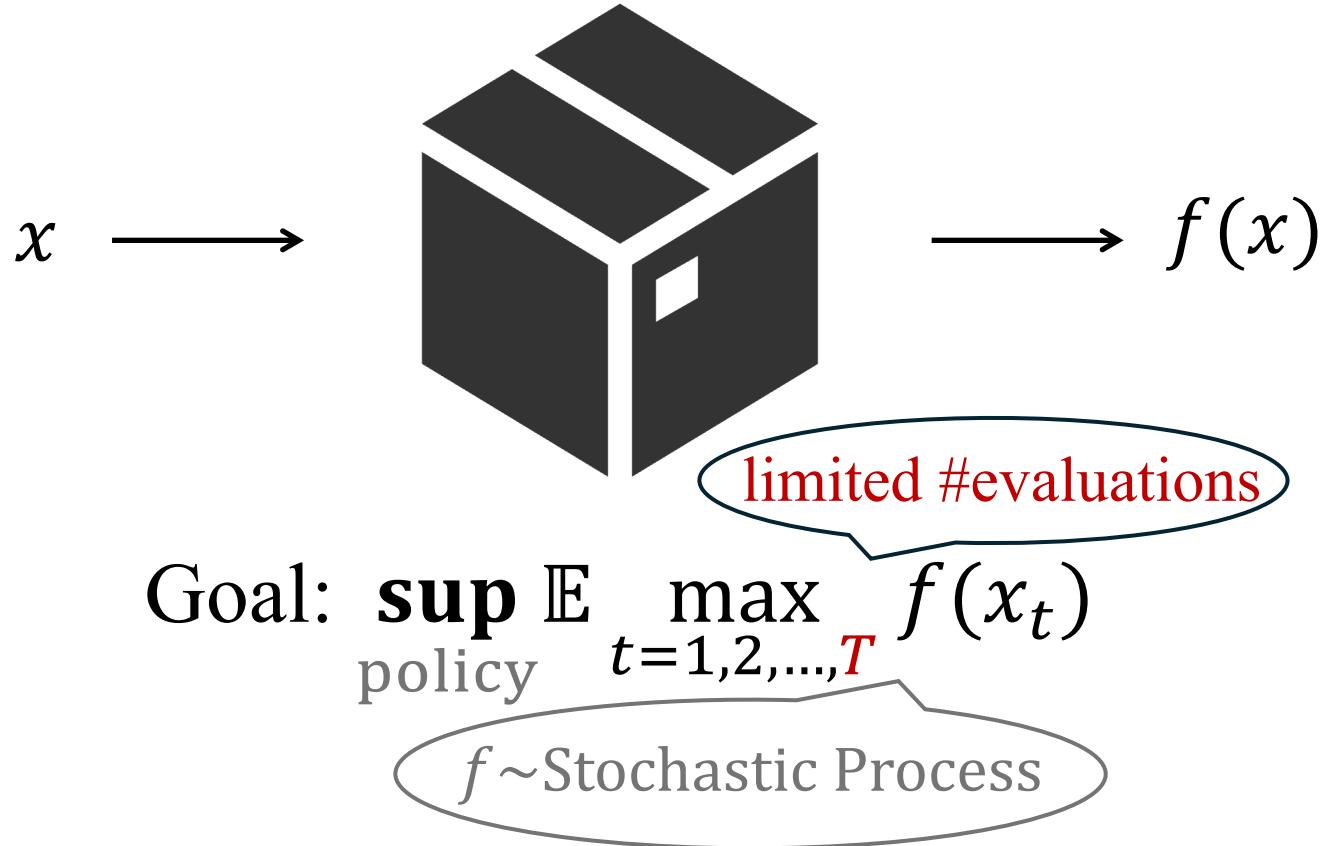
Optimizing Black-box Functions



$f \sim$ Stochastic Process

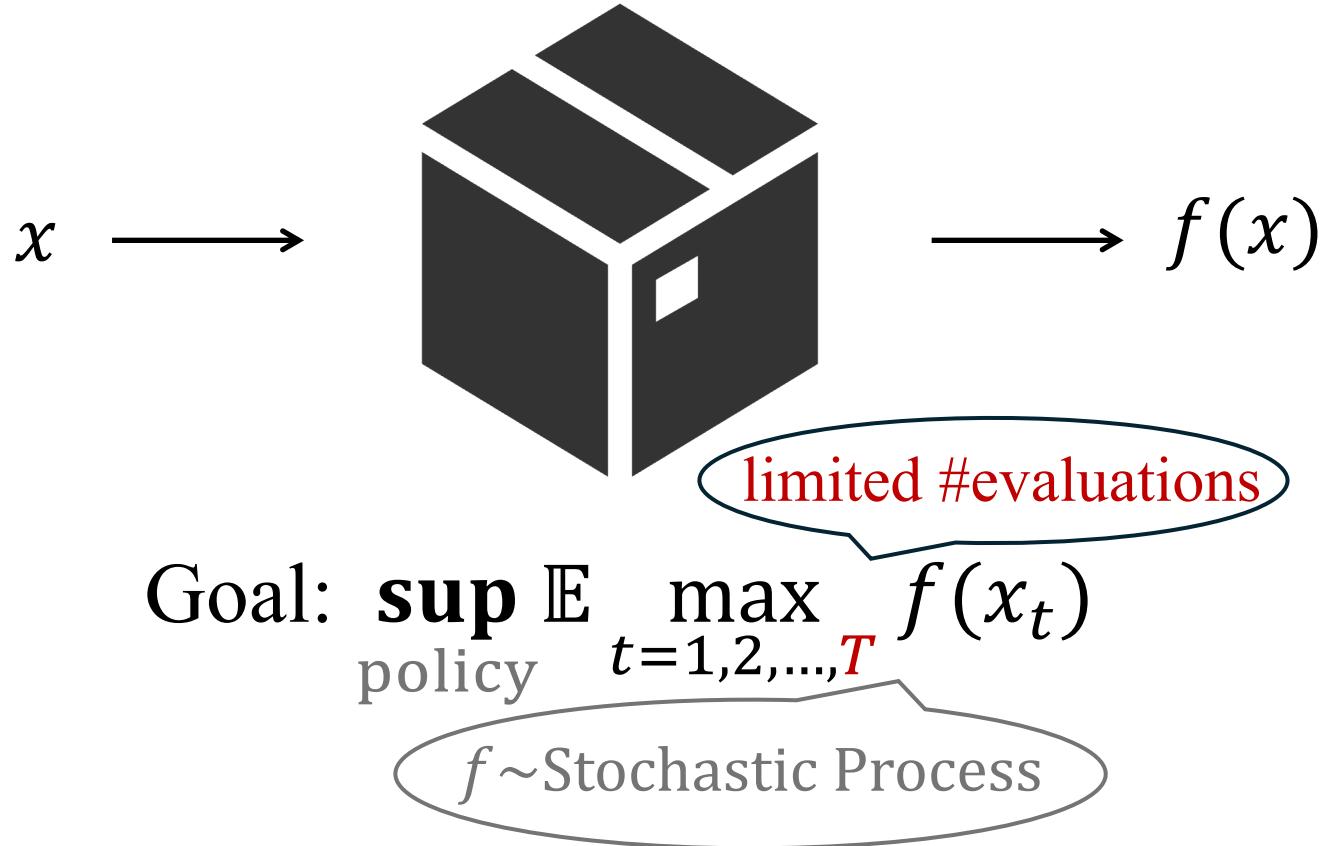
Grid Search? Random search?

Optimizing Black-box Functions



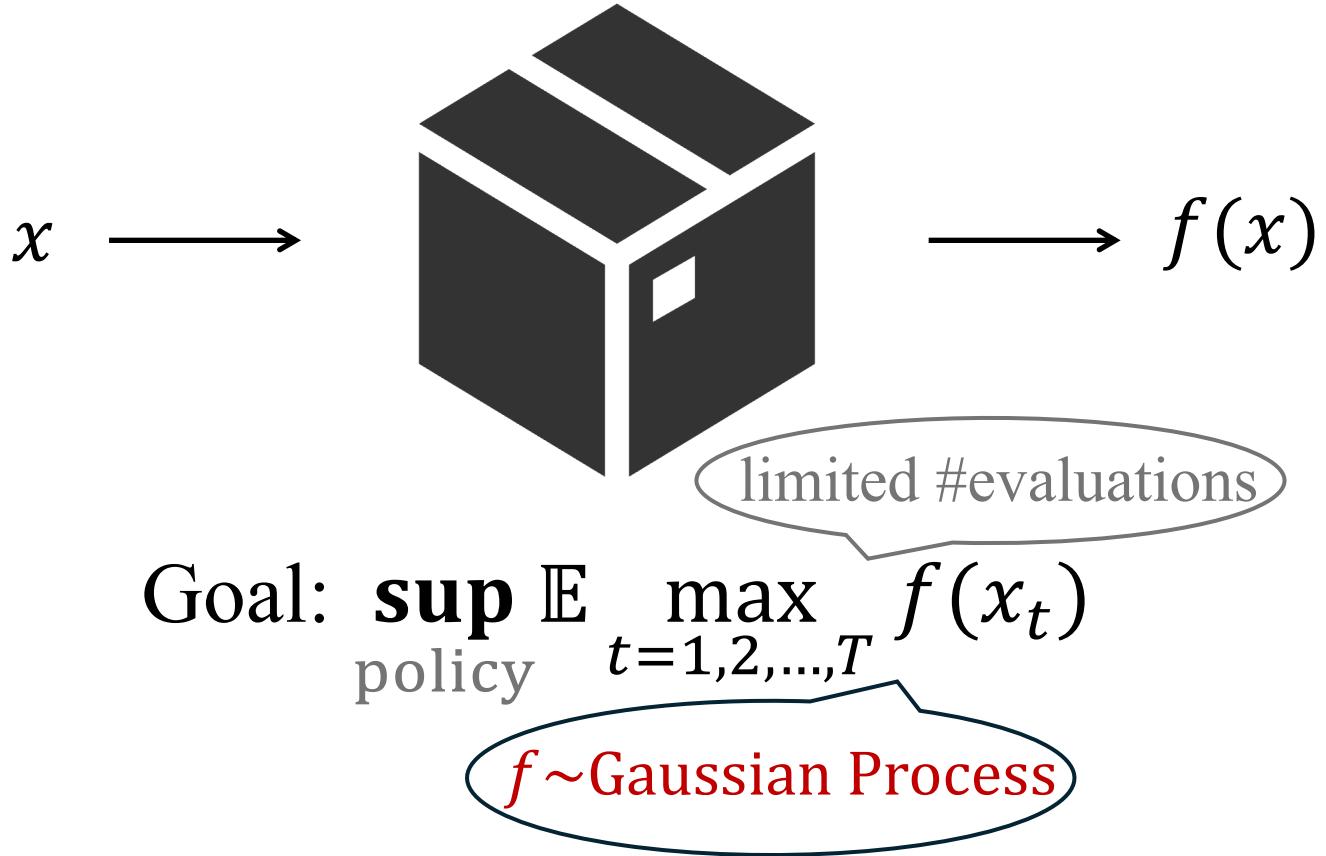
Grid Search? Random search?

Optimizing Black-box Functions



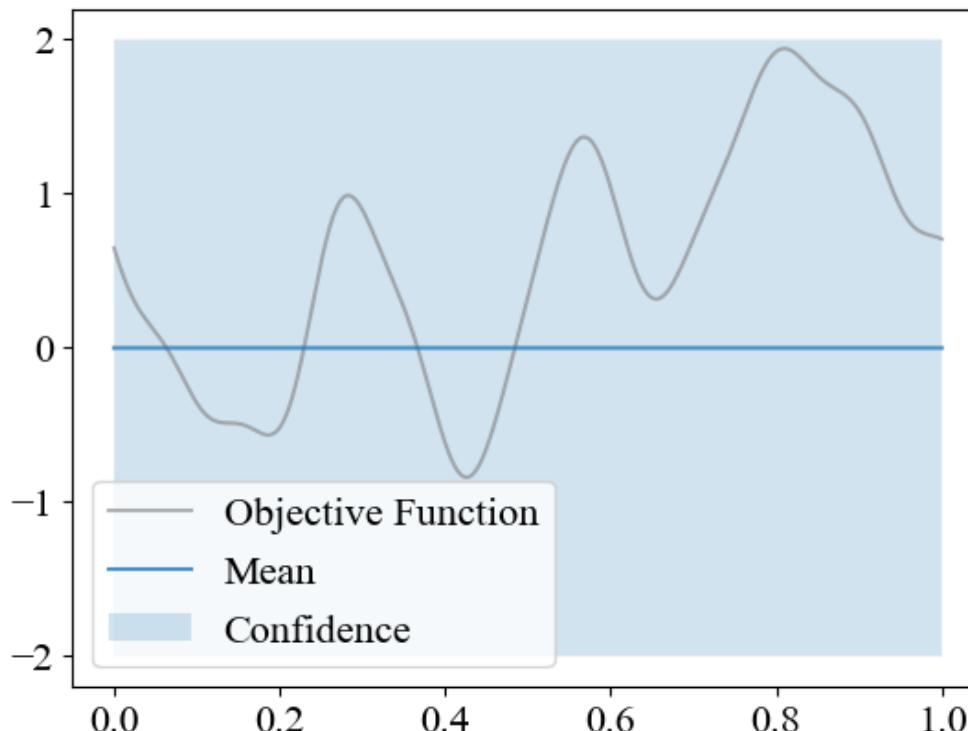
More efficient: Bayesian optimization!

Bayesian Optimization



Key idea: maintain posterior belief about f

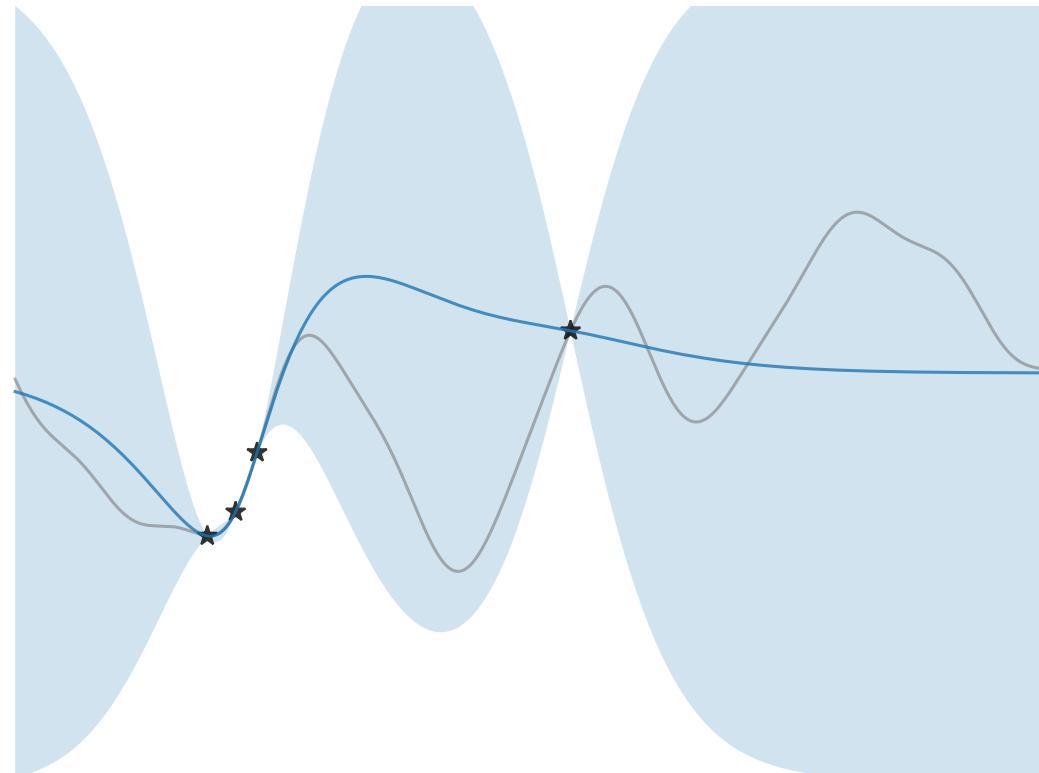
Bayesian Optimization



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

$f \sim \text{Gaussian Process}$

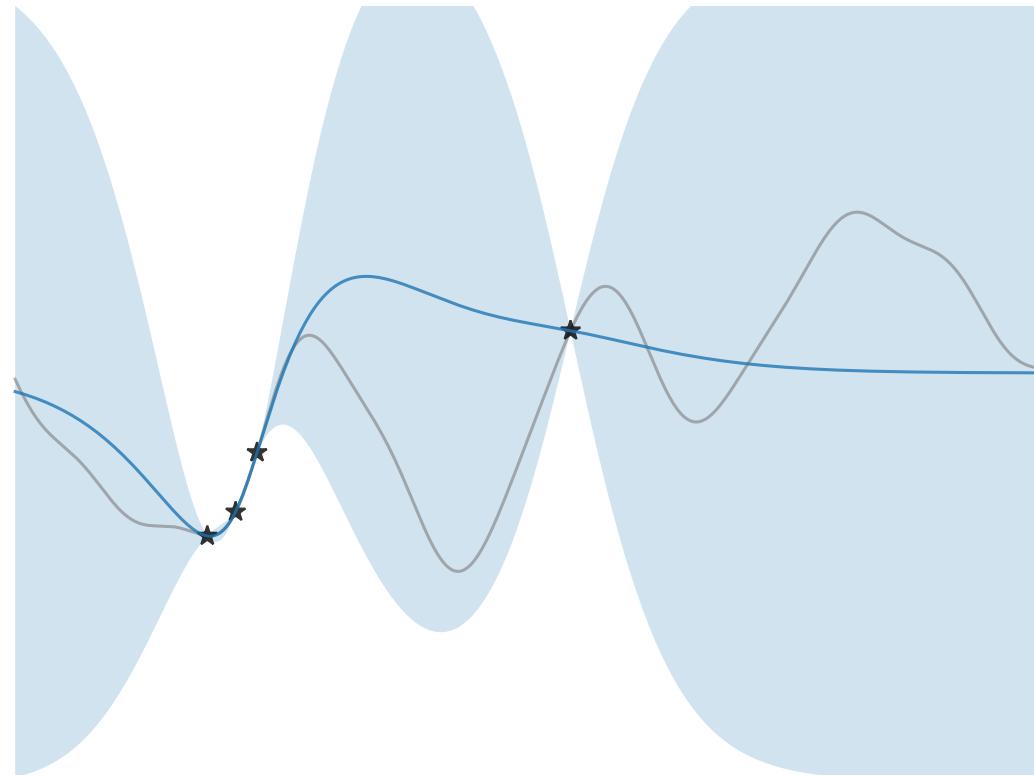
Bayesian Optimization



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

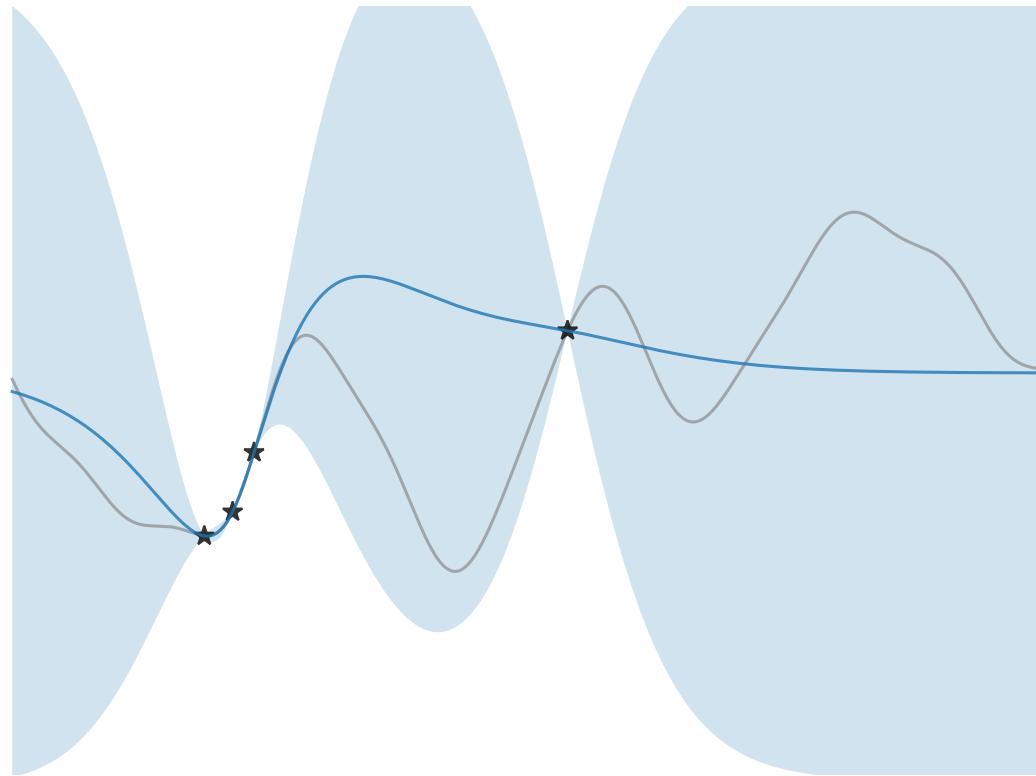
$f \sim \text{Gaussian Process}$

Bayesian Optimization



What to evaluate next?

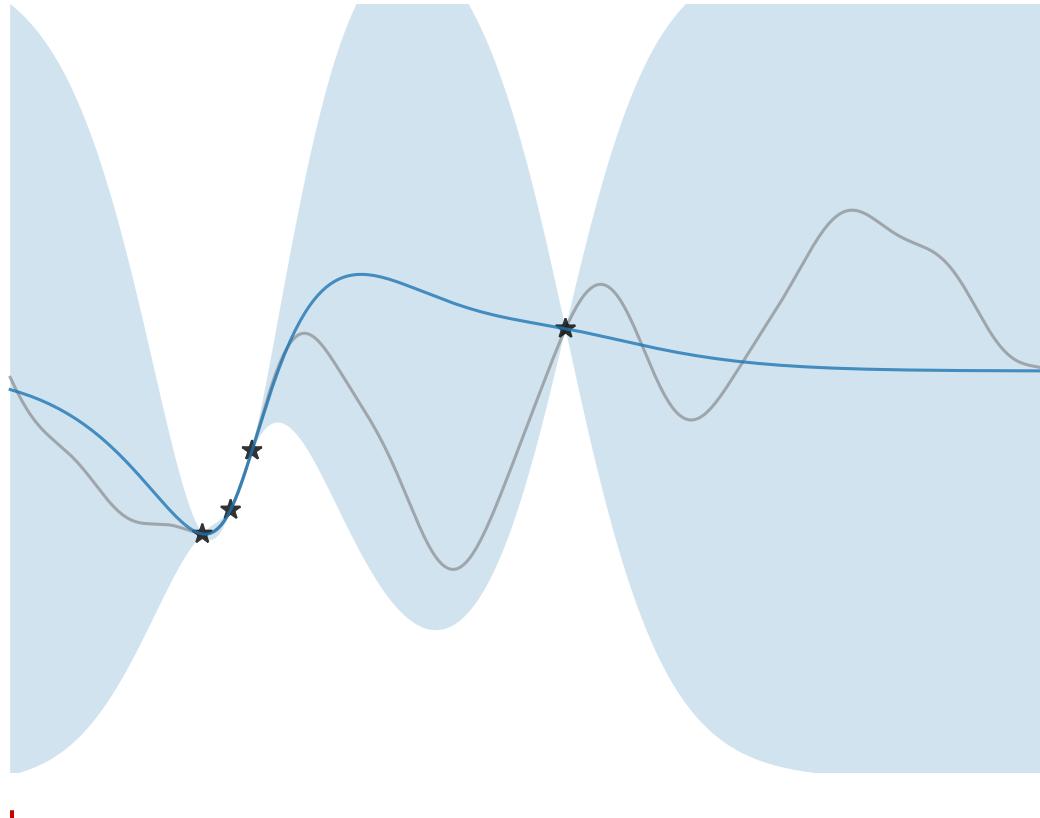
Bayesian Optimization



What to evaluate next?

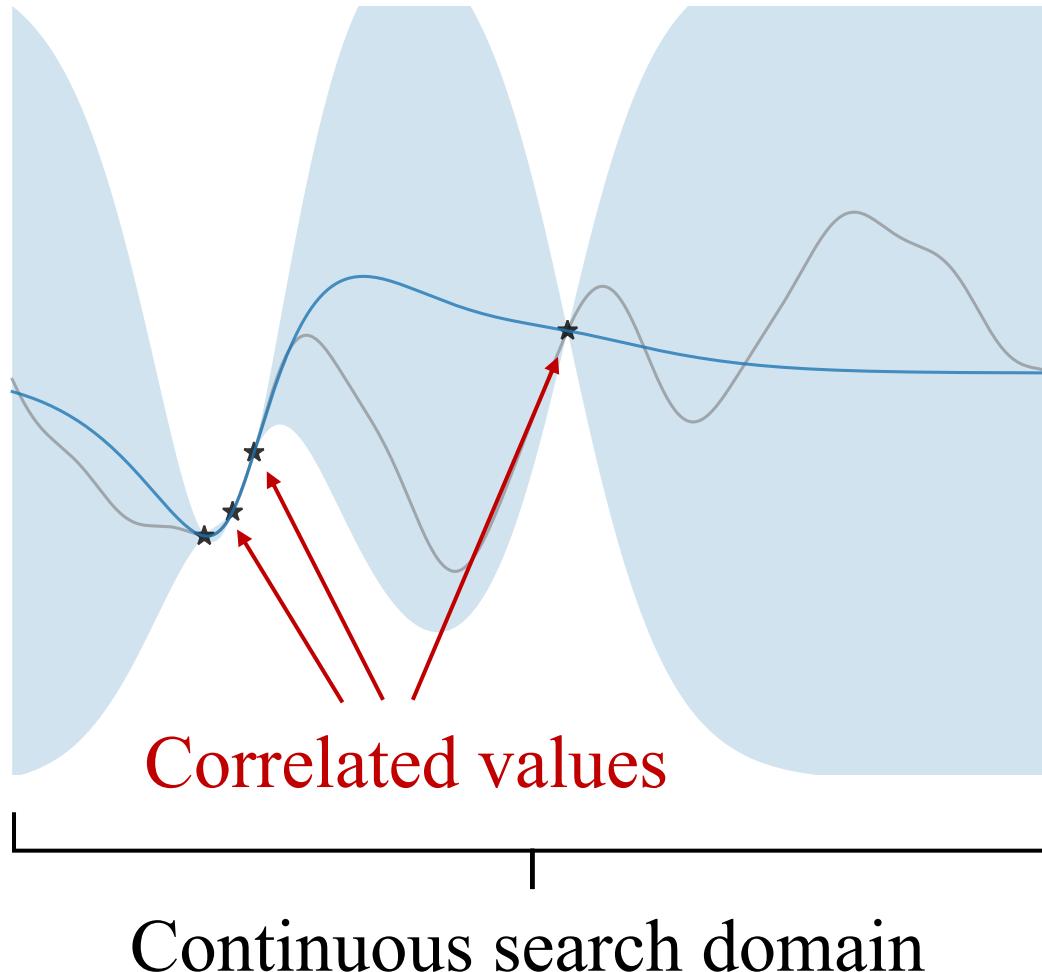
Optimal policy?

Properties of Bayesian Optimization

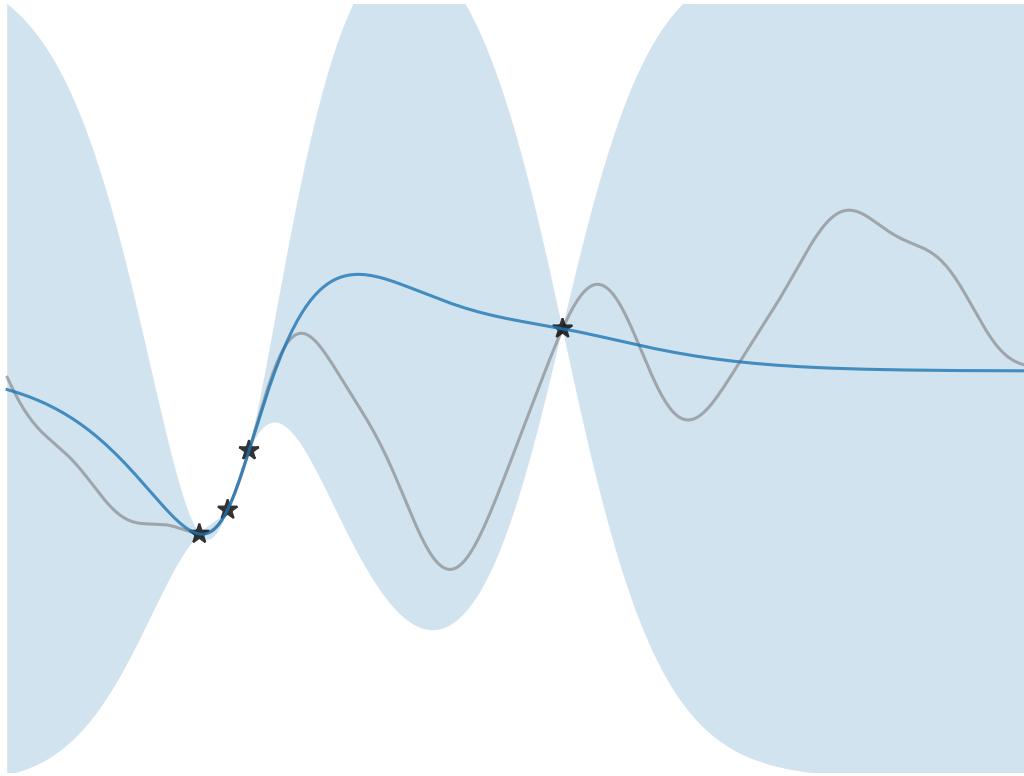


Continuous search domain

Properties of Bayesian Optimization

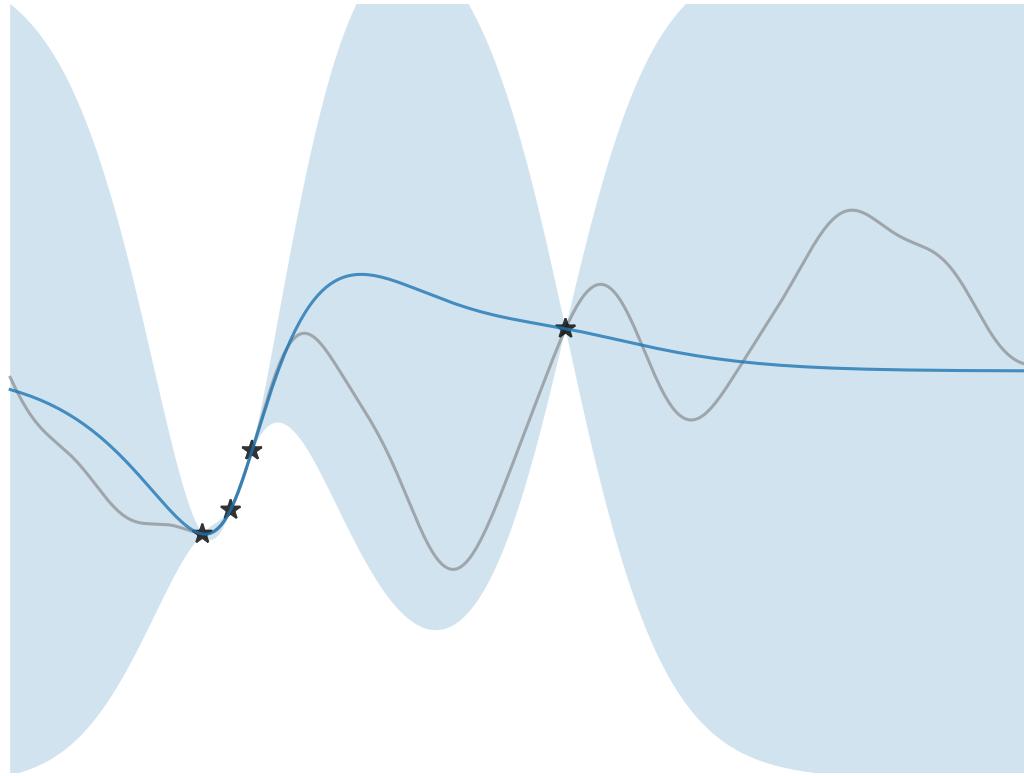


Properties of Bayesian Optimization



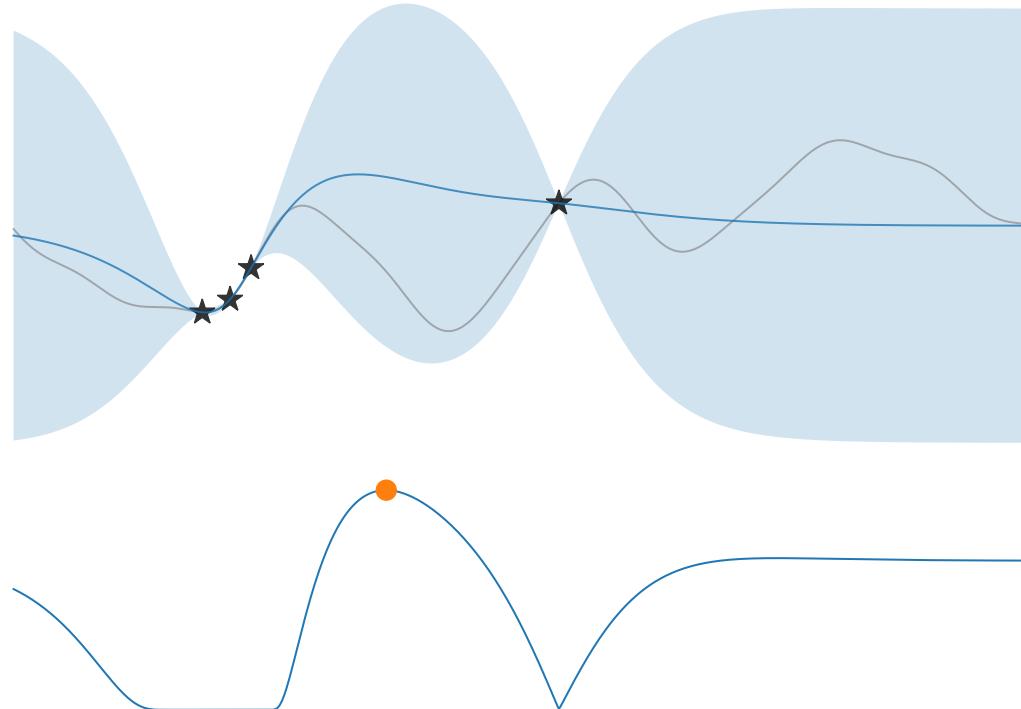
Correlation & continuity \Rightarrow Intractable MDP

Properties of Bayesian Optimization



Intractable MDP \Rightarrow Optimal policy unknown

Popular Policy: Expected Improvement



$$EI(x) = \mathbb{E}[\max(f(x) - y_{\text{best}}, 0) | D]$$

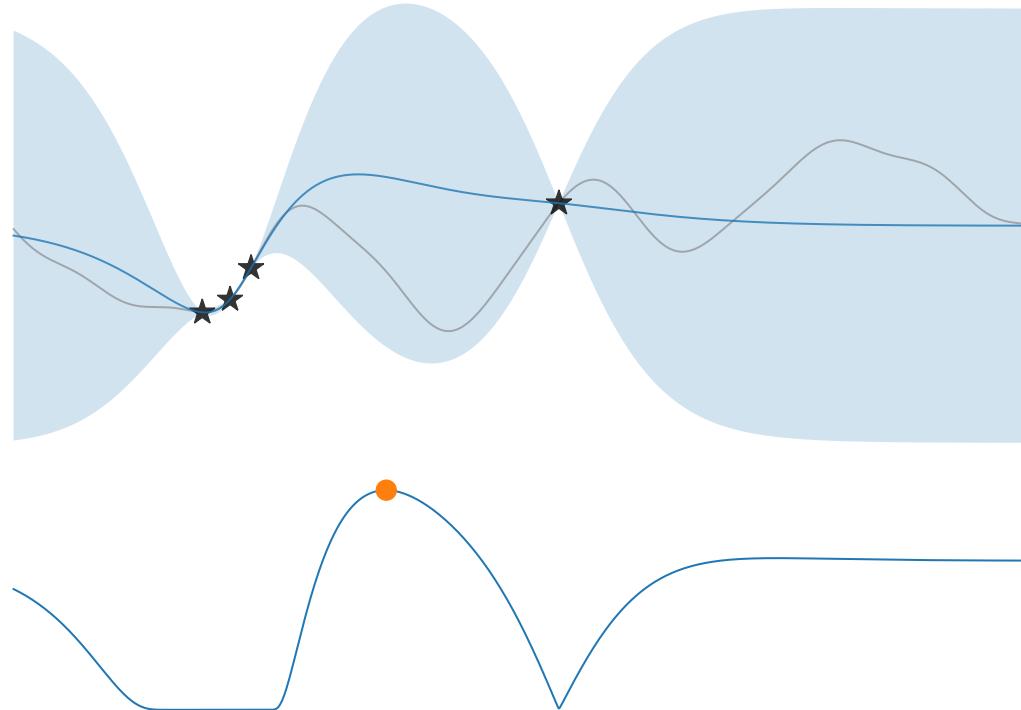
"improvement"

$$\max_x EI_{f|D}(x; y_{\text{best}})$$

posterior distribution

One-step approximation to MDP

Popular Policy: Expected Improvement



Other improvement-based policies:

- Probability of Improvement
- Knowledge Gradient
- Multi-step Lookahead EI
- :

multi-step approximation to MDP

Approaches to Bayesian Optimization

- Improvement-based:
 - Expected Improvement
 - Probability of Improvement
 - Knowledge Gradient
 - Multi-step Lookahead EI

Approaches to Bayesian Optimization

- Improvement-based:
 - Expected Improvement
 - Probability of Improvement
 - Knowledge Gradient
 - Multi-step Lookahead EI
- Entropy-based:
 - Max-value Entropy Search
 - Predictive Entropy Search

Approaches to Bayesian Optimization

- Improvement-based:
 - Expected Improvement
 - Probability of Improvement
 - Knowledge Gradient
 - Multi-step Lookahead EI
- Entropy-based:
 - Max-value Entropy Search
 - Predictive Entropy Search
- Upper Confidence Bound

Approaches to Bayesian Optimization

- Improvement-based:
 - Expected Improvement
 - Probability of Improvement
 - Knowledge Gradient
 - Multi-step Lookahead EI
- Entropy-based:
 - Max-value Entropy Search
 - Predictive Entropy Search
- Upper Confidence Bound
- Thompson Sampling

Approaches to Bayesian Optimization

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

Approaches to Bayesian Optimization

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

Approaches to Bayesian Optimization

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

Why another approach?

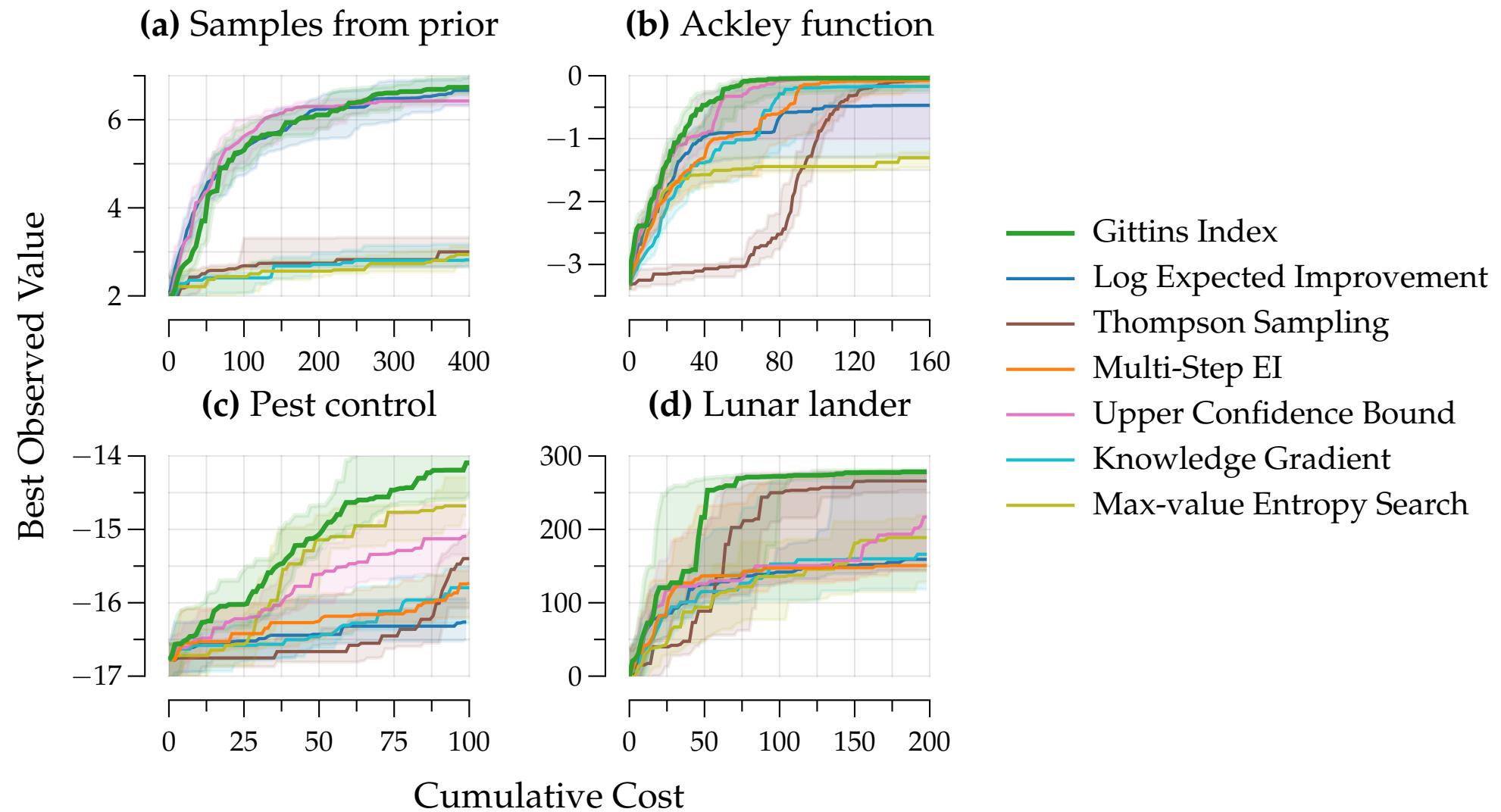
Approaches to Bayesian Optimization

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

Why another approach?

1. Competitive empirical performance

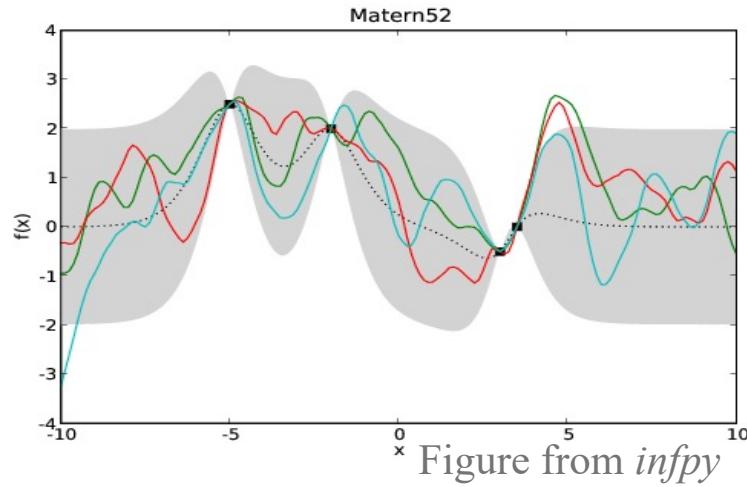
Experiment Result: Gittins Index vs Baselines



Experiment Setup: Objective Functions

Synthetic

Samples from prior



Ackley function

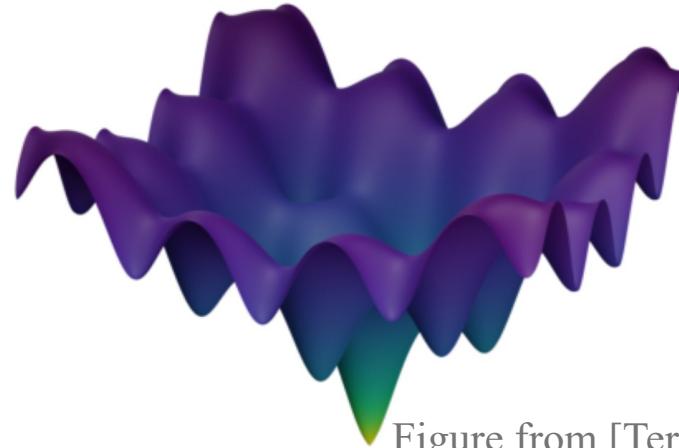


Figure from [Terenin'22]

Empirical

Pest Control



Figure from ChatGPT

Lunar Lander

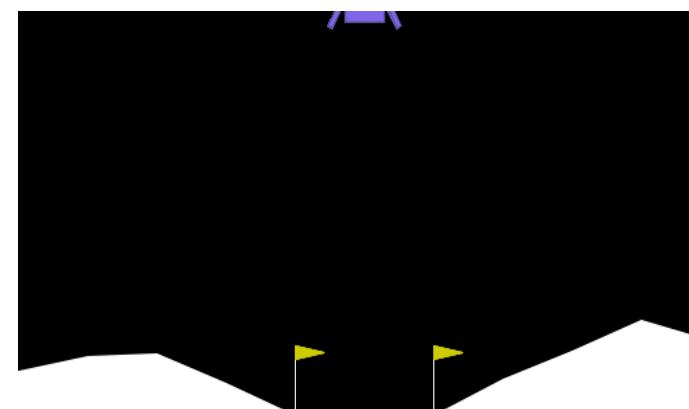


Figure from OpenAI Gym

Approaches to Bayesian Optimization

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

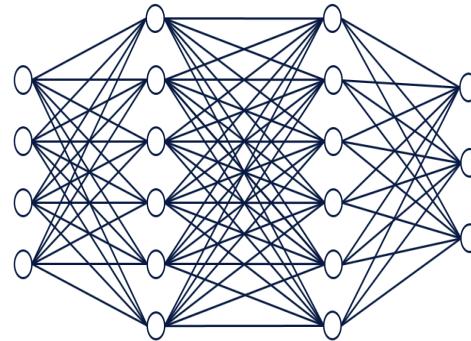
Why another approach?

1. Competitive empirical performance
2. Applicable to varying evaluation costs

Varying Evaluation Costs

Hyperparameter tuning:

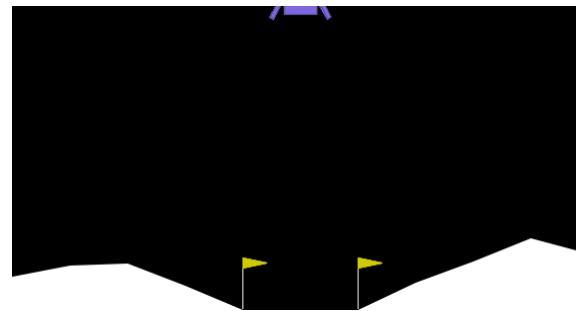
Training parameters →



→ Accuracy

Control optimization:

Control variables →



→ Reward

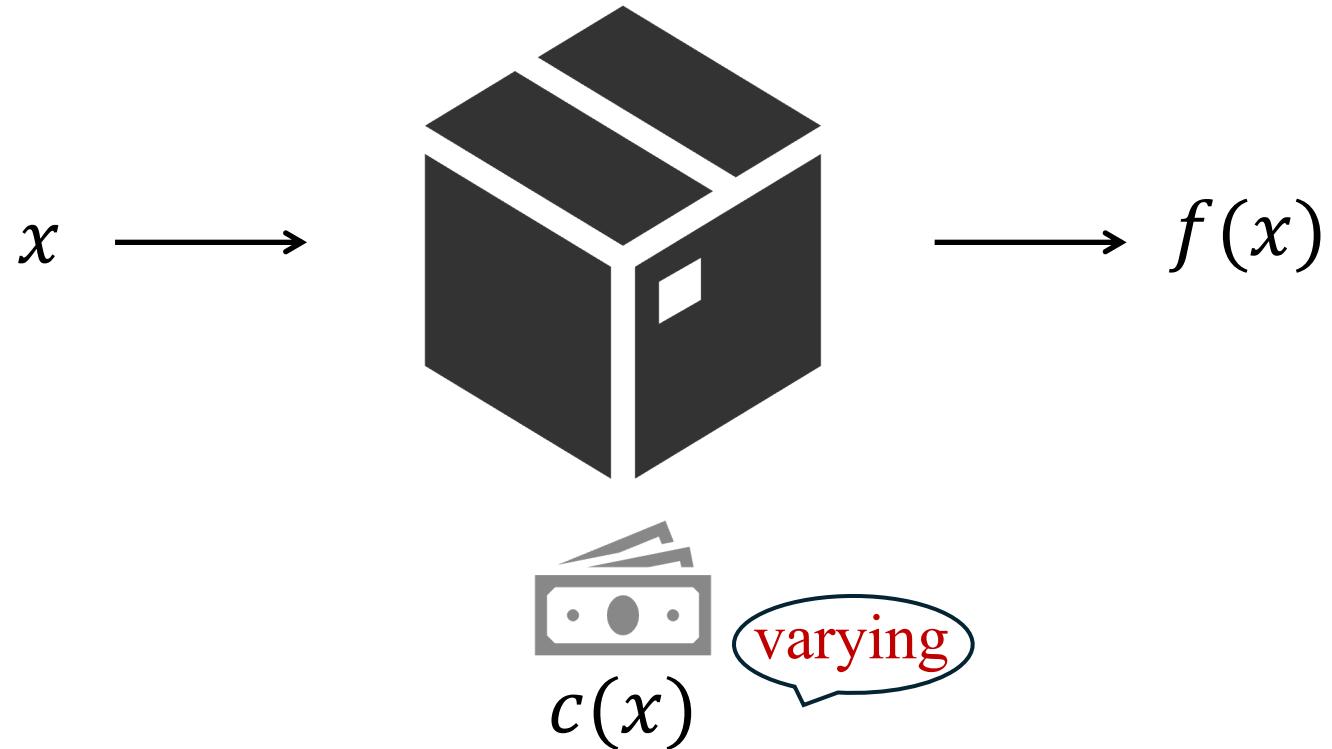
Adaptive experimentation:

Decision variables →



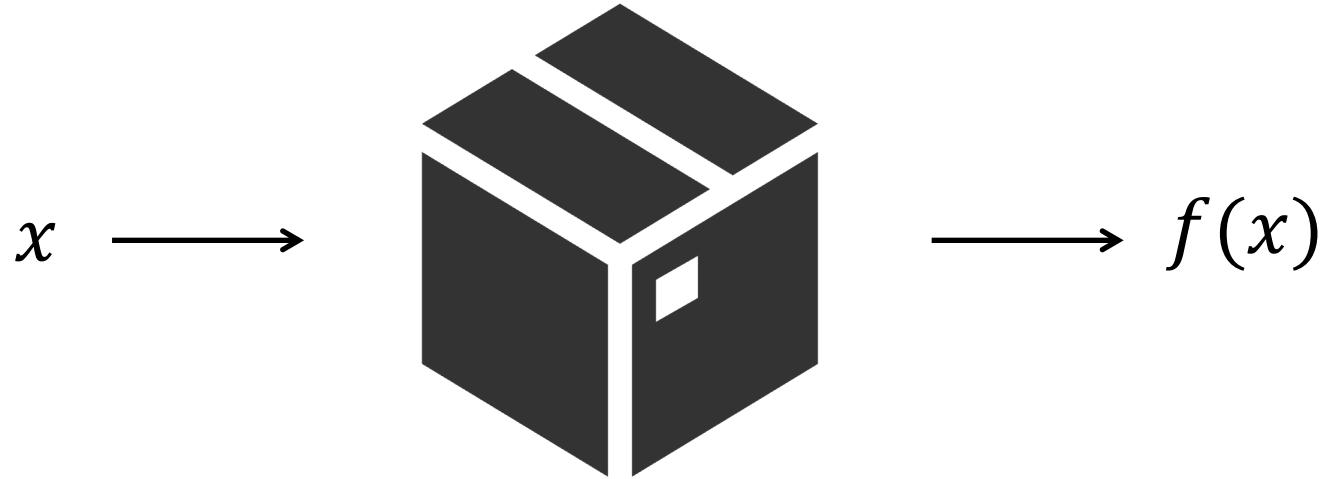
→ Revenue

Varying Evaluation Costs



Cost-aware Bayesian Optimization

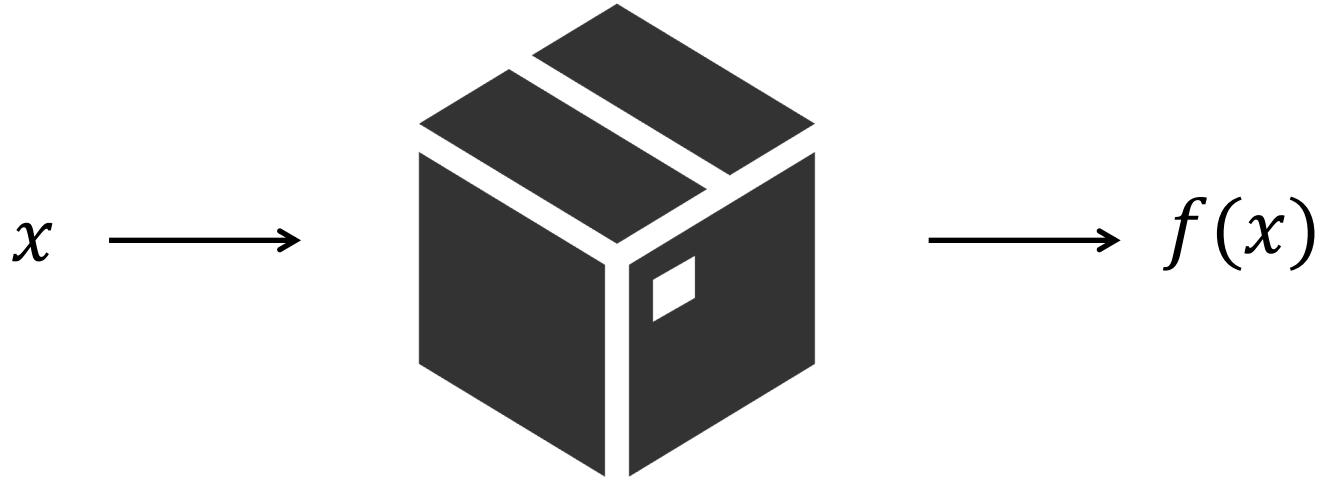
[Lee, Perrone, Archambeau, Seeger'21]



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

“Multi-step Budgeted Bayesian Optimization with Unknown Evaluation Costs”
[Astudillo, Jiang, Balandat, Bakshy, Frazier'21]

Cost-aware Bayesian Optimization



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

s.t. $\mathbb{E} \sum_{t=1}^T c(x_t) \leq B$ Expected budget constraint

Cost-aware Bayesian Optimization

	Uniform costs	Varying costs
One-step	Expected improvement $\max_x \text{EI}_{f D}(x; y_{\text{best}})$	

Cost-aware Bayesian Optimization

	Uniform costs	Varying costs
One-step	Expected improvement $\max_x \text{EI}_{f D}(x; y_{\text{best}})$	Expected improvement per cost [Snoek, Larochelle, Adams'21]

Cost-aware Bayesian Optimization

	Uniform costs	Varying costs
One-step	Expected improvement $\max_x \text{EI}_{f D}(x; y_{\text{best}})$	Expected improvement per cost $\max_x \text{EI}_{f D}(x; y_{\text{best}}) / c(x)$

Cost-aware Bayesian Optimization

	Uniform costs	Varying costs
One-step	Expected improvement $\max_x \text{EI}_{f D}(x; y_{\text{best}})$	Expected improvement per cost $\max_x \text{EI}_{f D}(x; y_{\text{best}})/c(x)$

Why divide?

Cost-aware Bayesian Optimization

	Uniform costs	Varying costs
One-step	Expected improvement $\max_x \text{EI}_{f D}(x; y_{\text{best}})$	Expected improvement per cost $\max_x \text{EI}_{f D}(x; y_{\text{best}})/c(x)$
		 arbitrarily bad

[Astudillo, Jiang, Balandat, Bakshy, Frazier'21]

Cost-aware Bayesian Optimization

	Uniform costs	Varying costs
One-step	Expected improvement	Expected improvement per cost
Multi-step	Multi-step Lookahead EI	Budgeted Multi-step Lookahead EI

slow

[Astudillo, Jiang, Balandat, Bakshy, Frazier'21]

Cost-aware Bayesian Optimization

Uniform costs

One-step Expected improvement

Multi-step Multi-step Lookahead EI

Upper Confidence Bound

Thompson Sampling

Varying costs

Expected improvement per cost

Budgeted Multi-step Lookahead EI

?

?

Cost-aware Bayesian Optimization

	Uniform costs	Varying costs
One-step	Expected improvement	Expected improvement per cost
Multi-step	Multi-step Lookahead EI Upper Confidence Bound Thompson Sampling	Budgeted Multi-step Lookahead EI ?
	:	:

Our view: lack of a guidance to incorporate costs

Cost-aware Bayesian Optimization

	Uniform costs	Varying costs
One-step	Expected improvement	Expected improvement per cost
Multi-step	Multi-step Lookahead EI Upper Confidence Bound Thompson Sampling	Budgeted Multi-step Lookahead EI ?
	:	:

New design principle: Gittins Index

Cost-aware Bayesian Optimization

Uniform costs

One-step Expected improvement

Multi-step Multi-step Lookahead EI

Upper Confidence Bound

Thompson Sampling

:

Varying costs

Expected improvement per cost

Budgeted Multi-step Lookahead EI

?

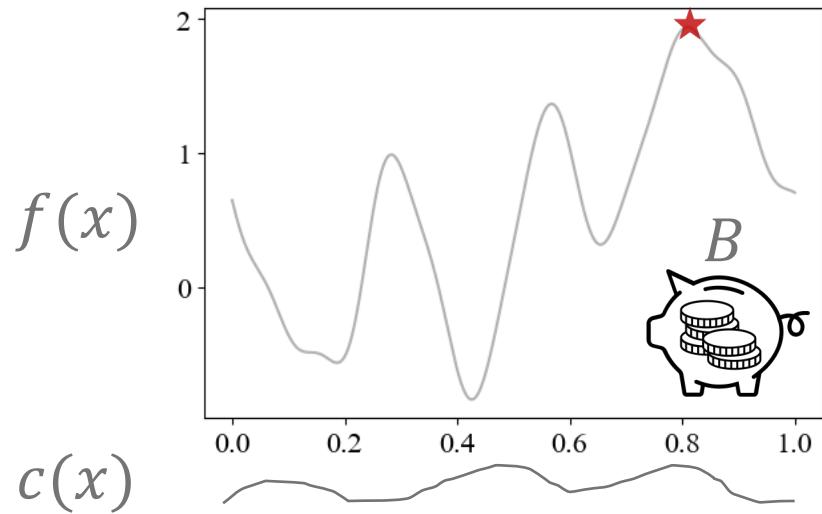
?

:

New design principle: Gittins Index

inherently cost-aware

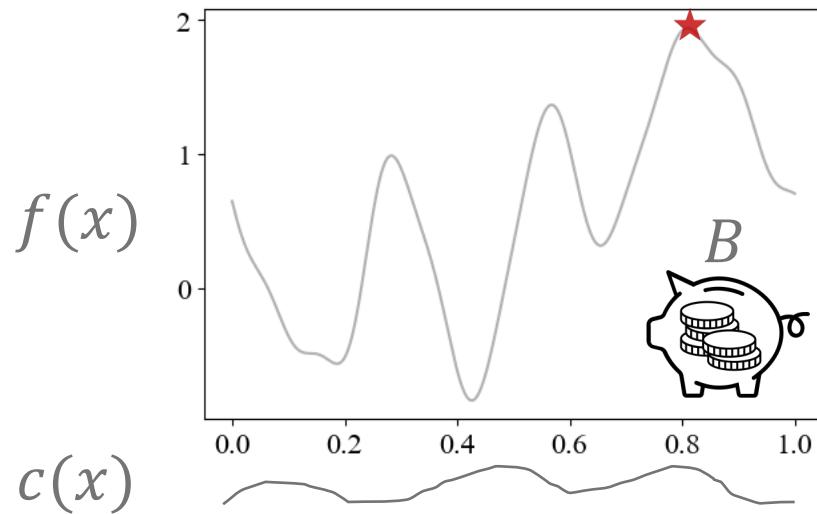
Cost-aware Bayesian Optimization



Continuous

Correlated

Cost-aware Bayesian Optimization



Continuous

Correlated



Pandora's Box

[Weitzman'79]

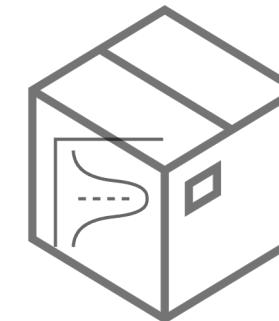
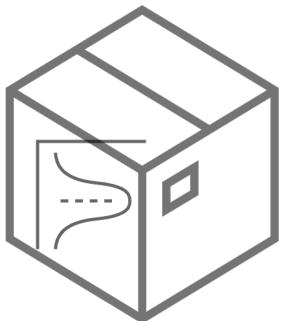


Discrete

Independent

Pandora's Box

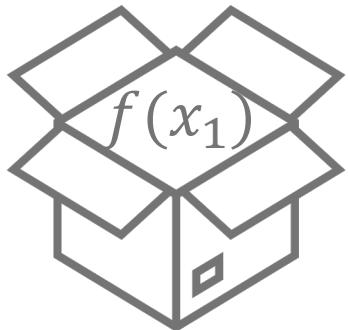
$t = 0$



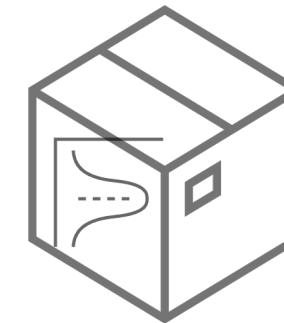
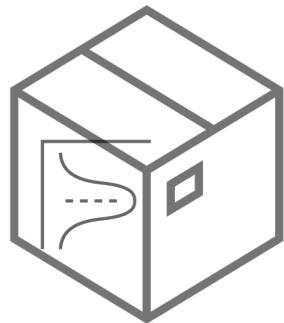
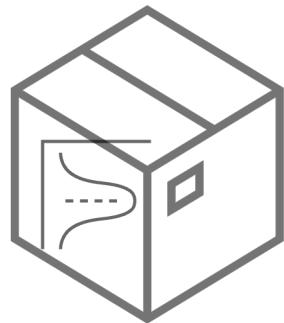
$$\sup_{\text{policy}} \mathbb{E} \left(\max_{t=1,2,\dots,T} f(x_t) - \sum_{t=1}^T c(x_t) \right)$$

Pandora's Box

$t = 1$



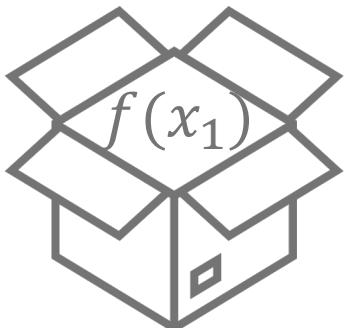
$c(x_1)$



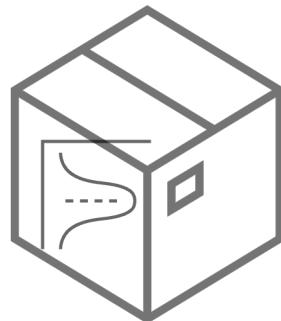
$$\sup_{\text{policy}} \mathbb{E} \left(\max_{t=1,2,\dots,T} f(x_t) - \sum_{t=1}^T c(x_t) \right)$$

Pandora's Box

$t = 2$



$c(x_1)$

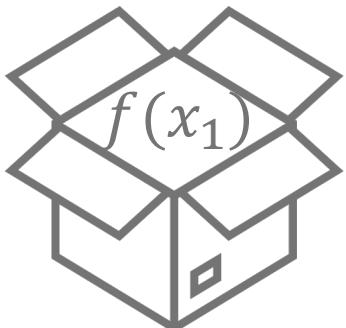


$c(x_2)$

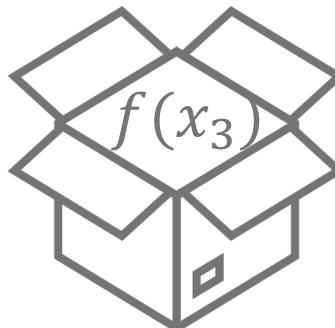
$$\sup_{\text{policy}} \mathbb{E} \left(\max_{t=1,2,\dots,T} f(x_t) - \sum_{t=1}^T c(x_t) \right)$$

Pandora's Box

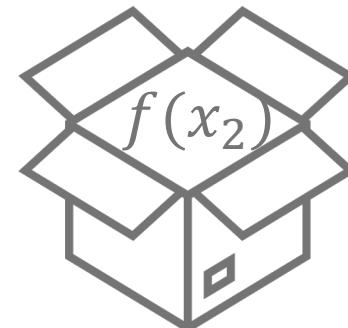
$t = 3$



$c(x_1)$



$c(x_3)$

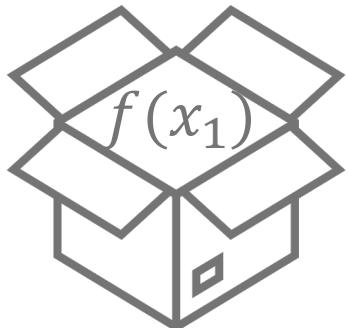


$c(x_2)$

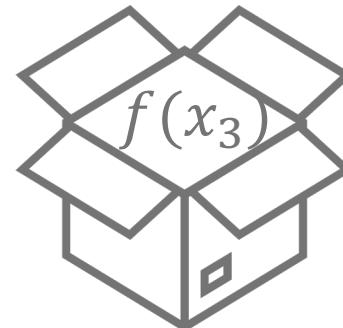
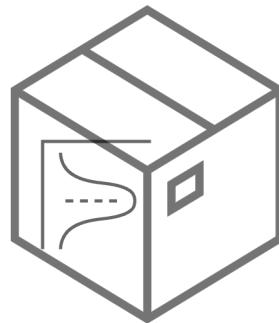
$$\sup_{\text{policy}} \mathbb{E} \left(\max_{t=1,2,\dots,T} f(x_t) - \sum_{t=1}^T c(x_t) \right)$$

Pandora's Box

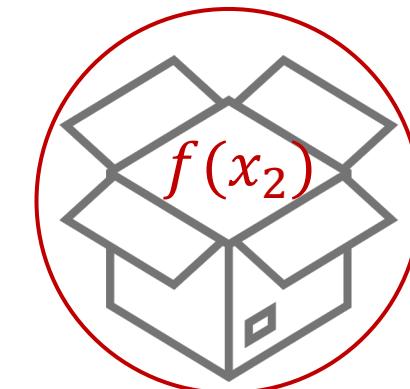
$t = T, \text{stop}$



$c(x_1)$



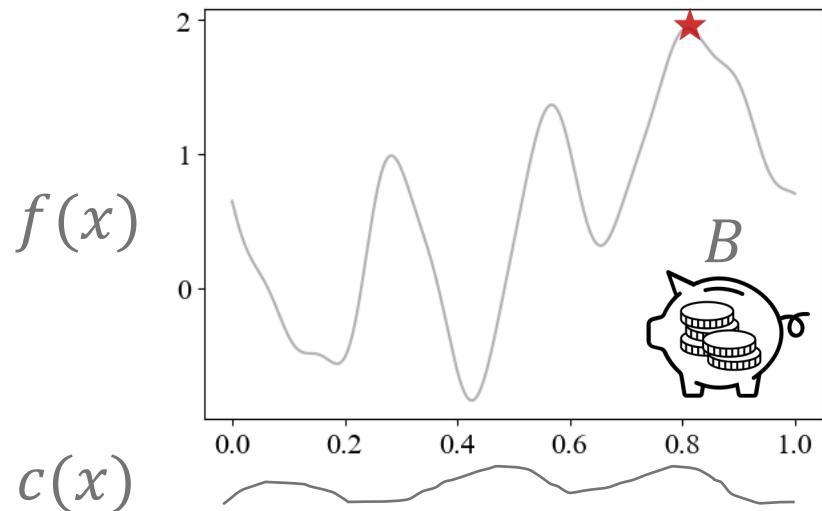
$c(x_3)$



$c(x_2)$

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

Cost-aware Bayesian Optimization



Continuous

Correlated

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]



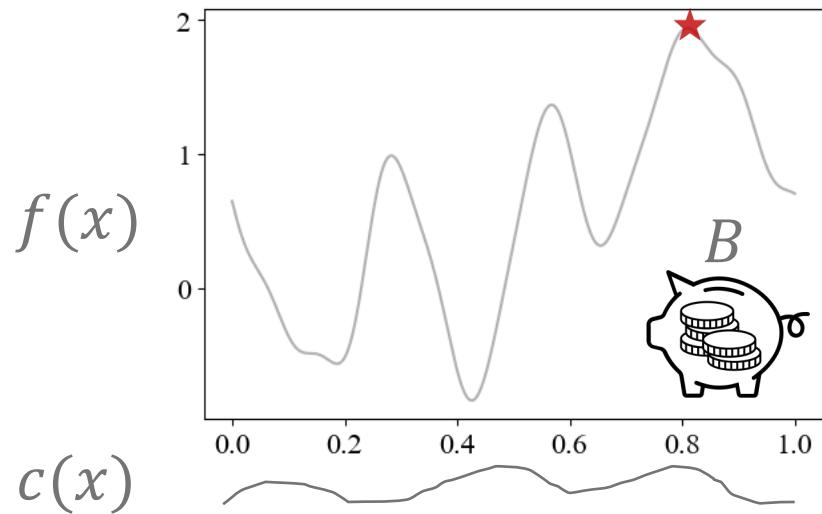
Discrete

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

Expected-budget-constrained

Pandora's Box

[Weitzman'79]



Discrete

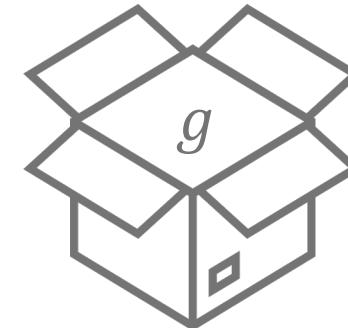
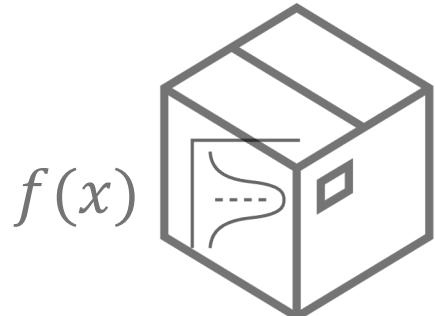
Independent

Cost-per-sample

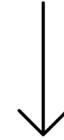
Optimal policy: Gittins index

Optimal Policy: Gittins Index

Step 1: Assign each box a Gittins index (**higher is better**)



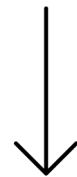
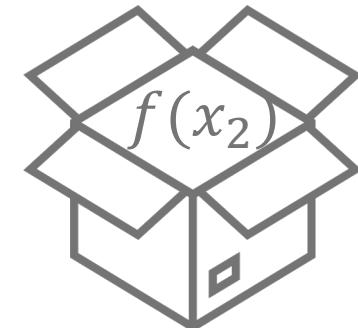
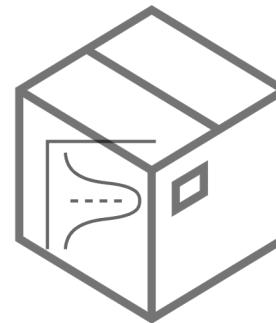
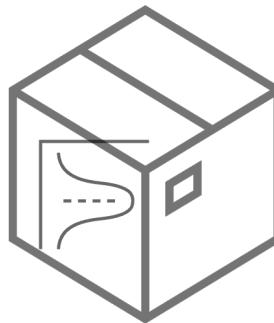
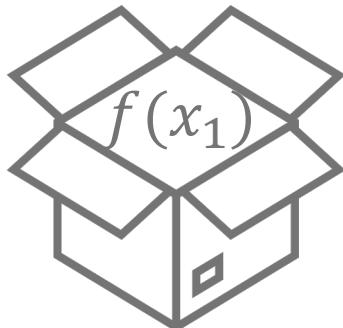
$$\text{GI}_f(x; c(x))$$



$$g$$

Optimal Policy: Gittins Index

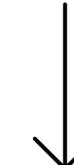
Step 2: Open the box with highest index if it is closed



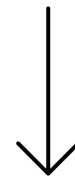
$$f(x_1)$$



$$\text{GI}_f(x; c(x))$$



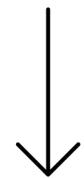
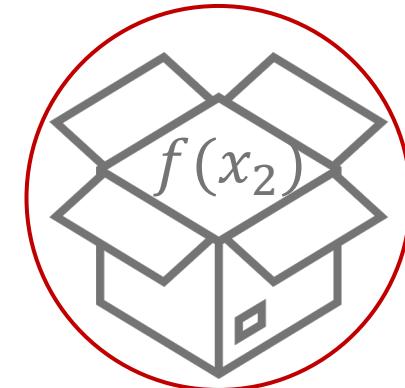
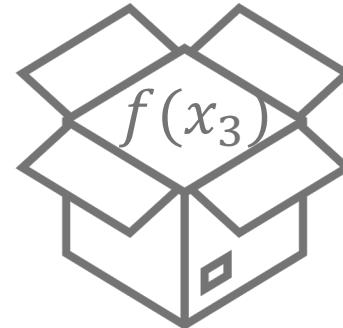
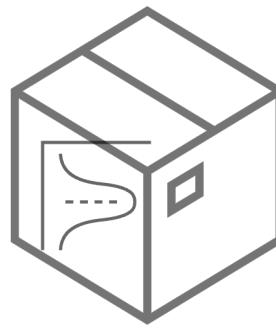
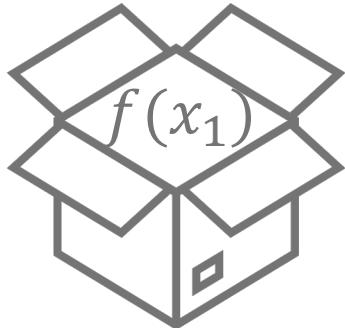
$$\text{GI}_f(x'; c(x'))$$



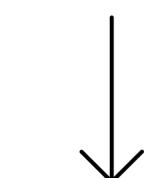
$$f(x_2)$$

Optimal Policy: Gittins Index

Step 2': **Select** the box with highest index if it is opened and **stop**



$$f(x_1)$$



$$\text{GI}_f(x; c(x))$$

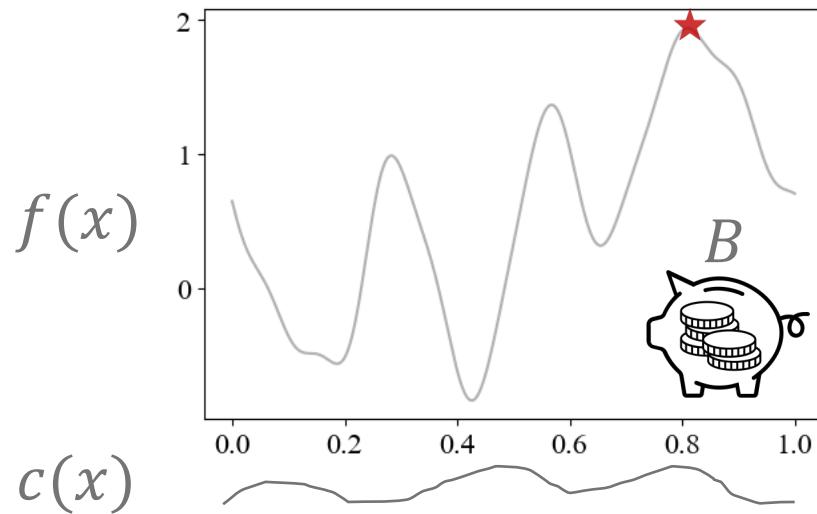


$$f(x_3)$$



$$f(x_2)$$

Cost-aware Bayesian Optimization



Continuous

Correlated

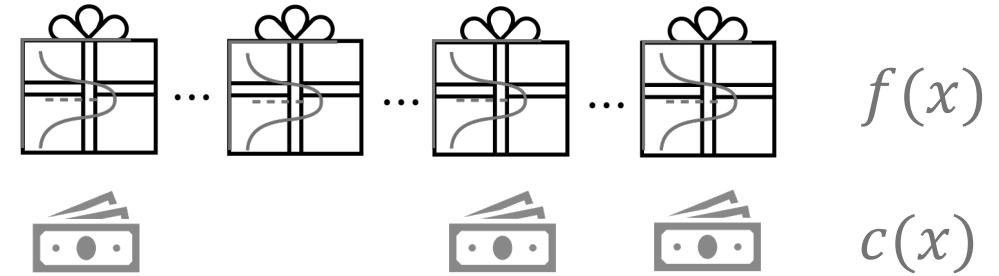
Expected-budget-constrained

How to translate?

Optimal policy: $\text{GI}_f(x; c(x))$

Pandora's Box

[Weitzman'79]



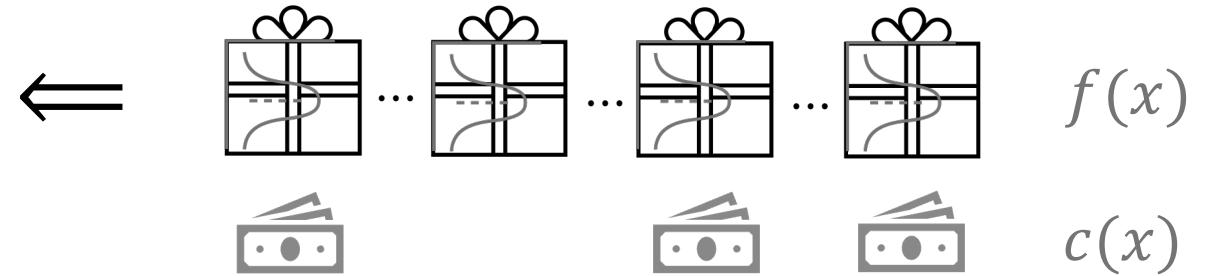
Discrete

Independent

Cost-per-sample

Pandora's Box

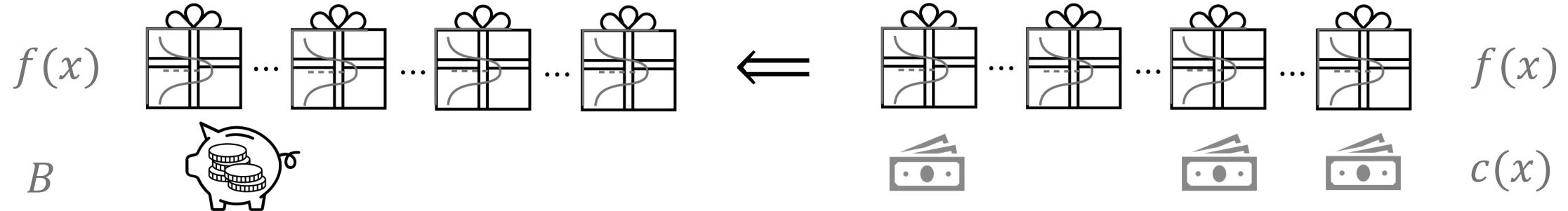
[Weitzman'79]



How to convert?
Expected-budget-constrained ← Cost-per-sample

Budgeted Pandora's Box

[Aminian, Manshadi, Niazadeh'24]

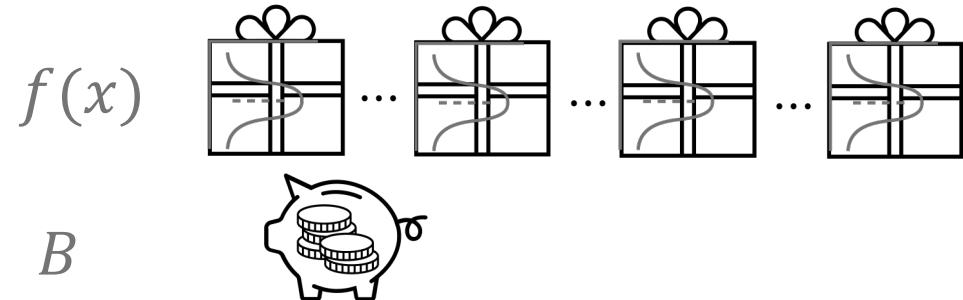


Expected-budget-constrained

Cost-per-sample

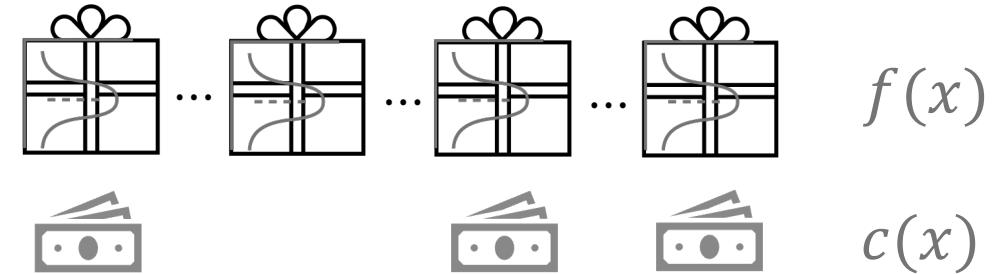
Budgeted Pandora's Box

[Aminian, Manshadi, Niazadeh'24]



Pandora's Box

[Weitzman'79]



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

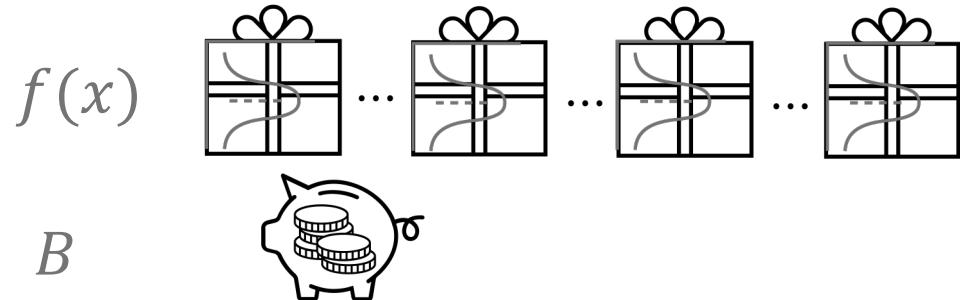
Expected-budget-constrained

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

Cost-per-sample

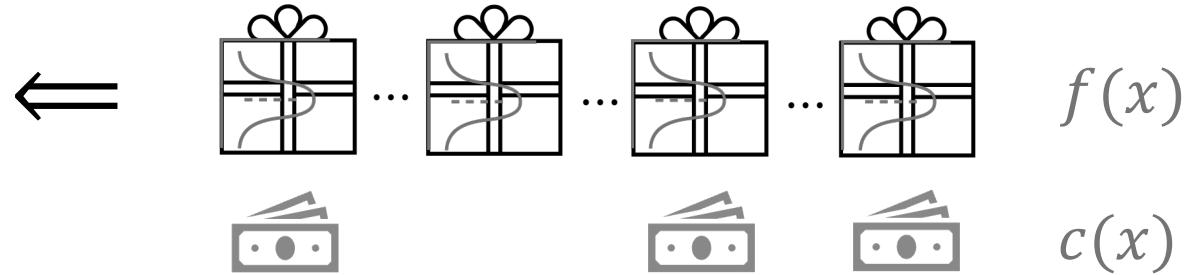
Budgeted Pandora's Box

[Aminian, Manshadi, Niazadeh'24]



Pandora's Box

[Weitzman'79]



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

Expected-budget-constrained \Leftrightarrow Lagrangian relaxation

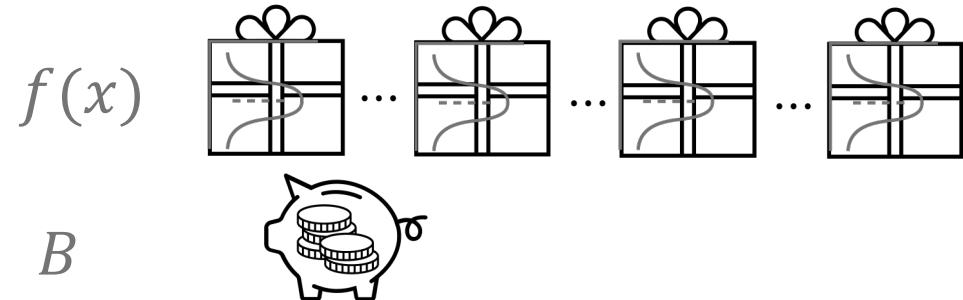
$$\sup_{\text{policy}} \mathbb{E} \left(\max_{t=1,2,\dots,T} f(x_t) - \lambda_B \sum_{t=1}^T c(x_t) \right)$$

scaling factor

Cost-per-sample

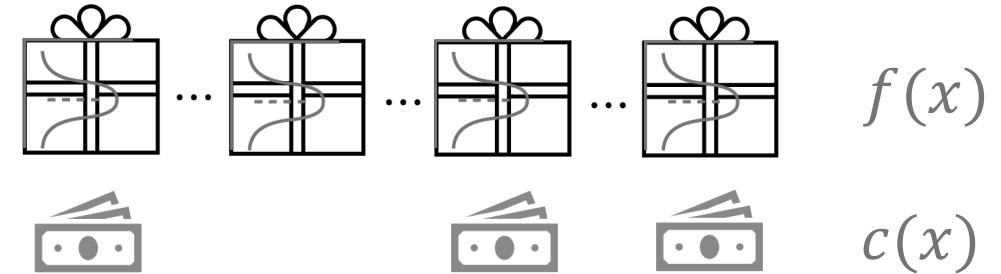
Budgeted Pandora's Box

[Aminian, Manshadi, Niazadeh'24]



Pandora's Box

[Weitzman'79]



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

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

Expected-budget-constrained

Cost-per-sample

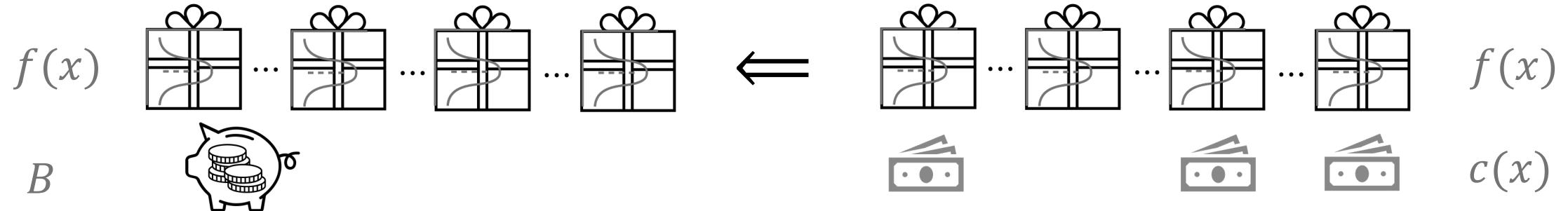
How to translate?



Optimal policy: $\text{GI}_f(x; c)$

Budgeted Pandora's Box

[Aminian, Manshadi, Niazadeh'24]



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

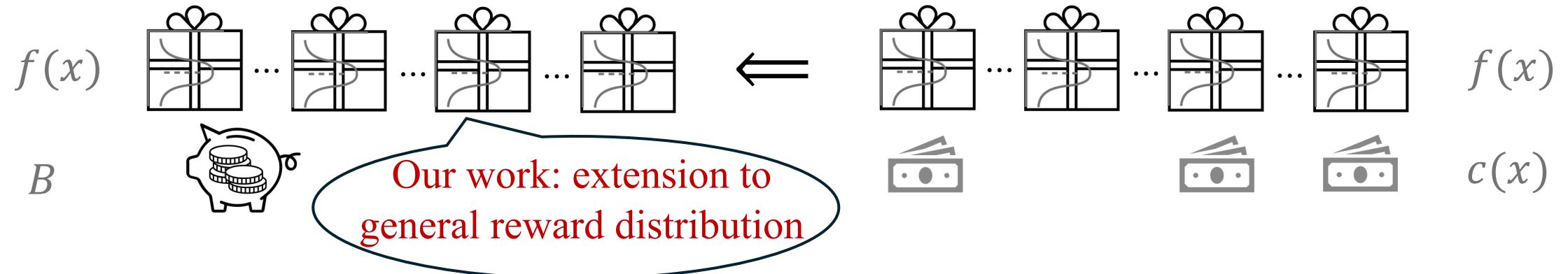
$$\sup_{\text{policy}} \mathbb{E} \left(\max_{t=1,2,\dots,T} f(x_t) - \sum_{t=1}^T c(x_t) \right)$$

Expected-budget-constrained \Leftrightarrow
 Optimal policy: $\text{GI}_f(x; \lambda_B c(x)) \Leftarrow$ scale costs

Cost-per-sample
 Optimal policy: $\text{GI}_f(x; c(x))$

Budgeted Pandora's Box

[Aminian, Manshadi, Niazadeh'24]



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

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

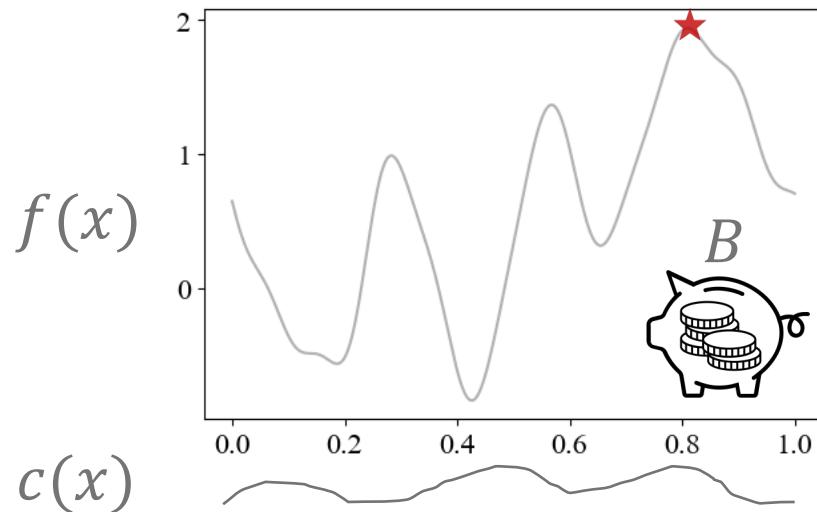
Expected-budget-constrained

Cost-per-sample

Optimal policy: $\text{GI}_f(x; \lambda_B c(x)) \stackrel{\text{scale costs}}{\Leftarrow}$

Optimal policy: $\text{GI}_f(x; c(x))$

Cost-aware Bayesian Optimization



Continuous

Correlated

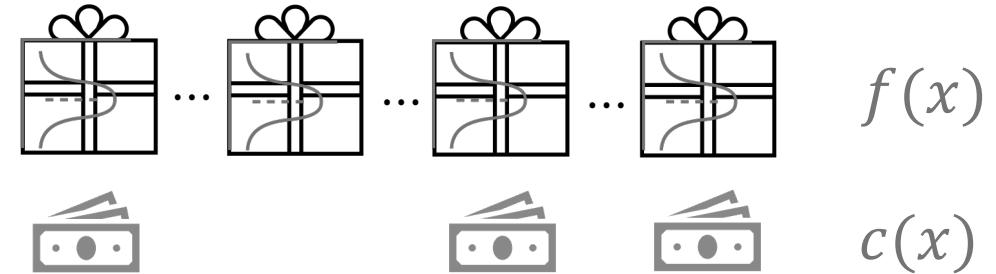
Expected-budget-constrained

How to translate?

Optimal policy: $\text{GI}_f(x; c(x))$

Pandora's Box

[Weitzman'79]



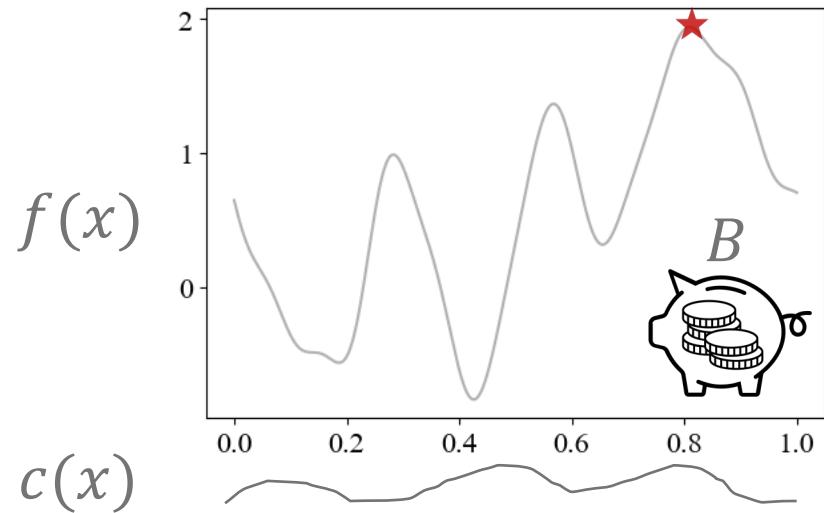
Discrete

Independent



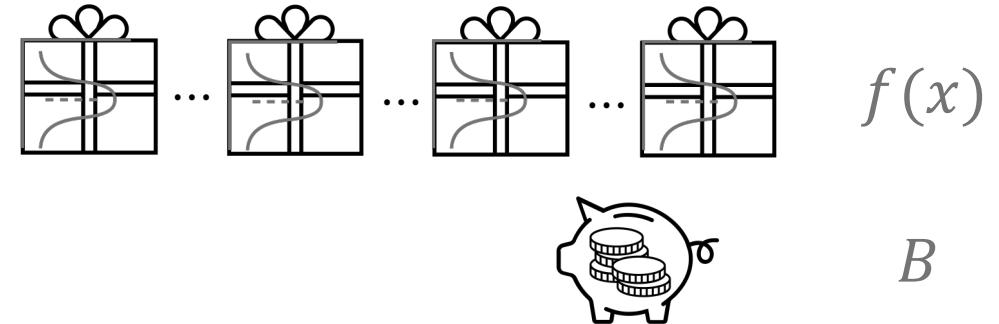
Cost-per-sample

Cost-aware Bayesian Optimization Budgeted Pandora's Box



Continuous

Correlated



Discrete

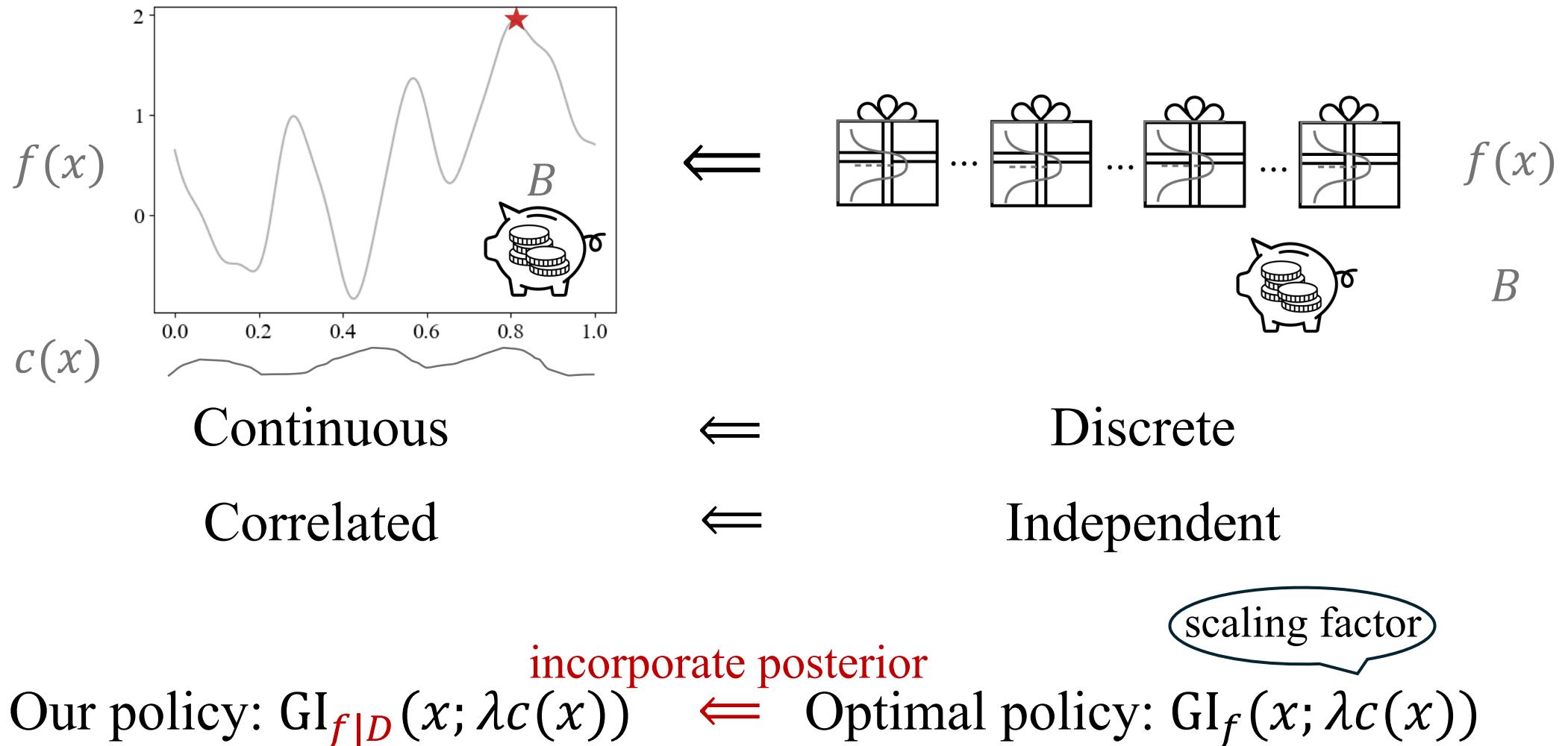
Independent

How to translate?

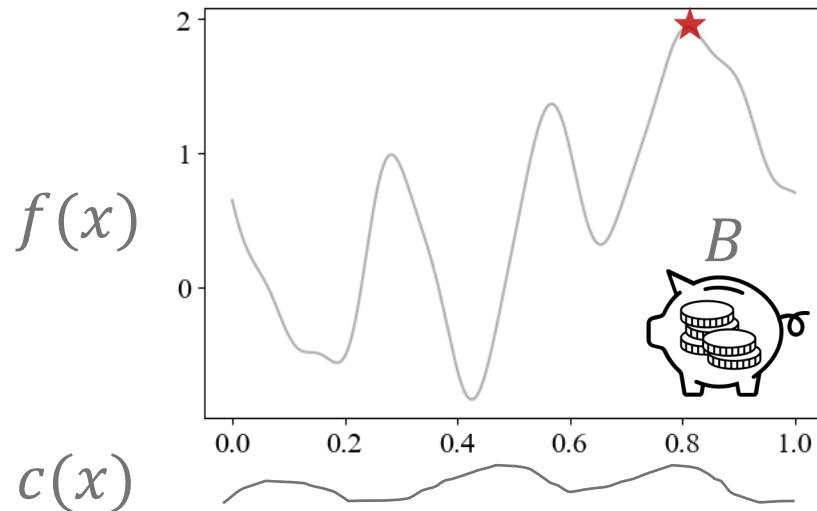
\Leftarrow Optimal policy: $\text{GI}_f(x; \lambda c(x))$

scaling factor

Cost-aware Bayesian Optimization Budgeted Pandora's Box



Cost-aware Bayesian Optimization



Pandora's Box



Continuous



Discrete



Correlated



Independent



Expected-budget-constrained

incorporate posterior

Cost-per-sample

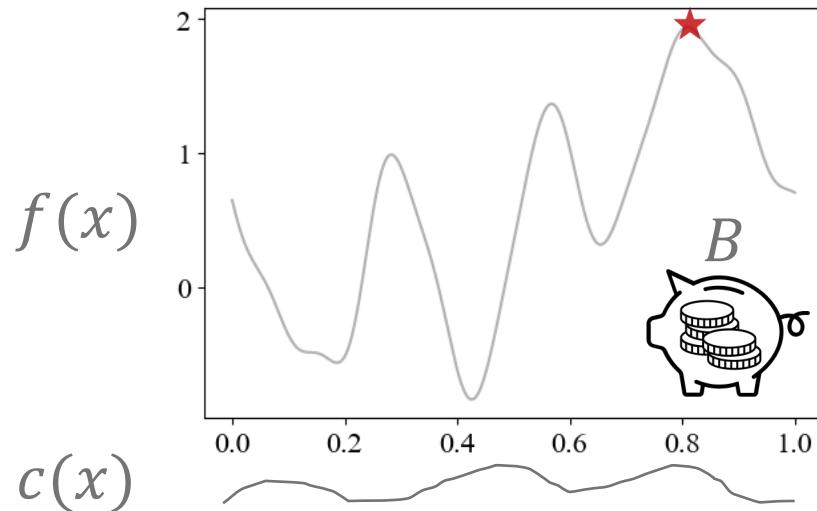


Our policy: $\text{GI}_{f|D}(x; \lambda c(x))$

scale costs

\Leftarrow Optimal policy: $\text{GI}_f(x; c(x))$

Cost-aware Bayesian Optimization



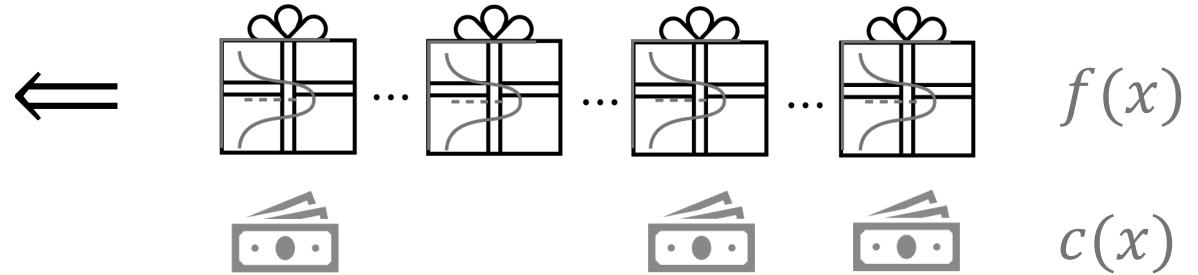
Continuous

Correlated

Expected-budget-constrained
incorporate posterior
Our policy: $\text{GI}_{f|D}(x; \lambda c(x))$

How to compute?

Pandora's Box



Discrete

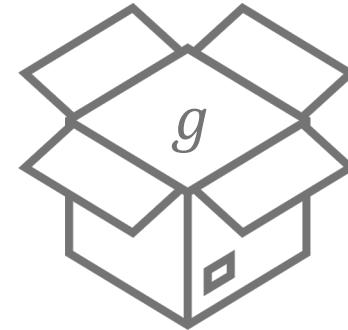
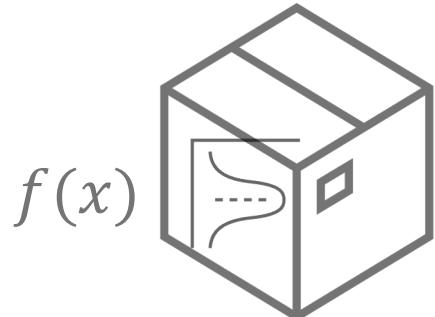
Independent

Cost-per-sample

scale costs

Optimal policy: $\text{GI}_f(x; c(x))$

How to compute Gittins index?



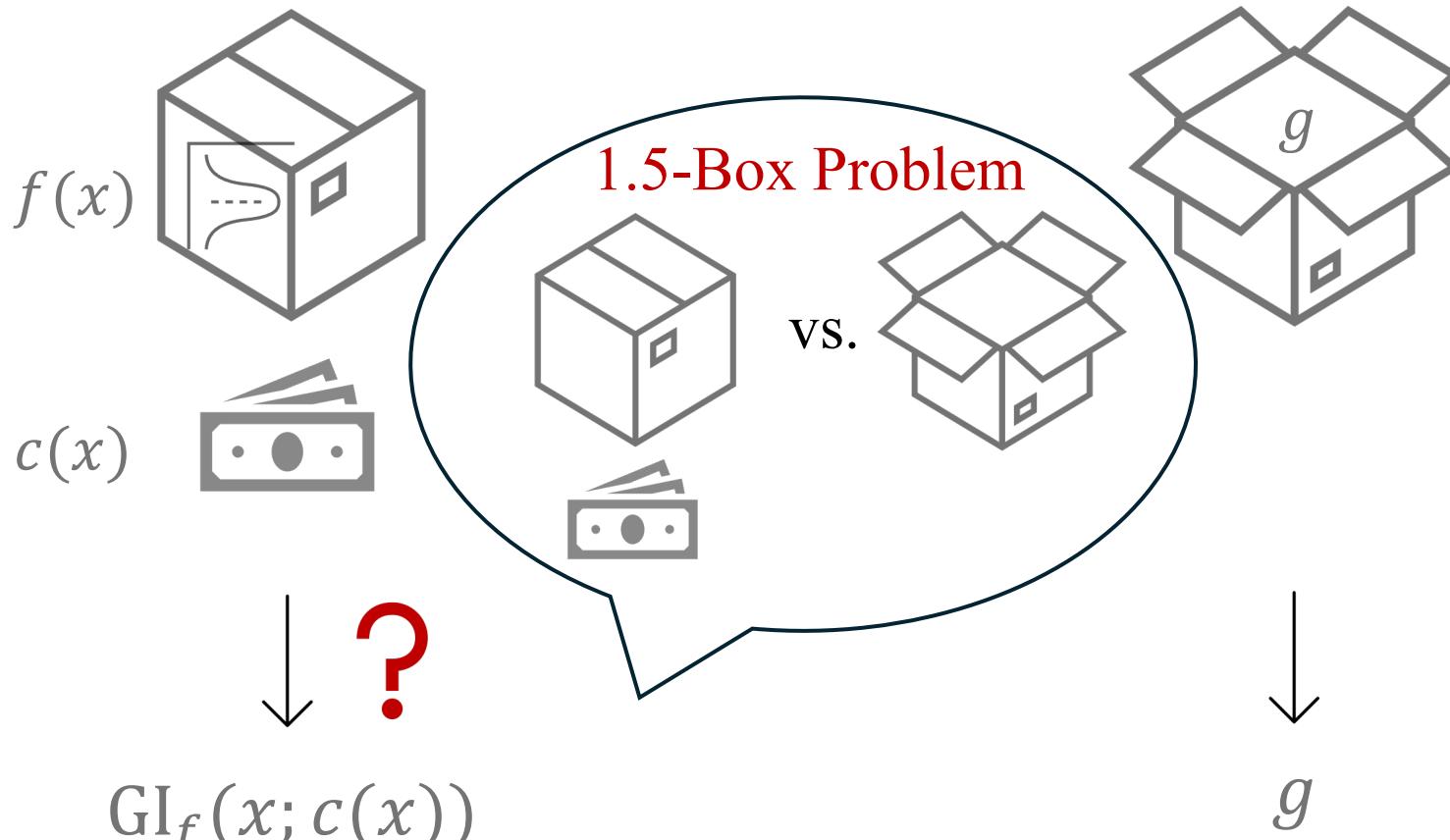
↓ ?

$\text{GI}_f(x; c(x))$

↓

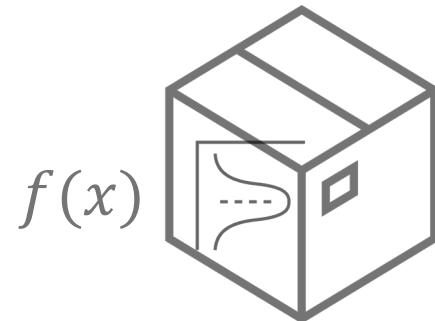
g

How to compute Gittins index?

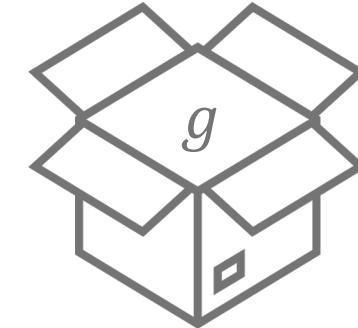


Whether to open a new box or take current best?

Gittins Index Computation: 1.5-Box Problem



vs.

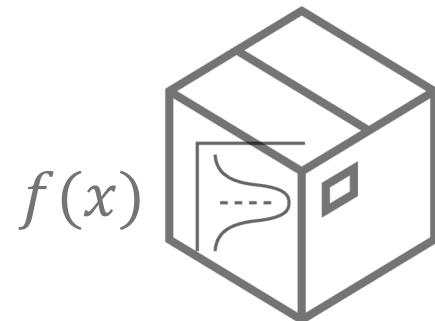


Open closed box

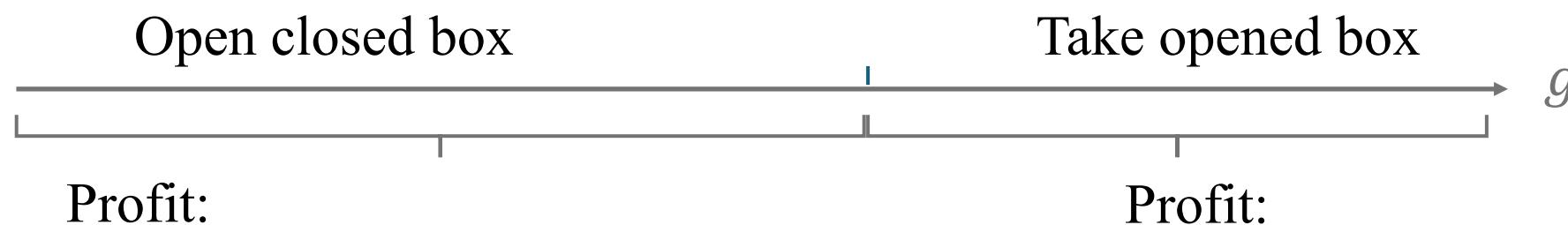
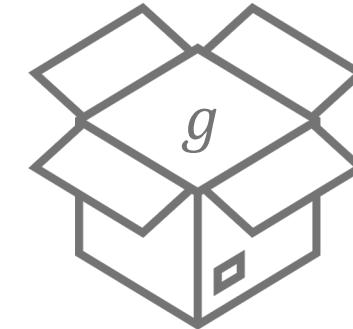
Take opened box

Whether to open a new box or take current best?

Gittins Index Computation: 1.5-Box Problem

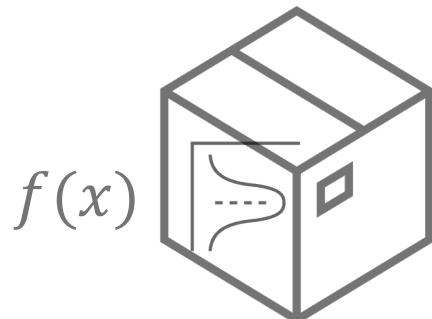


vs.

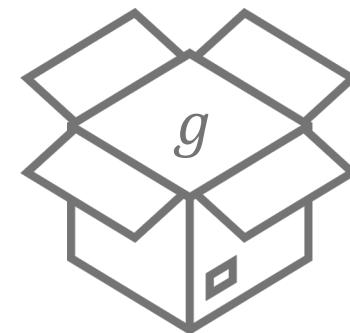


$$\begin{cases} \text{take inside value, } f(x) > g \\ \text{take outside option, } f(x) \leq g \end{cases}$$

Gittins Index Computation: 1.5-Box Problem



vs.

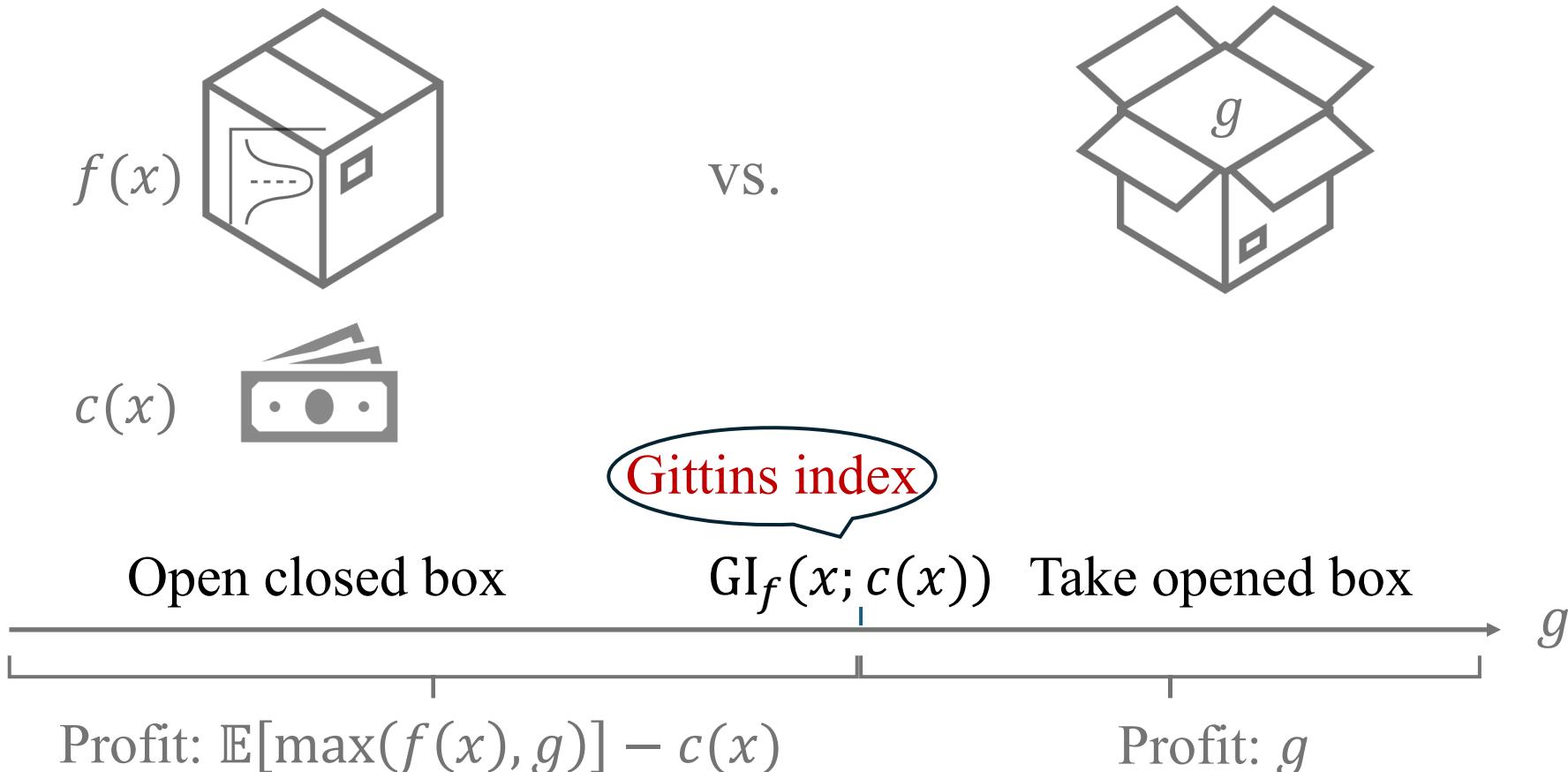


Profit: $\mathbb{E}[\max(f(x), g)] - c(x)$

Profit: g

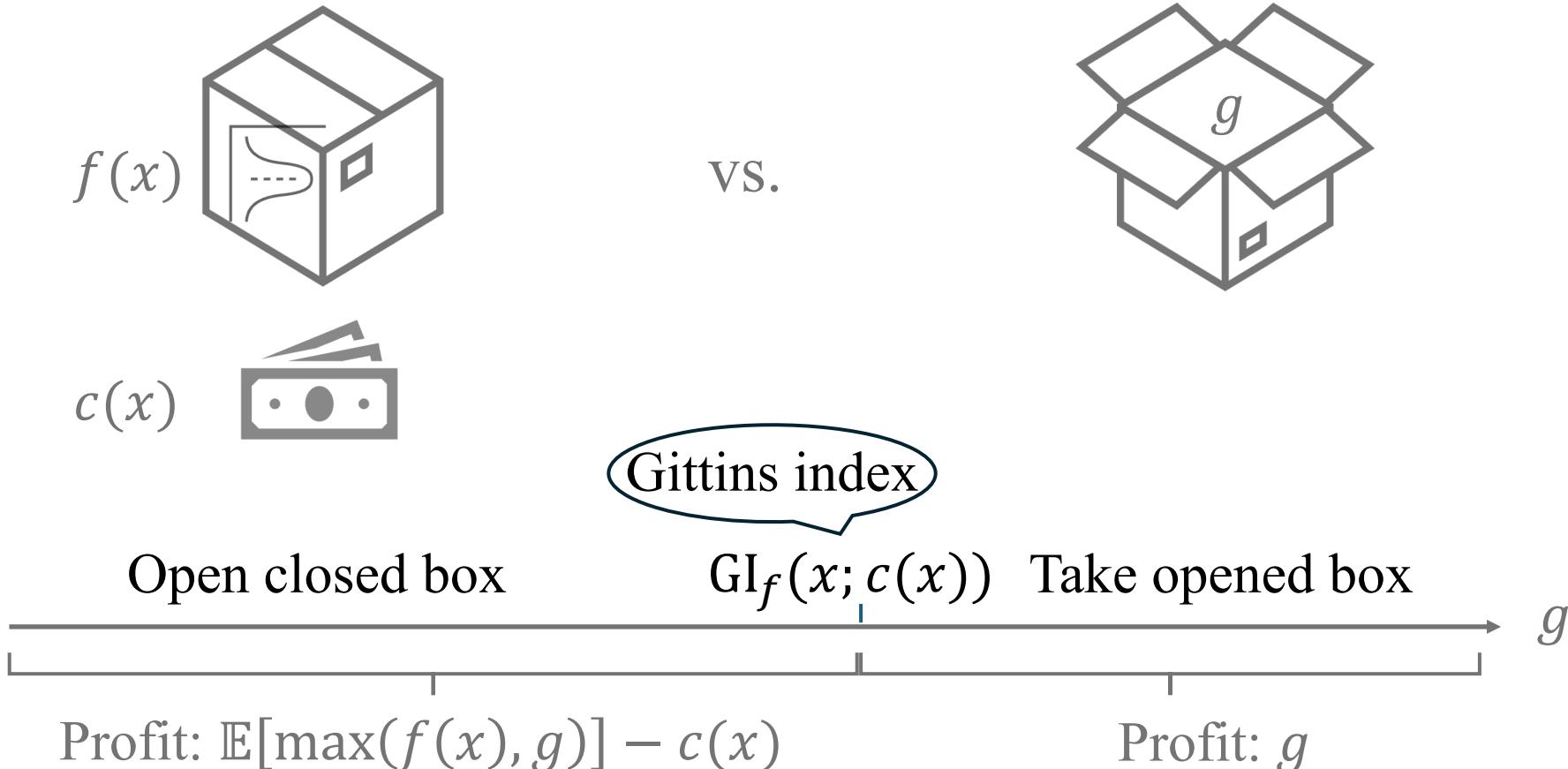
$\begin{cases} \text{take inside value, } f(x) > g \\ \text{take outside option, } f(x) \leq g \end{cases}$

Gittins Index Computation: 1.5-Box Problem



$\text{GI}_f(x; c(x))$: solution g to $\mathbb{E}[\max(f(x), g)] - c(x) = g$

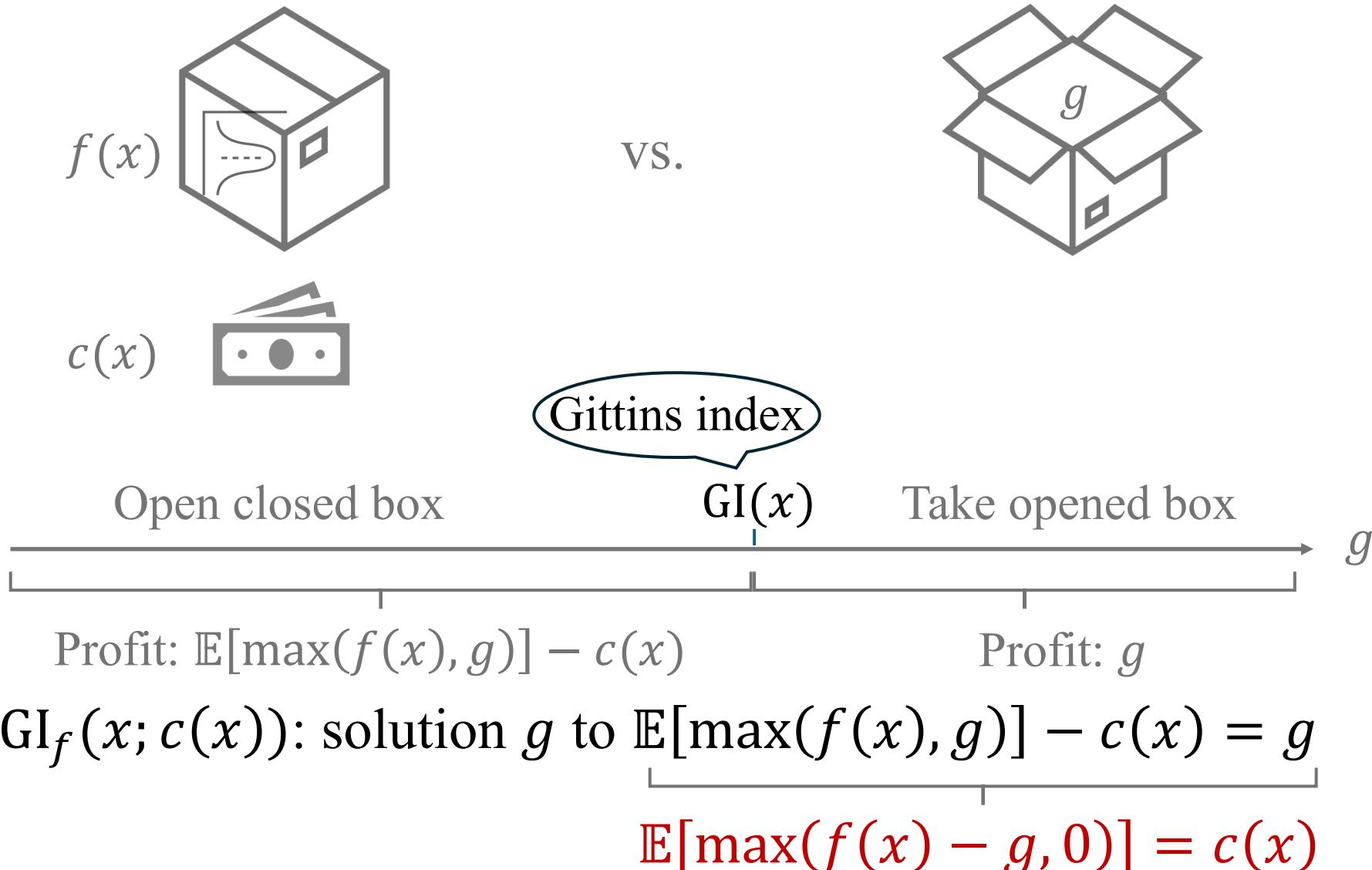
Gittins Index Computation: 1.5-Box Problem



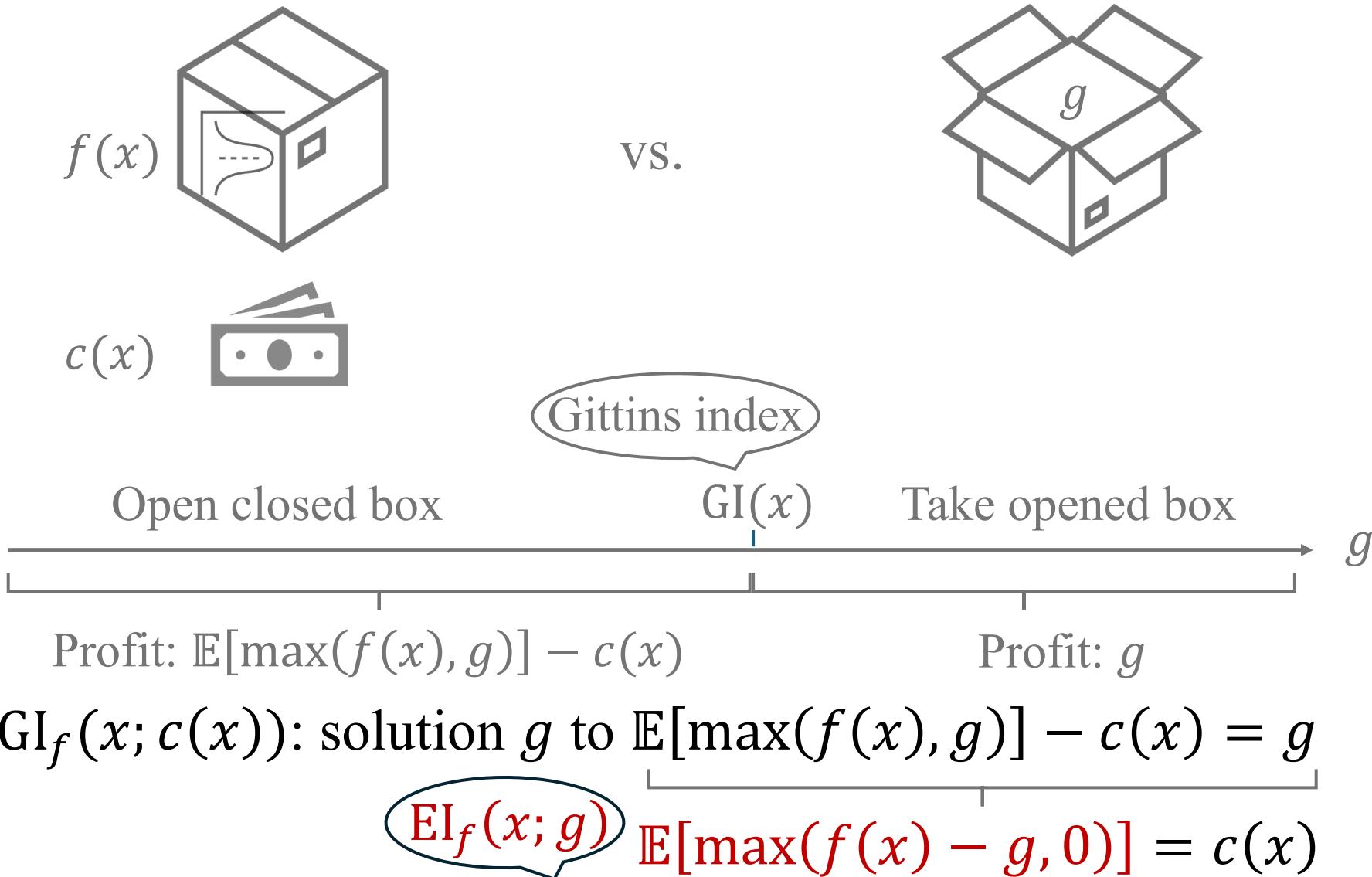
$\text{GI}_f(x; c(x))$: solution g to $\mathbb{E}[\max(f(x), g)] - c(x) = g$

Larger the cost, smaller the Gittins index

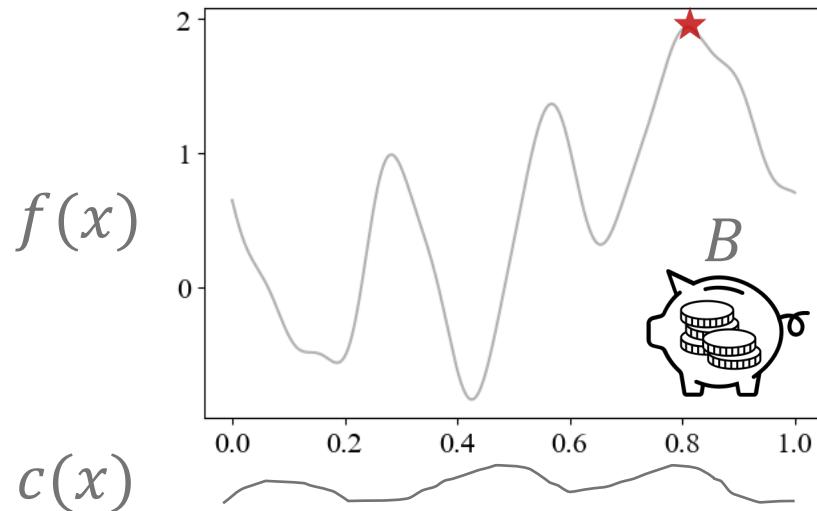
Gittins Index Computation: 1.5-Box Problem



Gittins Index Computation: 1.5-Box Problem



Cost-aware Bayesian Optimization



Continuous

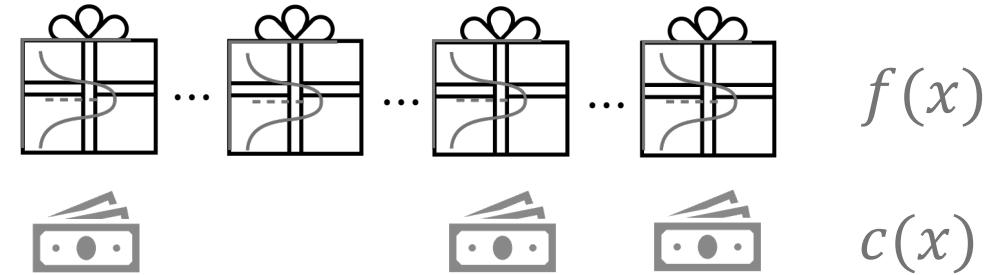
Correlated

Expected-budget-constrained
incorporate posterior

Is Gittins index good?

←
scale costs

Pandora's Box



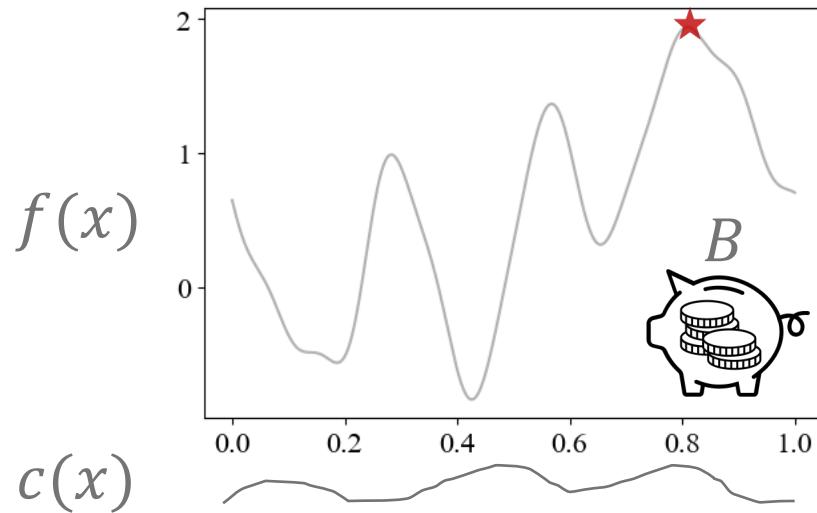
Discrete

Independent

Cost-per-sample

Gittins index is optimal

Cost-aware Bayesian Optimization



Continuous

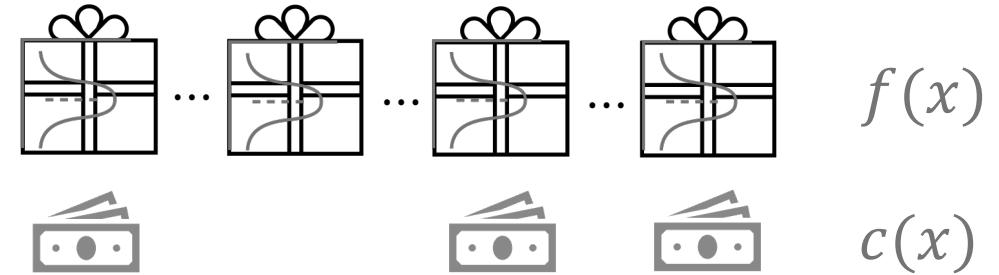
Correlated

Expected-budget-constrained

Is Gittins index **empirically** good?

this work

Pandora's Box



Discrete

Independent

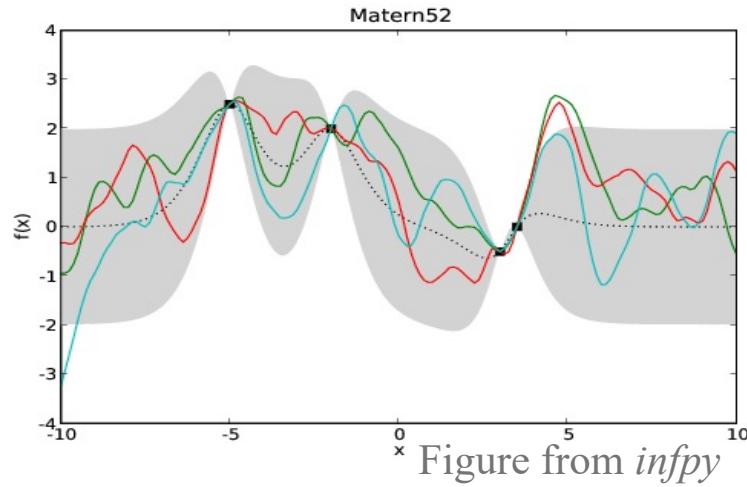
Cost-per-sample

Gittins index is optimal

Experiment Setup: Objective Functions

Synthetic

Samples from prior



Ackley function

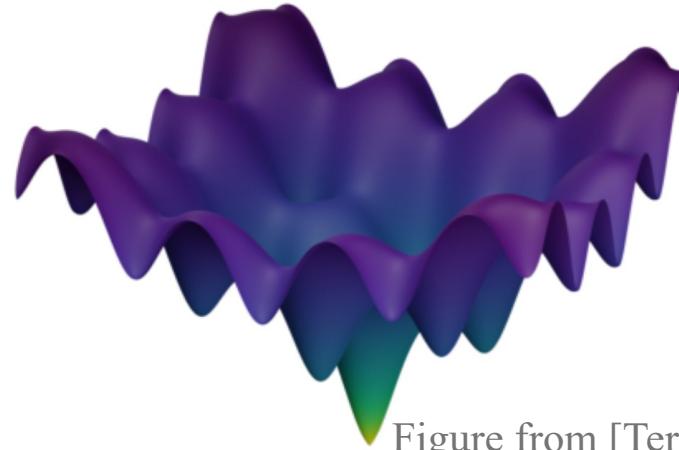


Figure from [Terenin'22]

Empirical

Pest Control



Figure from ChatGPT

Lunar Lander

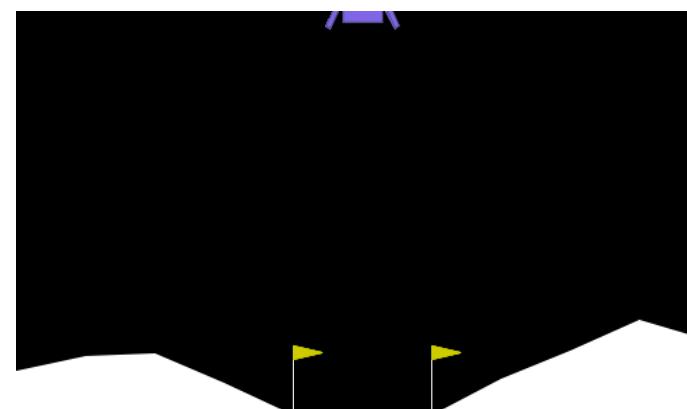
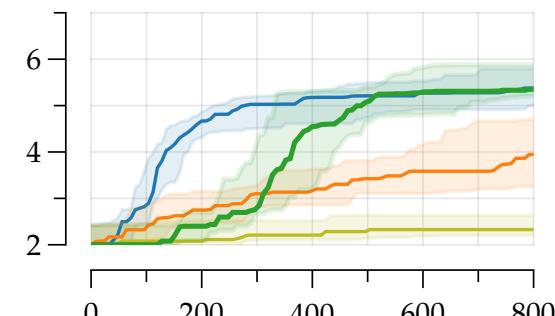
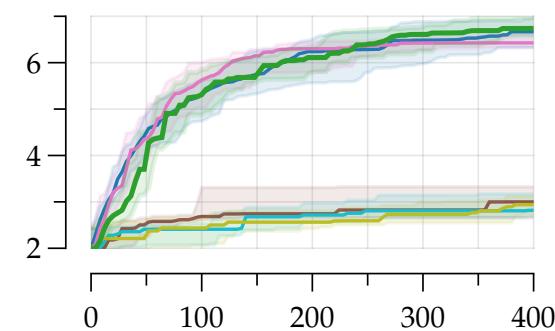


Figure from OpenAI Gym

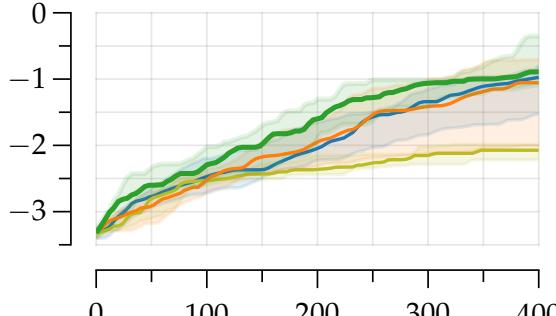
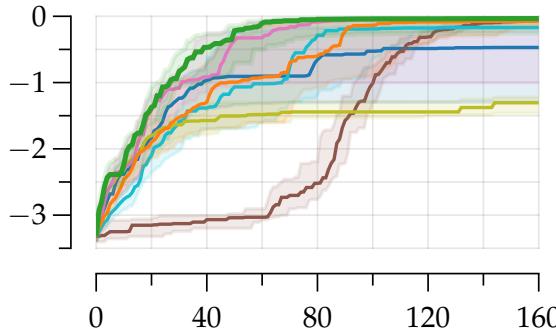
Experiment Results: PBGI vs Baselines

Synthetic

(a) Samples from prior

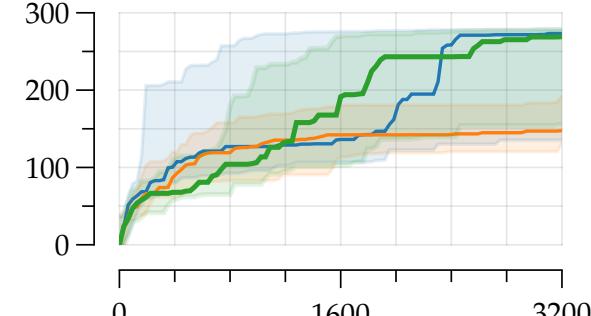
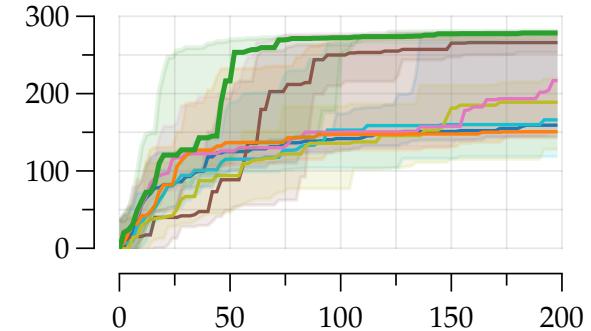


(b) Ackley function



Empirical

(d) Lunar lander



Cumulative Cost

— Pandora's Box Gittins Index (PBGI)
— Upper Confidence Bound (UCB)

— Log Expected Improvement (LogEI)
— Log EI Per Unit Cost (LogEIPC)

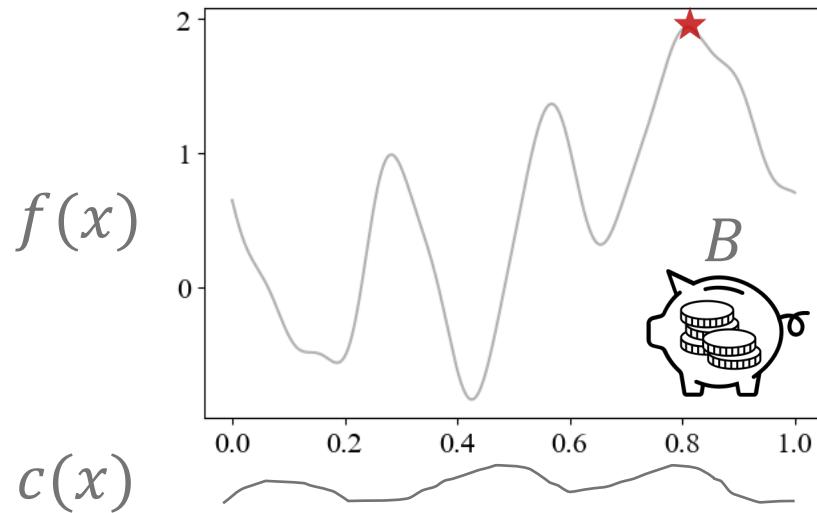
— Thompson Sampling
— Knowledge Gradient

— (Budgeted) Multi-Step EI
— (Multi-fidelity) Max-value Entropy Search

Uniform Costs

Varying Costs

Cost-aware Bayesian Optimization



Continuous

Correlated

Expected-budget-constrained

Is Gittins index **theoretically** good?

Pandora's Box



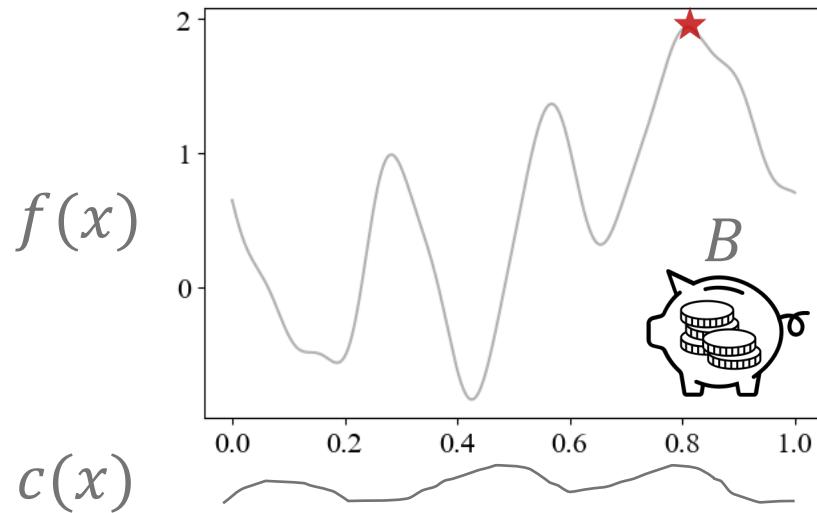
Discrete

Independent

Cost-per-sample

Gittins index is optimal

Cost-aware Bayesian Optimization



Continuous

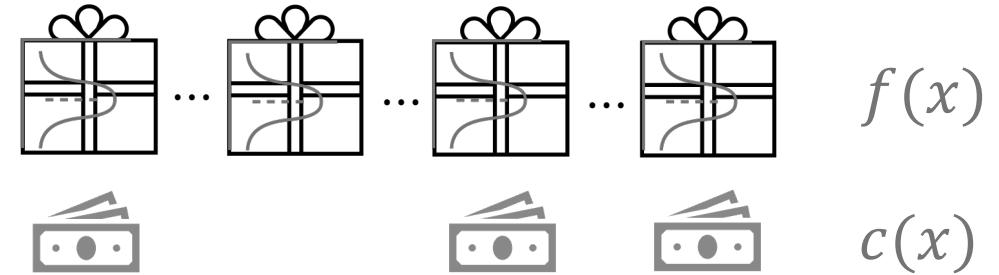
Correlated

Expected-budget-constrained

Is Gittins index **theoretically** good?

ongoing work

Pandora's Box



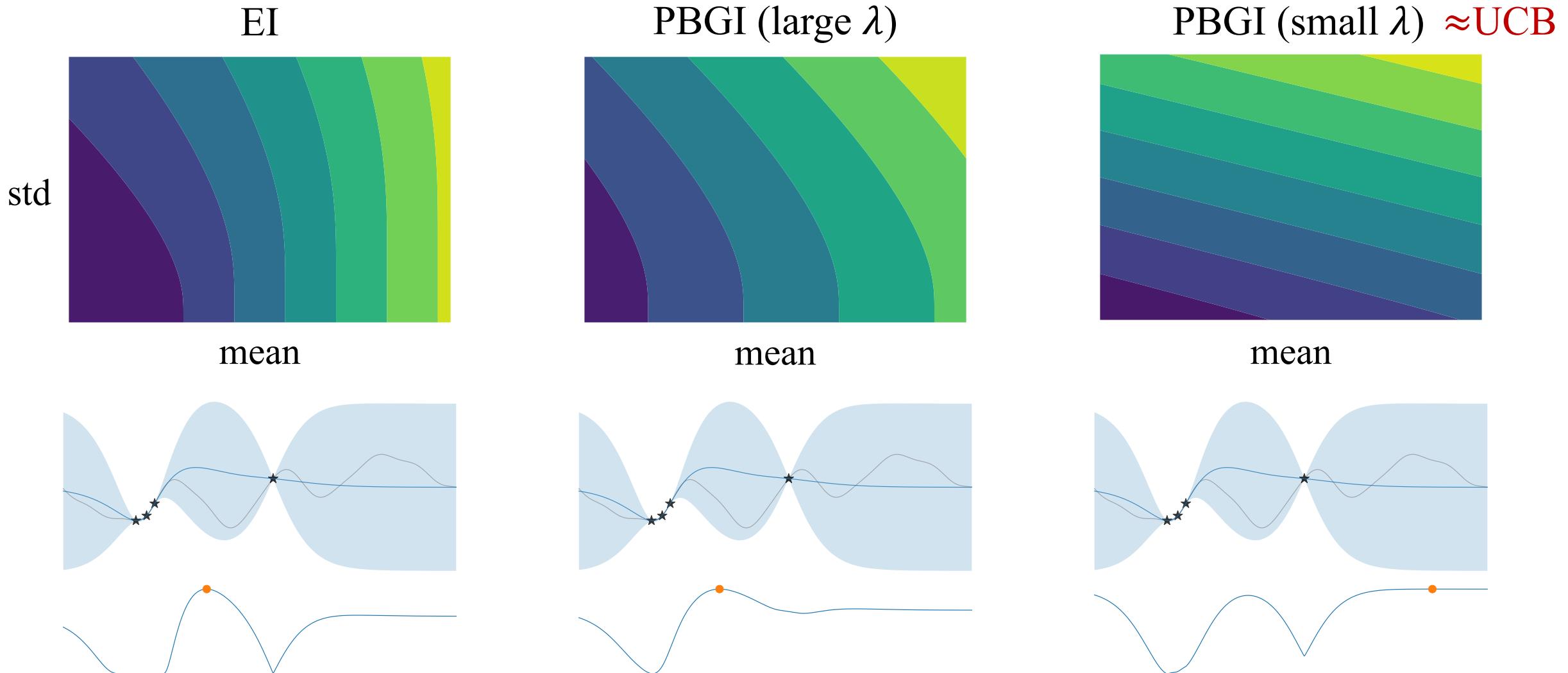
Discrete

Independent

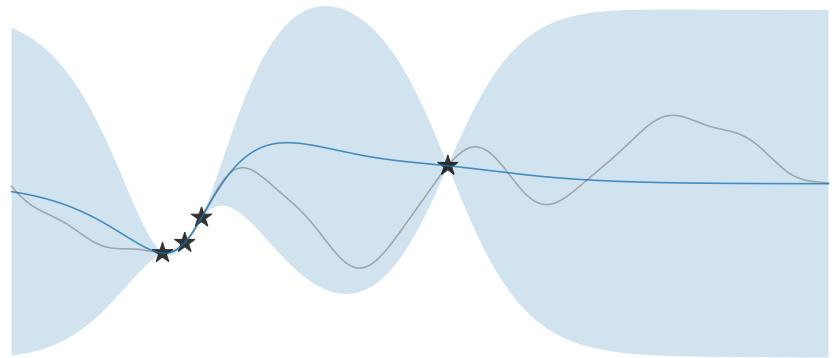
Cost-per-sample

Gittins index is optimal

Discussion 1: EI vs PBGI



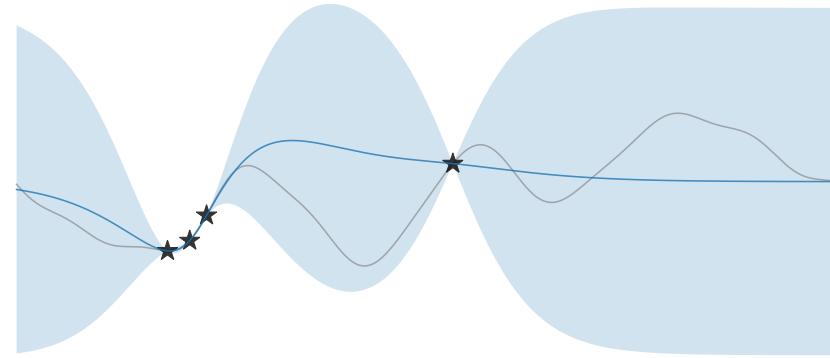
Expected Improvement



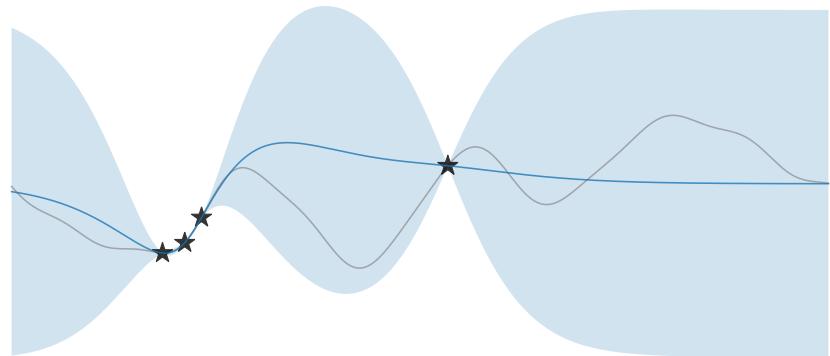
$$\text{EI}_{f|D}(x; g) := \mathbb{E}[\max(f(x) - g, 0) | D] \quad \text{GI}_{f|D}(x) = g \text{ s.t. } \text{EI}_{f|D}(x; g) = \lambda c(x)$$

Gaussian distribution

Gittins Index



Expected Improvement

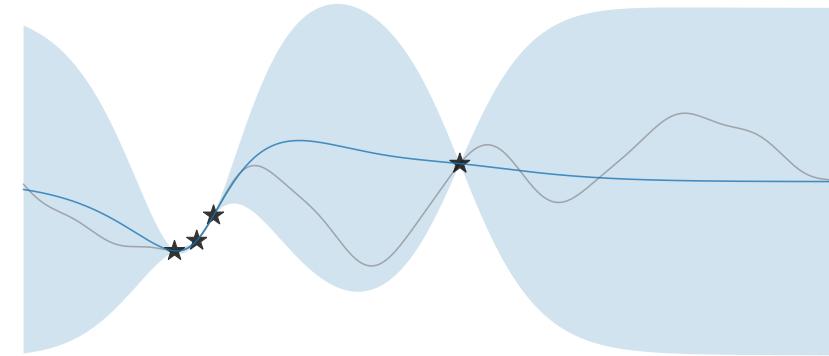


$$\text{EI}_{f|D}(x; g) := \mathbb{E}[\max(f(x) - g, 0) | D]$$

Gaussian distribution

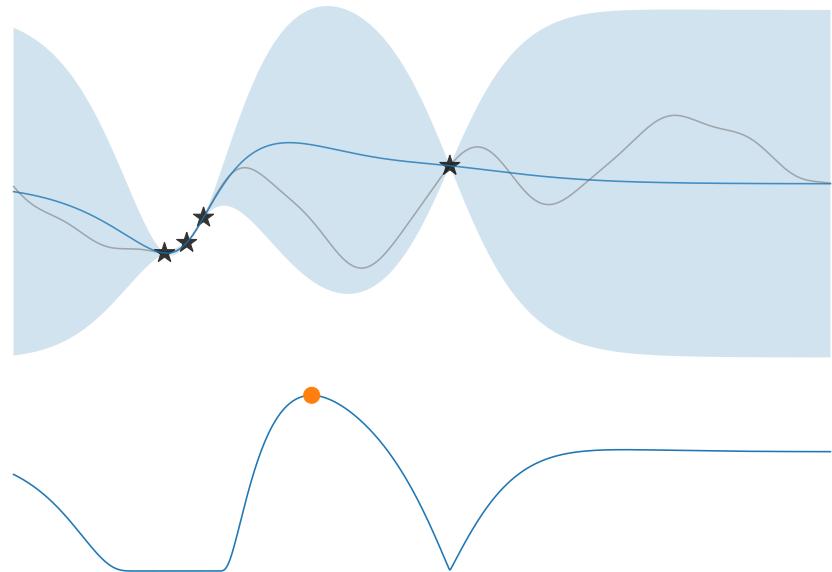
analytical expression

Gittins Index



$$\text{GI}_{f|D}(x) = g \text{ s.t. } \text{EI}_{f|D}(x; g) = \lambda c(x)$$

Expected Improvement

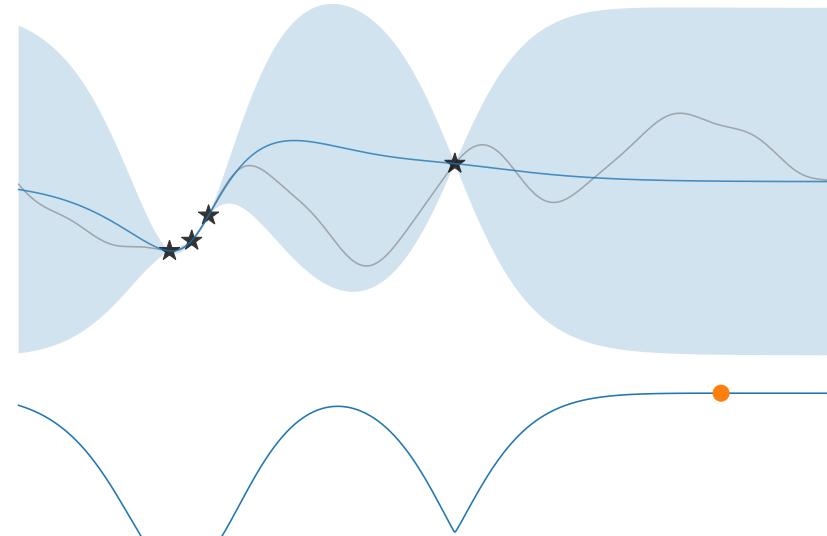


$$\text{EI}_{f|D}(x; g) := \mathbb{E}[\max(f(x) - g, 0) | D]$$

Gaussian distribution

analytical expression

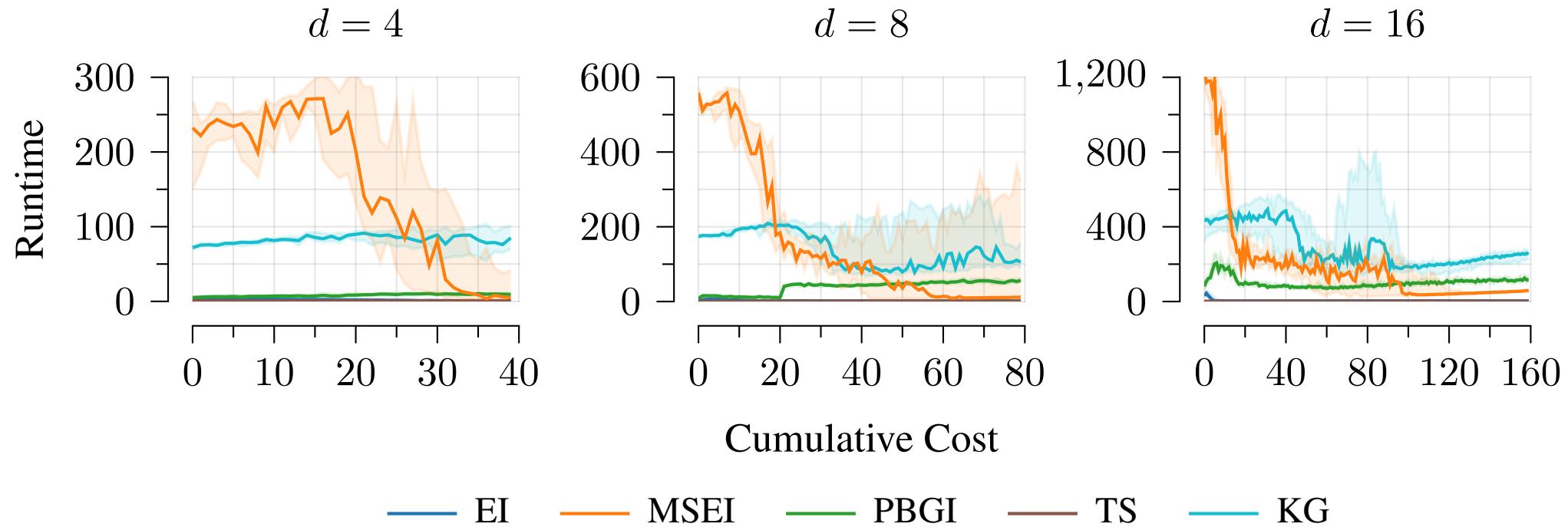
Gittins Index



$$\text{GI}_{f|D}(x) = g \text{ s.t. } \text{EI}_{f|D}(x; g) = \lambda c(x)$$

$\text{GI}_{f|D}$ is easy to compute using $\text{EI}_{f|D}$ + bisection search!

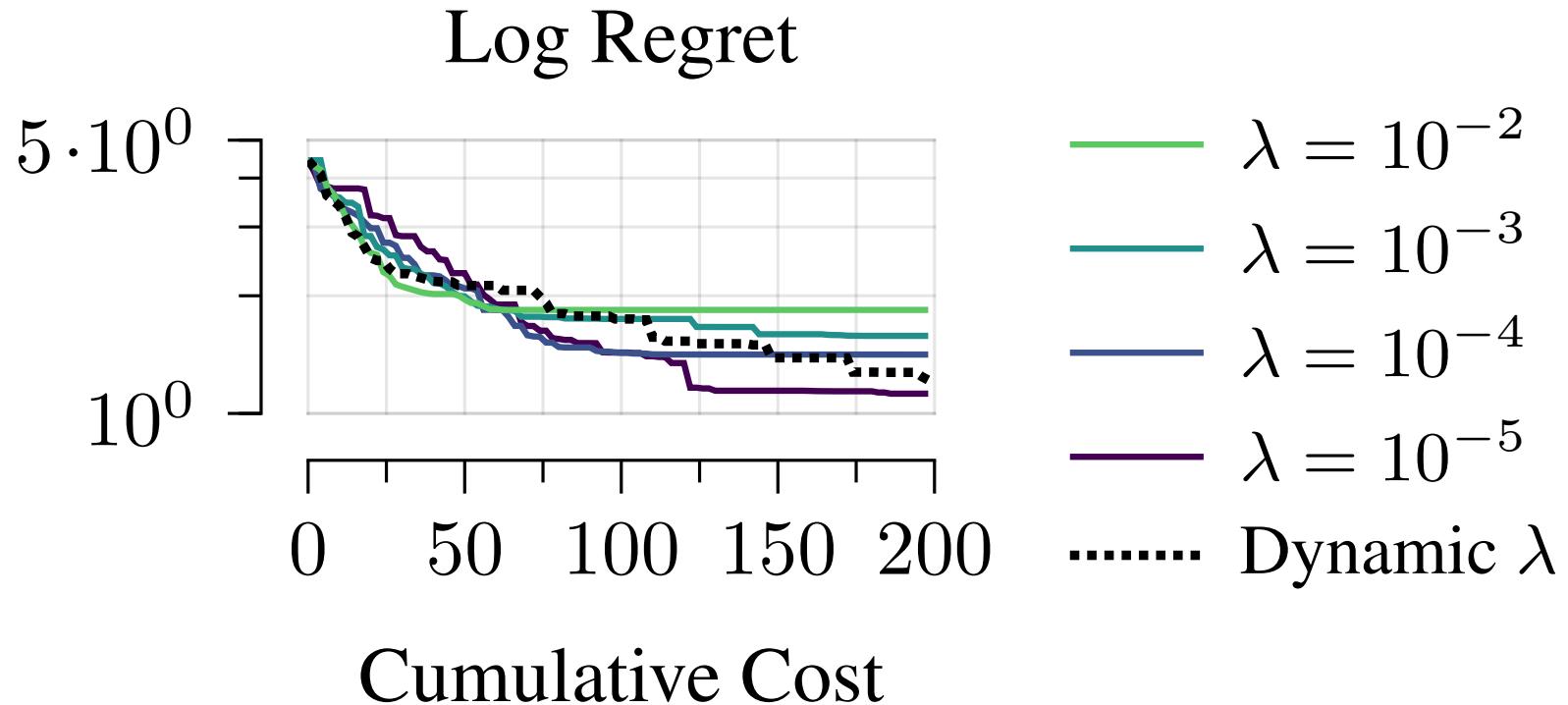
Discussion 2: Timing Experiment



$$\text{EI}_{f|D}(x; g) := \mathbb{E}[\max(f(x) - g, 0) \mid D] \quad \text{GI}_{f|D}(x) = g \text{ s.t. } \text{EI}_{f|D}(x; g) = \lambda c(x)$$

PBGI is **easy to compute using EI + bisection search!**

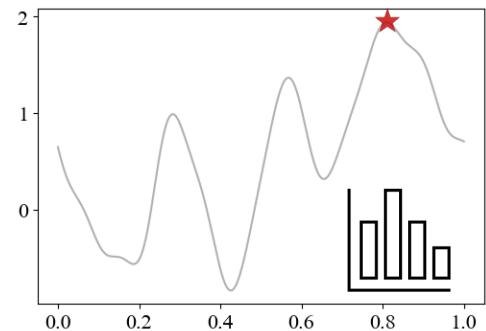
Discussion 3: Effect of λ



Larger budget, smaller λ , higher exploration

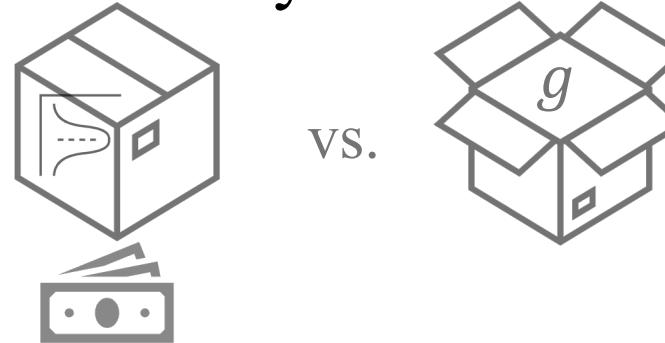
New Design Principle: Gittins Index

Studied Problem



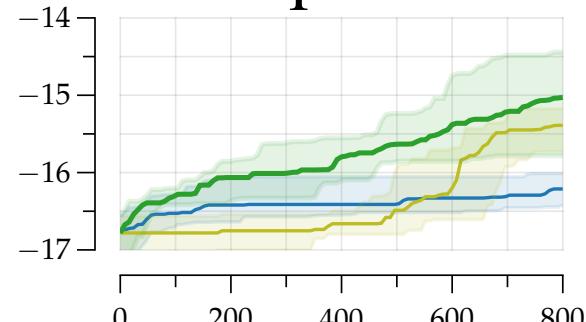
Cost-aware Bayesian optimization

Key idea



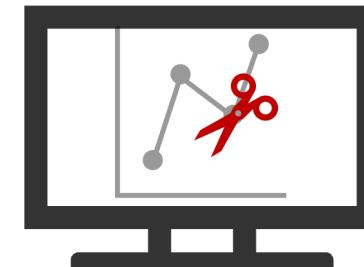
Link to Pandora's box
and Gittins index theory

Impact



Competitive empirical performance

Ongoing work



Cost-aware experiment-level &
trial-level early-stopping

Find our paper on ArXiv!



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