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

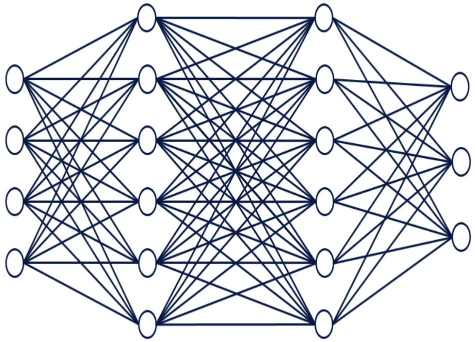
Peter Frazier

Ziv Scully

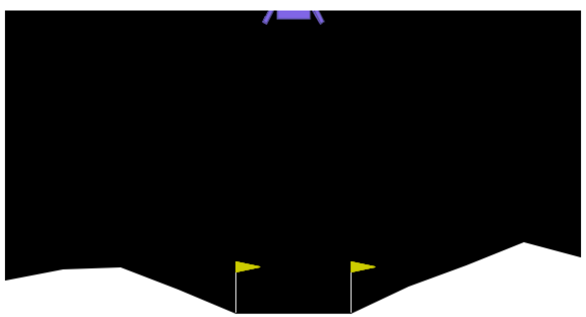
Alexander Terenin

World of Hyperparameter Optimization

Hyperparameter tuning:

Training hyperparameters →  → Accuracy

Control optimization:

Control variables →  → Reward

Adaptive experimentation:

Decision variables → 

→ Revenue

World of Hyperparameter Optimization

Black-box optimization:

Input hyperparameters →  → Performance metric

World of Hyperparameter Optimization

Black-box function:

$x \longrightarrow$

$\longrightarrow f(x)$

$f \sim$Stochastic Process

Optimizing Black-box Functions

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

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

$f \sim$Stochastic Process

Optimizing Black-box Functions

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

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

$f \sim$Stochastic Process

Grid Search? Random search?

Optimizing Black-box Functions

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

expensive-to-evaluate

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

$f \sim$ Stochastic Process

Grid Search? Random search?

Optimizing Black-box Functions

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

limited #evaluations

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

$f \sim$Stochastic Process

Grid Search? Random search?

Optimizing Black-box Functions

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

limited #evaluations

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

$f \sim$Stochastic Process

More efficient: Bayesian optimization!

Bayesian Optimization

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

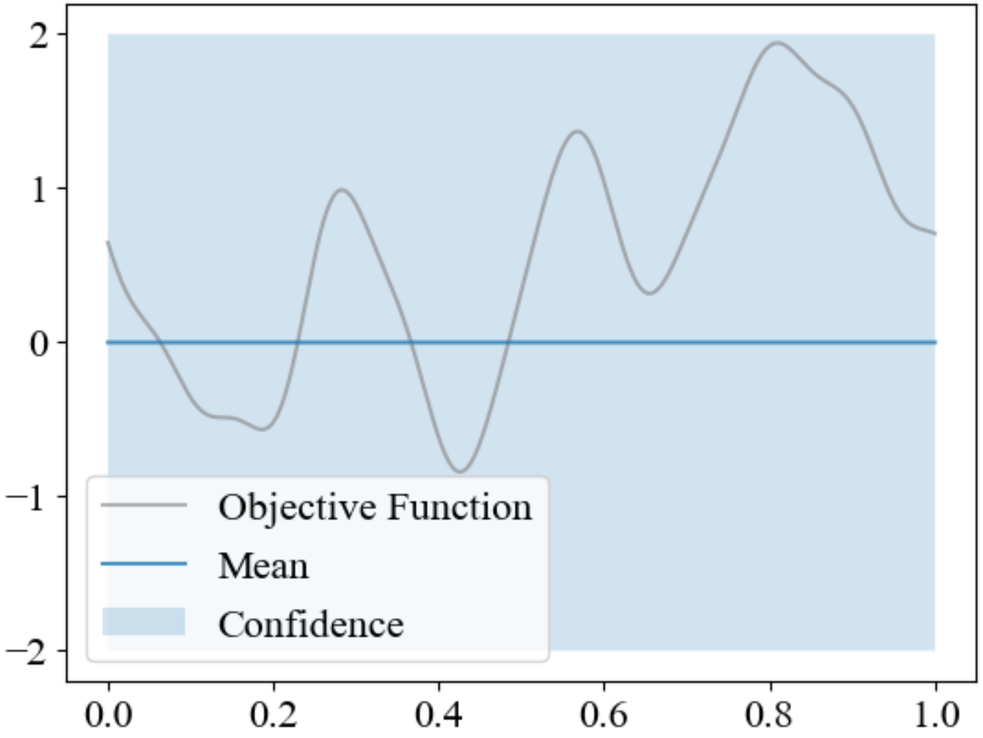
limited #evaluations

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

$f \sim$Gaussian Process

Key idea: maintain posterior belief about f

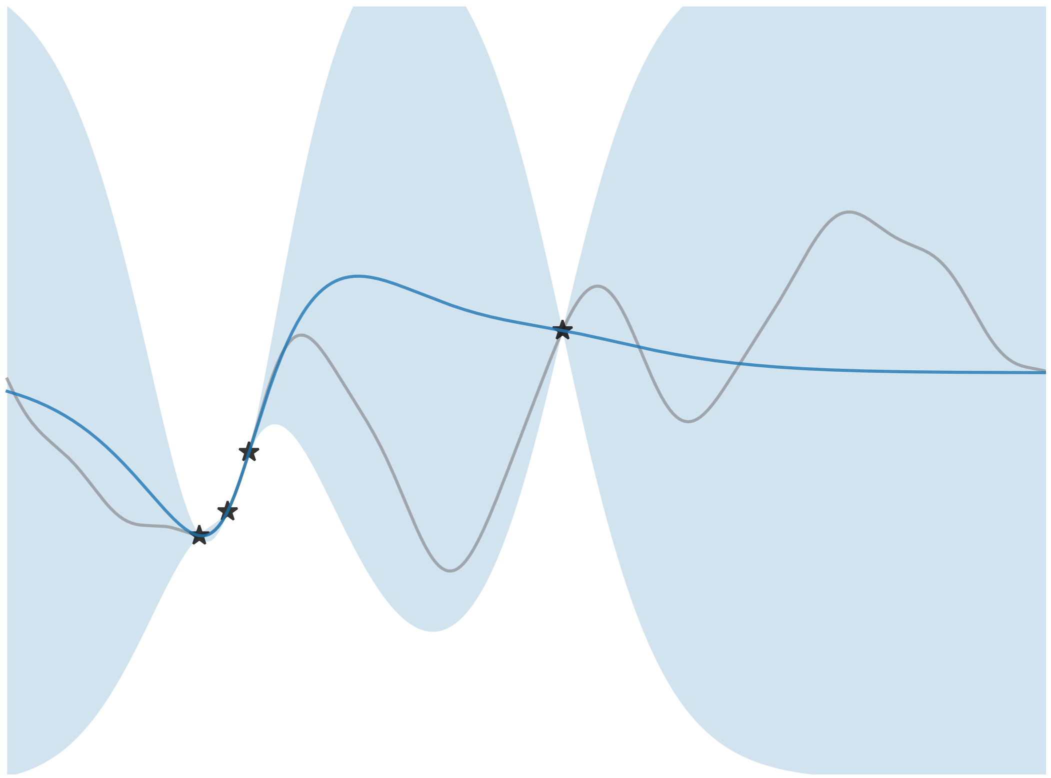
Bayesian Optimization



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

$f \sim$Gaussian Process

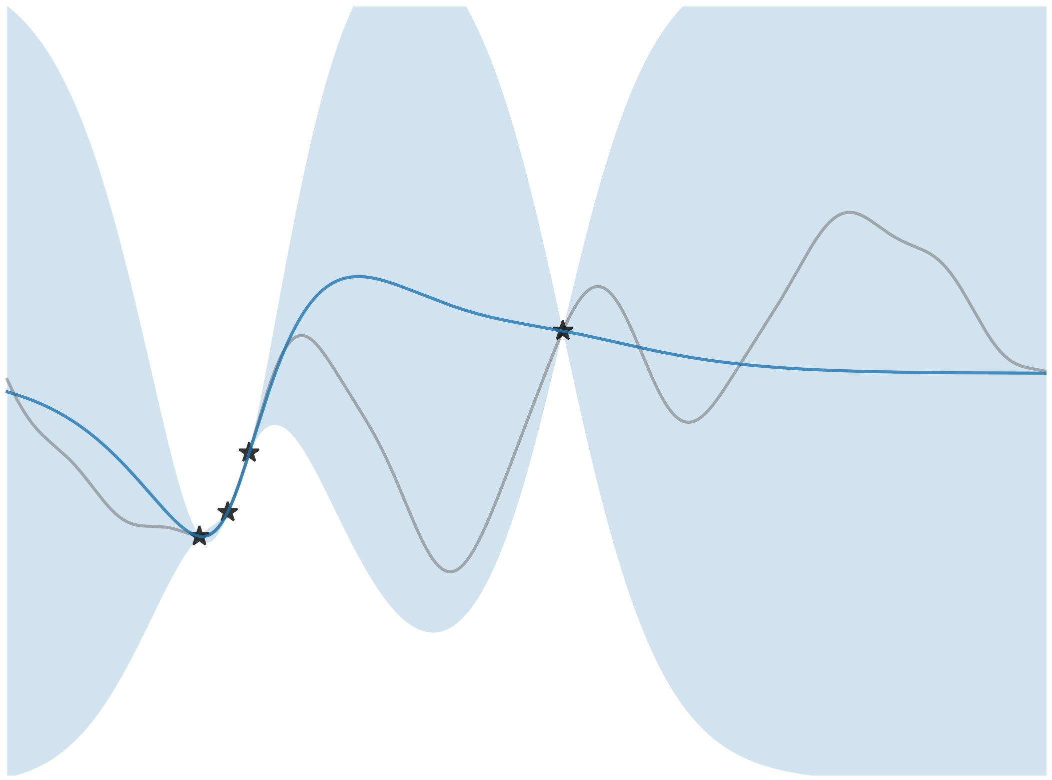
Bayesian Optimization



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

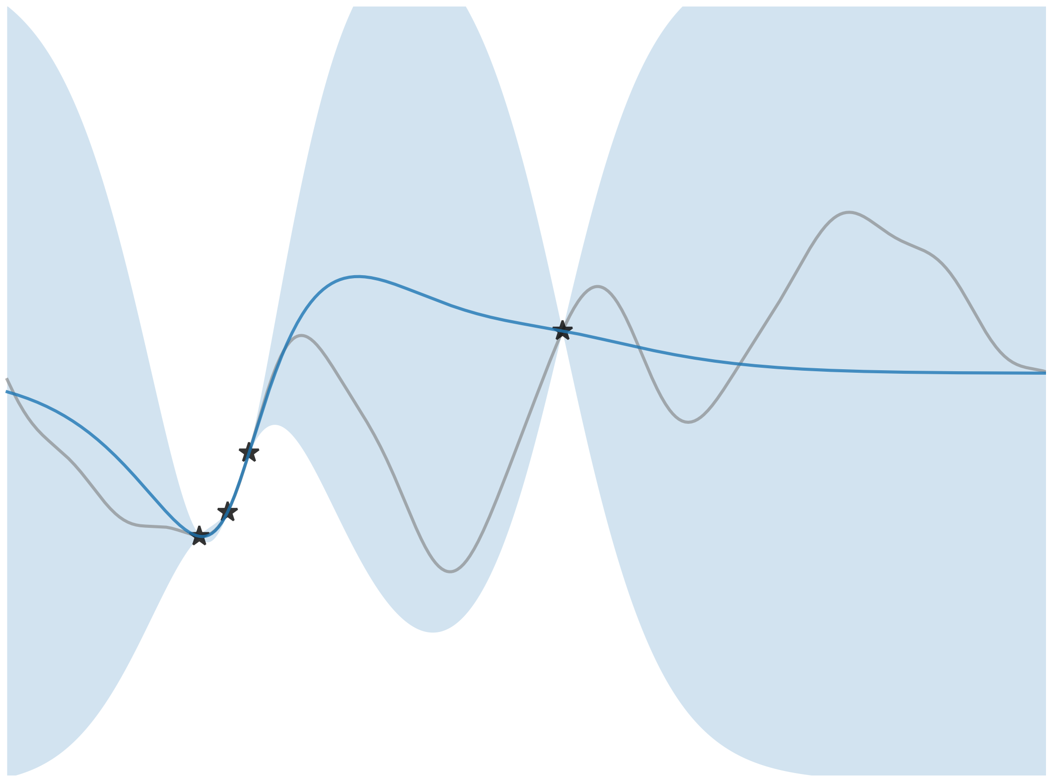
$f \sim$Gaussian Process

Bayesian Optimization



What to evaluate next?

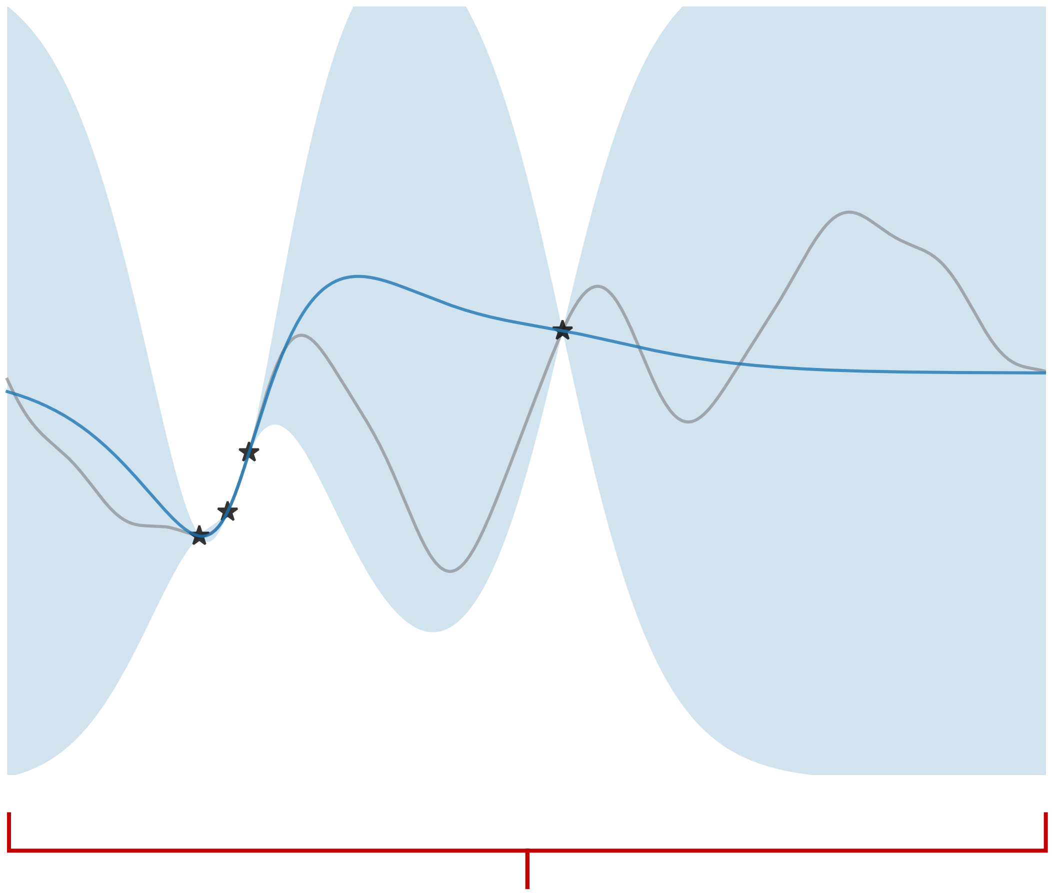
Bayesian Optimization



What to evaluate next?

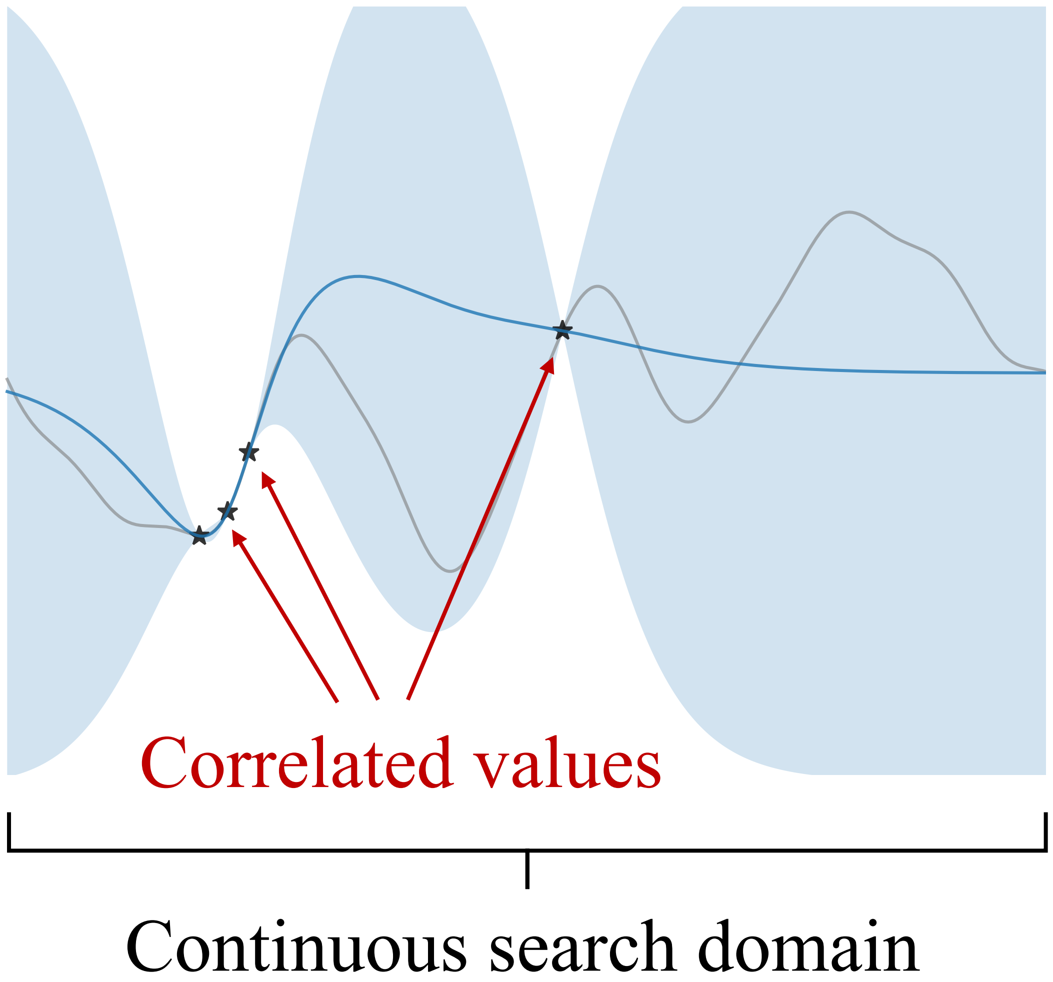
Optimal policy?

Properties of Bayesian Optimization

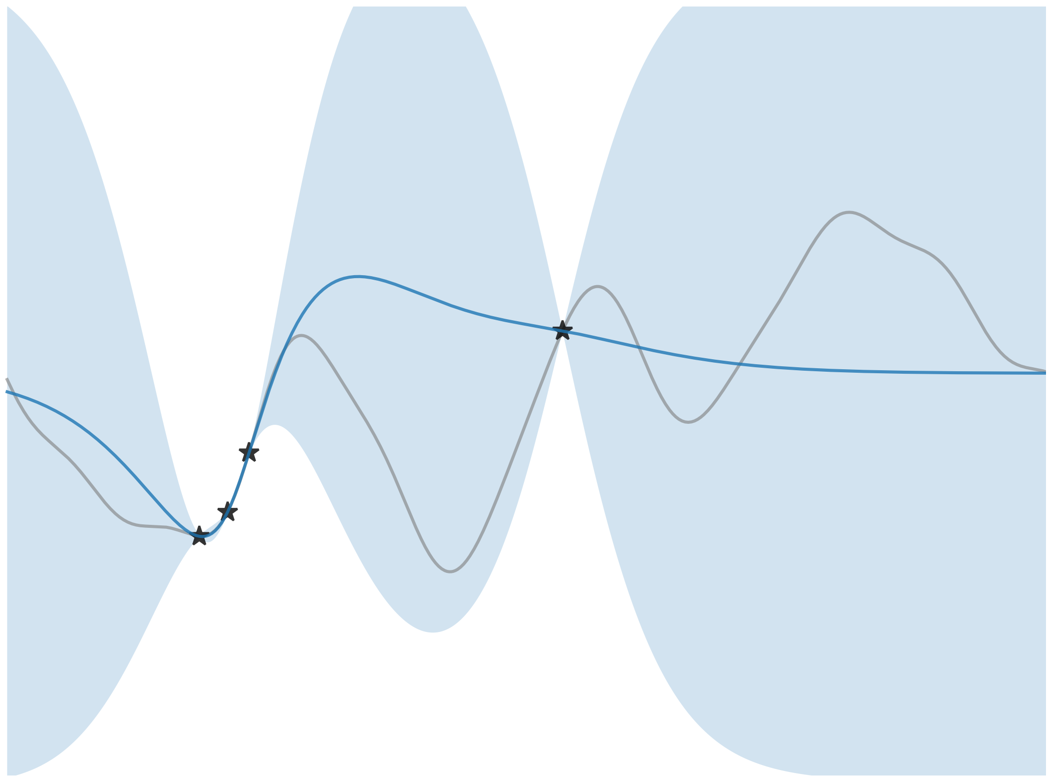


Continuous search domain

Properties of Bayesian Optimization

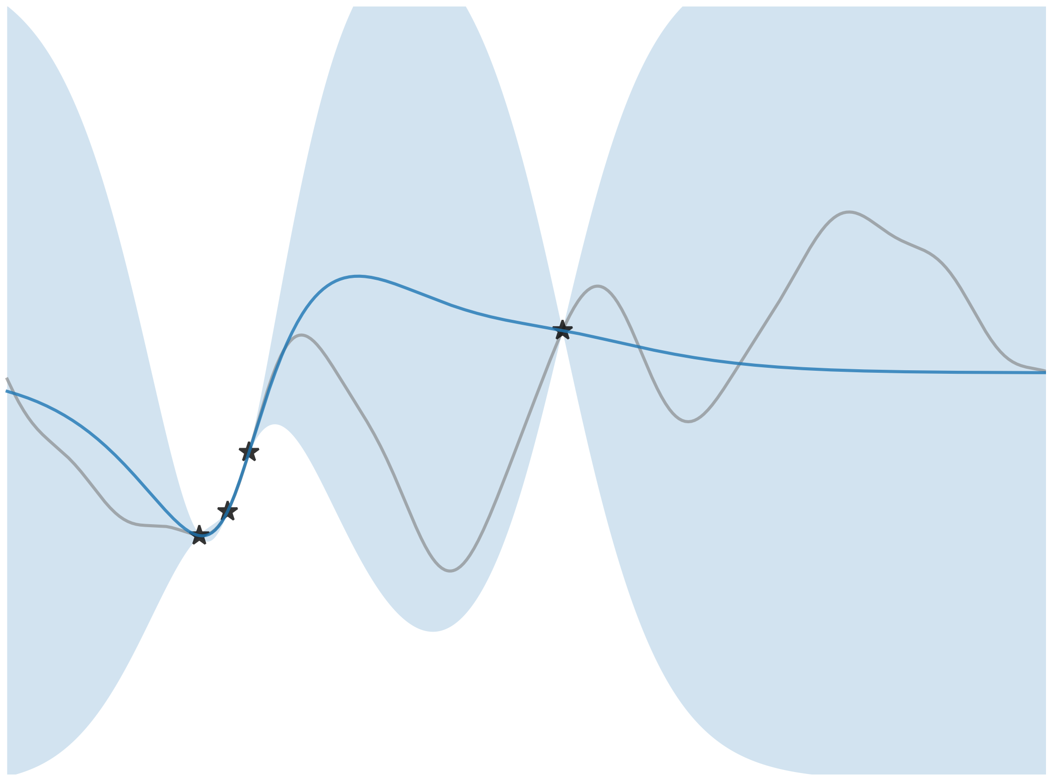


Properties of Bayesian Optimization



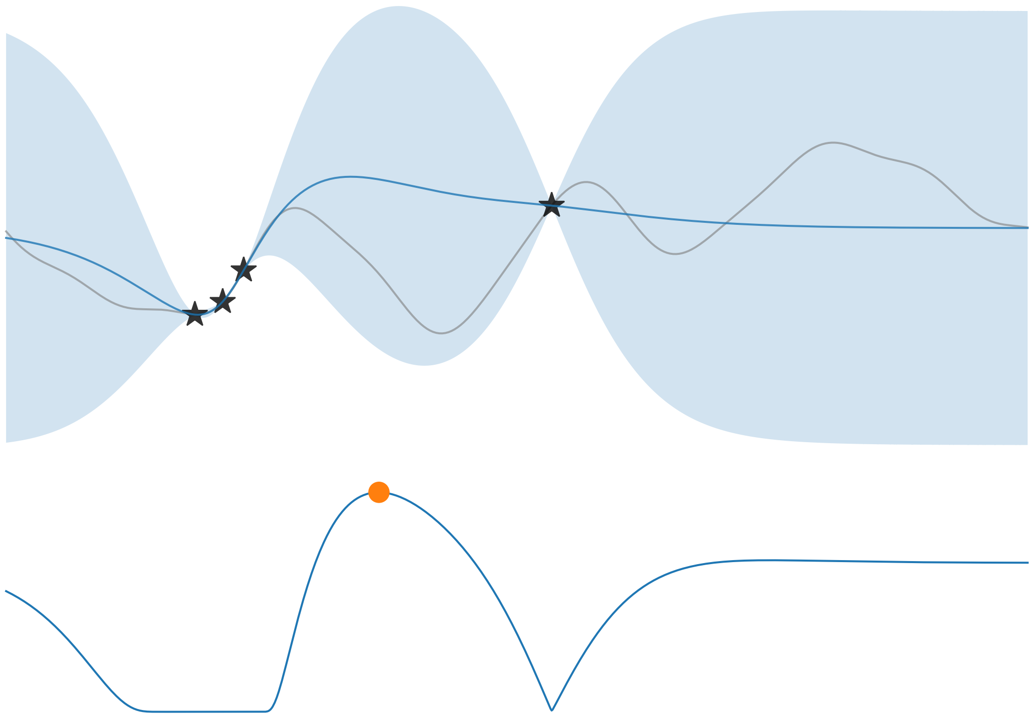
Correlation & continuity $\Longrightarrow$ Intractable MDP

Properties of Bayesian Optimization



Intractable MDP $\Longrightarrow$ Optimal policy unknown

Popular Policy: Expected Improvement

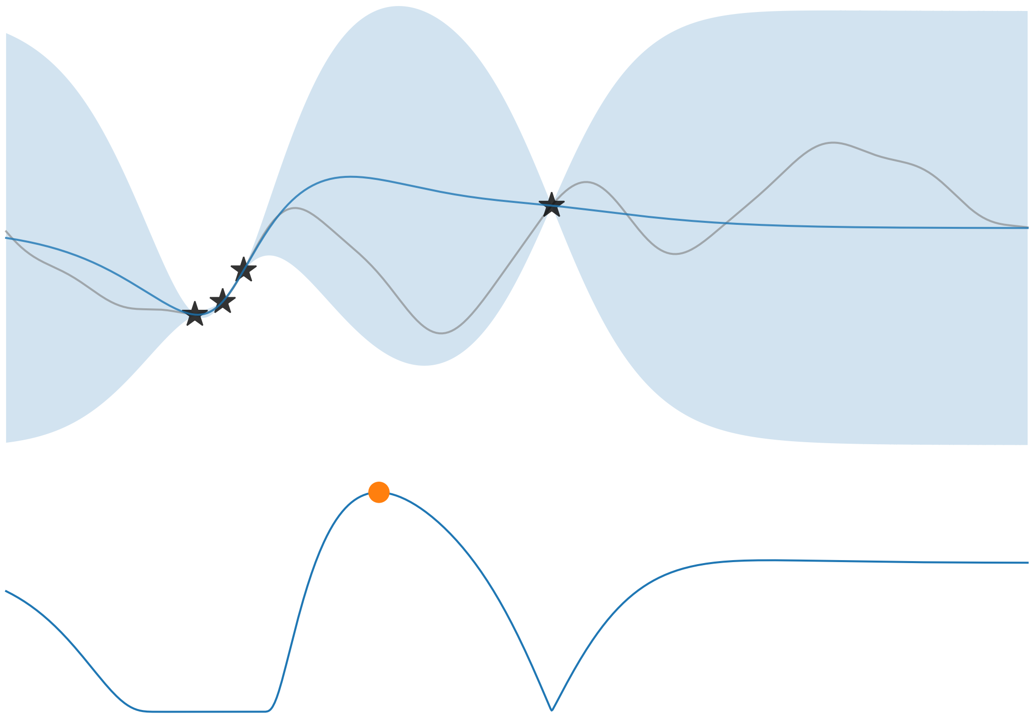


$$\text{EI}(x) = \mathbb{E}[\underbrace{\max(f(x) - \overbrace{y_{\text{best}}}^{\text{current best observed}}, 0)}_{\text{“improvement”}} \mid \overbrace{D}^{\text{data}}]$$

$$\max_x \text{EI}_{\underbrace{f|D}_{\text{posterior distribution}}}(x; y_{\text{best}})$$

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?

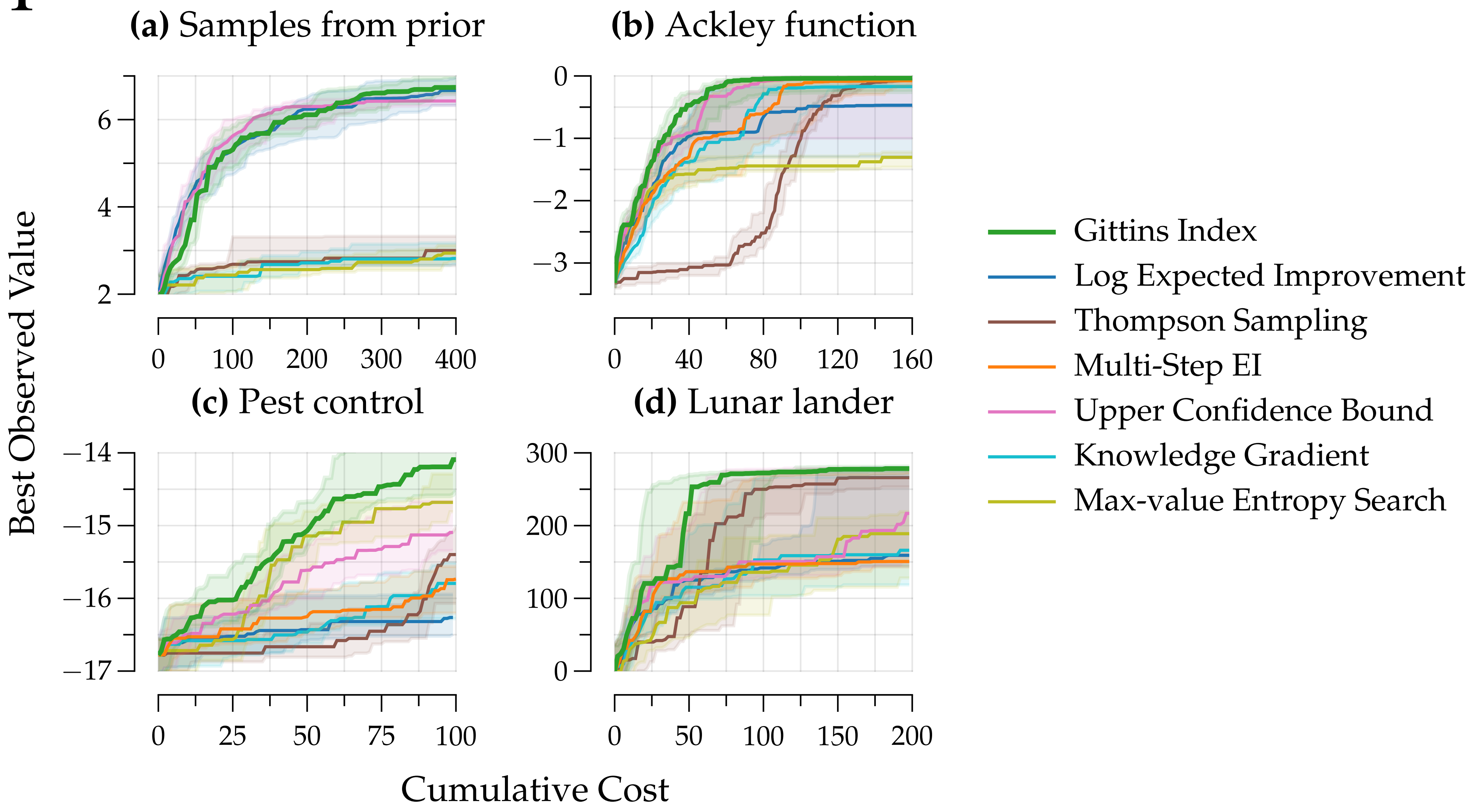
Approaches to Bayesian Optimization

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

Why another approach?

1. Competitive empirical performance

Experiment Results: Gittins Index vs Baselines



Experiment Setup: Objective Functions

Synthetic

Samples from prior

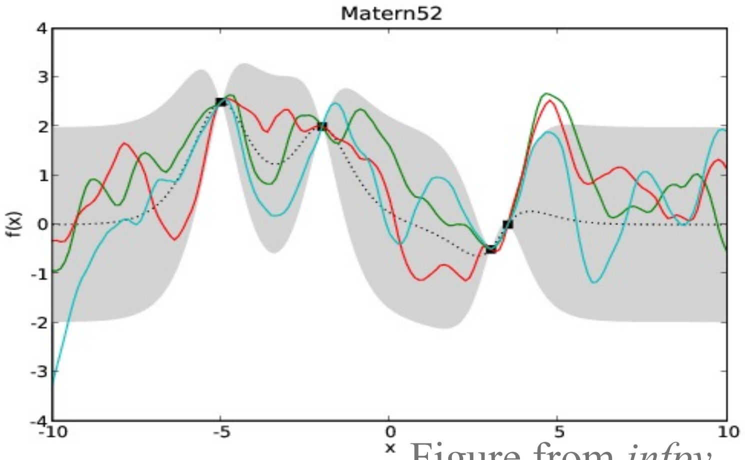


Figure from *infpy*

Ackley function

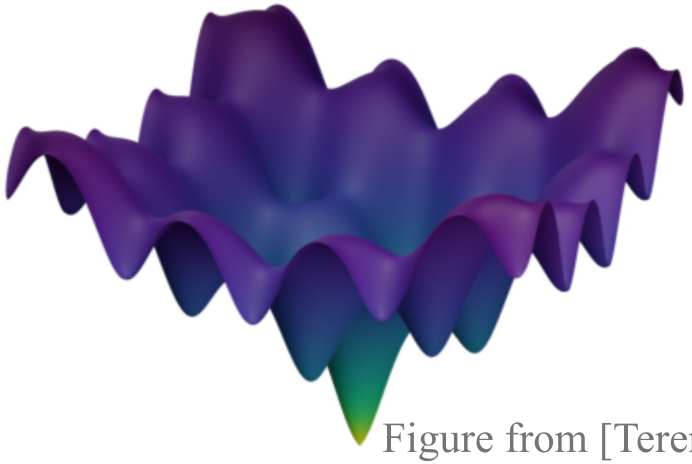


Figure from [Terenin'22]

Empirical

Pest Control

Figure from ChatGPT

Lunar Lander

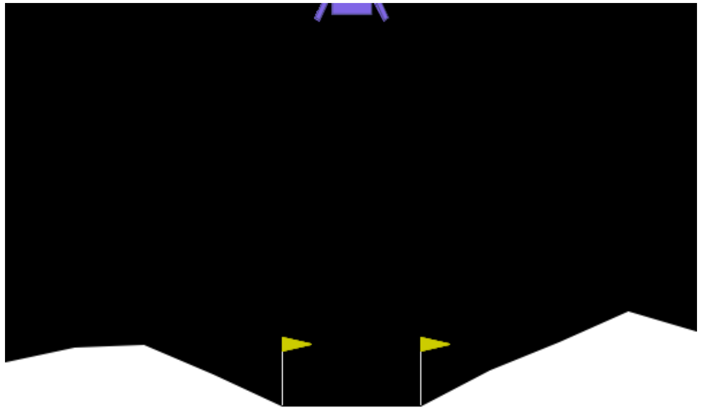


Figure from OpenAI Gym

Approaches to Bayesian Optimization

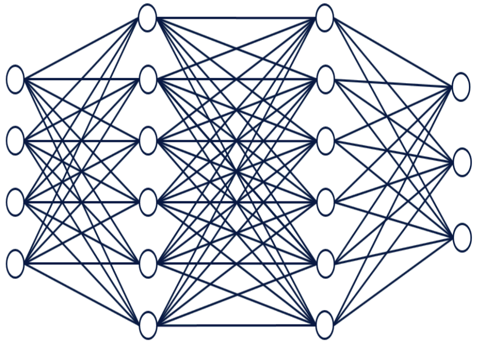
- Improvement-based
- Entropy-based
- Upper Confidence Bound
- Thompson Sampling
- Our work: Gittins Index

Why another approach?

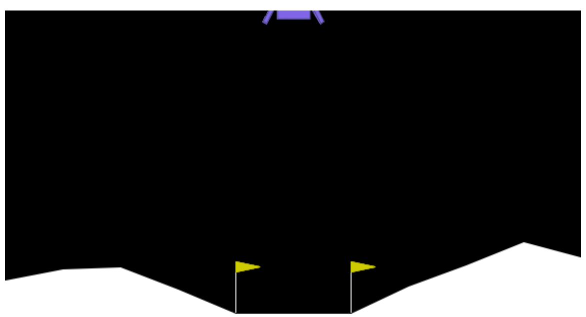
1. Competitive empirical performance
2. Applicable to varying evaluation costs

Examples: Varying Evaluation Time

Hyperparameter tuning:

Training hyperparameters →  → Accuracy

Control optimization:

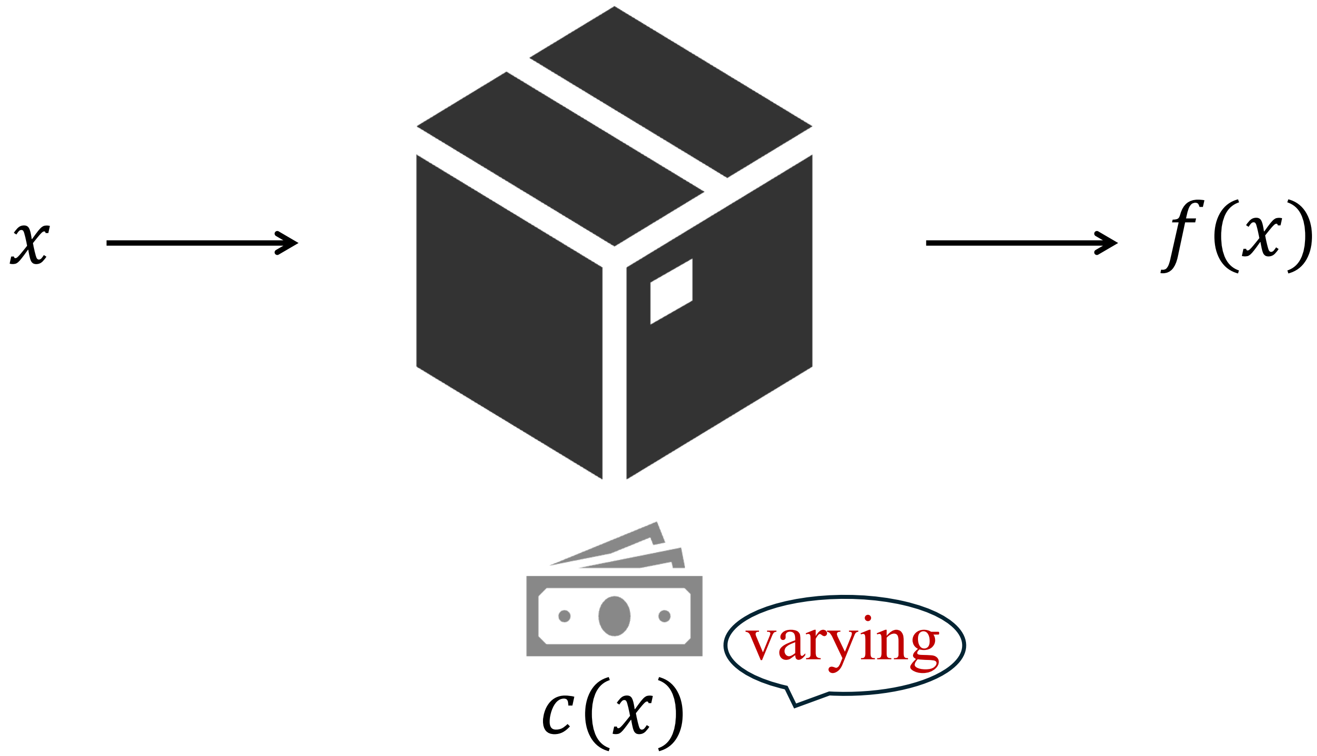
Control variables →  → Reward

Adaptive experimentation:

Decision variables → 

→ Revenue

Varying Evaluation Costs



Cost-aware Bayesian Optimization

[Lee, Perrone, Archambeau, Seeger'21]

$x \longrightarrow$

$\longrightarrow f(x)$

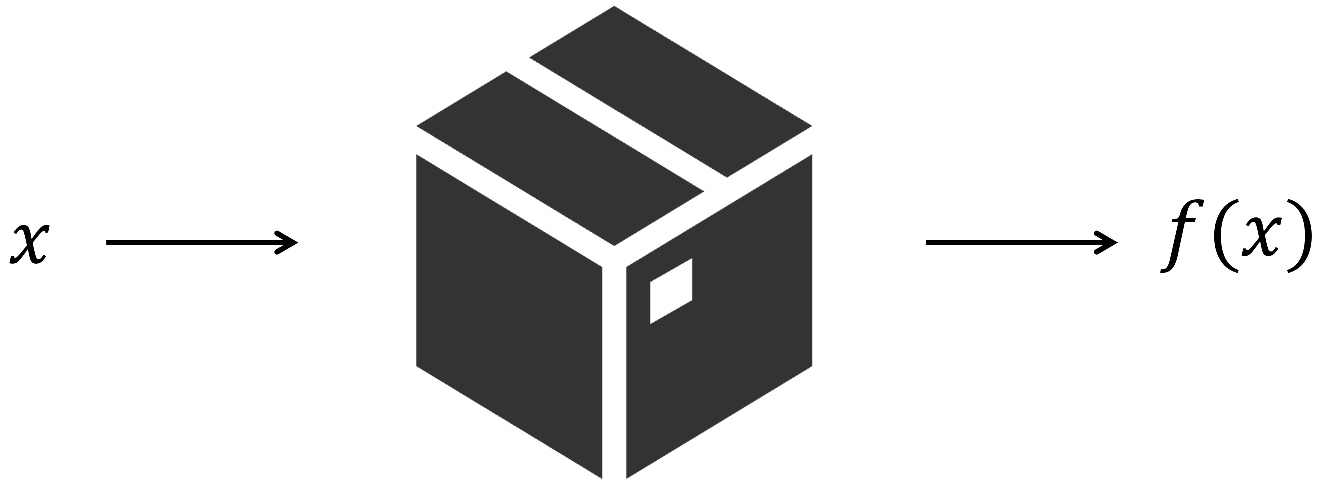
$$\text{Goal: } \sup_{\text{policy}} \mathbb{E} \max_{t=1,2,\dots,T} f(x_t)$$

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

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

Cost-aware Bayesian Optimization

[Lee, Perrone, Archambeau, Seeger'21]
[Astudillo, Jiang, Balandat, Bakshy, Frazier'21]



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

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

Our work studies expected budget constraint

Cost-aware Bayesian Optimization

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

Cost-aware Bayesian Optimization

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

Cost-aware Bayesian Optimization

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

Cost-aware Bayesian Optimization

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

Why divide?

Cost-aware Bayesian Optimization

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

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

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
	Upper Confidence Bound	?
	Thompson Sampling	?
	⋮	⋮

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	Budgeted Multi-step Lookahead EI
	Upper Confidence Bound	?
	Thompson Sampling	?
	⋮	⋮

New design principle: Gittins Index

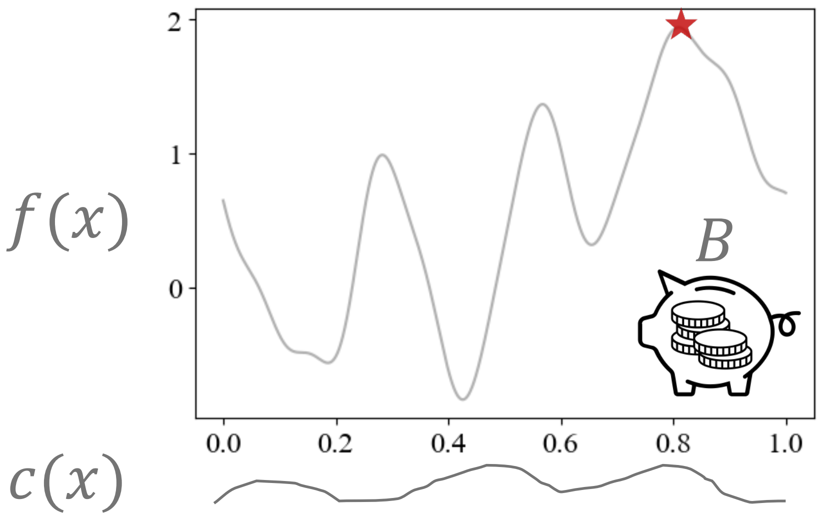
Cost-aware Bayesian Optimization

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

New design principle: Gittins Index

inherently cost-aware

Cost-aware Bayesian Optimization



Continuous

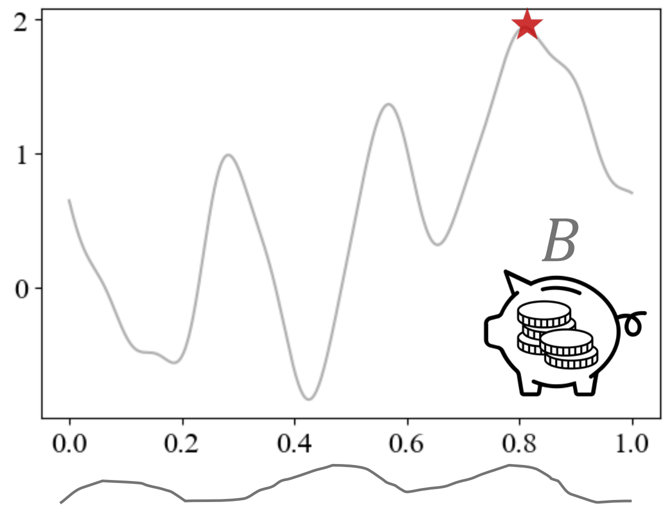
Correlated

Cost-aware Bayesian Optimization

Pandora's Box

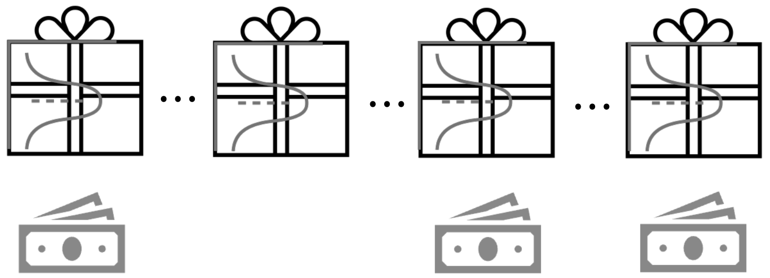
[Weitzman'79]

$f(x)$



$c(x)$

$\Longrightarrow$



$f(x)$

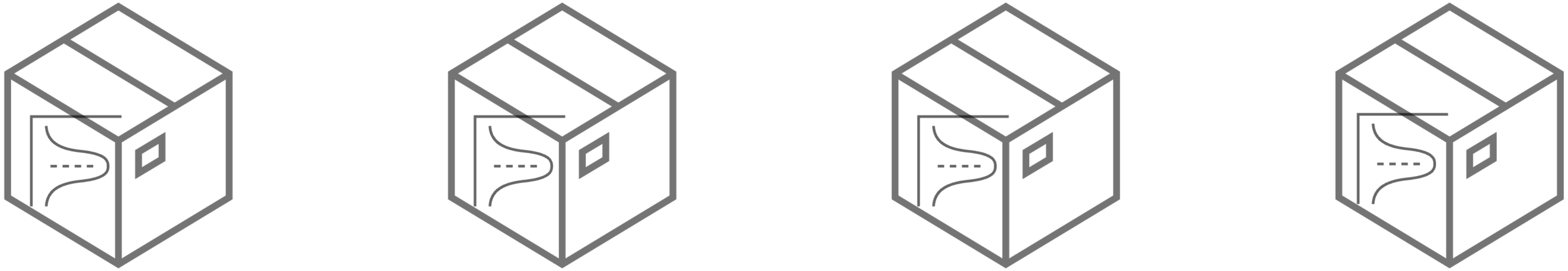
$c(x)$

Continuous $\Longrightarrow$ Discrete

Correlated $\Longrightarrow$ Independent

Pandora’s Box

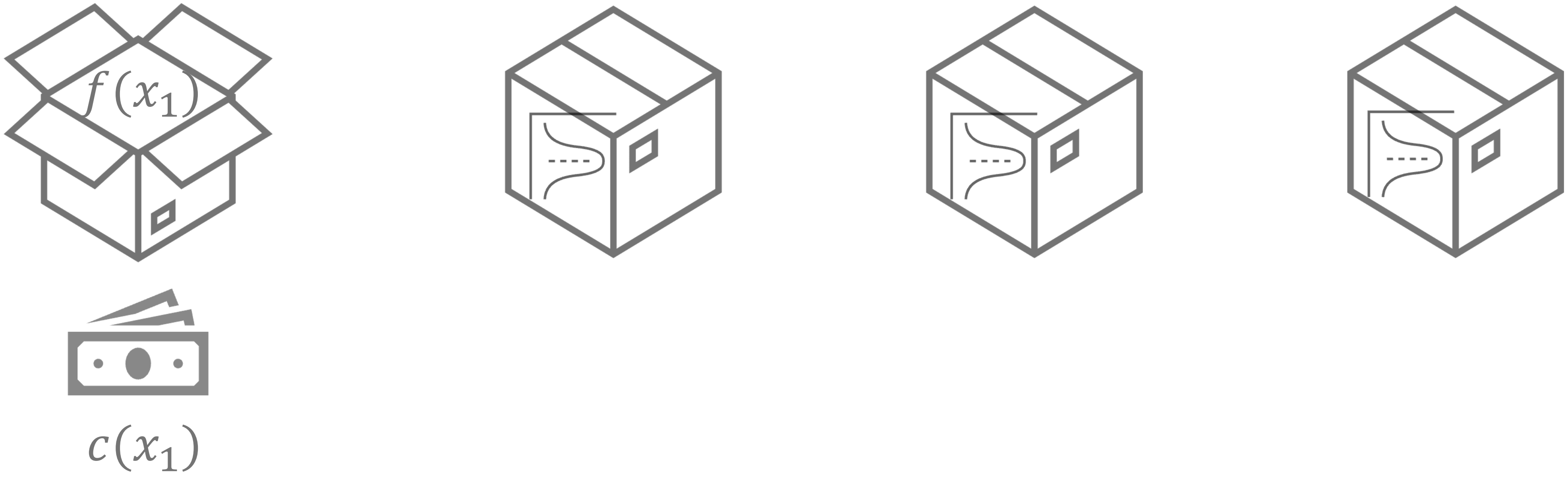
$t = 0$



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

Pandora’s Box

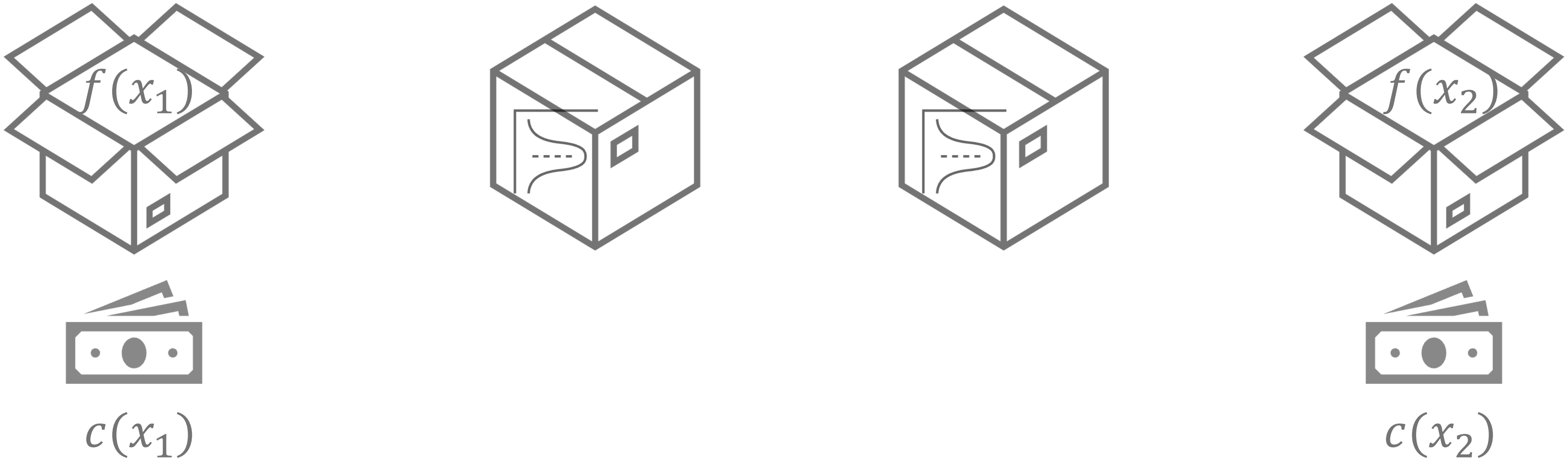
$t = 1$



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

Pandora's Box

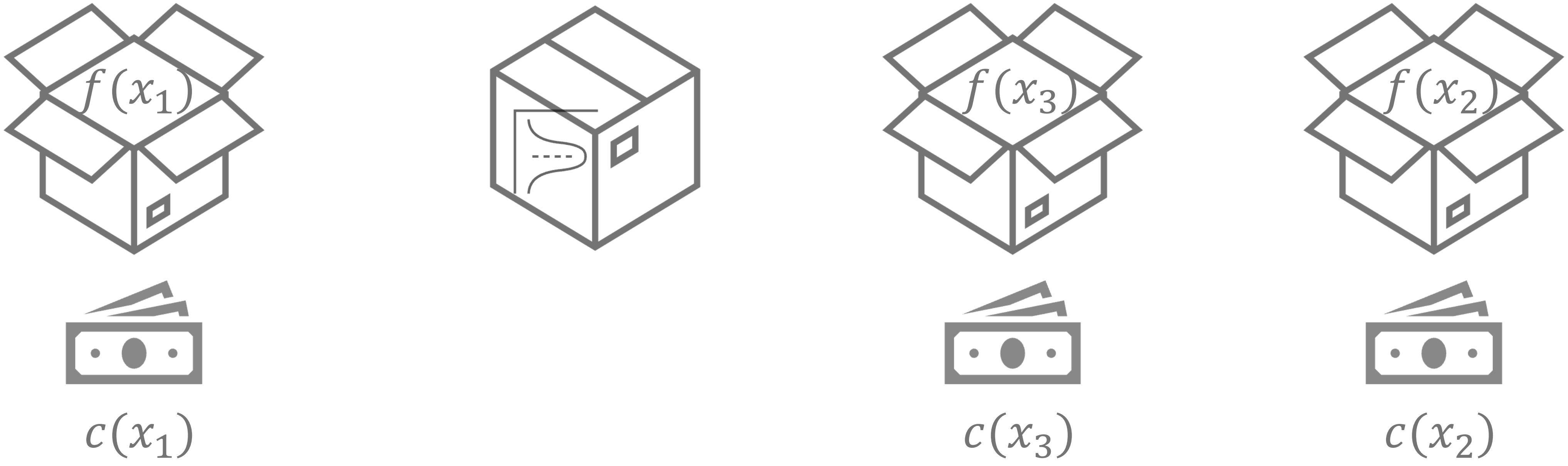
$t = 2$



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

Pandora's Box

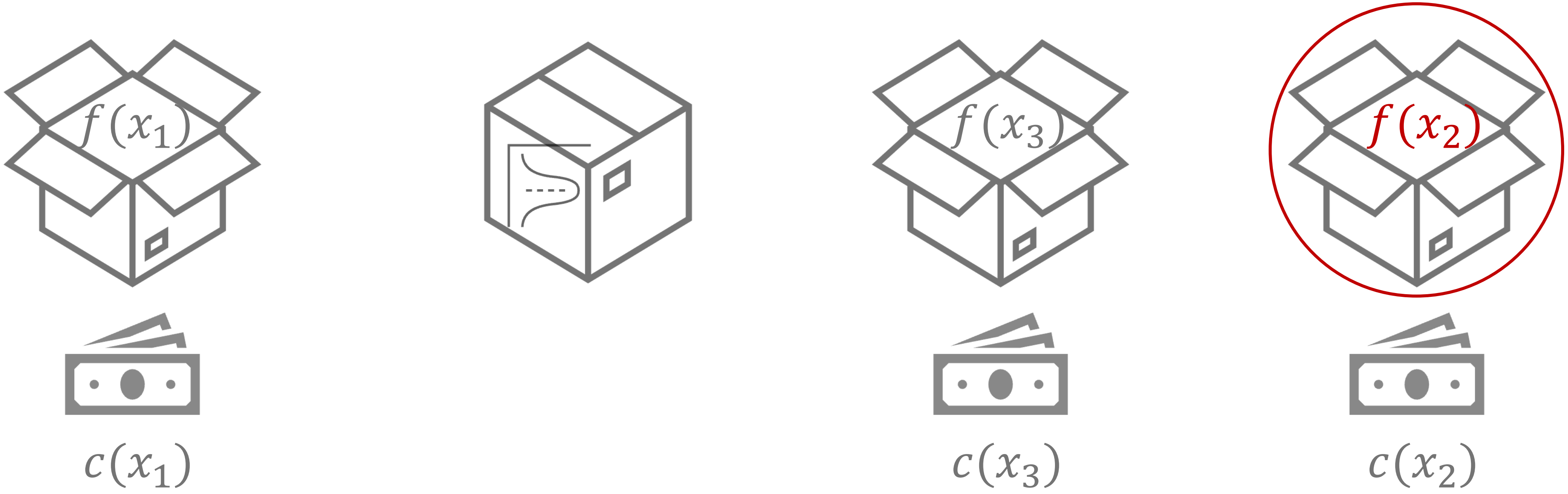
$t = 3$



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

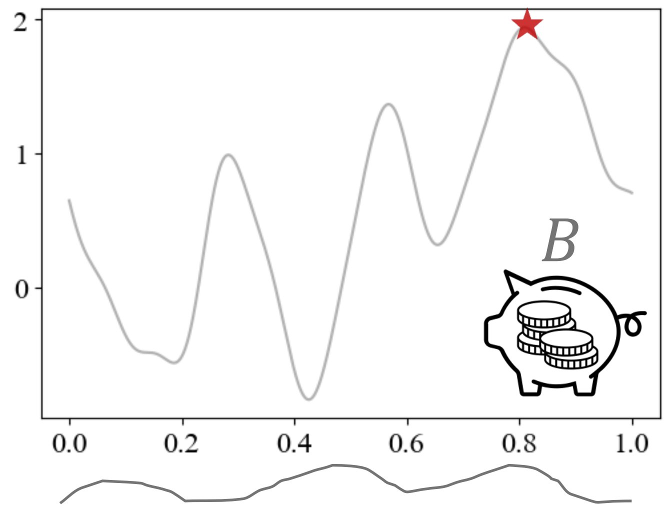
Pandora's Box

$t = T$, 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



Continuous

Correlated

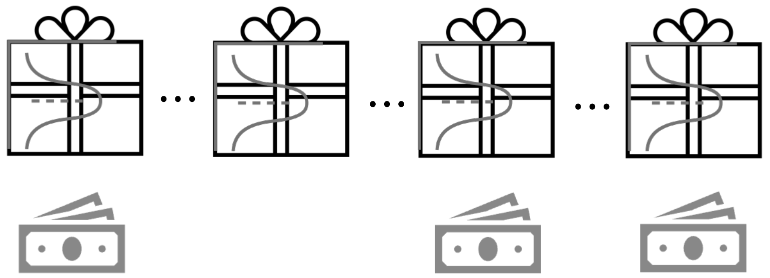
Expected-budget-constrained

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

$$\text{s.t. } \mathbb{E} \sum_{t=1}^{T} c(x_t) \le B$$

Pandora's Box

[Weitzman'79]



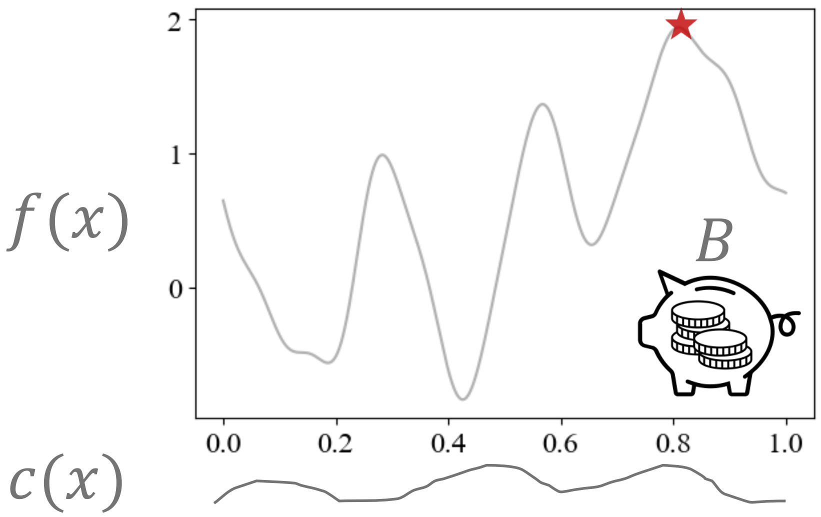
Discrete

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



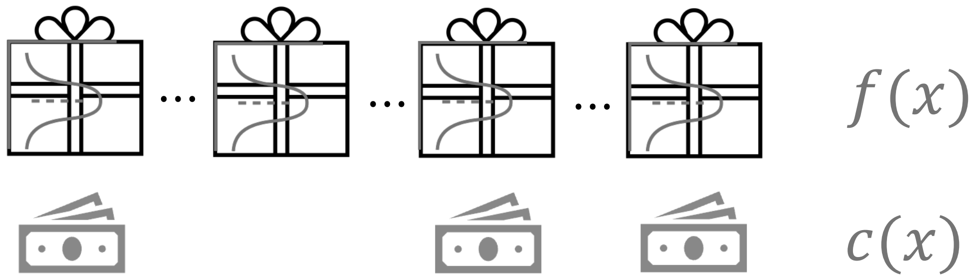
Continuous

Correlated

Expected-budget-constrained

Pandora's Box

[Weitzman'79]



Discrete

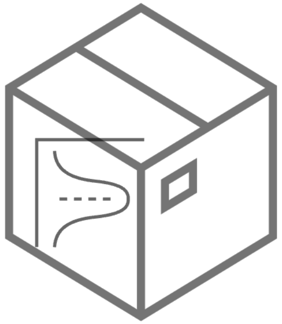
Independent

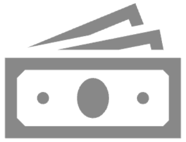
Cost-per-sample

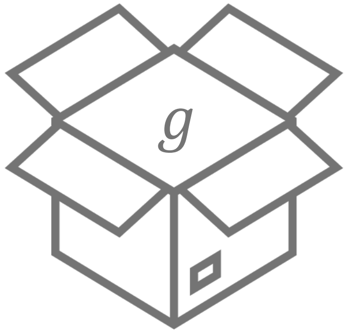
Optimal policy: Gittins index

Optimal Policy: Gittins Index

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

$f(x)$ 

$c(x)$ 



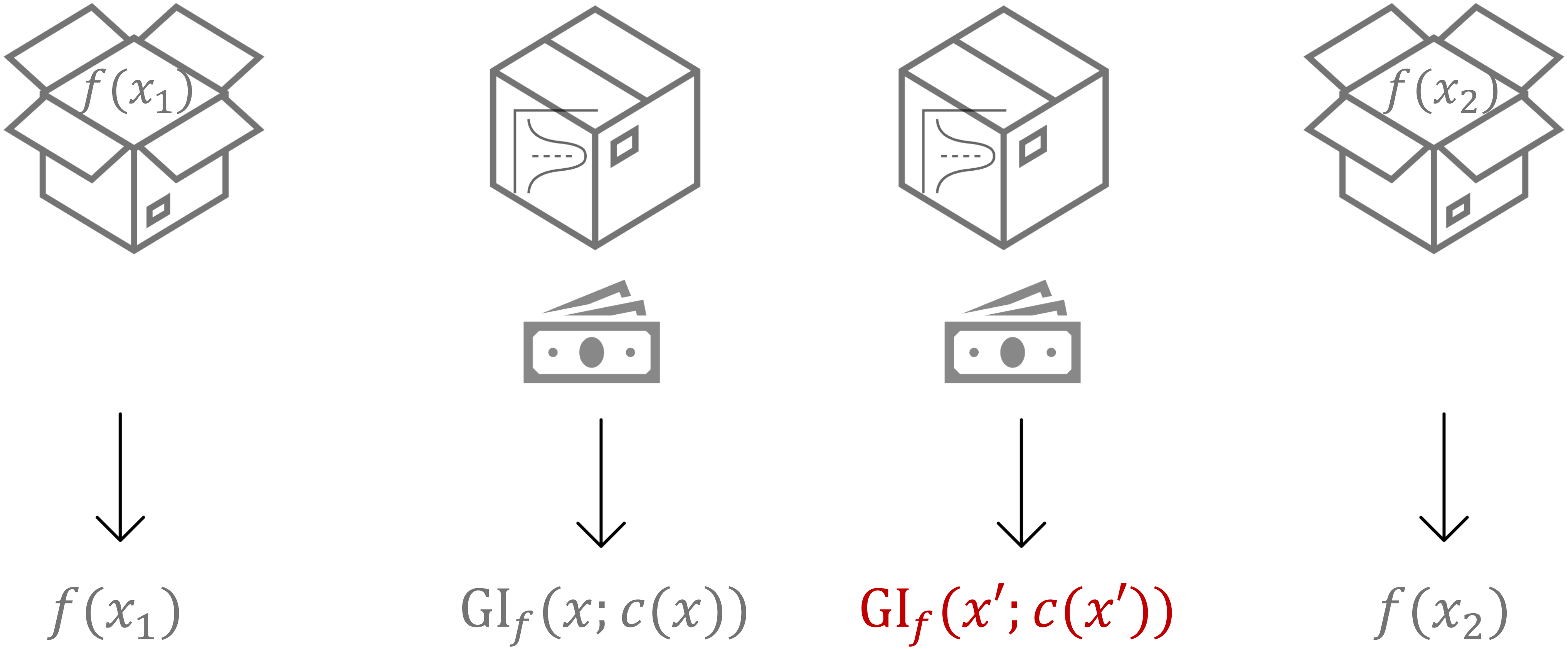
$$\downarrow \qquad\qquad \downarrow$$

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

g

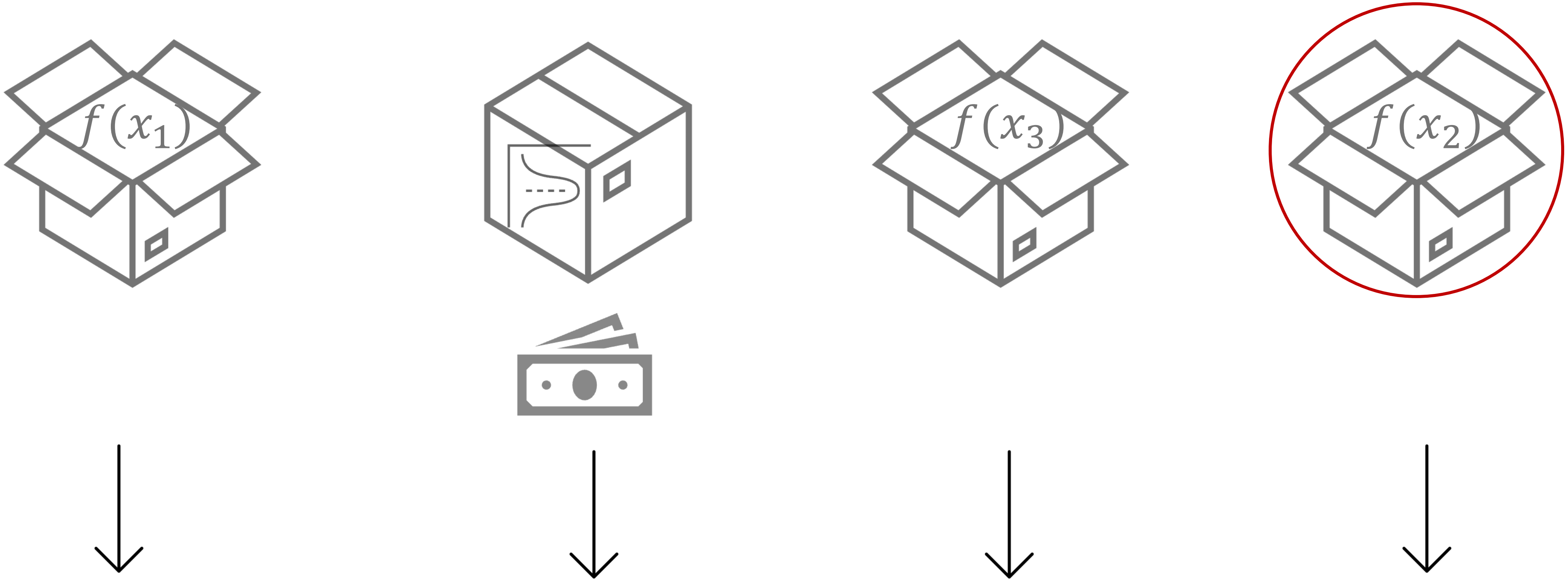
Optimal Policy: Gittins Index

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



Optimal Policy: Gittins Index

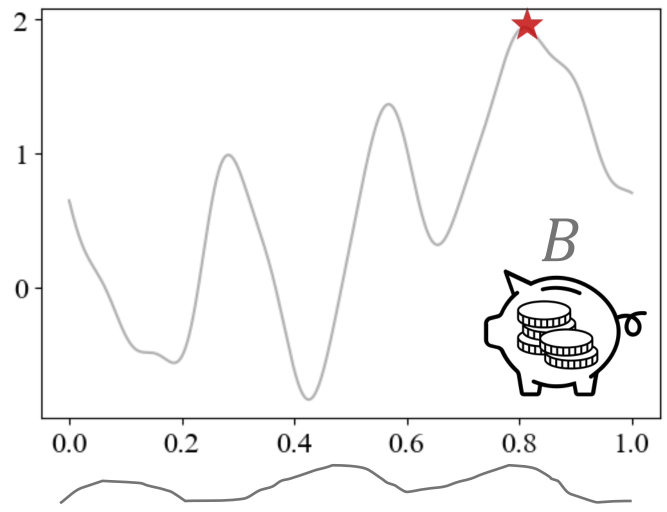
Step 2’: Select the box with highest index if it is opened and stop



$f(x_1)$ $\mathrm{GI}_f(x; c(x))$ $f(x_3)$ $f(x_2)$

Cost-aware Bayesian Optimization

$f(x)$



$c(x)$

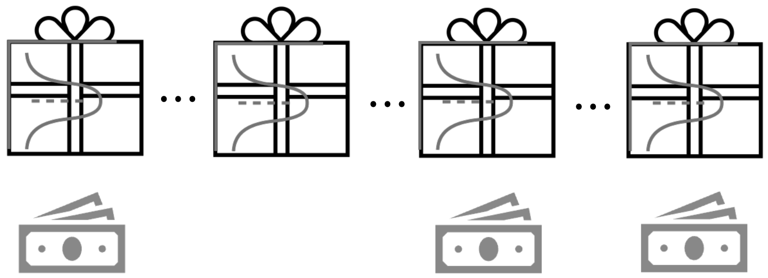
Continuous

Correlated

Expected-budget-constrained

Pandora's Box

[Weitzman'79]



$f(x)$

$c(x)$

Discrete

Independent

Cost-per-sample

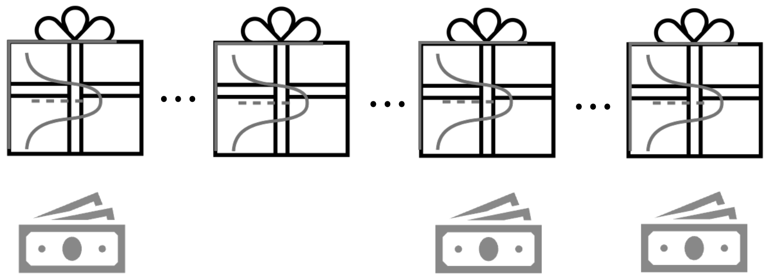
How to translate?

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

Pandora's Box

[Weitzman'79]

$\Longleftarrow$



$f(x)$

$c(x)$

How to convert?

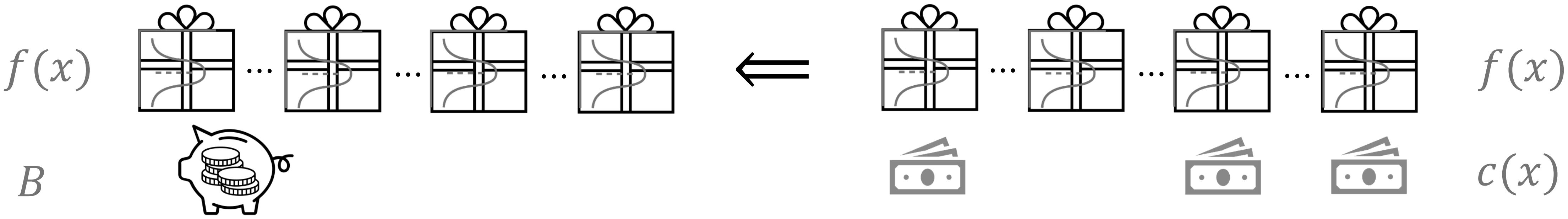
Expected-budget-constrained $\Longleftarrow$ Cost-per-sample

Budgeted Pandora's Box

[Aminian, Manshadi, Niazadeh'24]

Pandora's Box

[Weitzman'79]



Expected-budget-constrained

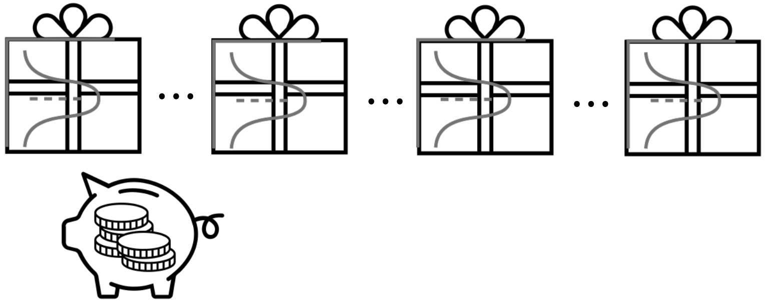
Cost-per-sample

Budgeted Pandora's Box

[Aminian, Manshadi, Niazadeh'24]

$f(x)$

B



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

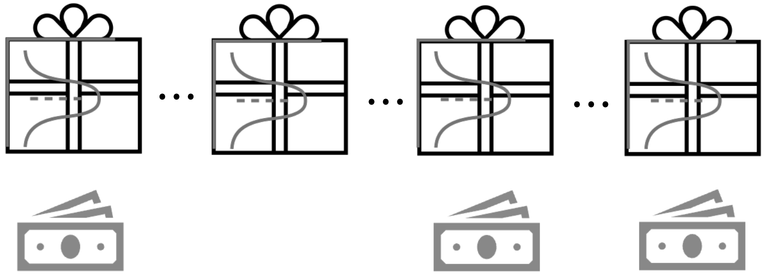
$$\textbf{s.t. } \mathbb{E} \textstyle\sum_{t=1}^{T} c(x_t) \leq B$$

Expected-budget-constrained

$\Leftarrow$

Pandora's Box

[Weitzman'79]



$f(x)$

$c(x)$

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

Cost-per-sample

Budgeted Pandora's Box

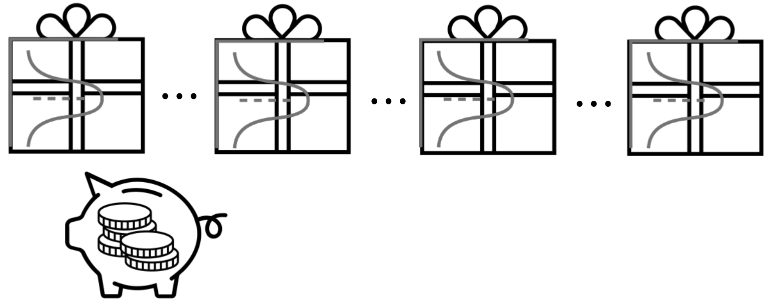
[Aminian, Manshadi, Niazadeh'24]

Pandora's Box

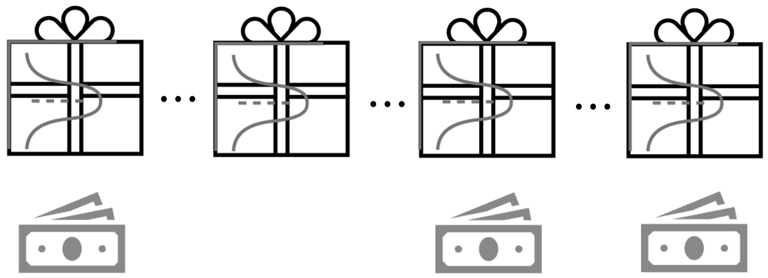
[Weitzman'79]

$f(x)$

B



$\Longleftarrow$



$f(x)$

$c(x)$

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

scaling factor

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

Lagrangian relaxation

Expected-budget-constrained $\Longleftrightarrow$ Cost-per-sample

Budgeted Pandora's Box

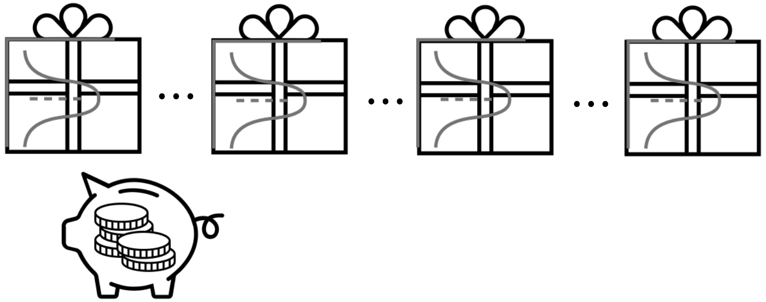
[Aminian, Manshadi, Niazadeh'24]

Pandora's Box

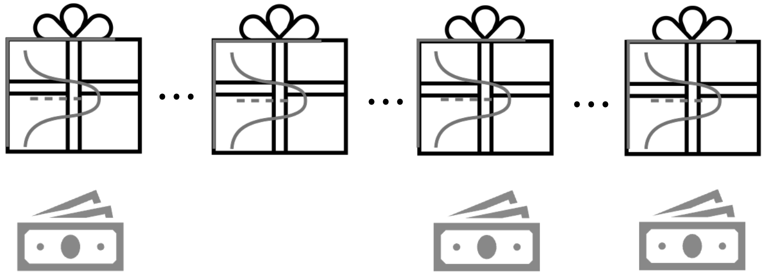
[Weitzman'79]

$f(x)$

B



$\Longleftarrow$



$f(x)$

$c(x)$

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

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

Expected-budget-constrained $\Longleftrightarrow$ Cost-per-sample

How to translate?

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

Budgeted Pandora's Box

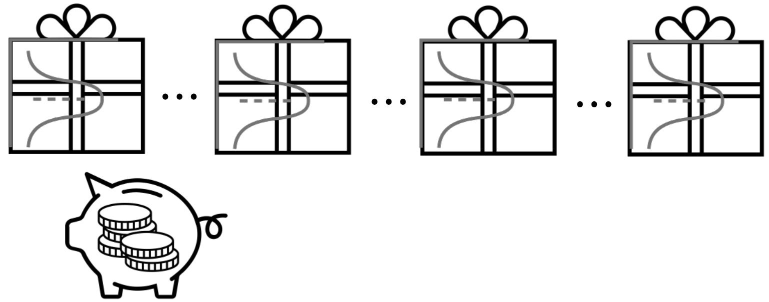
[Aminian, Manshadi, Niazadeh'24]

Pandora's Box

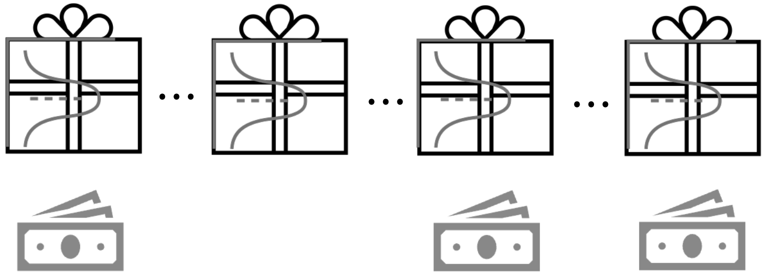
[Weitzman'79]

$f(x)$

B



$\Longleftarrow$



$f(x)$

$c(x)$

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

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

Expected-budget-constrained $\Longleftrightarrow$ Cost-per-sample

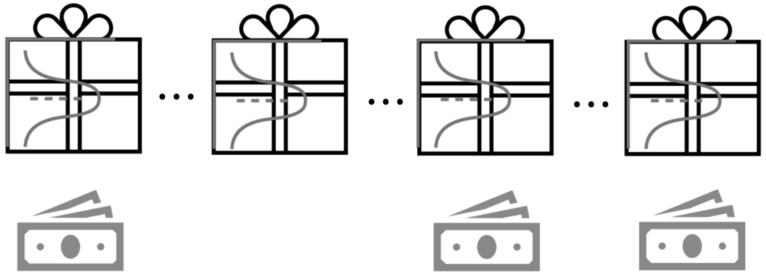
Optimal policy: $\mathrm{GI}_f(x; \lambda_B c(x))$ $\overset{\text{scale costs}}{\Longleftarrow}$ Optimal policy: $\mathrm{GI}_f(x; c(x))$

Budgeted Pandora's Box

[Aminian, Manshadi, Niazadeh'24]

Pandora's Box

[Weitzman'79]

$f(x)$  $\Longleftarrow$  $f(x)$

B $\qquad\qquad$ $c(x)$

Our work: extension to general reward distribution

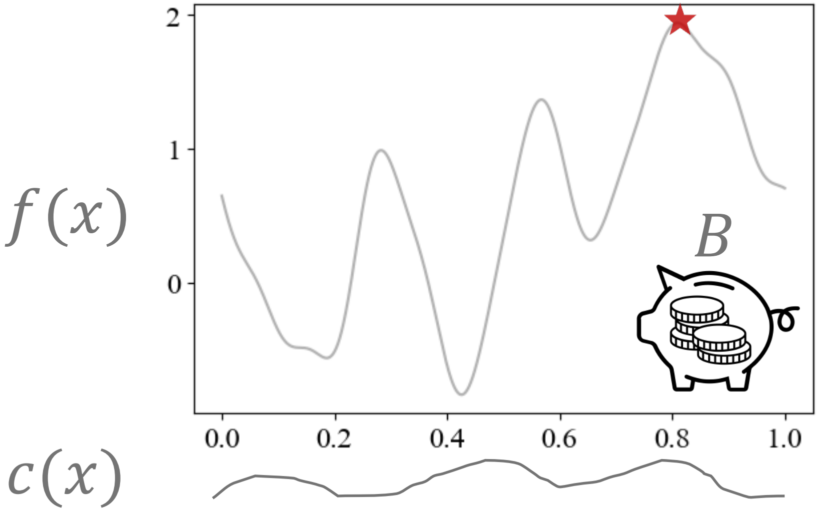
$$\sup_{\text{policy}} \mathbb{E} \max_{t=1,2,\ldots,T} f(x_t)$$
$$\textbf{s.t. } \mathbb{E} \textstyle\sum_{t=1}^{T} c(x_t) \leq B$$

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

Expected-budget-constrained $\Longleftrightarrow$ Cost-per-sample

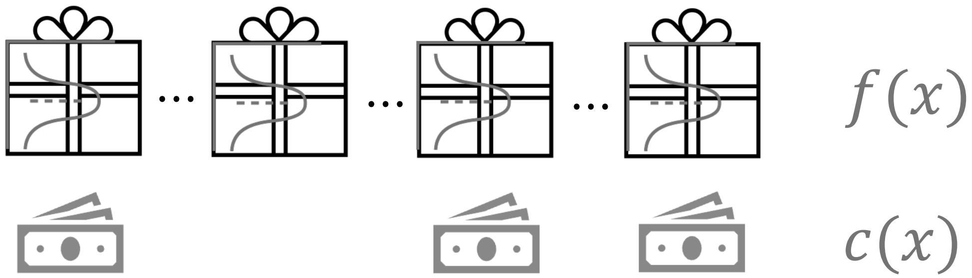
Optimal policy: $\mathrm{GI}_f(x; \lambda_B c(x))$ $\overset{\text{scale costs}}{\Longleftarrow}$ Optimal policy: $\mathrm{GI}_f(x; c(x))$

Cost-aware Bayesian Optimization



Pandora's Box

[Weitzman'79]



Continuous — Discrete

Correlated — Independent

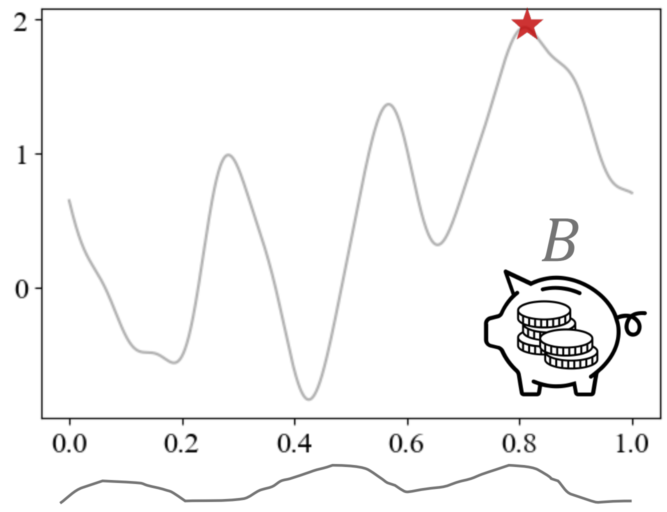
Expected-budget-constrained $\Longleftarrow$ Cost-per-sample ✓

How to translate?

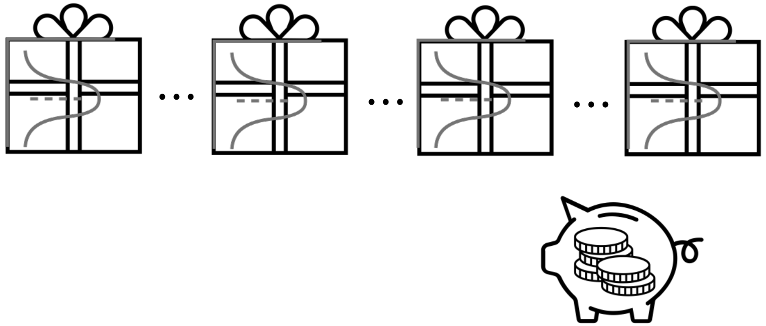
$\Longleftarrow$ Optimal policy: $\mathrm{GI}_f(x; c(x))$

Cost-aware Bayesian Optimization   Budgeted Pandora's Box

$f(x)$



$c(x)$



$f(x)$

B

Continuous

Correlated

Discrete

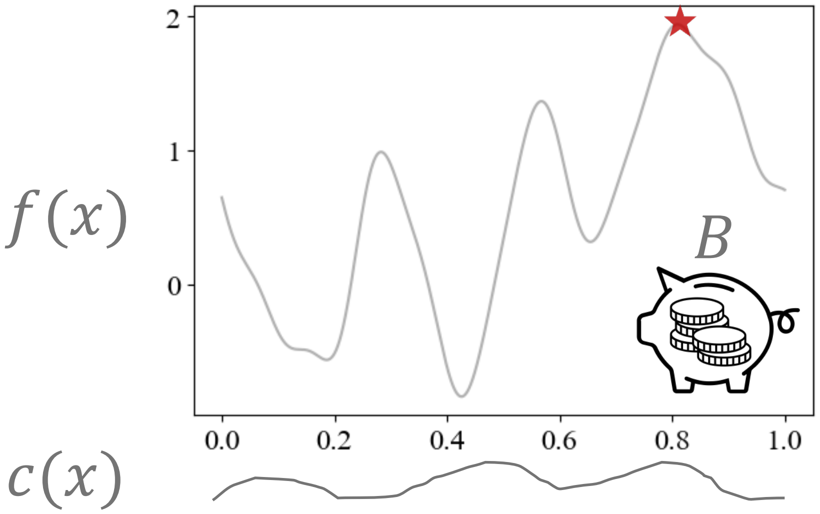
Independent

scaling factor

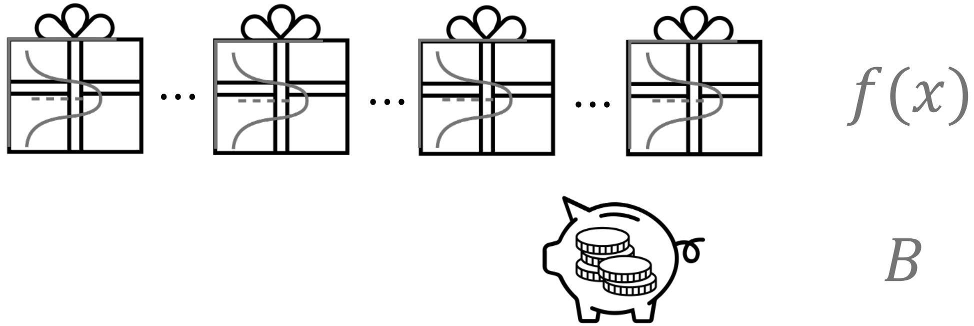
How to translate?

$\Longleftarrow$ Optimal policy: $\mathrm{GI}_f(x; \lambda c(x))$

Cost-aware Bayesian Optimization Budgeted Pandora's Box



$\Longleftarrow$



Continuous $\Longleftarrow$ Discrete

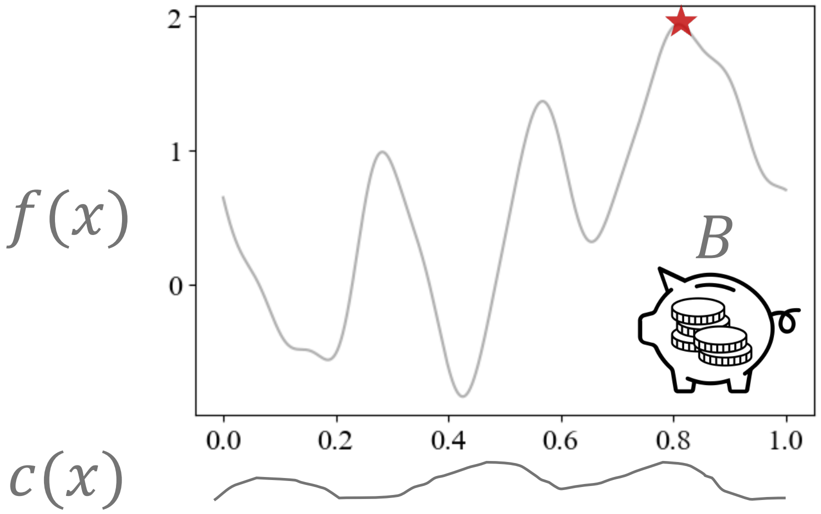
Correlated $\Longleftarrow$ Independent

scaling factor

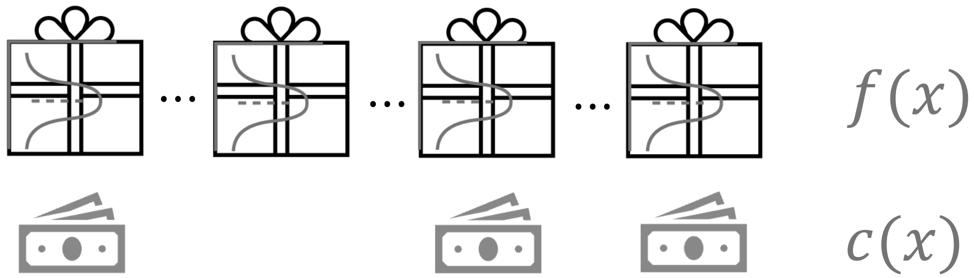
incorporate posterior

Our policy: $\mathrm{GI}_{f|D}(x;\lambda c(x))$ $\Longleftarrow$ Optimal policy: $\mathrm{GI}_f(x;\lambda c(x))$

Cost-aware Bayesian Optimization   Pandora's Box



$\Leftarrow$

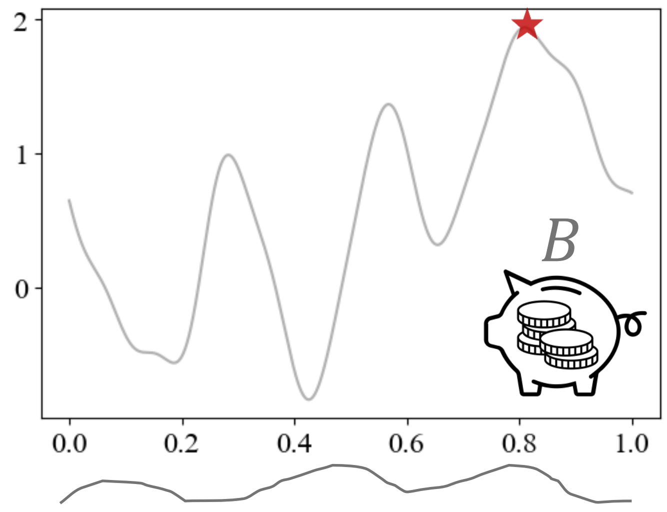


Continuous	$\Leftarrow$	Discrete	✓
Correlated	$\Leftarrow$	Independent	✓
Expected-budget-constrained	$\Leftarrow$	Cost-per-sample	✓

Our policy: $\mathrm{GI}_{f|D}(x; \lambda c(x))$ $\xLeftarrow[\text{scale costs}]{\text{incorporate posterior}}$ Optimal policy: $\mathrm{GI}_f(x; c(x))$

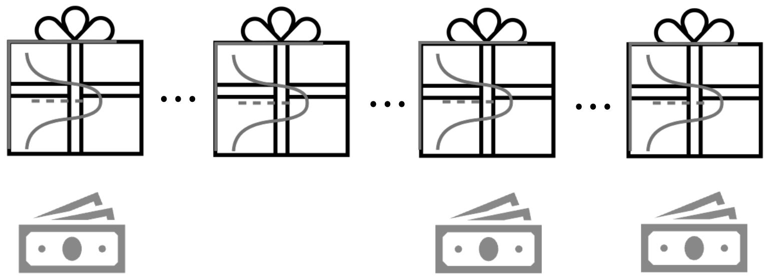
Cost-aware Bayesian Optimization $\quad$ Pandora's Box

$f(x)$



$c(x)$

$\Longleftarrow$



$f(x)$

$c(x)$

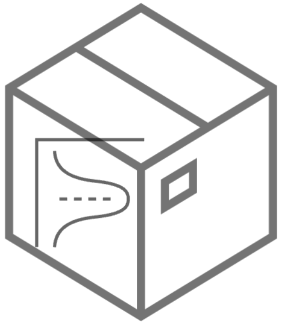
Cost-aware Bayesian Optimization	Pandora's Box
Continuous	Discrete
Correlated	Independent
Expected-budget-constrained	Cost-per-sample

Our policy: $\mathrm{GI}_{f|D}(x; \lambda c(x))$ $\underset{\text{scale costs}}{\overset{\text{incorporate posterior}}{\Longleftarrow}}$ Optimal policy: $\mathrm{GI}_f(x; c(x))$

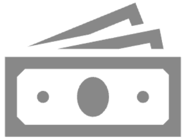
How to compute?

How to compute Gittins index?

$f(x)$

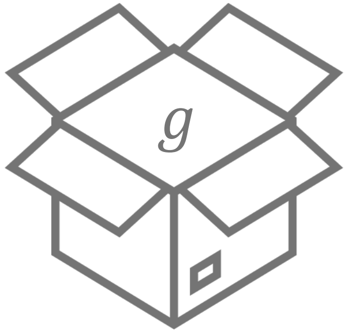


$c(x)$



$\downarrow$?

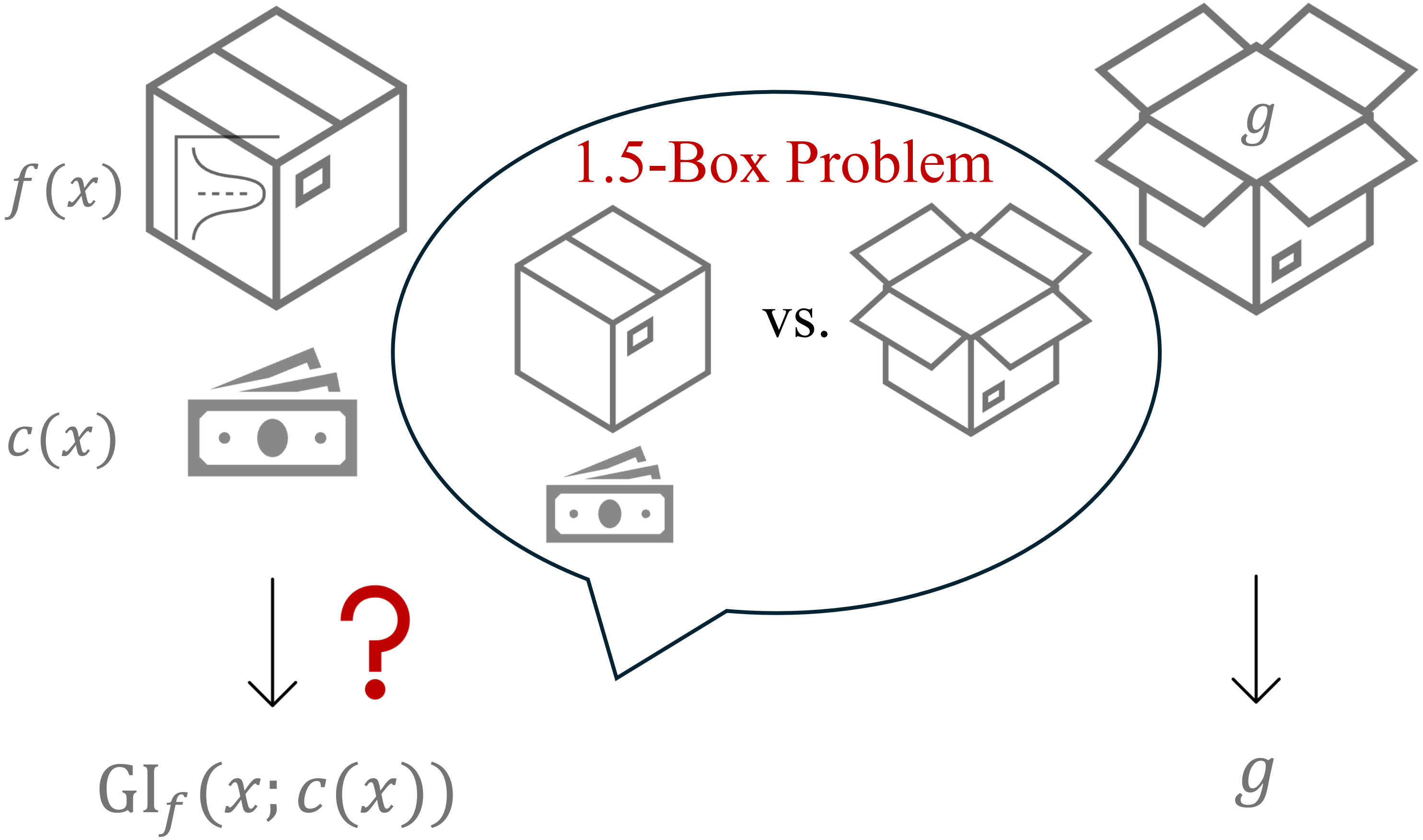
$\mathrm{GI}_f(x; c(x))$



$\downarrow$

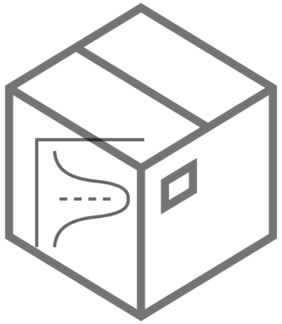
g

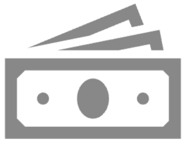
How to compute Gittins index?



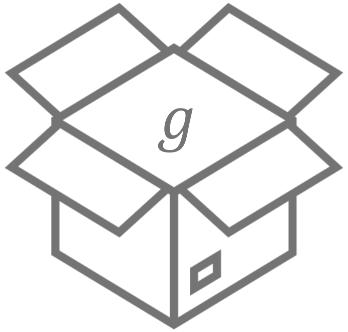
Whether to open a new box or take current best?

Gittins Index Computation: 1.5-Box Problem

$f(x)$ 

$c(x)$ 

vs.

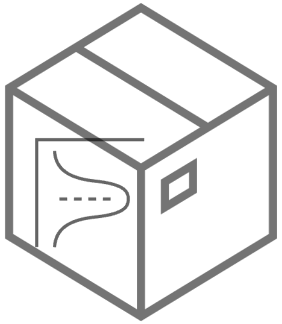


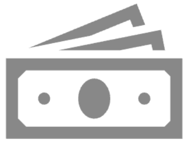
Open closed box

Take opened box

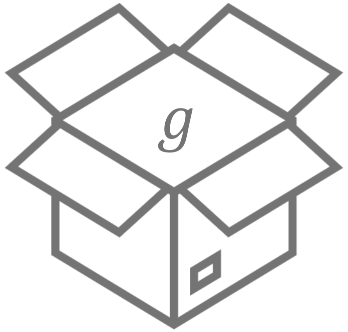
Whether to open a new box or take current best?

Gittins Index Computation: 1.5-Box Problem

$f(x)$ 

$c(x)$ 

vs.



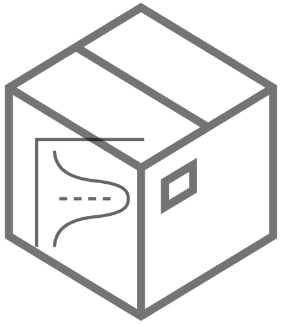
Open closed box — Take opened box — g

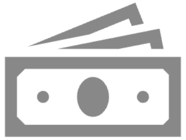
Profit:

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

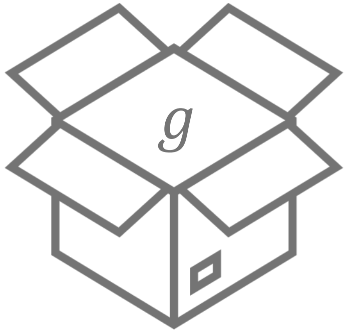
Profit:

Gittins Index Computation: 1.5-Box Problem

$f(x)$ 

$c(x)$ 

vs.



Open closed box | Take opened box → g

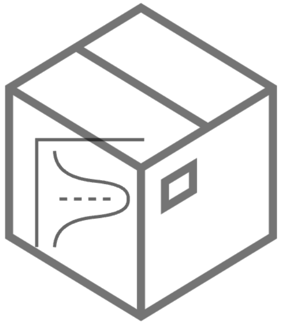
Profit: $\mathbb{E}[\underbrace{\max(f(x), g)}]-c(x)$

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

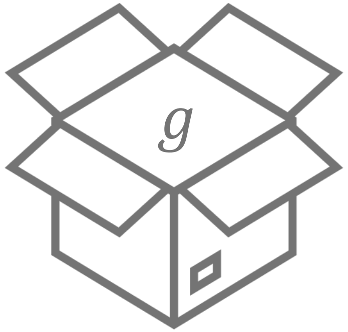
Profit: g

Gittins Index Computation: 1.5-Box Problem

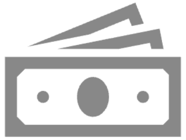
$f(x)$

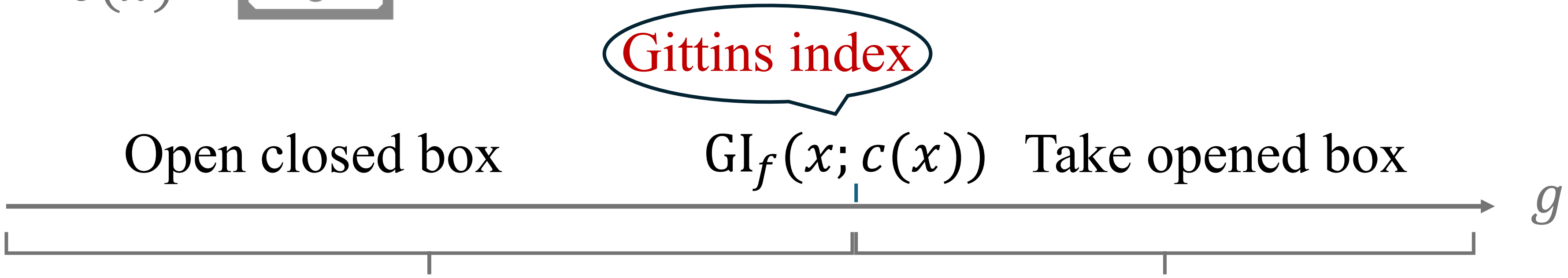


vs.



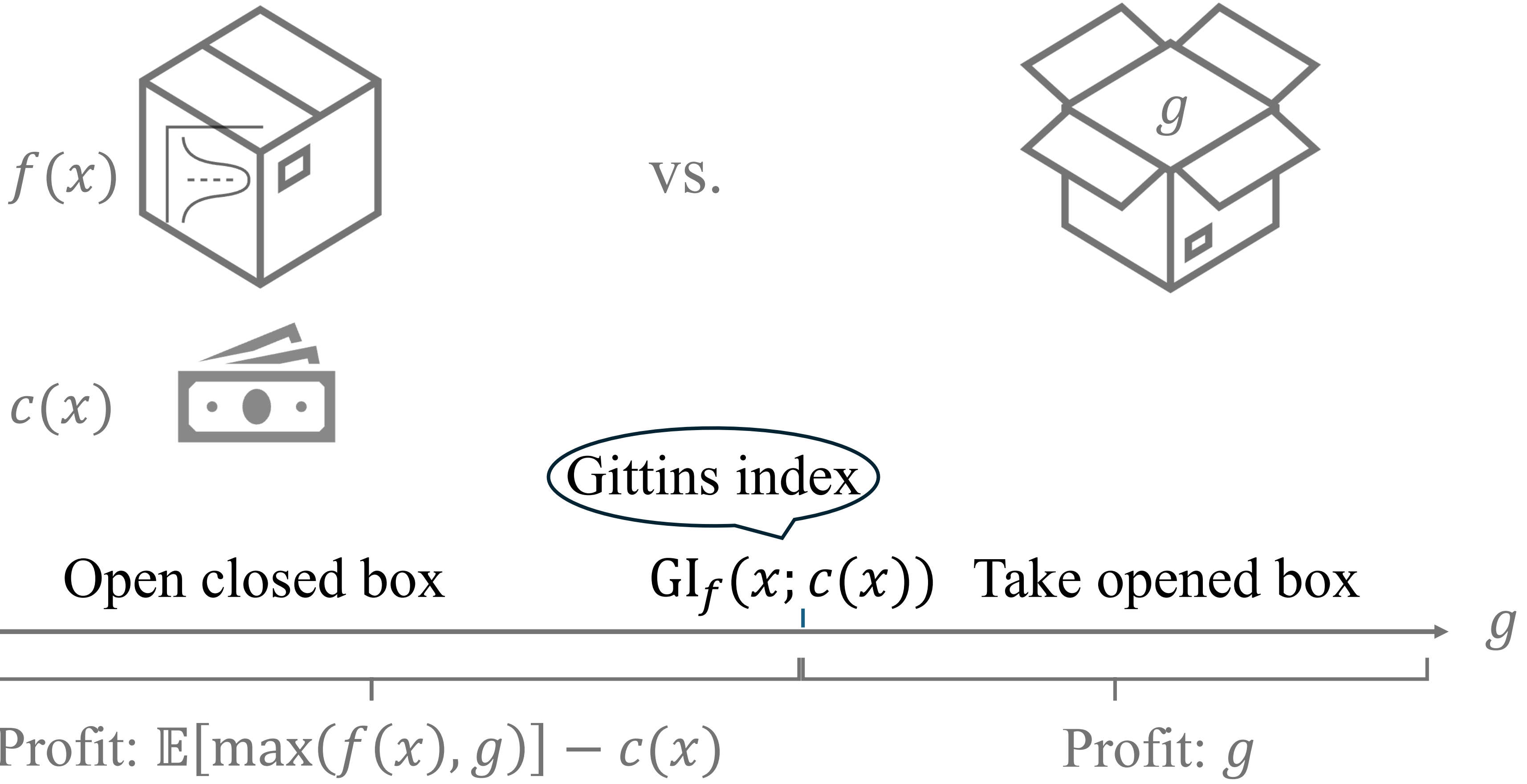
$c(x)$





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

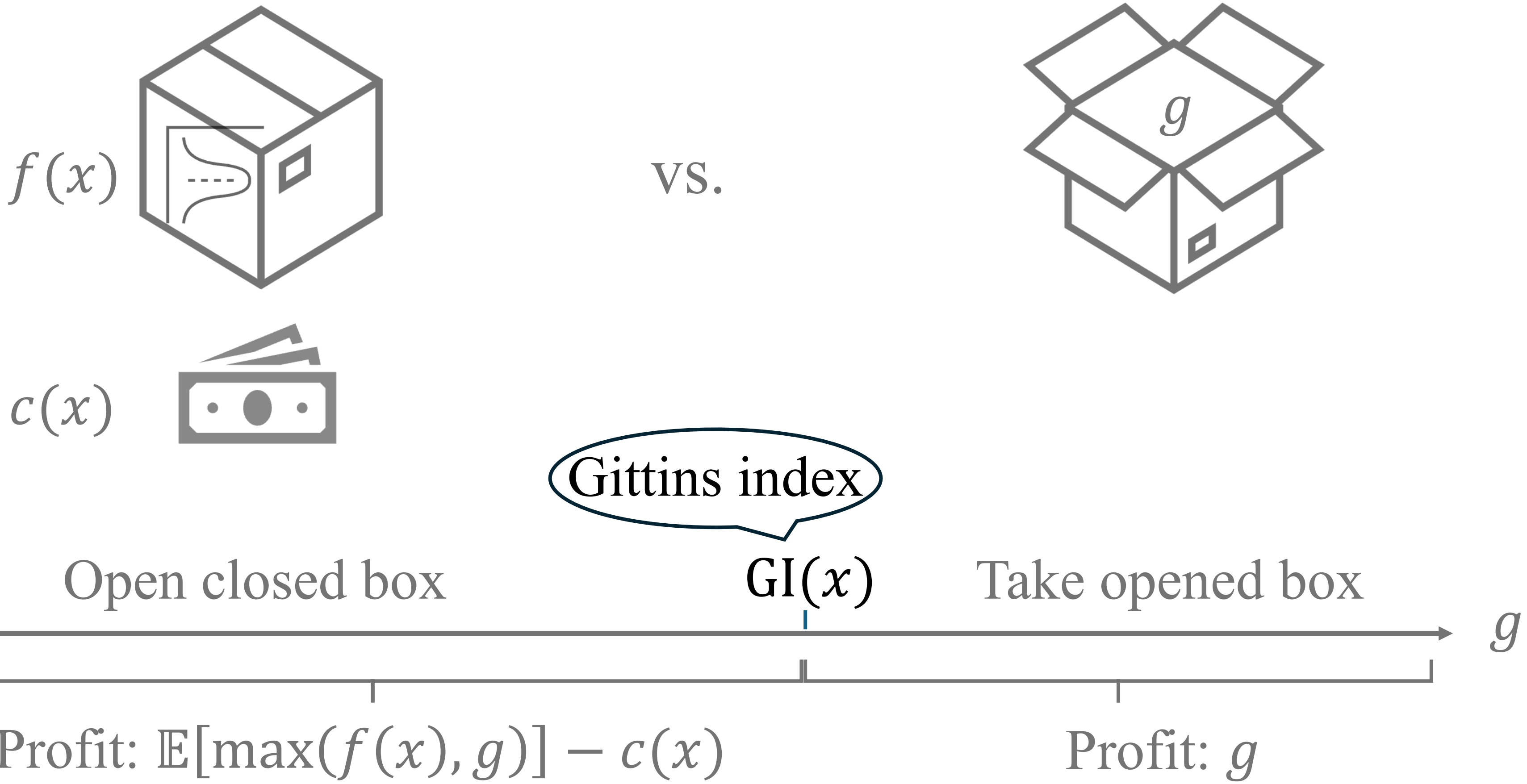
Gittins Index Computation: 1.5-Box Problem



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

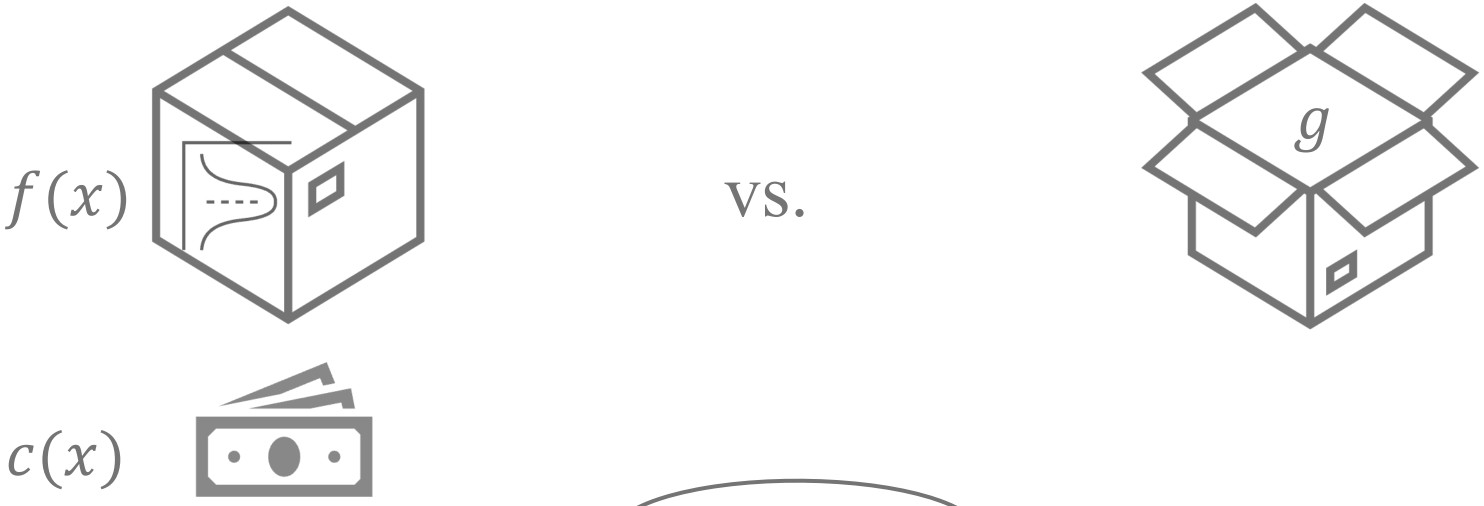
Larger the cost, smaller the Gittins index

Gittins Index Computation: 1.5-Box Problem



$$\mathrm{GI}_f(x; c(x))\text{: solution } g \text{ to } \underbrace{\mathbb{E}[\max(f(x), g)] - c(x) = g}_{\color{red}{\mathbb{E}[\max(f(x) - g, 0)] = c(x)}}$$

Gittins Index Computation: 1.5-Box Problem



Gittins index

$\mathrm{GI}(x)$

Open closed box | Take opened box → g

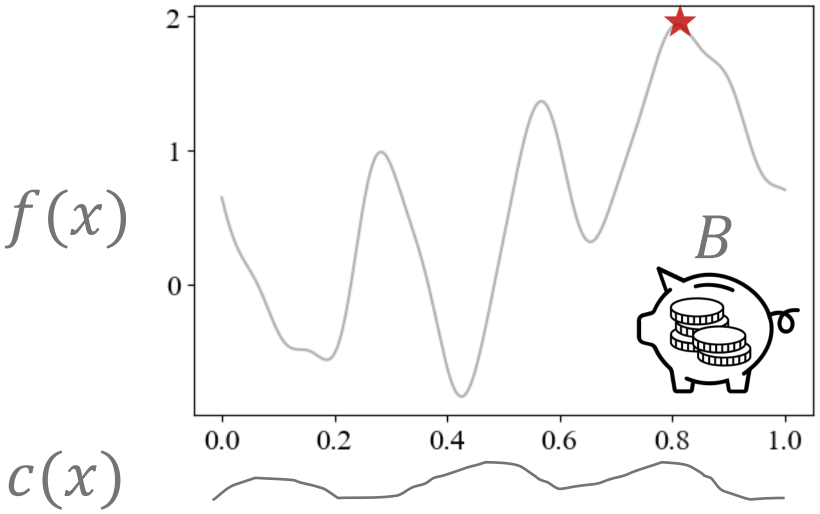
Profit: $\mathbb{E}[\max(f(x), g)] - c(x)$ | Profit: g

$\mathrm{GI}_f(x; c(x))$: solution g to $\underbrace{\mathbb{E}[\max(f(x), g)] - c(x) = g}_{\mathbb{E}[\max(f(x) - g, 0)] = c(x)}$

$\mathrm{EI}_f(x; g)$

$\mathbb{E}[\max(f(x) - g, 0)] = c(x)$

Cost-aware Bayesian Optimization



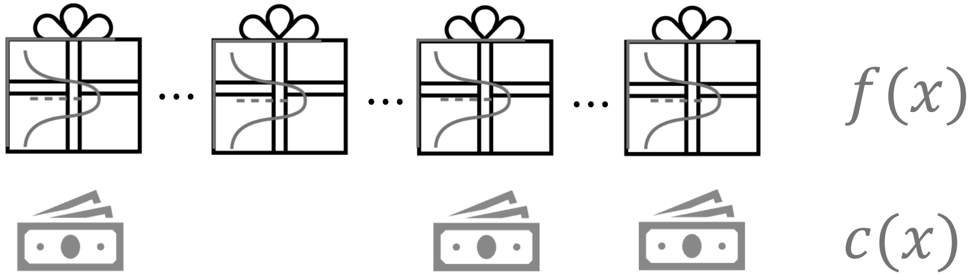
Continuous

Correlated

Expected-budget-constrained

Is Gittins index good?

Pandora's Box



Discrete

Independent

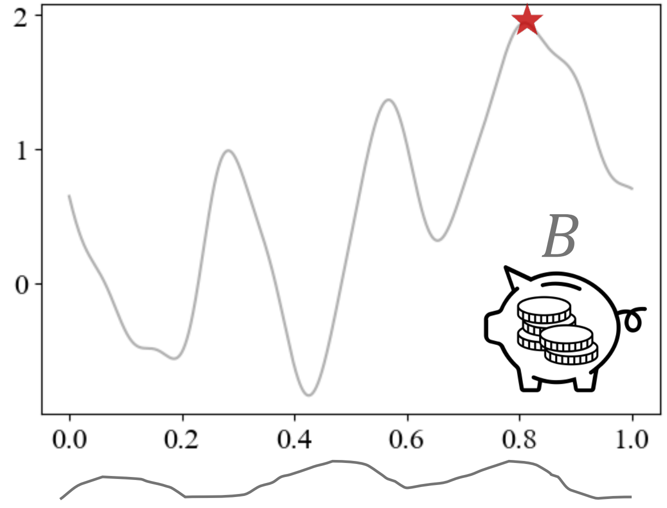
Cost-per-sample

Gittins index is optimal

$\xLeftarrow[\text{scale costs}]{\text{incorporate posterior}}$

Cost-aware Bayesian Optimization

$f(x)$



$c(x)$

Continuous

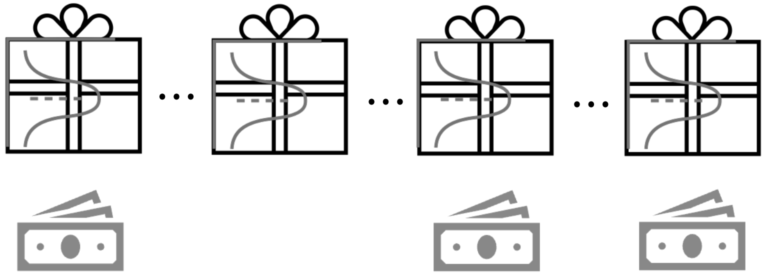
Correlated

Expected-budget-constrained

Is Gittins index empirically good?

this work

Pandora's Box



$f(x)$

$c(x)$

Discrete

Independent

Cost-per-sample

Gittins index is optimal

Experiment Setup: Objective Functions

Synthetic

Samples from prior

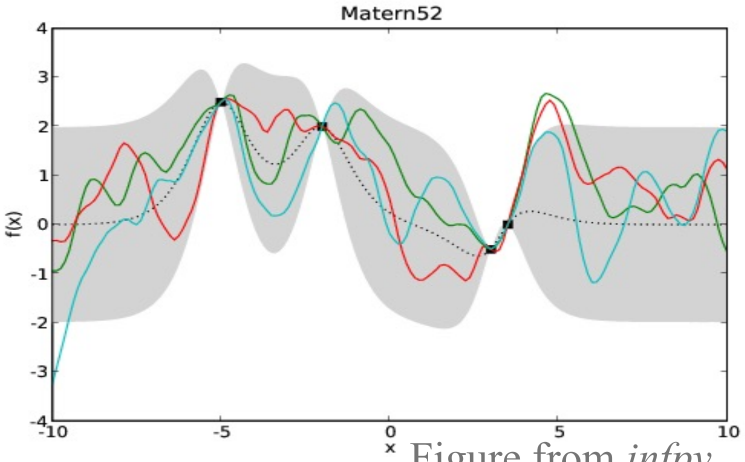


Figure from *infpy*

Ackley function

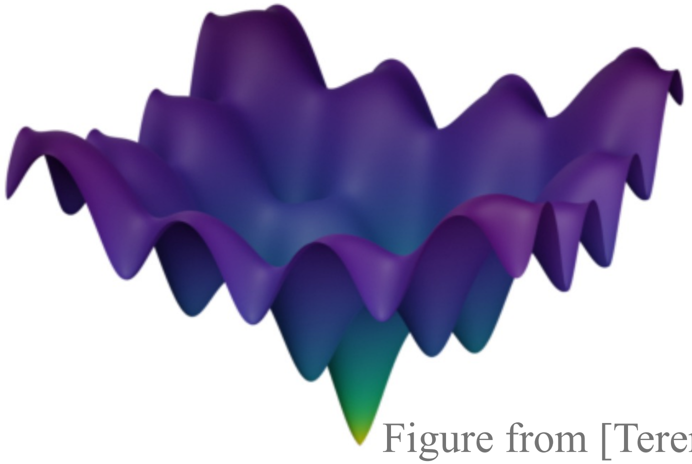


Figure from [Terenin'22]

Empirical

Pest Control

Figure from ChatGPT

Lunar Lander

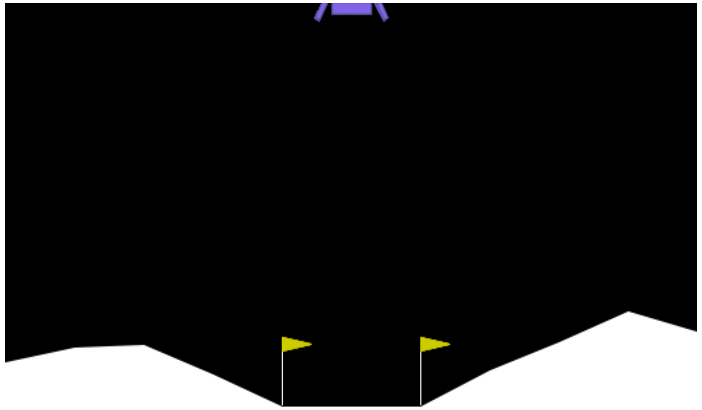
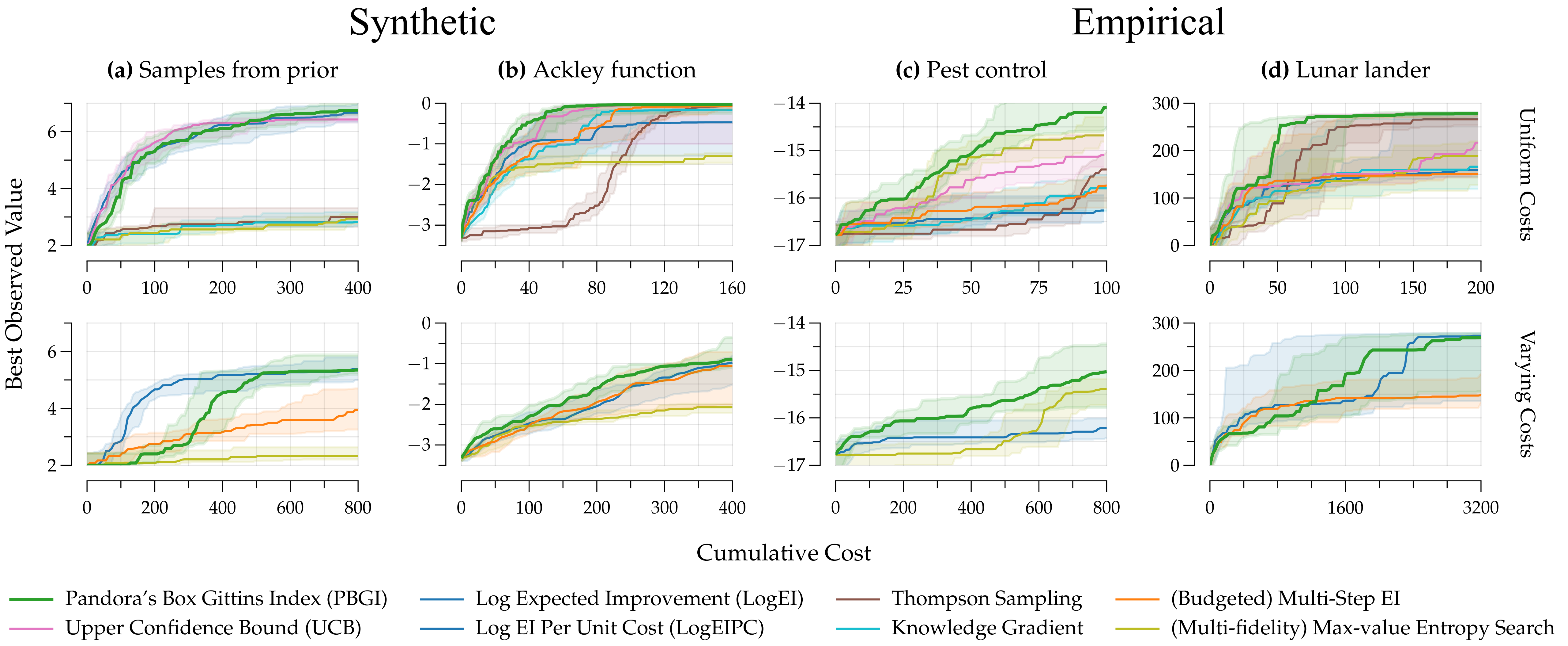


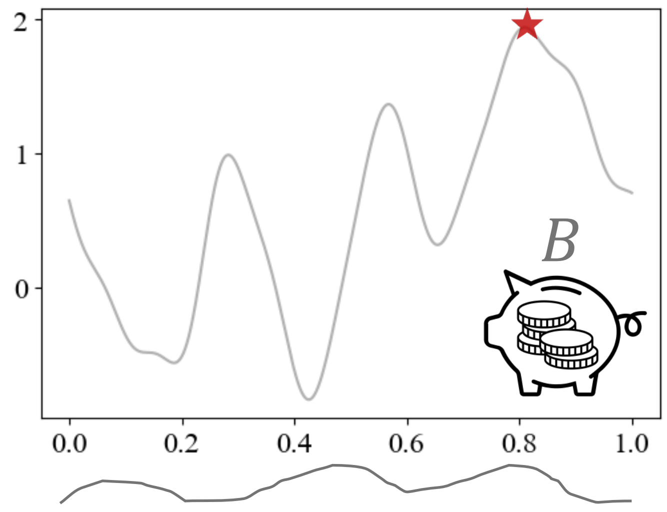
Figure from OpenAI Gym

Experiment Results: Gittins Index vs Baselines



Cost-aware Bayesian Optimization

$f(x)$



$c(x)$

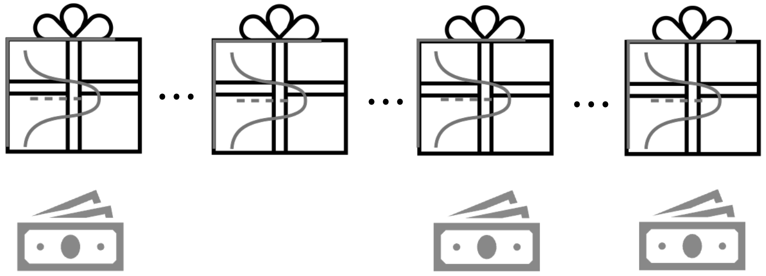
Continuous

Correlated

Expected-budget-constrained

Is Gittins index theoretically good?

Pandora's Box



$f(x)$

$c(x)$

Discrete

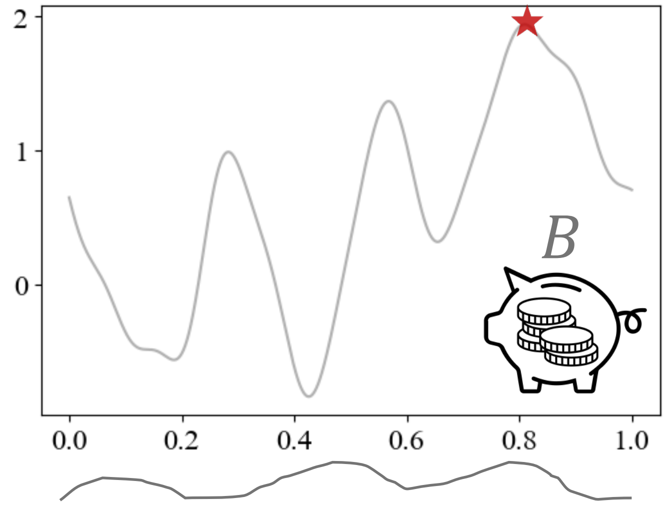
Independent

Cost-per-sample

Gittins index is optimal

Cost-aware Bayesian Optimization

$f(x)$



$c(x)$

Continuous

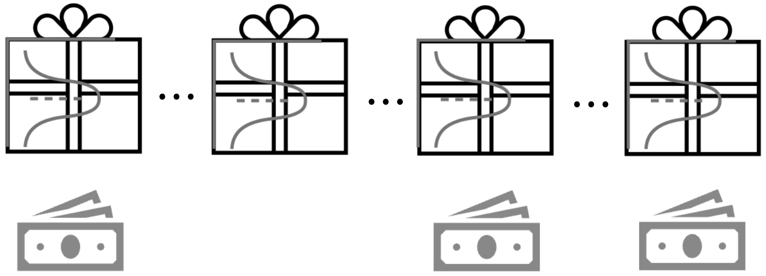
Correlated

Expected-budget-constrained

Is Gittins index theoretically good?

ongoing work

Pandora’s Box



$f(x)$

$c(x)$

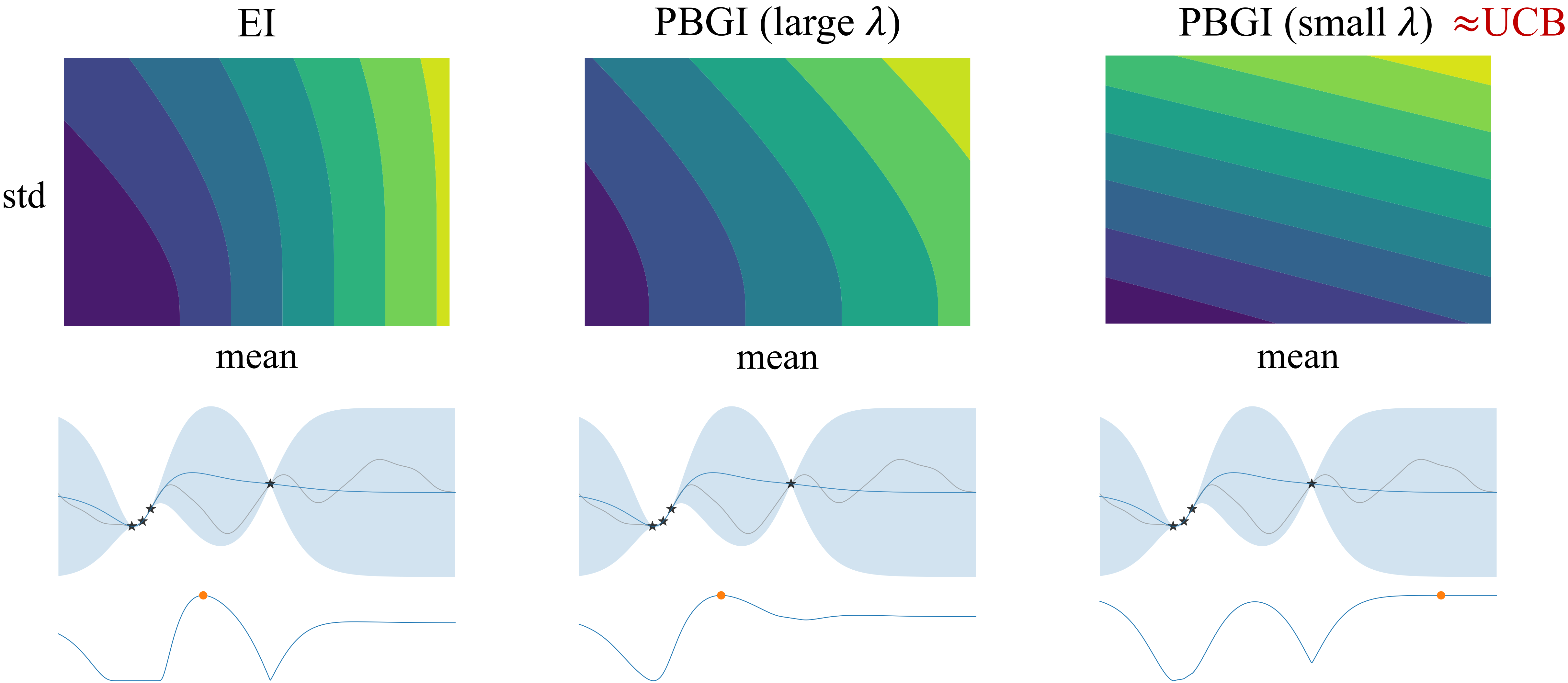
Discrete

Independent

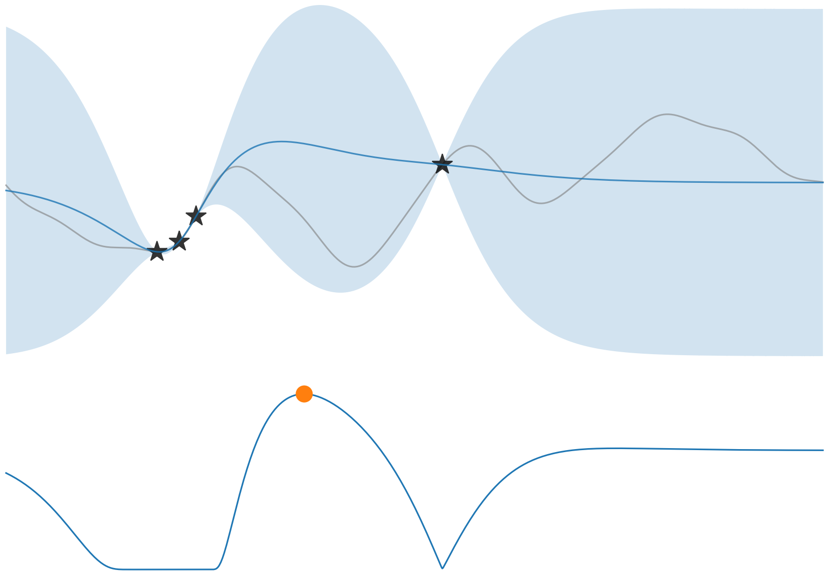
Cost-per-sample

Gittins index is optimal

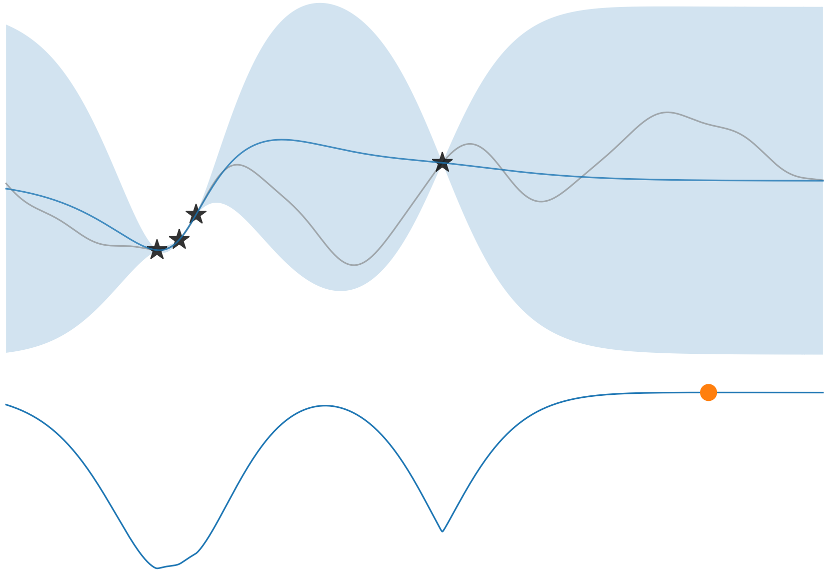
Discussion 1: EI vs PBGI



Expected Improvement



Gittins Index

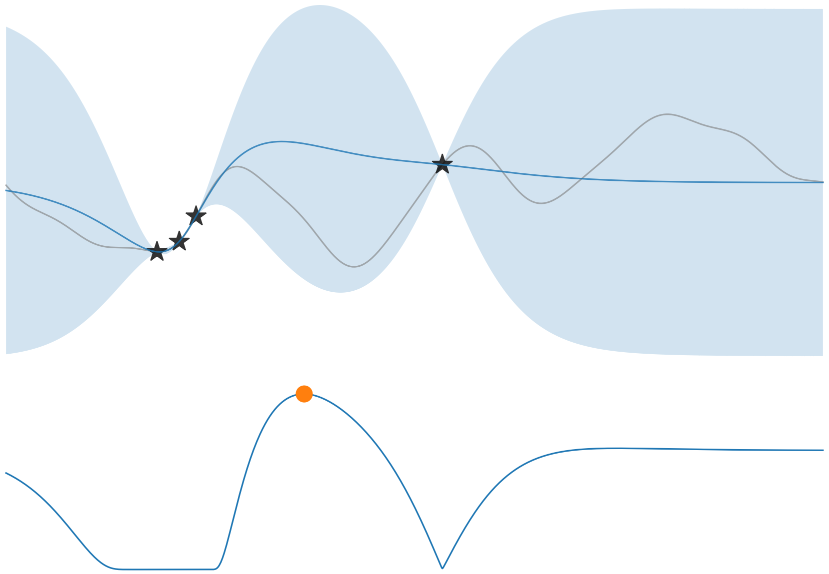


$\mathrm{EI}_{f|D}(x; g) := \mathbb{E}[\max(f(x) - g, 0) \mid D]$

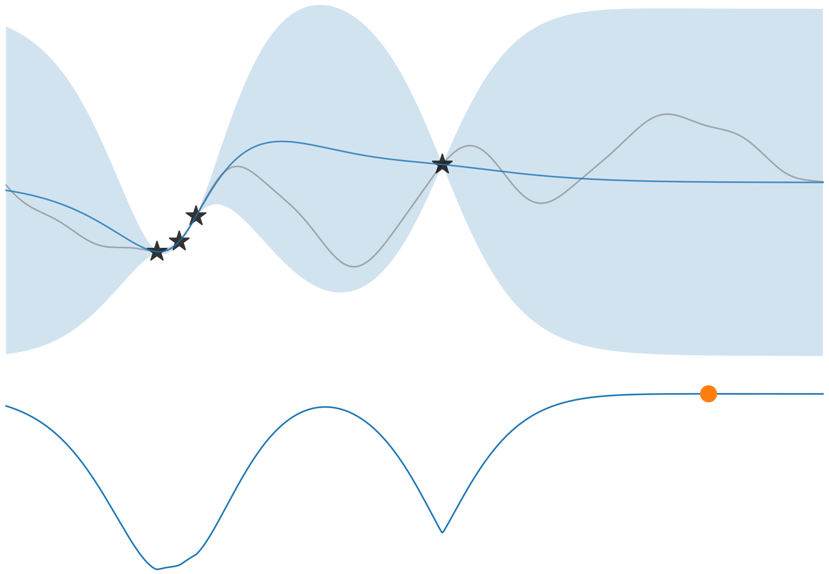
Gaussian distribution

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

Expected Improvement



Gittins Index

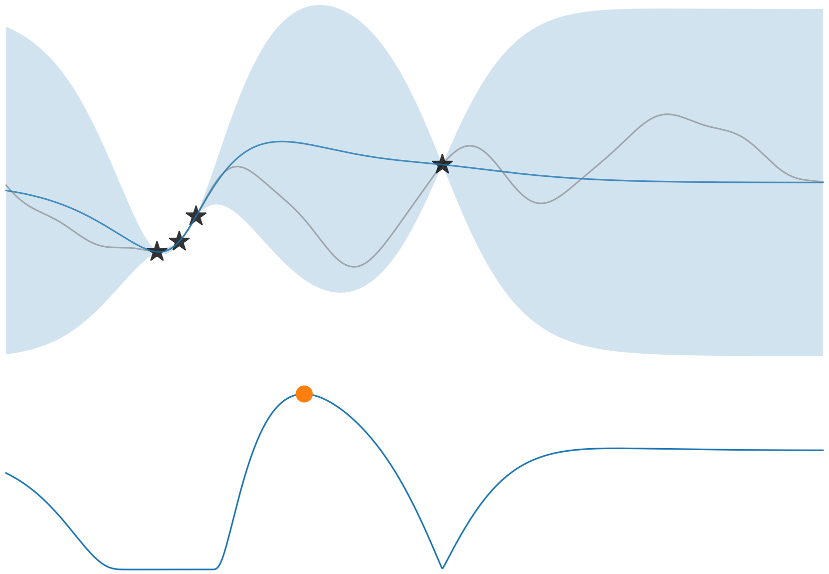


$$\mathrm{EI}_{f|D}(x;g) := \underbrace{\mathbb{E}[\max(f(x)-g,0) \mid D]}_{\text{analytical expression}}$$

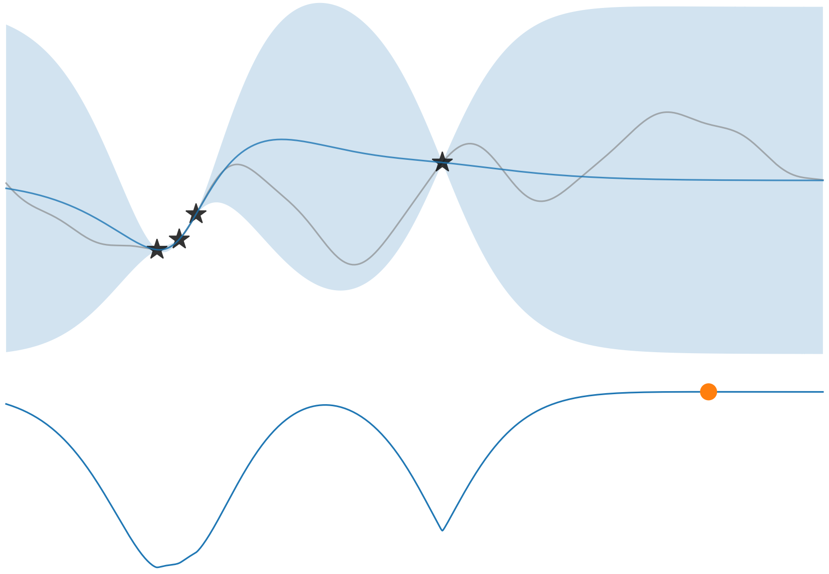
Gaussian distribution

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

Expected Improvement



Gittins Index



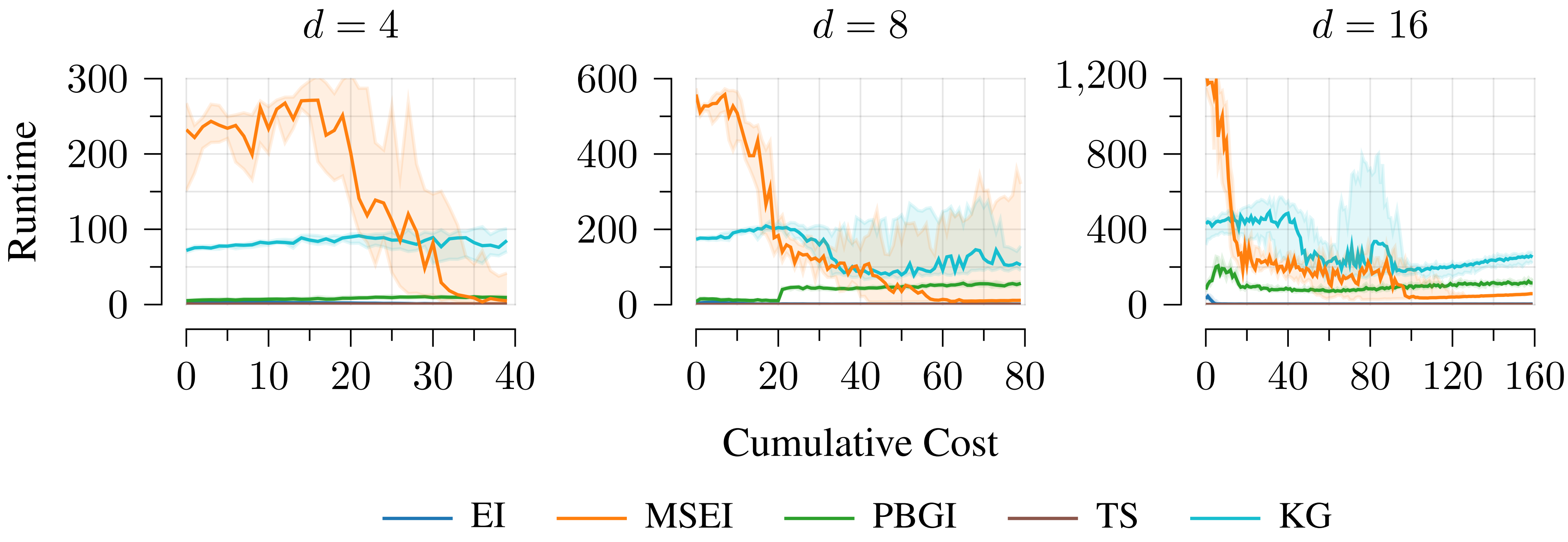
$$\mathrm{EI}_{f|D}(x;g) := \underbrace{\mathbb{E}[\max(f(x)-g,0)\mid D]}_{\text{analytical expression}}$$

Gaussian distribution

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

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

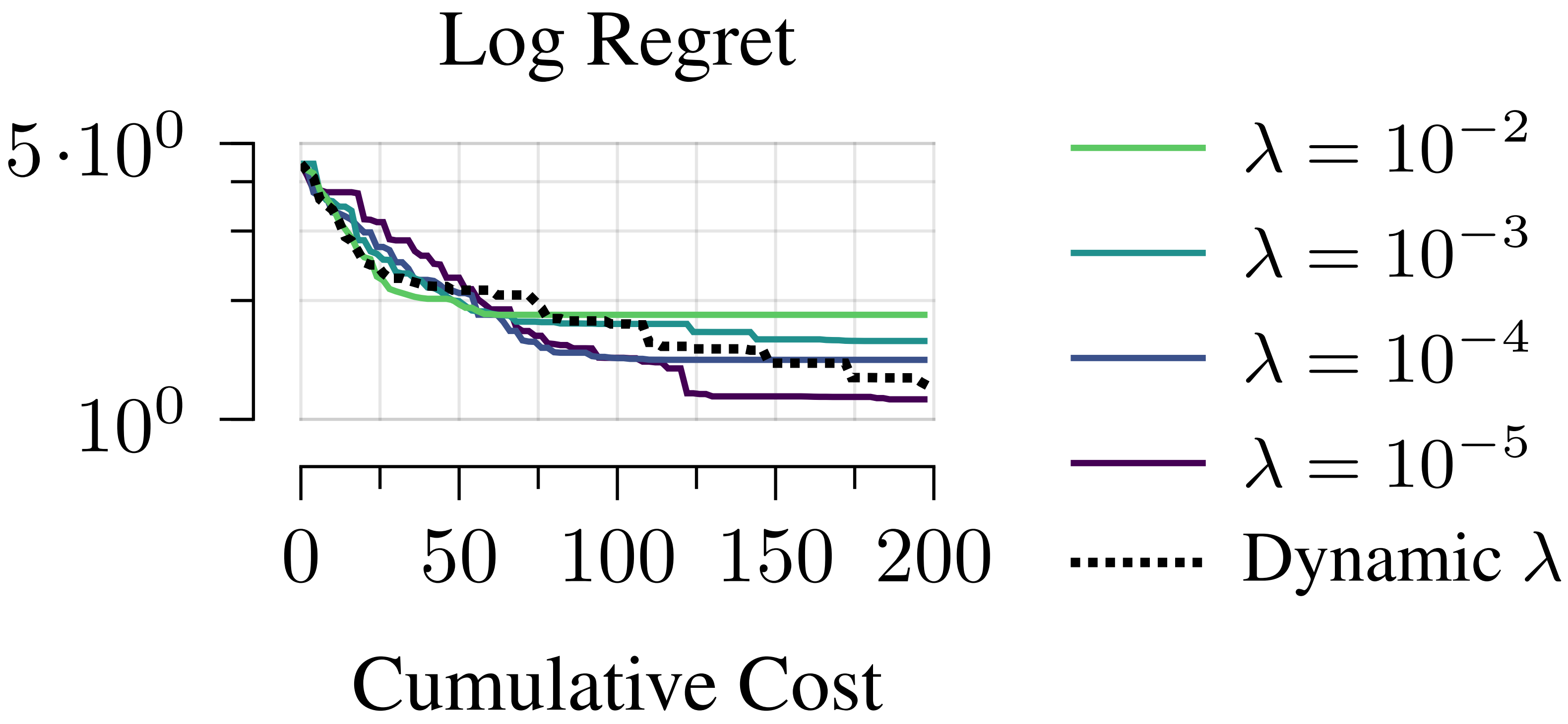
Discussion 2: Timing Experiment



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

PBGI is easy to compute using EI + bisection search!

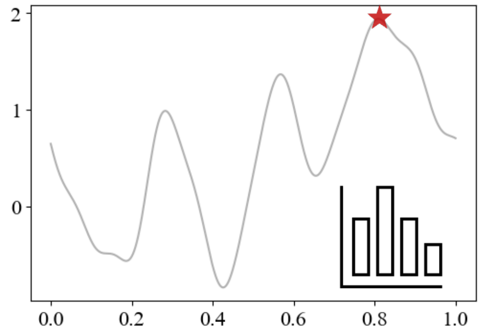
Discussion 3: Effect of λ



Larger budget, smaller λ, higher exploration

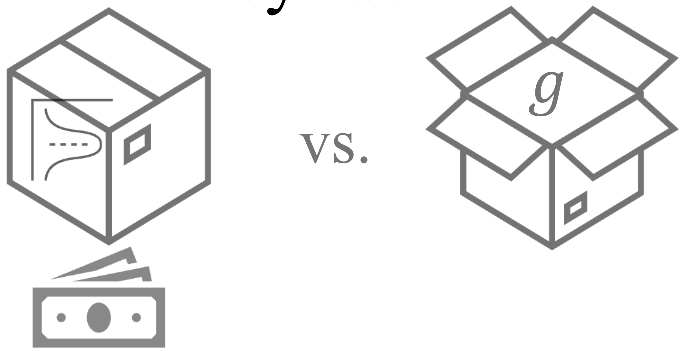
New Design Principle: Gittins Index

Studied Problem



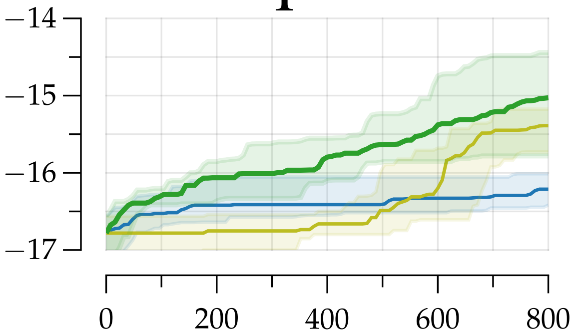
Cost-aware Bayesian optimization

Key idea



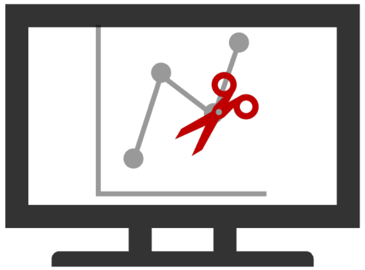
Link to Pandora's box
and Gittins index theory

Impact



Competitive empirical performance

Ongoing work



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

Check our NeurIPS’24 paper on ArXiv!

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