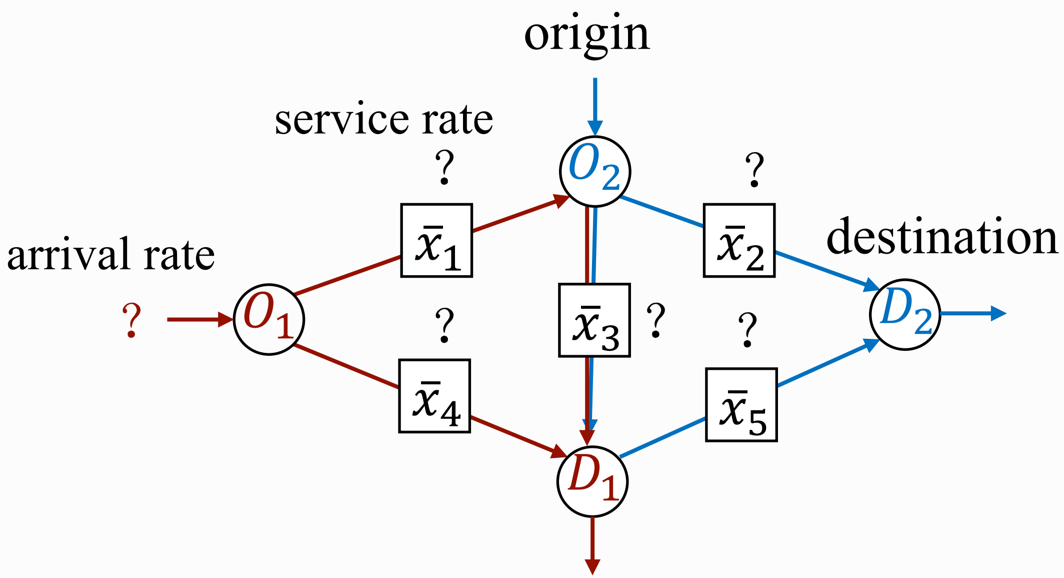
Data-Efficient Decision-making under Uncertainty: AutoML and Beyond

Qian Xie 谢倩 (Cornell ORIE)

Job talk for engineering school audiences

Research Overview: Decision-making under Uncertainty

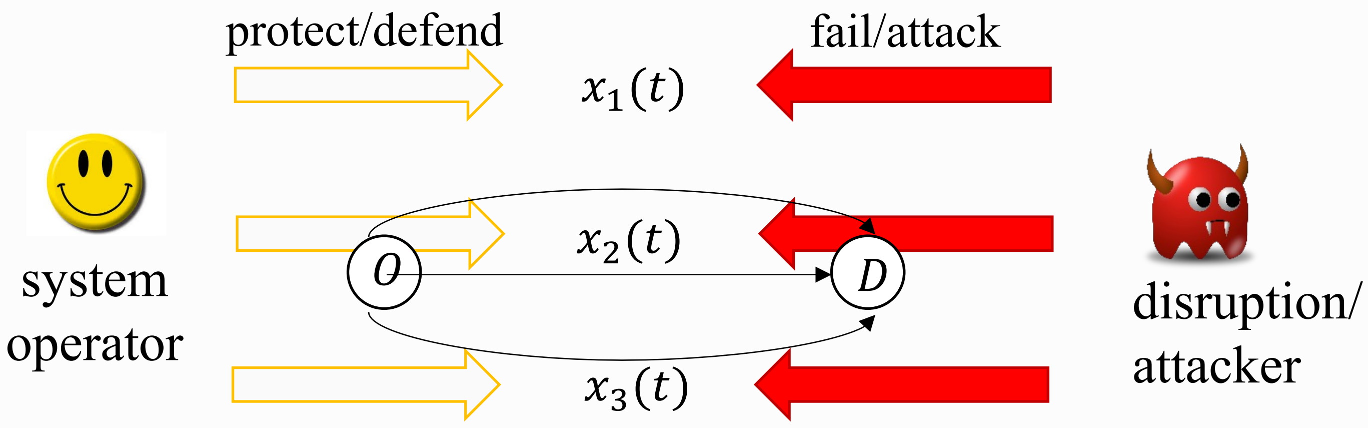
- Earlier: stochastic control (actions chosen to regulate system)
  - Unknown system rates (robust/model-free control) [IEEE TCNS]



"Stabilizing Queuing Networks with Model Data-Independent Control." IEEE TCNS.

Research Overview: Decision-making under Uncertainty

- Earlier: stochastic control (actions chosen to regulate system)
  - Unknown system rates (robust/model-free control) [IEEE TCNS]
  - Random failure/attack realizations (optimal & robust control) [Automatica]



"Stabilizing Queuing Networks with Model Data-Independent Control." IEEE TCNS.

"Cost-aware Defense for Parallel Server Systems against Reliability and Security Failures." Automatica.

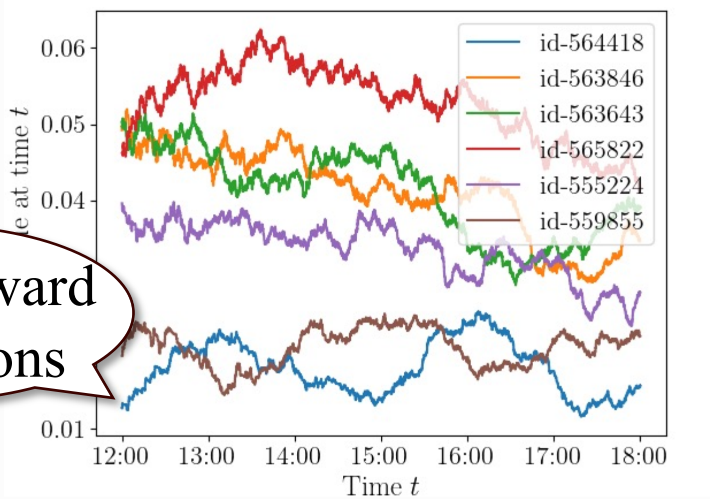
Research Overview: Decision-making under Uncertainty

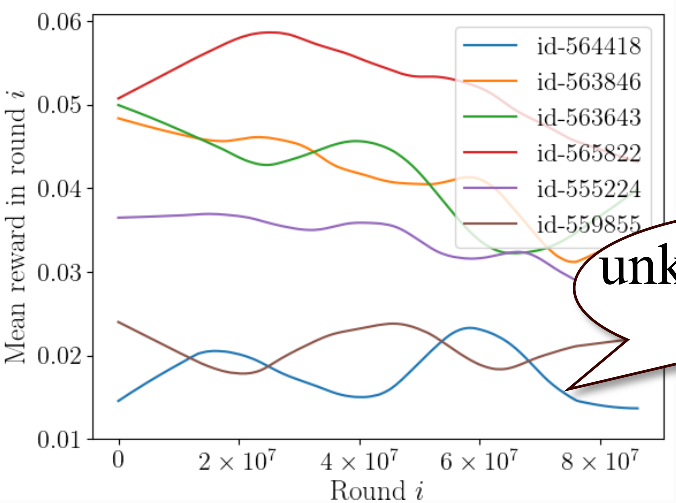
- Earlier: stochastic control (actions chosen to regulate system)
  - Unknown system rates (robust/model-free control) [IEEE TCNS]
  - Random failure/attack realizations (optimal & robust control) [Automatica]
- Recent: learning-based decision-making (actions chosen to reduce uncertainty due to limited feedback)

Research Overview: Decision-making under Uncertainty

- Earlier: stochastic control (actions chosen to regulate system)
  - Unknown system rates (robust/model-free control) [IEEE TCNS]
  - Random failure/attack realizations (optimal & robust control) [Automatica]
- Recent: learning-based decision-making (actions chosen to reduce uncertainty due to limited feedback)
  - Unknown reward function; random observations (bandits) [ICML'23 + OR (*major revision*)]

random reward observations





unknown reward function

"Smooth Nonstationary Bandits." ICML'23

Research Overview: Decision-making under Uncertainty

- Recent: learning-based decision-making (actions chosen to reduce uncertainty due to limited feedback)
  - Unknown reward function; random observations (bandits) [ICML'23 + *OR*]
  - Unknown objective functions (black-box optimization) with
    - varying evaluation costs [NeurIPS'24 + INFORMS DM Paper Finalist]
    - adaptive stopping time [Under review + AutoML'25 (non-archival)]
    - multi-stage feedback [To be submitted]
    - multi-source environment information [NeurIPS'25 LAW]

"Cost-aware Bayesian Optimization via the Pandora's Box Gittins Index." NeurIPS'24.

"Cost-aware Stopping for Bayesian Optimization." Under review.

"Bayesian-optimal Decision-making under Cost-aware Multi-stage Feedback via the Gittins Index."

"LLM-Driven Composite NAS for Multi-Source RL State Encoding." NeurIPS'25 LAW.

Research Overview: Decision-making under Uncertainty

- Recent: learning-based decision-making (actions chosen to reduce uncertainty due to limited feedback)
  - Unknown reward function; random observations (bandits) [ICML'23 + *OR*]
  - Unknown objective functions (black-box optimization) with

This talk's focus

varying evaluation costs [NeurIPS'24 + INFORMS DM Paper Finalist]

adaptive stopping time [Under review + AutoML'25 (non-archival)]

multi-stage feedback [To be submitted]

multi-source environment information [NeurIPS'25 LAW]

"Cost-aware Bayesian Optimization via the Pandora's Box Gittins Index." NeurIPS'24.

"Cost-aware Stopping for Bayesian Optimization." Under review.

"Bayesian-optimal Decision-making under Cost-aware Multi-stage Feedback via the Gittins Index."

"LLM-Driven Composite NAS for Multi-Source RL State Encoding." NeurIPS'25 LAW.

Motivation: World of Optimization under Uncertainty

ML model training:

≠ parameters
(weights, biases)

Training hyperparameters
(e.g., learning rate, # layers) ⟶

⟶ Accuracy

Motivation: World of Optimization under Uncertainty

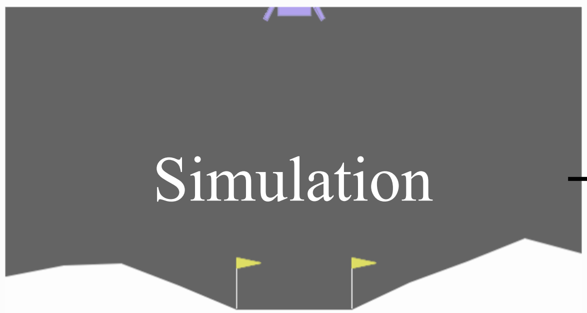
≠ parameters
(weights, biases)

ML model training:

Training hyperparameters (e.g., learning rate, # layers) → 

→ Accuracy

Control design:

Control variables (e.g., position, angle, velocity) → 

→ Reward

Motivation: World of Optimization under Uncertainty

non-analytical & no gradient info

Design choices x →  → Performance metric $f(x)$

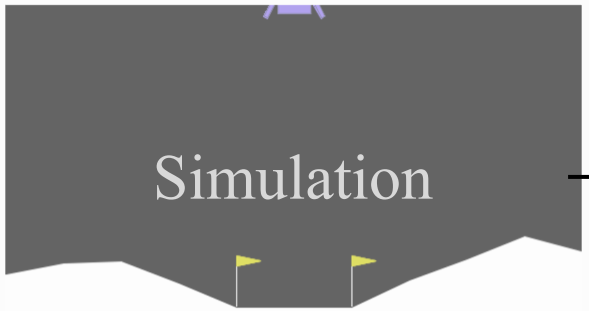
ML model training:

≠ parameters (weights, biases)

Training hyperparameters (e.g., learning rate, # layers) → 

→ Accuracy

Control design:

Control variables (e.g., position, angle, velocity) → 

→ Reward

Motivation: World of Optimization under Uncertainty

Black-box optimization:

(gradient-based methods not applicable)

non-analytical & no gradient info

Input x ⟶  ⟶ Observed outcome $f(x)$

ML model training:

≠ parameters (weights, biases)

Training hyperparameters (e.g., learning rate, # layers) ⟶ 

⟶ Accuracy

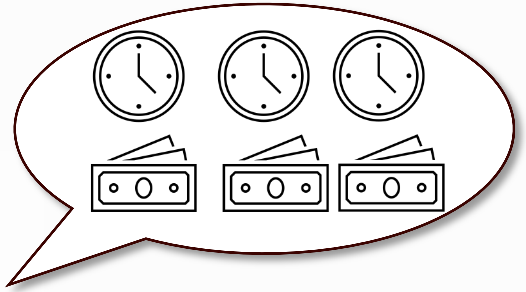
Background: Black-Box Optimization

Black-box optimization:

(gradient-based methods not applicable)

Input x →  → Observed outcome $f(x)$

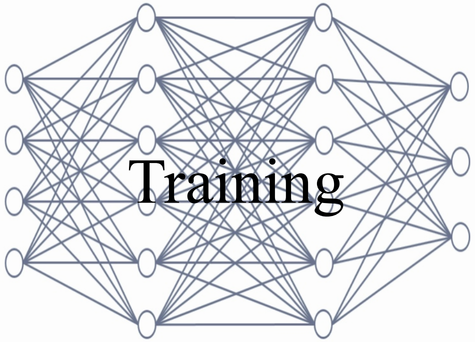
expensive-to-evaluate



Training time

Compute credits

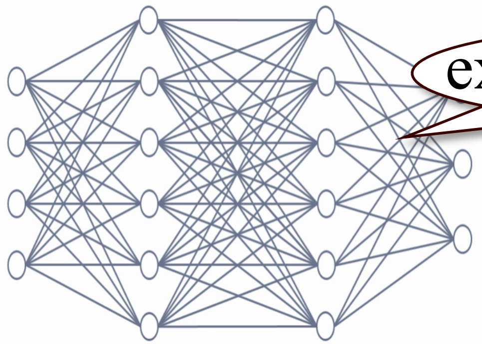
ML model training:

Training hyperparameters (e.g., learning rate, # layers) → 

→ Accuracy

Naïve (Non-Adaptive) Approach: Grid Search

ML model training:

Training hyperparameters ⟶  ⟶ Accuracy

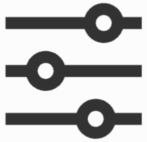
expensive-to-evaluate

Training hyperparameter	Range	Number of Options
Batch size	[16, 512]	10
Learning rate	[1e-4, 1e-1]	10
Momentum	[0.1, 0.99]	10
Weight decay	[1e-5, 1e-1]	10
Number of layers	{1, 2, 3, 4}	4
Max units per layer	[64, 1024]	10
Dropout	[0.0, 1.0]	10

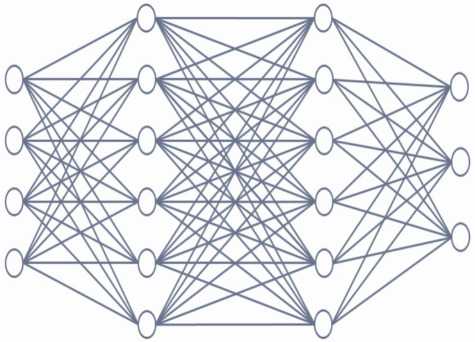
40,000,000 combinations!

Data-Driven (Adaptive) Approach

Automated machine learning:
(AutoML)

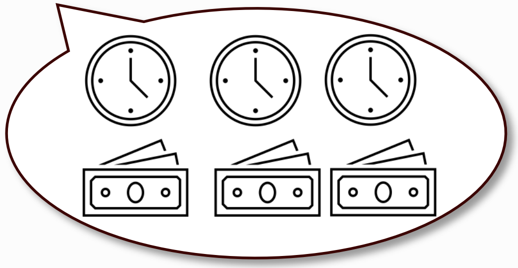


Hyperparameters



Training Pipeline

Performance metric
(e.g., accuracy)



Run next evaluation

Provide feedback

Data-efficient decision rule
(What to try next, when to stop)

Existing Umbrellas of Black-Box Optimization

Naïve (non-adaptive) approaches:	Data-driven (adaptive) approaches:
• Grid search	• Local search (e.g., simulated annealing)
• Random search	• Evolutionary algorithms (e.g., genetic algo)
• Manual tuning	• Bayesian optimization (e.g., EI, UCB, TS)
	• Reinforcement learning (e.g., PPO, ENAS)
	• LLM-based search agent (e.g., GENIUS)

Existing Umbrellas of Black-Box Optimization

Naïve (non-adaptive) approaches:	Data-driven (adaptive) approaches:
• Grid search	• Local search (e.g., simulated annealing)
• Random search	• Evolutionary algorithms (e.g., genetic algo)
• Manual tuning	• Bayesian optimization (e.g., EI, UCB, TS)
	• Reinforcement learning (e.g., PPO, ENAS)
	• LLM-based search agent (e.g., GENIUS)

Limitations in practice:

1. Limited principled guidance (e.g., naïve)

2. Often data-inefficient (e.g., naïve, local search, evolutionary algo)

3. No or ad-hoc incorporation of additional side info (most) (e.g., varying eval costs)

Overview of Contributions Across My Work

Naïve (non-adaptive) approaches:	Data-driven (adaptive) approaches:
• Grid search	• Local search
• Random search	• Evolutionary algorithms
• Manual tuning	• Bayesian optimization ★ Part I
	• Reinforcement learning ★ Part II
	• LLM-based search agent ★

Contributions of methods in my work:

1. Principled guidance

2. Competitive empirical performance

3. Principled incorporation of additional side info

★ New methods under this umbrella

Outline

Part I (Recent):

Bayesian Optimization via Gittins Index Design Principle

Part II (Ongoing):

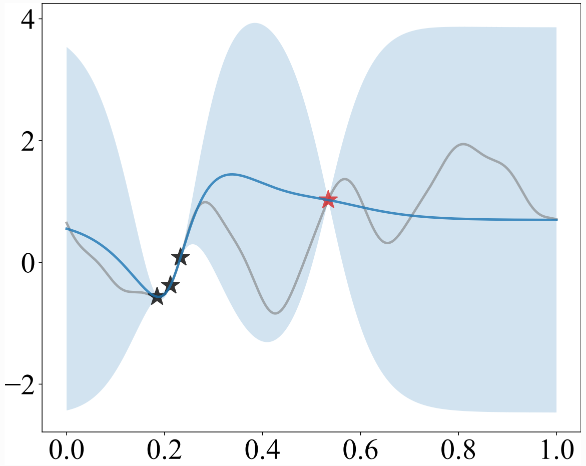
Black-box Optimization Beyond AutoML

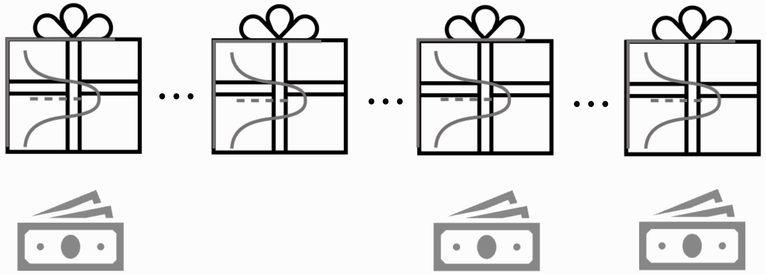
Reinforcement Learning · *Engineering Design* · *Scientific Discovery* · *LLM*

Part I (Recent):

Bayesian Optimization via Gittins Index

Design Principle





Bayesian Optimization

$f \sim$ Stochastic Process

$x \longrightarrow$

$\longrightarrow f(x)$

Black-box function

Bayesian Optimization

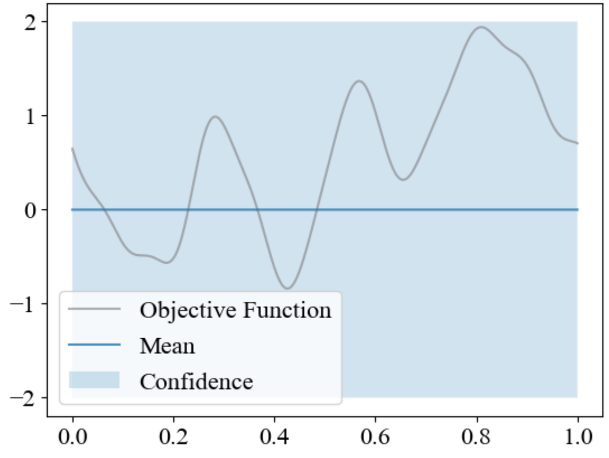
Time 0

$f \sim$ Stochastic Process

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

Black-box function

Model belief



Probabilistic model

(e.g., Gaussian process)

Bayesian Optimization

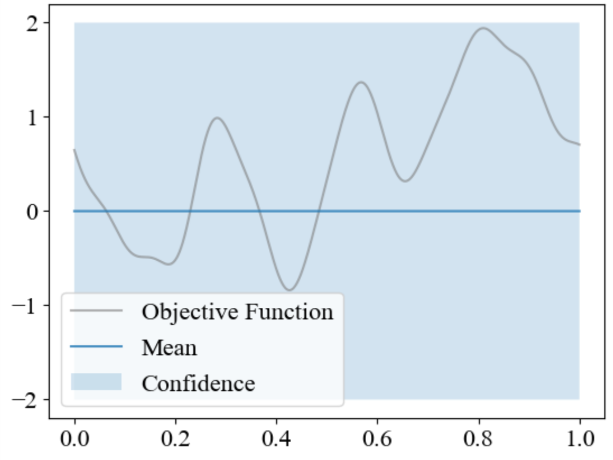
Time t

$f \sim$Stochastic Process

$x_1, \ldots, x_t \longrightarrow$  $\longrightarrow f(x_1), \ldots, f(x_t)$

Black-box function

Model belief



Probabilistic model
(e.g., Gaussian process)

Bayesian Optimization

Time t

$f \sim$ Stochastic Process

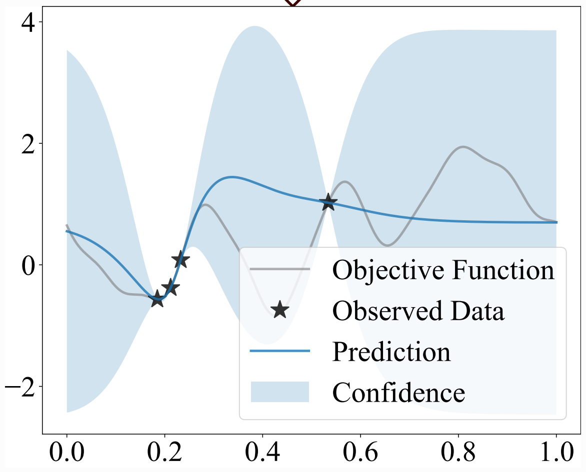
$x_1, \dots, x_t \longrightarrow$

$\longrightarrow f(x_1), \dots, f(x_t)$

Black-box function

Update belief

(Bayes' rule)



Probabilistic model

(e.g., Gaussian process)

Bayesian Optimization

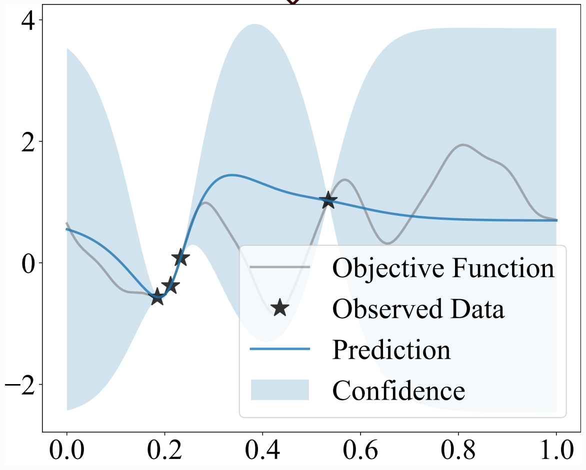
Time t

$x_1, \ldots, x_t \longrightarrow$

Black-box function

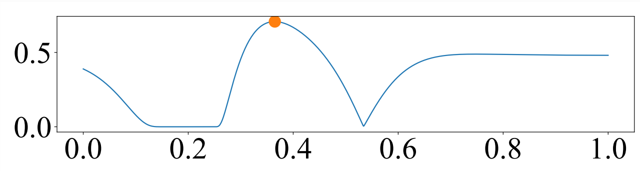
$\longrightarrow f(x_1), \ldots, f(x_t)$

Update belief
(Bayes' rule)



Probabilistic model
(e.g., Gaussian process)

scoring (worth of each point



Decision rule
(e.g., EI, UCB, TS)

Bayesian Optimization

Time t

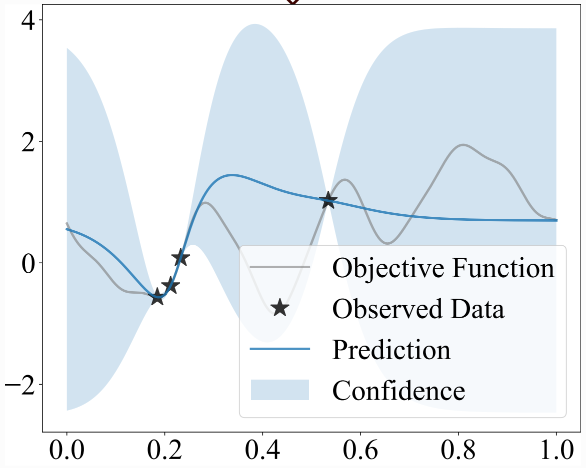
$x_1, \dots, x_t \longrightarrow$

$\longrightarrow f(x_1), \dots, f(x_t)$

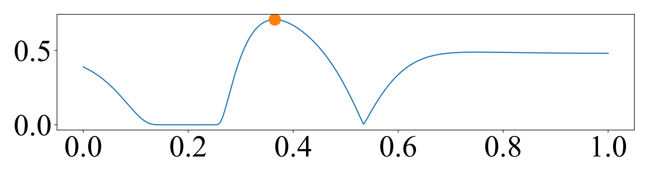
Black-box function

Decide next input
(gradient-based optimization)

Update belief
(Bayes' rule)



Probabilistic model
(e.g., Gaussian process)



Decision rule
(e.g., EI, UCB, TS)

scoring (worth of each point

Bayesian Optimization

Time $t + 1$

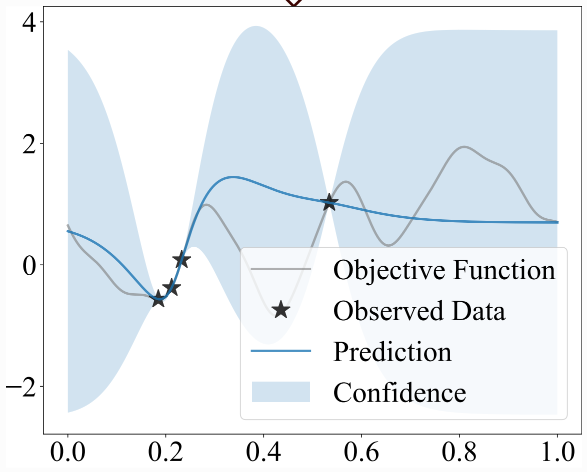
$x_1, \dots, x_{t+1} \longrightarrow$

Black-box function

$\longrightarrow f(x_1), \dots, f(x_{t+1})$

Update belief

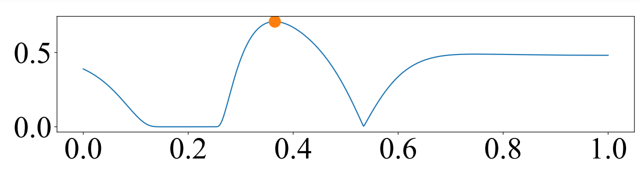
(Bayes' rule)



Probabilistic model

(e.g., Gaussian process)

scoring (worth of each point



Decision rule

(e.g., EI, UCB, TS)

Decide next input

(gradient-based optimization)

Bayesian Optimization

Time $t + 1$

$x_1, \ldots, x_{t+1}$

Black-box function

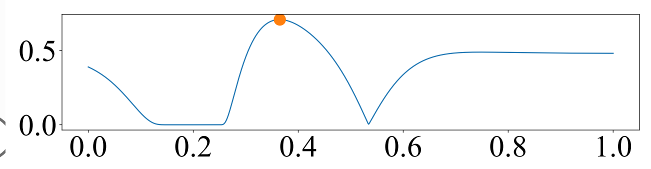
$f(x_1), \ldots, f(x_{t+1})$

Update belief
(Bayes' rule)

Probabilistic model
(e.g., Gaussian process)

scoring (worth of each point

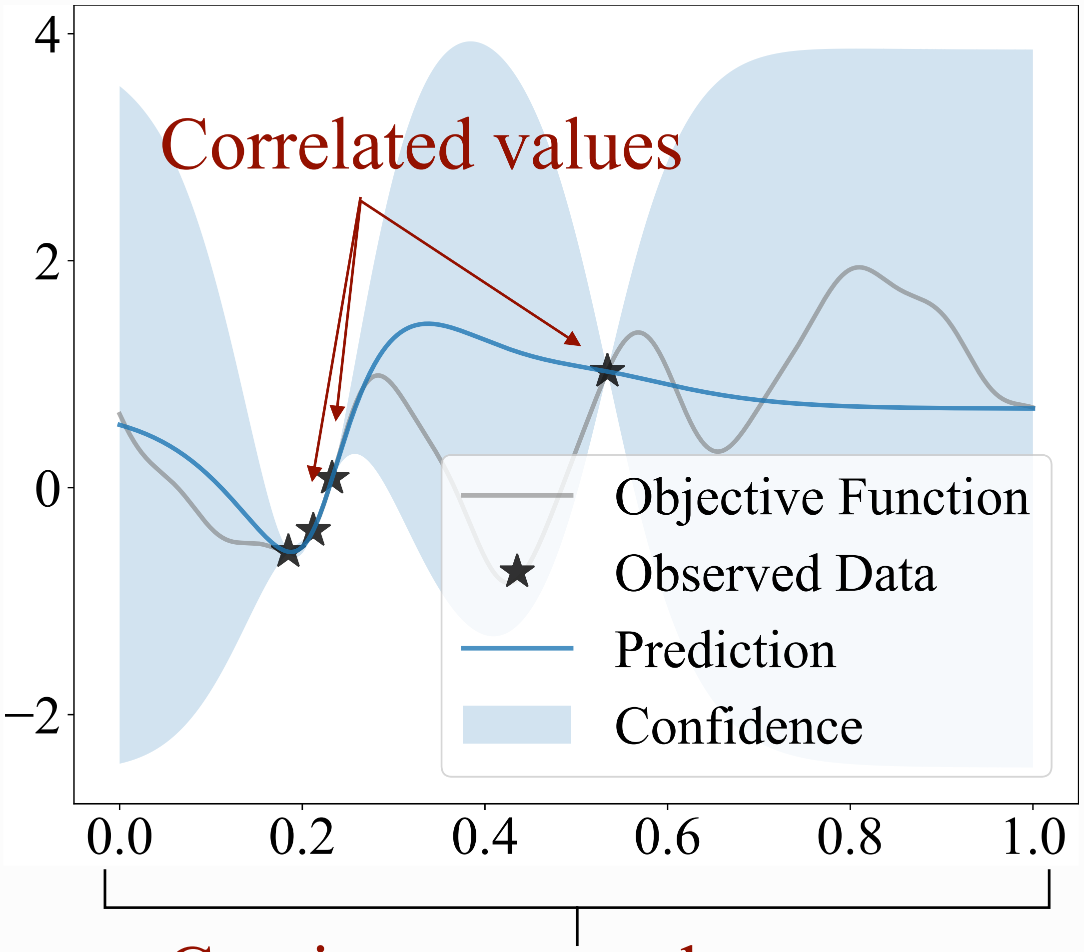
My focus



Decision rule
(e.g., EI, UCB, TS)

Decide next input
(gradient-based optimization)

Challenges in Decision Rule Design



Correlation & continuity $\Longrightarrow$ Intractable MDP $\Longrightarrow$ Optimal policy unknown

Popular Decision Rule: Expected Improvement

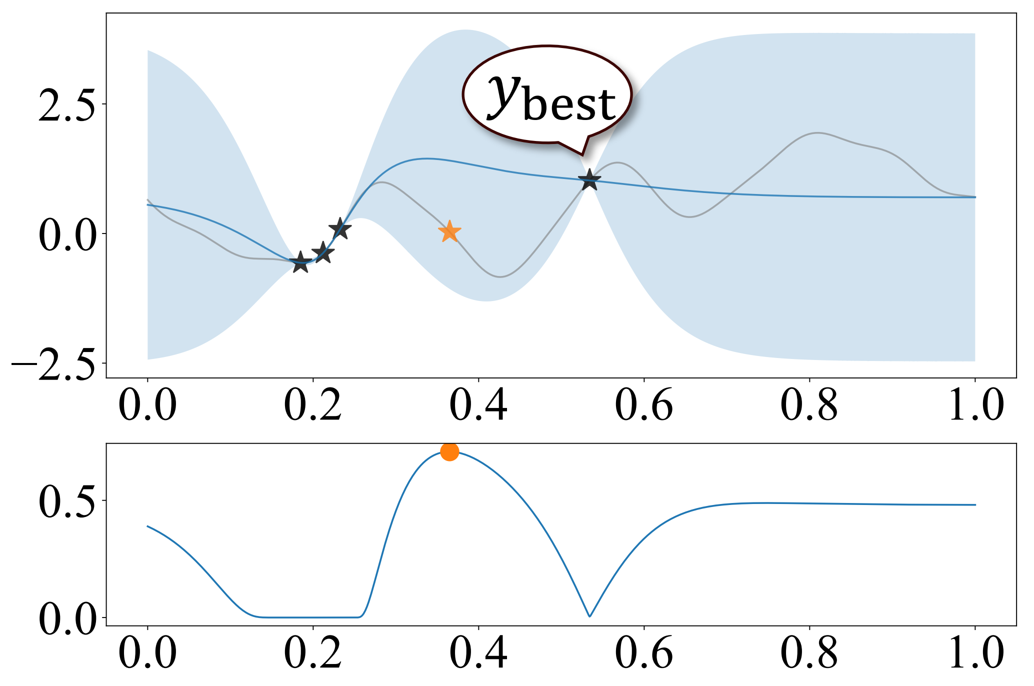
current best observed

data D

$$\mathrm{EI}(x) = \mathbb{E}[\underbrace{\max(f(x) - y_{\text{best}}, 0) \mid}_{\text{“improvement”}} x_1, \ldots, x_t]$$

$$x_{t+1} = \max_x \mathrm{EI}_{f|D}(x)$$

posterior distribution



One-step approximation to MDP

Expected improvement $\mathrm{EI}(x)$

Popular Decision Rule: Expected Improvement

current best observed

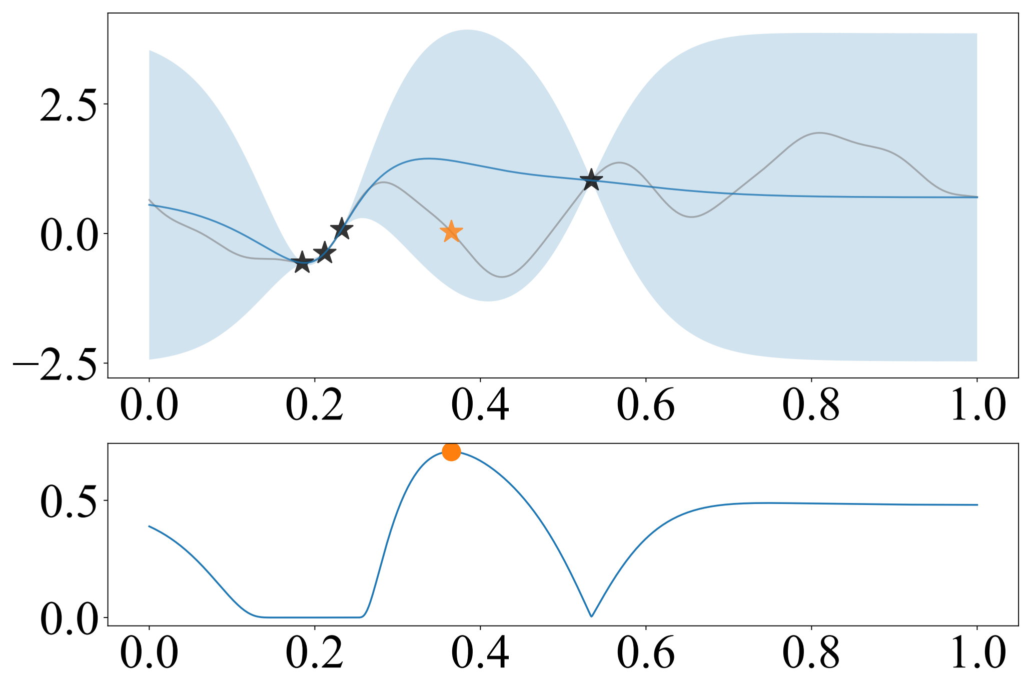
data D

$$\mathrm{EI}(x) = \mathbb{E}[\underbrace{\max(f(x) - y_{\text{best}}, 0)}_{\text{“improvement”}} \mid x_1, \ldots, x_t]$$

$$x_{t+1} = \max_x \mathrm{EI}_{f|D}(x; y_{\text{best}})$$

posterior distribution

One-step approximation to MDP

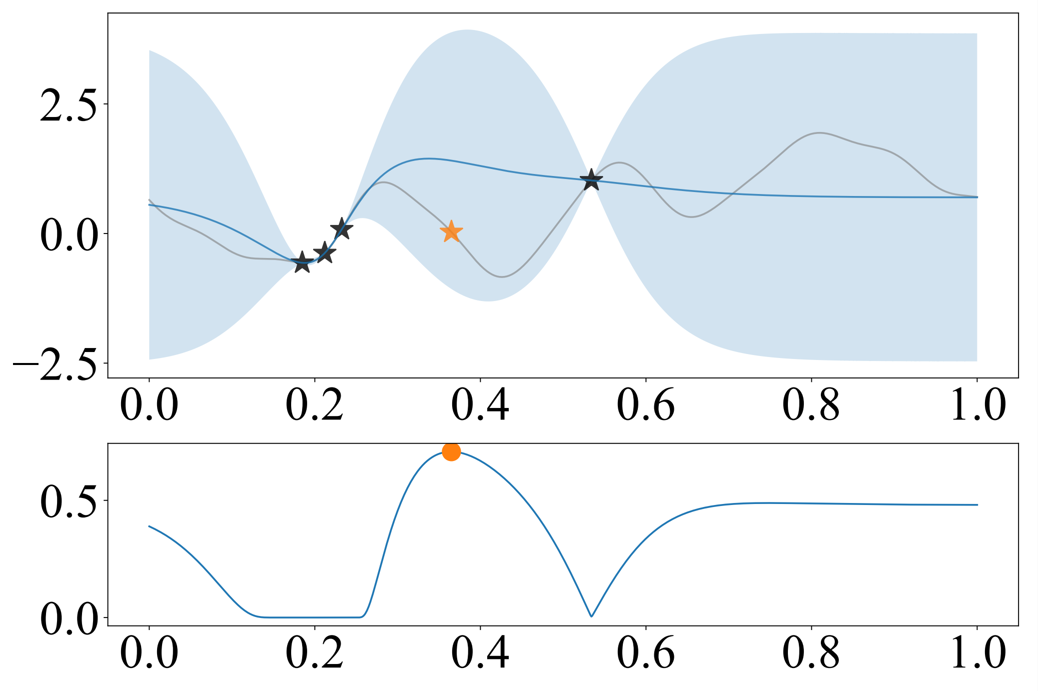


Expected improvement $\mathrm{EI}(x)$

Improvement-based design principle

Existing Design Principles

- Improvement-based (e.g., EI)
- Entropy-based
- Confidence bounds (UCB/LCB)
- Thompson sampling (TS)

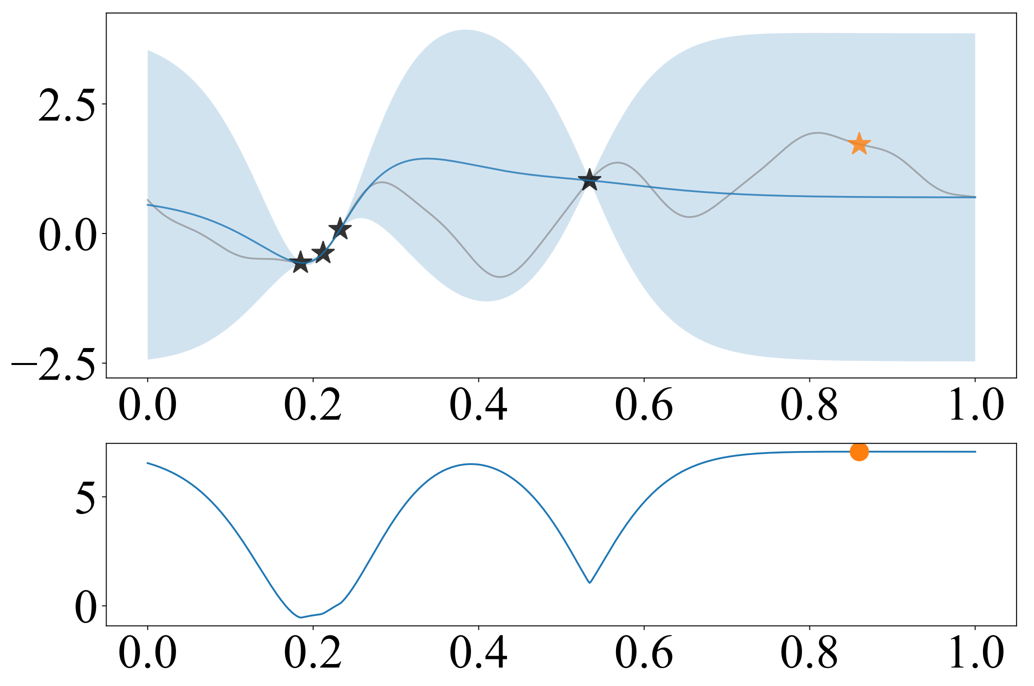


Expected improvement $\mathrm{EI}(x)$

Improvement-based design principle

New Design Principle: Gittins Index

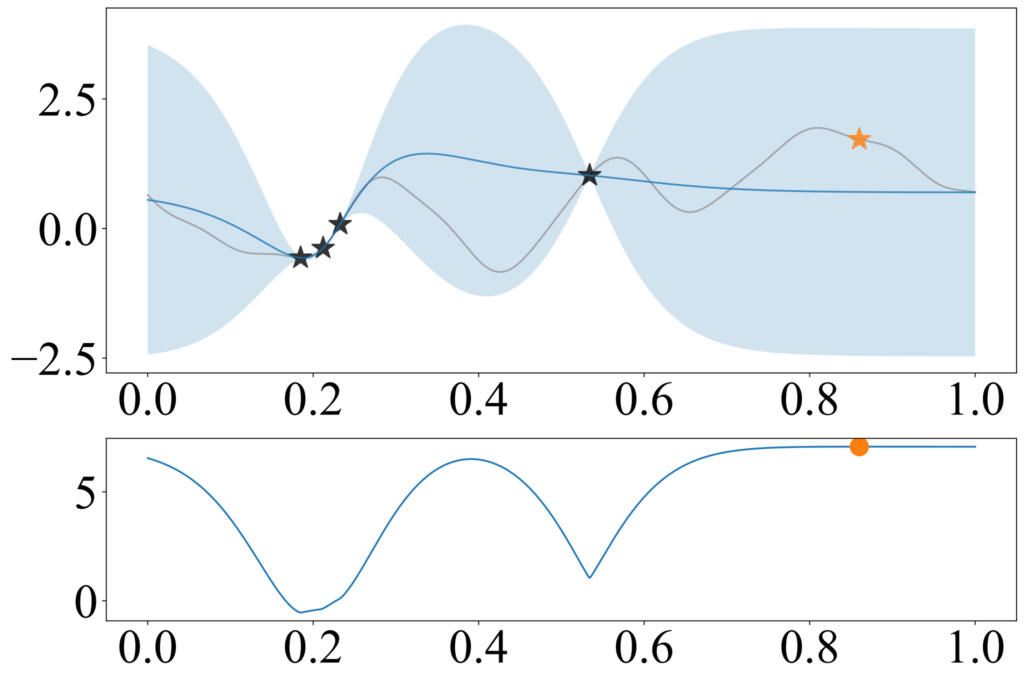
- Improvement-based (e.g., EI)
- Entropy-based
- Confidence bounds (UCB/LCB)
- Thompson sampling (TS)
- Gittins Index



Gittins index $\mathrm{GI}(x)$

New Design Principle: Gittins Index

- Improvement-based (e.g., EI)
- Entropy-based
- Confidence bounds (UCB/LCB)
- Thompson sampling (TS)
- Gittins Index



Gittins index $\mathrm{GI}(x)$

? Why another principle?

Our Contribution: Gittins Index Principle

Joint work with Ziv Scully and Alexander Terenin et al.

1. Principled easy-to-compute decision rules

2. Natural incorporation of side info and flexibility

3. Competitive performance on benchmarks

4. Theoretical guarantees

Our Contribution: Gittins Index Principle

Joint work with Ziv Scully and Alexander Terenin et al.

	EI
1. Principled easy-to-compute decision rules	✓
2. Natural incorporation of side info and flexibility	✗
3. Competitive performance on benchmarks	✓✗
4. Theoretical guarantees	✓✗

Our Contribution: Gittins Index Principle

Joint work with Ziv Scully and Alexander Terenin et al.

EI

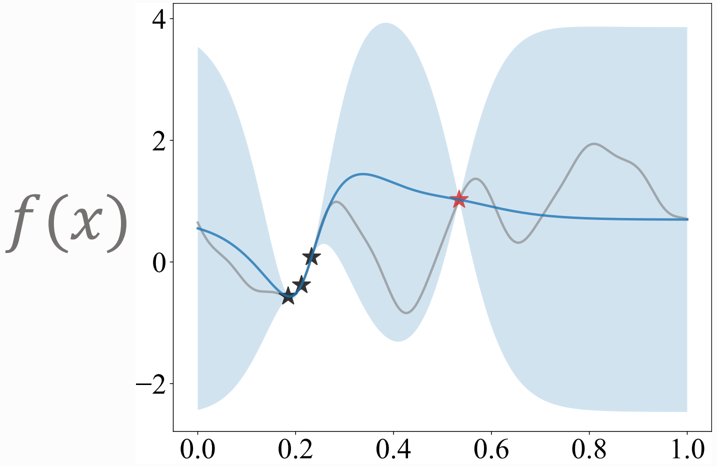
1. Principled easy-to-compute decision rules ✓

2. Natural incorporation of side info and flexibility ✗

3. Competitive performance on benchmarks

4. Theoretical guarantees

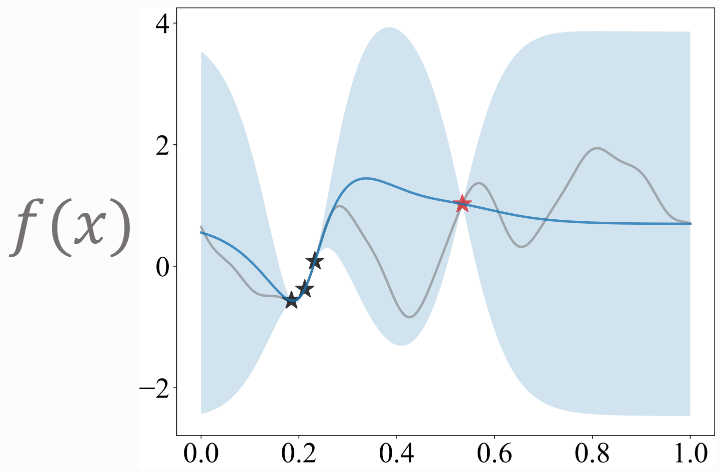
Bayesian Optimization



Continuous search space

Correlated function values

Bayesian Optimization

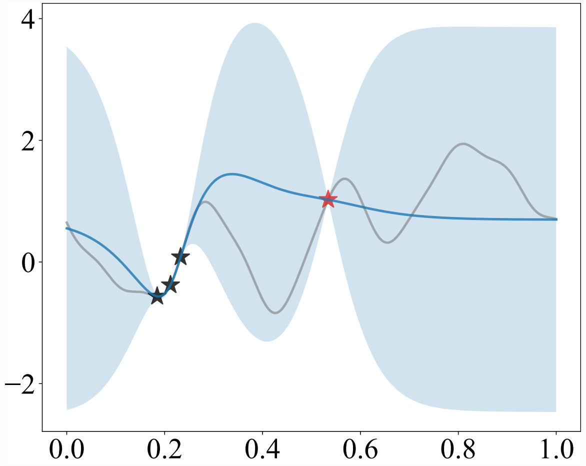


Continuous search space $\Longrightarrow$ Discrete

Correlated function values $\Longrightarrow$ Independent

Bayesian Optimization

$f(x)$



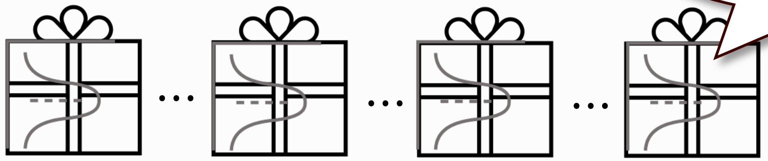
Continuous search space

Correlated function values

Pandora's Box

[Weitzman'79]

distribution of hidden reward



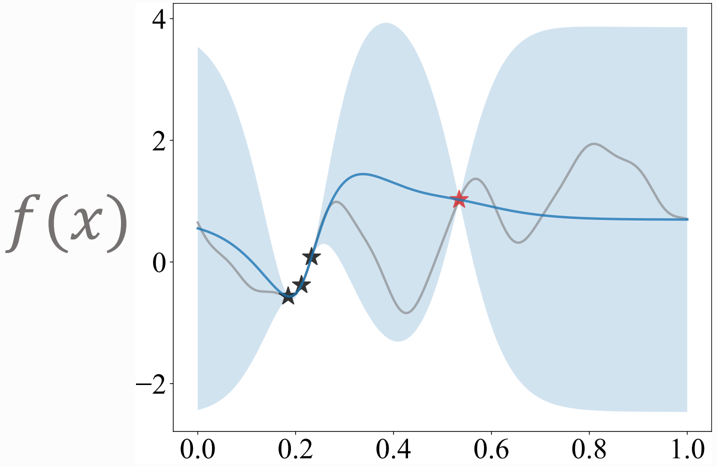
… … …

$f(x)$

Discrete

Independent

Bayesian Optimization

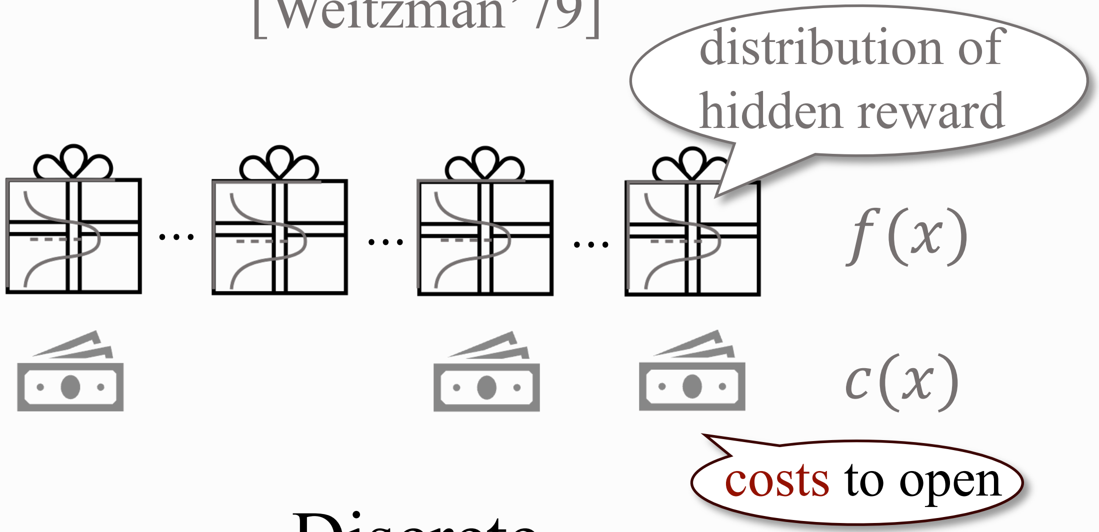


Continuous search space

Correlated function values

Pandora's Box

[Weitzman'79]

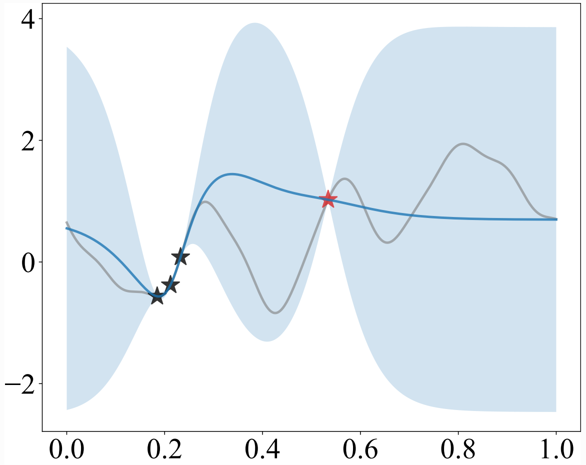


Discrete

Independent

Bayesian Optimization

$f(x)$

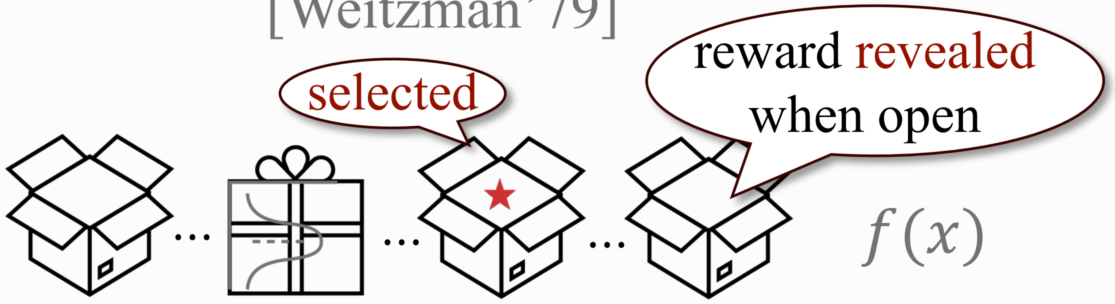


Continuous search space

Correlated function values

Pandora's Box

[Weitzman'79]



$f(x)$

$c(x)$

costs to open

Discrete

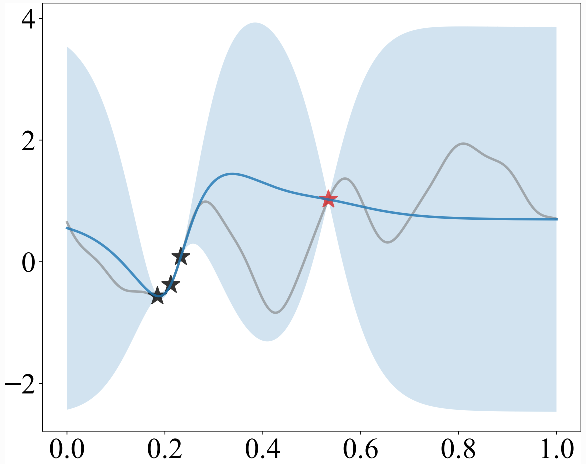
Independent

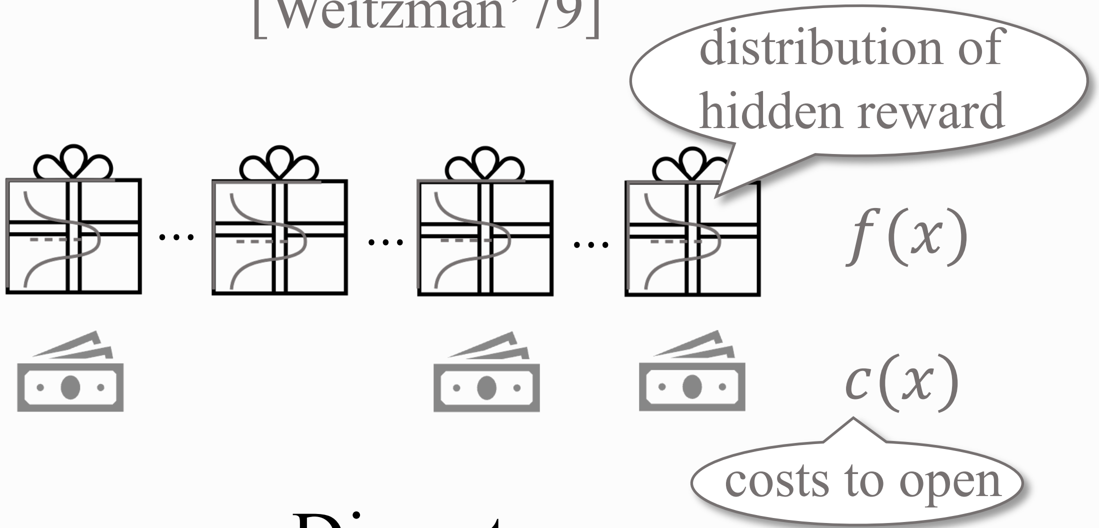
Bayesian Optimization

Pandora's Box

[Weitzman'79]

$f(x)$





$f(x)$

$c(x)$

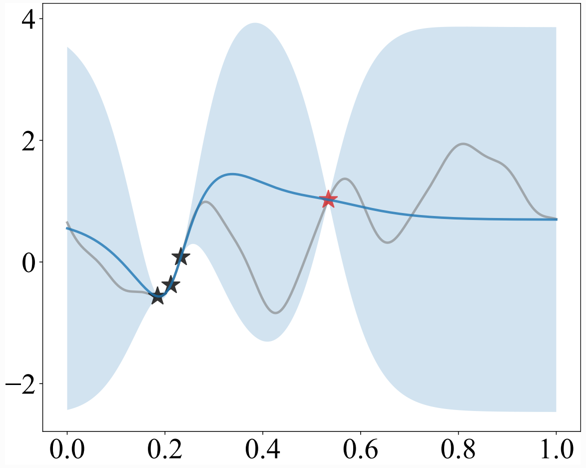
Continuous search space | Discrete

Correlated function values | Independent

Optimal policy: Gittins index

Bayesian Optimization

$f(x)$

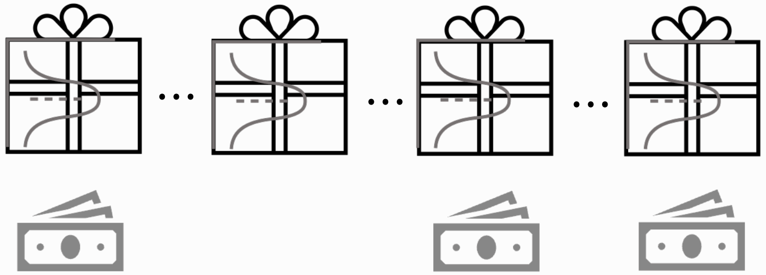


Continuous search space

Correlated function values

Pandora's Box

[Weitzman'79]



$f(x)$

$c(x)$

costs to open

Discrete

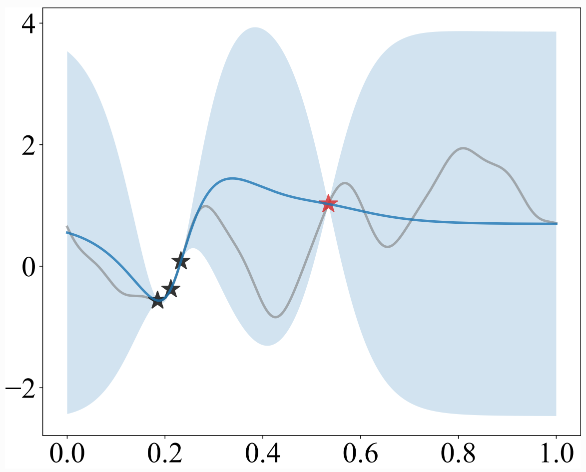
Independent

How to translate?

$\Longleftarrow$ Optimal policy: Gittins index

Bayesian Optimization

$f(x)$

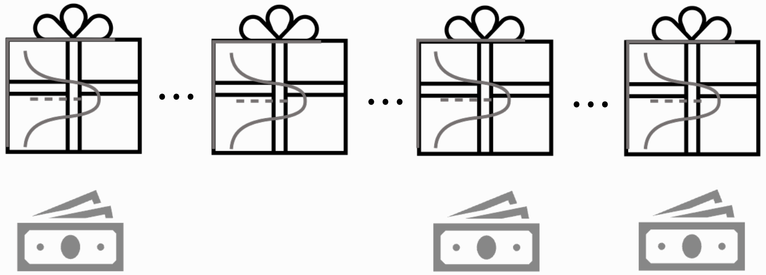


Continuous search space

Correlated function values

Pandora's Box

[Weitzman'79]



$f(x)$

$c(x)$

costs to open

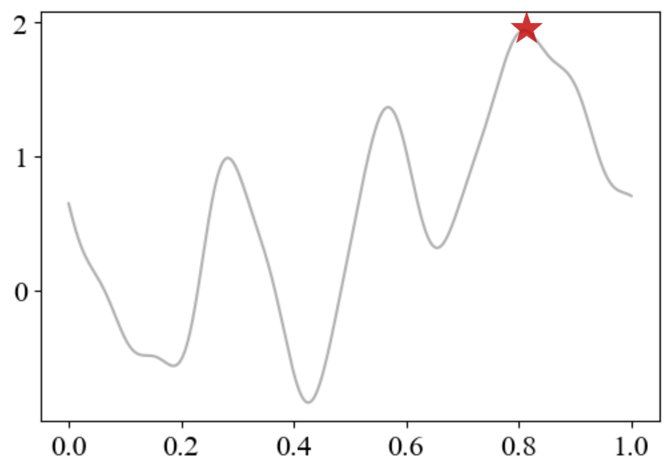
Discrete

Independent

Our policy: $\mathrm{GI}_{f|D}(x;c)$ $\xLeftarrow[\text{take continuum limit}]{\text{incorporate posterior}}$ Optimal policy: $\mathrm{GI}_{f}(x;c)$

Bayesian Optimization

$f(x)$

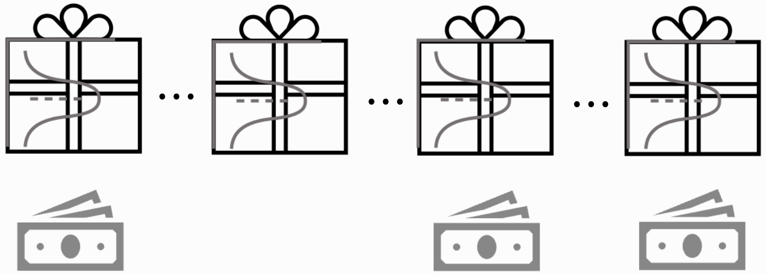


Continuous

Correlated

Pandora's Box

[Weitzman'79]



$f(x)$

$c(x)$

costs to open

Discrete

Independent

incorporate posterior

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

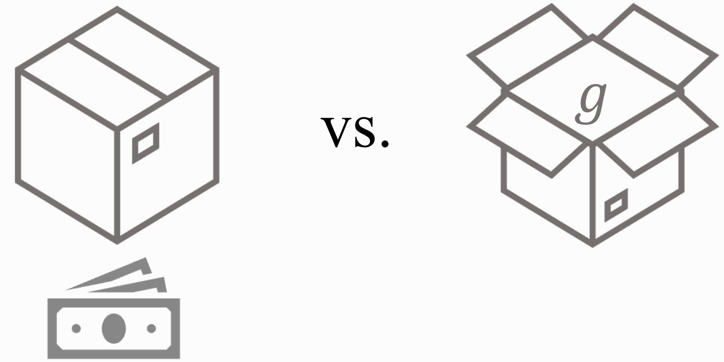
take continuum limit

How to compute?

Intuition

Exploration Exploitation

$f(x)$



vs.

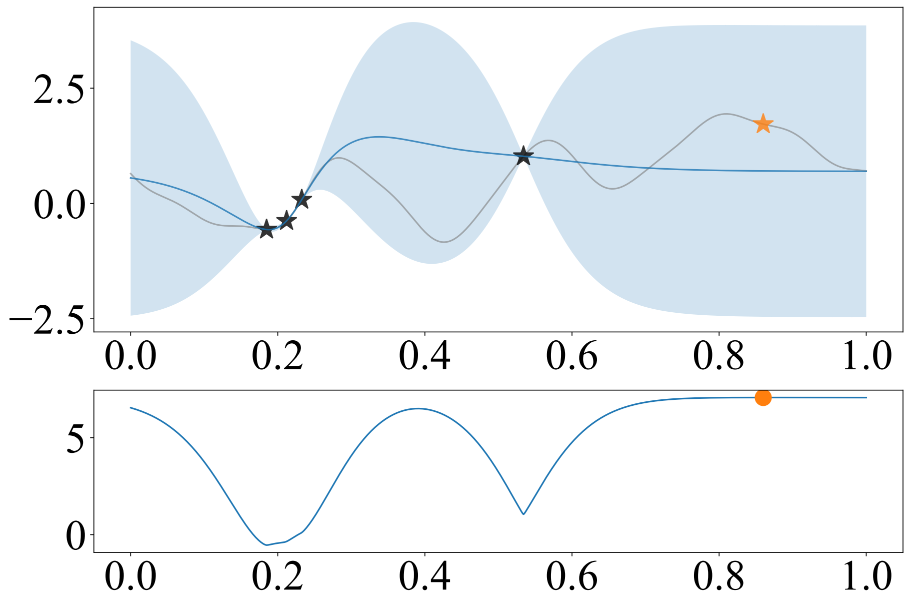
$c(x)$

Open closed box Take opened box

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

Should one open box? Depend on g!

Gittins Index



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

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

Intuition

Exploration | Exploitation

$f(x)$ vs.

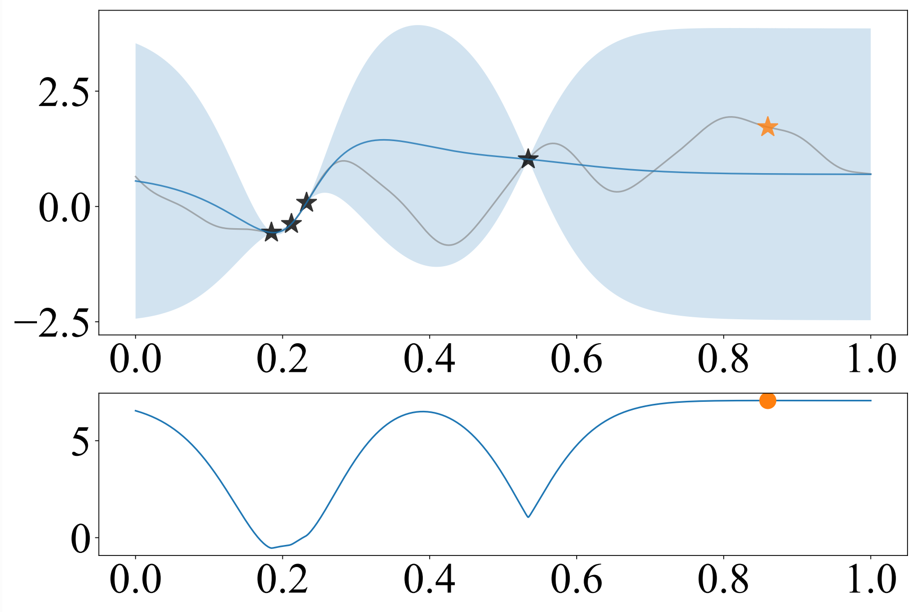
$c(x)$

Open closed box | Take opened box

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

Should one open box? Depend on g!

Gittins Index



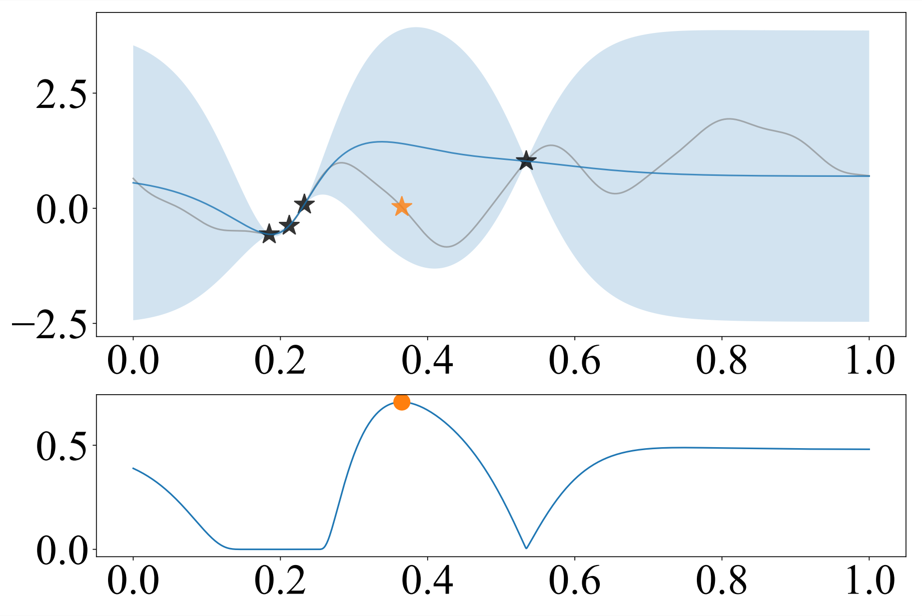
$$\mathrm{GI}_{f|D}(x; c) := \text{solution } g \text{ s.t.}$$

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

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

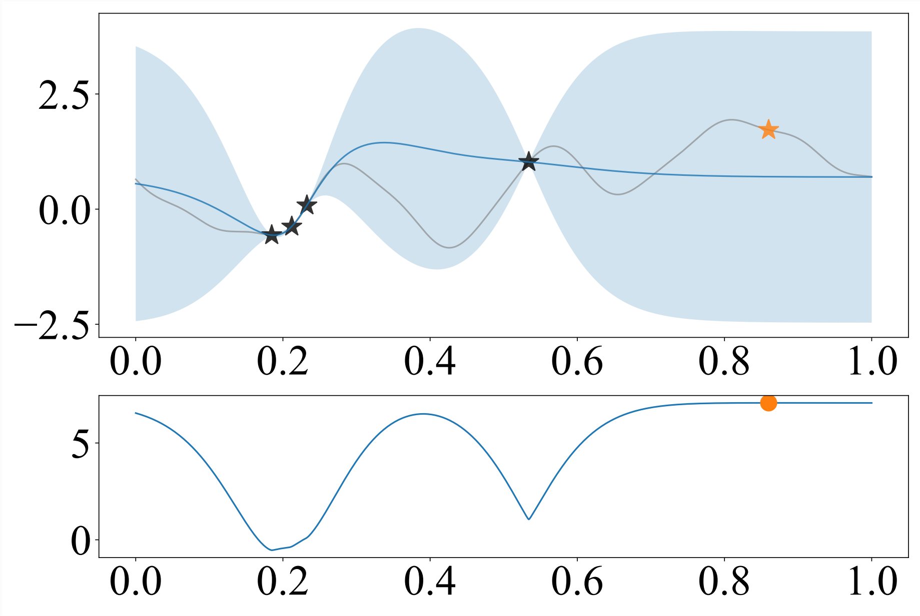
$$\Leftrightarrow \underbrace{\mathbb{E}[\max(f(x) - g, 0) \mid D]}_{\mathrm{EI}_{f|D}(x;\, g)} = c(x)$$

Expected Improvement



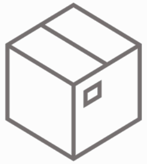
$$\mathrm{EI}_{f|D}(x) := \mathbb{E}[\max(f(x) - y_{\text{best}}, 0) \mid D]$$

Gittins Index



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

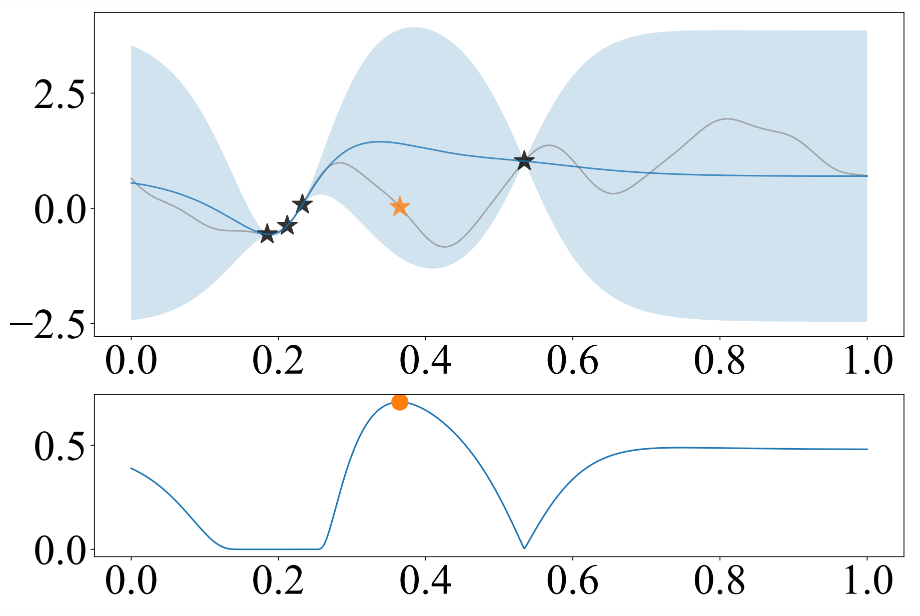
$$\text{where } \mathrm{EI}_{f|D}(x; g) := \mathbb{E}[\max(f(x) - g, 0) \mid D]$$

$f(x)$  vs. 

$c(x)$ 

Exploration Exploitation

Expected Improvement

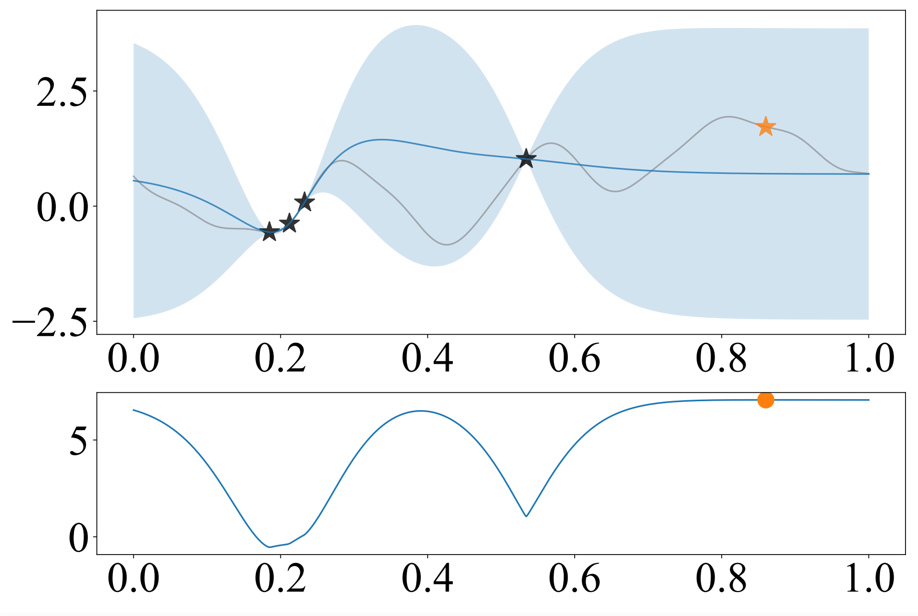


$\text{EI}_{f|D}(x) := \mathbb{E}[\max(f(x) - y_{\text{best}}, 0) \mid D]$

Temporal simplification to MDP

(One-step)

Gittins Index

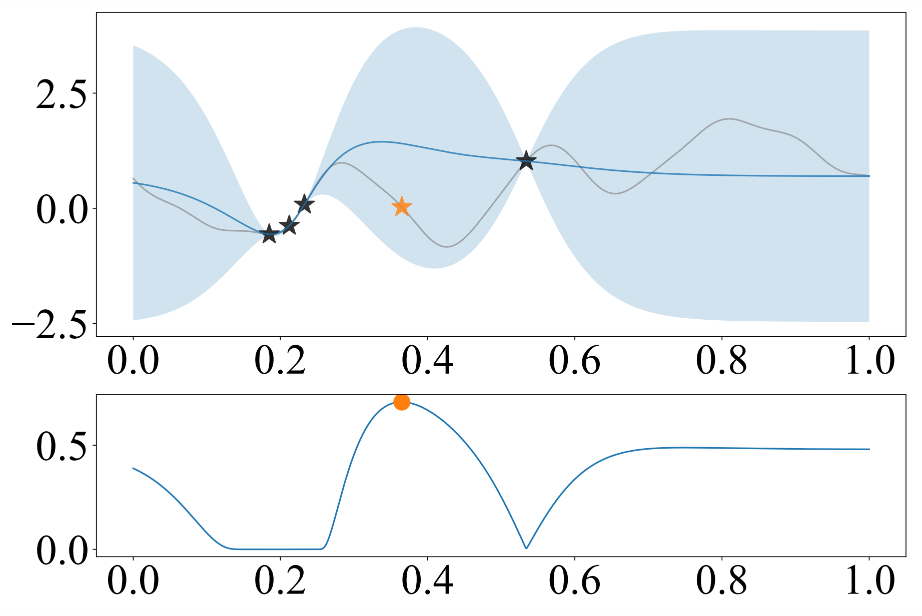


$\text{GI}_{f|D}(x; c) :=$ solution g s.t. $\text{EI}_{f|D}(x; g) = c(x)$

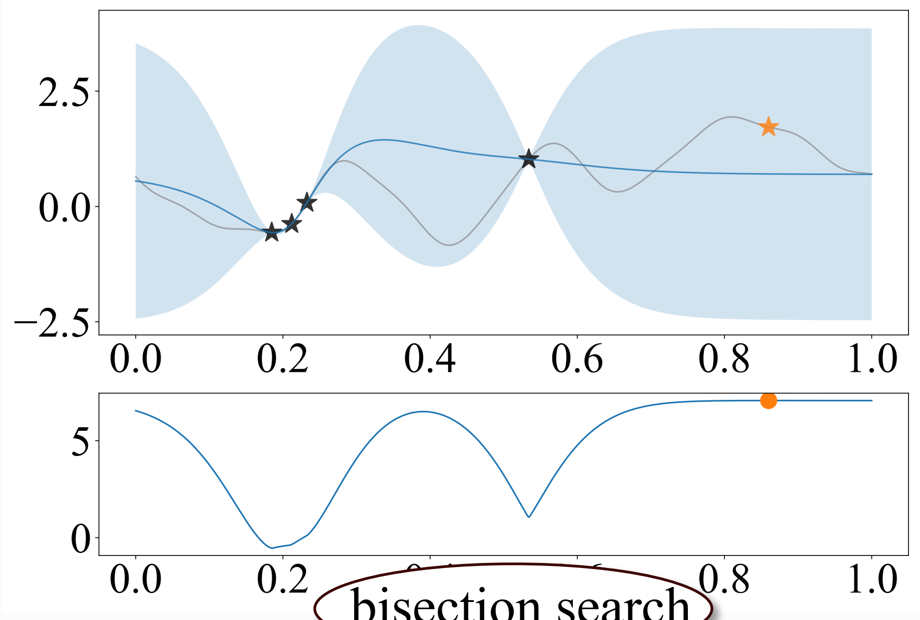
where $\text{EI}_{f|D}(x; g) := \mathbb{E}[\max(f(x) - g, 0) \mid D]$

Spatial simplification to MDP

Expected Improvement



Gittins Index



bisection search

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

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

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

analytical expression

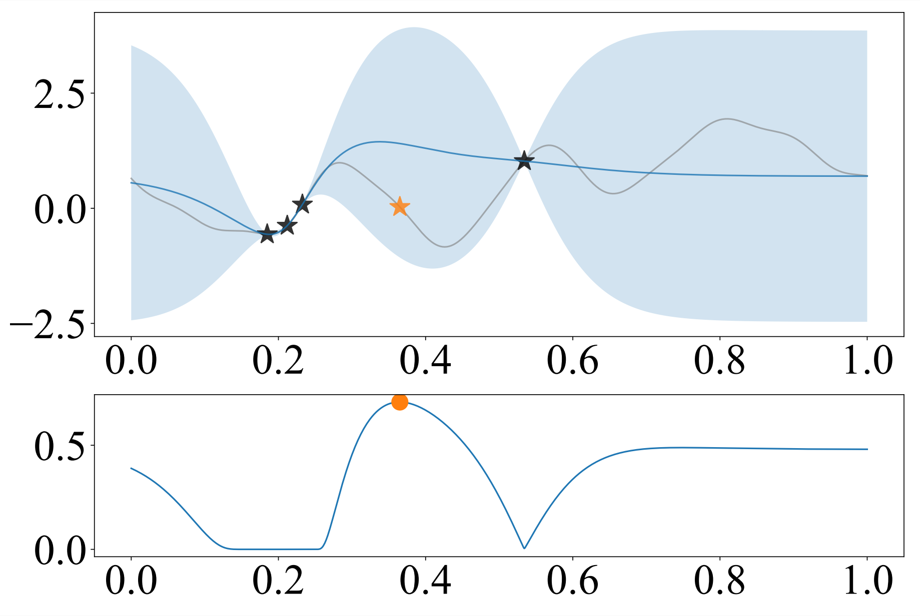
Temporal simplification to MDP

Spatial simplification to MDP

Both are principled and easy-to-compute!

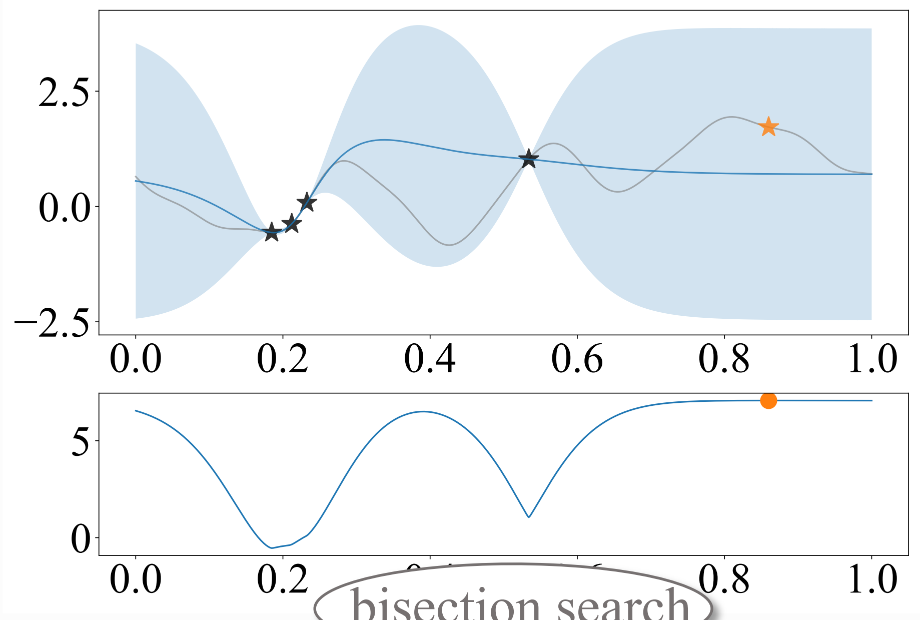
"Cost-aware Bayesian Optimization via the Pandora's Box Gittins Index." NeurIPS'24.

Expected Improvement



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

Gittins Index



bisection search

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

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

Google DeepMind

FunBO: Discovering new acquisition functions for Bayesian Optimization with FunSearch

hard to discover GI

Our Contribution: Gittins Index Principle

Joint work with Ziv Scully and Alexander Terenin et al.

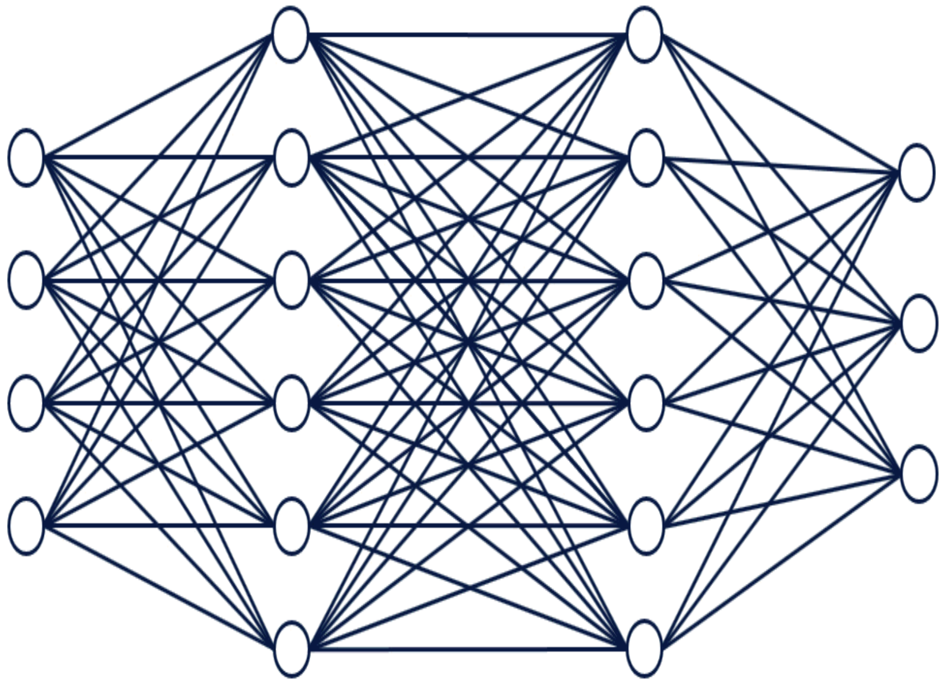
	EI
1. Principled easy-to-compute decision rules	✓
2. Natural incorporation of side info and flexibility	✗
3. Competitive performance on benchmarks	✓✗
4. Theoretical guarantees	✓✗

Under-explored Side Info and Practical Flexibility

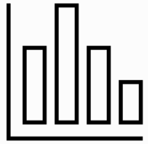
Varying evaluation costs

Smart stopping time

Observable multi-stage feedback



How does existing principle incorporate them?



Varying evaluation costs

$$\mathrm{EIPC}(x;c) = \mathrm{EI}(x)/c(x)$$

[Snoek et al.'12]

Arbitrarily bad

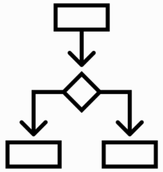
[Astudillo et al.'21]

Smart stopping time

$$\tau\colon\ \mathrm{EI}(x_\tau) \le \theta$$

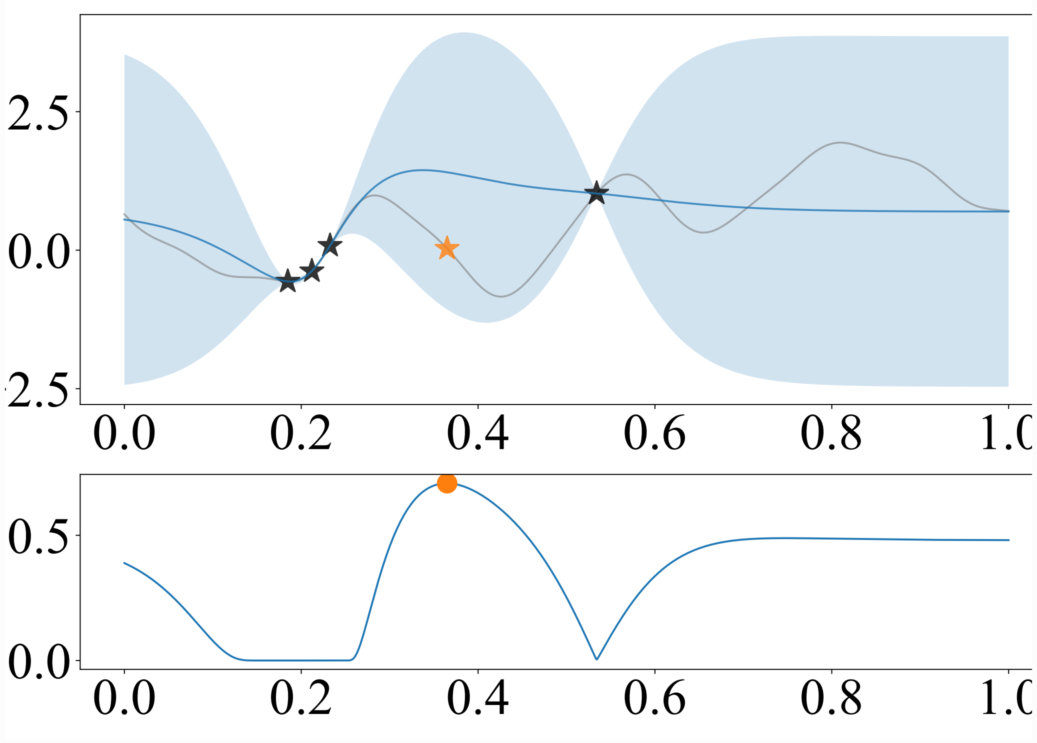
[Locatelli'97,
Nguyen et al.'17,
Ishibashi et al.'23]

Which threshold?



Observable multi-stage feedback

?



Expected improvement $\mathrm{EI}(x)$

Under-explored Side Info and Practical Flexibility

- Varying evaluation costs
- Smart stopping time
- Observable multi-stage feedback

New design principle:
Gittins index

Why Gittins index?

Varying evaluation costs

Smart stopping time

Observable multi-stage feedback

New design principle:
Gittins index

Optimal in related sequential decision problems

Why Gittins index?

Varying evaluation costs

Features in Pandora's box

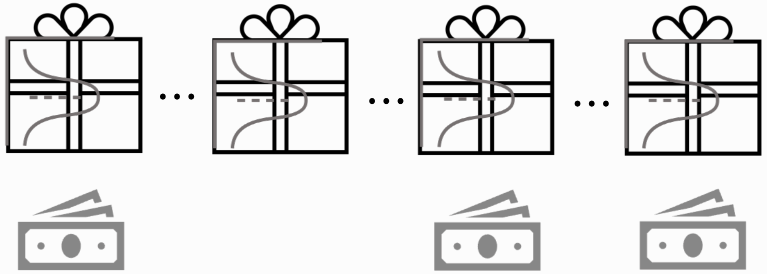
Smart stopping time

Features in Pandora's box

Observable multi-stage feedback

New design principle:
Gittins index

Optimal in related sequential decision problems



Why Gittins index?

Varying evaluation costs

Features in Pandora's box

Smart stopping time

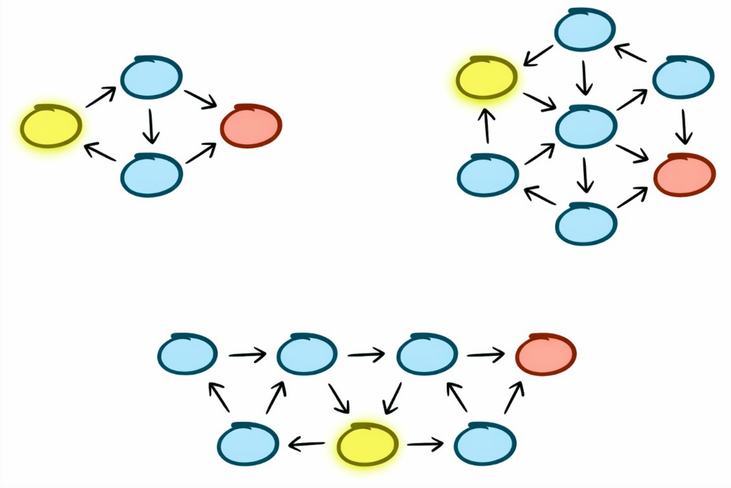
Features in Pandora's box

Observable multi-stage feedback

Features in Markov chain selection

New design principle:
Gittins index

Optimal in related sequential decision problems



Why Gittins index?

Varying evaluation costs

Features in Pandora's box

This talk's focus

Smart stopping time

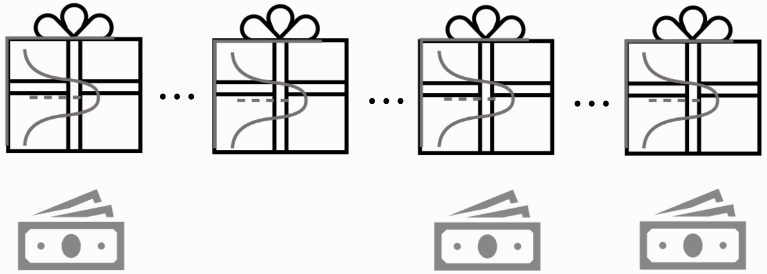
Features in Pandora's box

Observable multi-stage feedback

Features in Markov chain selection

New design principle:
Gittins index

Optimal in related sequential decision problems

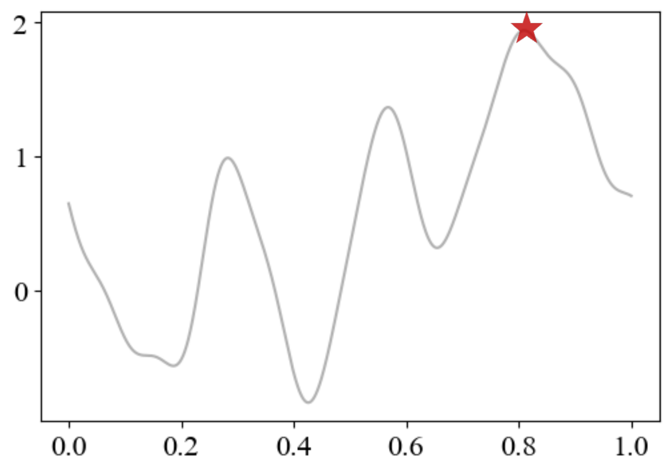


"Cost-aware Bayesian Optimization via the Pandora's Box Gittins Index." NeurIPS'24.

"Cost-aware Stopping for Bayesian Optimization." Under review.

Bayesian Optimization

$f(x)$



Continuous

Correlated

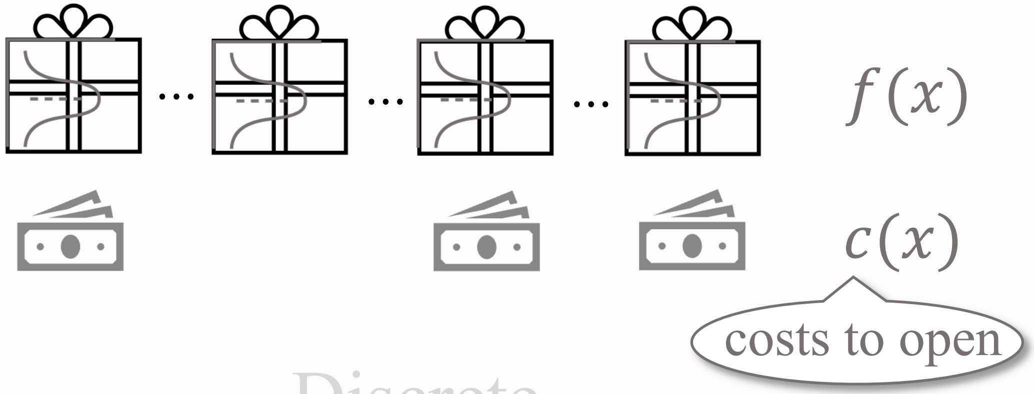
Fixed-iteration

Cost-unaware

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

Pandora's Box

[Weitzman'79]



Discrete

Independent

Flexible-stopping

Cost-aware

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

Expected Improvement vs Gittins Index

Varying evaluation costs

$\mathrm{EIPC}(x;c) = \mathrm{EI}(x)/c(x)$

Arbitrarily bad

$\mathrm{GI}(x;c) := \text{solution } g \text{ s.t. } \mathrm{EI}(x;g) = c(x)$

naturally incorporates costs

Smart stopping time

$\tau\colon\ \mathrm{EI}(x_\tau) \le \theta$

Which threshold?

$\tau\colon\ \mathrm{GI}(x_\tau;c) \le y_{\text{best}}$

$\Leftrightarrow \tau\colon \mathrm{EIPC}(x_\tau;c) \le 1$

derived shared stopping rule

"Cost-aware Bayesian Optimization via the Pandora's Box Gittins Index." NeurIPS'24.

"Cost-aware Stopping for Bayesian Optimization." Under review.

Our Contribution: Gittins Index Principle

Joint work with Ziv Scully and Alexander Terenin et al.

EI

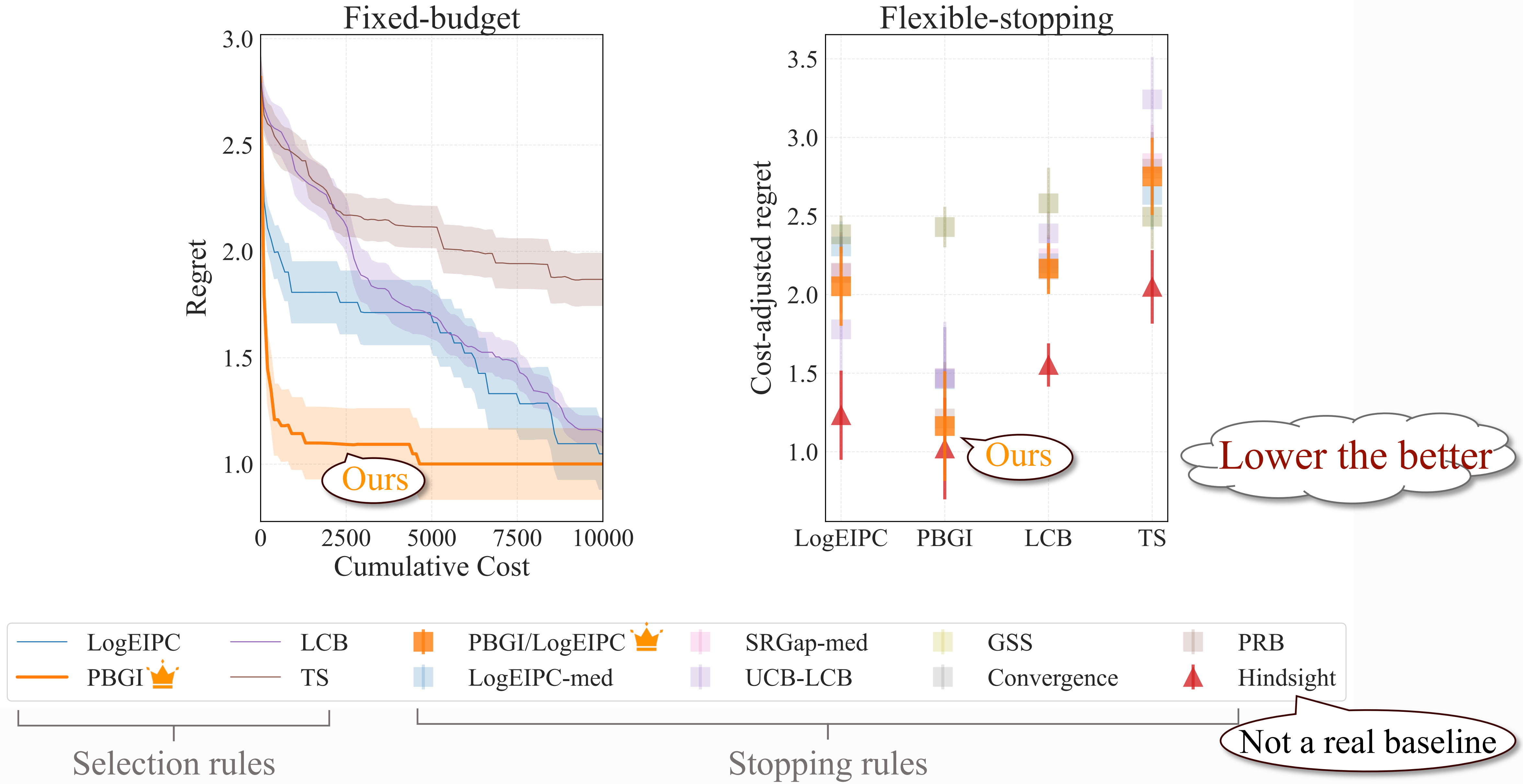
1. Principled easy-to-compute decision rules ✓

2. Natural incorporation of side info and flexibility ✗

3. Competitive performance on benchmarks ✓✗

4. Theoretical guarantees ✓✗

Gittins Index vs Baselines on AutoML Benchmark



Our Contribution: Gittins Index Principle

Joint work with Ziv Scully and Alexander Terenin et al.

EI

1. Principled easy-to compute decision rules ✓

2. Natural incorporation of side info and flexibility ✗

3. Competitive performance on benchmarks

4. Theoretical guarantees

Theoretical Guarantee and Empirical Validation

Theorem (Safeguard Guarantee)

$\mathbb{E}[R(\text{ours}; \text{PBGI})] \leq R[\text{stopping immediately}]$

or LogEIPC

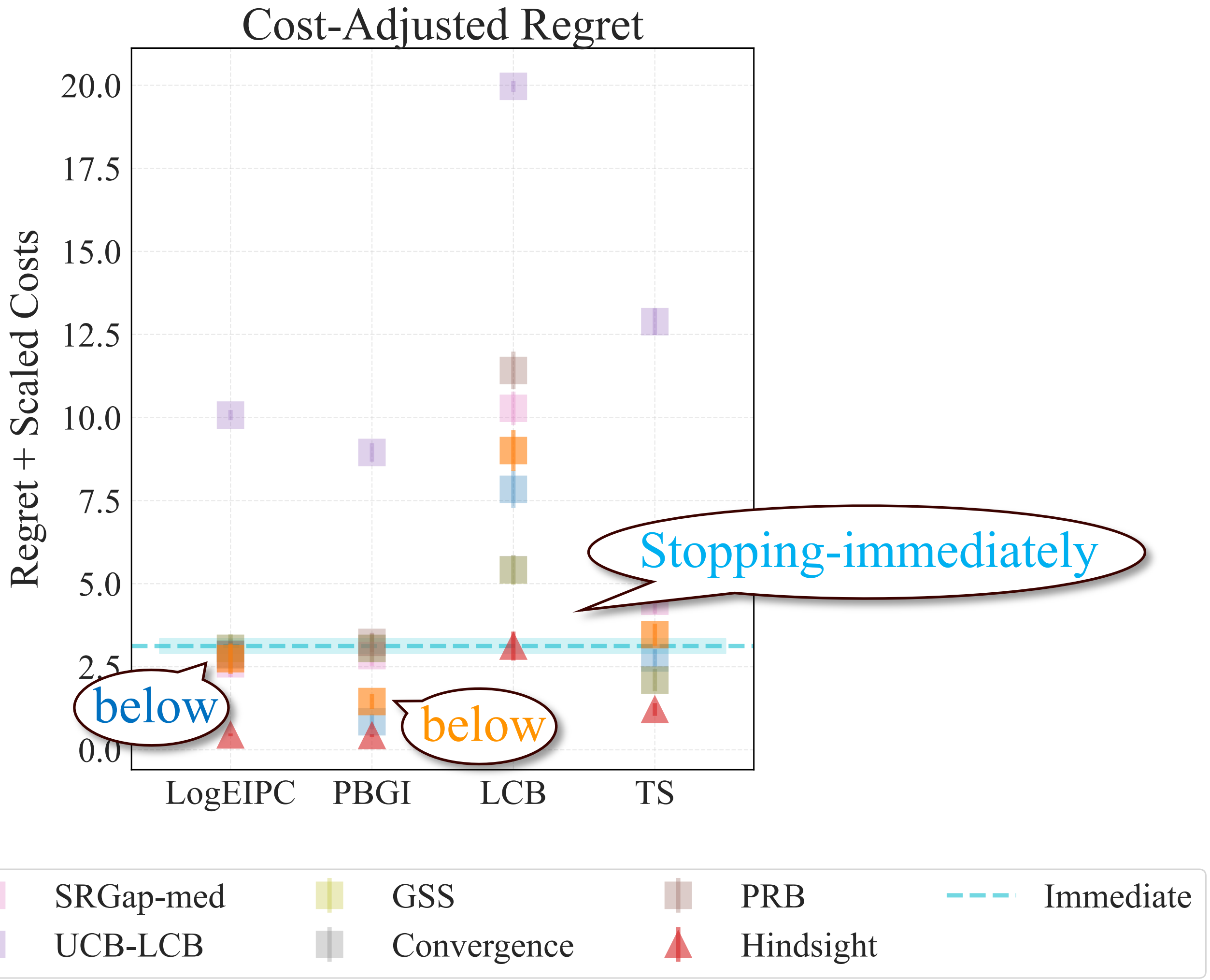
cost-adjusted regret

Implication:

- Matches the best achievable performance in the worst case (evaluations are all very costly).
- Avoids over-spending — a property many cost-unaware stopping rules lack.

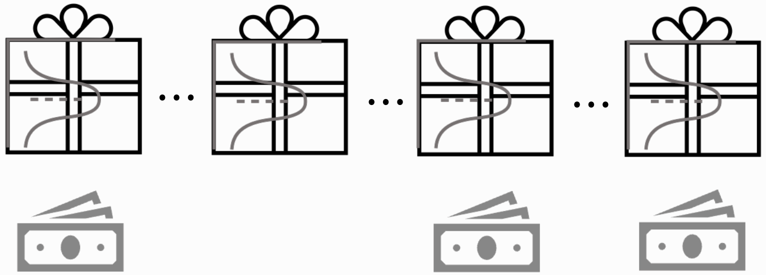
Proof idea: For all $t < \tau$, $\text{EI}(x_{t+1}) \geq c(x_{t+1})$.

stopping time



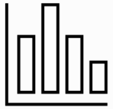
"Cost-aware Stopping for Bayesian Optimization." Under review.

Principled method design

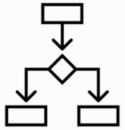


Link to Pandora's Box problem
& Gittins index theory

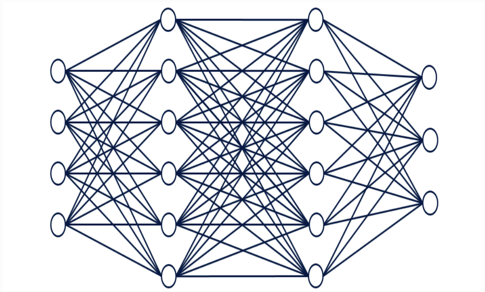
Natural incorporation of side info

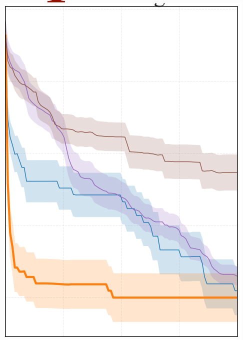

Varying evaluation costs


Adaptive stopping time


Multi-stage feedback

Competitive empirical performance





Interests from practitioners (e.g., Meta)


"Cost-aware Bayesian Optimization via the Pandora's Box Gittins Index." NeurIPS'24.

Theoretical guarantees

$$\mathbb{E}[R(\text{ours; PBGI})] \leq R[\text{stopping immediately}]$$

Ongoing: regret bound


"Cost-aware Stopping for Bayesian Optimization." Under review.