

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

To appear in NeurIPS'24

Qian Xie (Cornell ORIE)

Joint work with Raul Astudillo, Peter Frazier, Ziv Scully, and Alexander Terenin

INFORMS'24 Data Mining Best General Paper Competition

Coauthors



Raul Astudillo



Peter Frazier



Ziv Scully

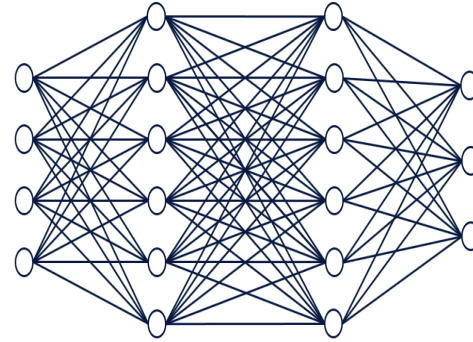


Alexander Terenin

World of Parameter Optimization

Hyperparameter tuning:

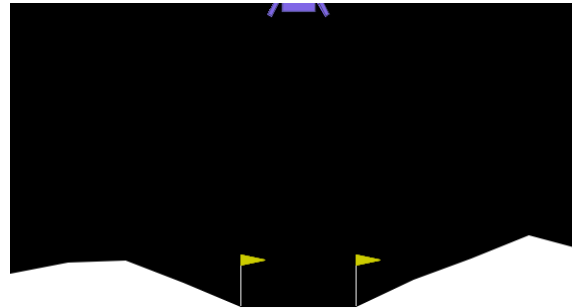
Training parameters →



→ Accuracy

Control optimization:

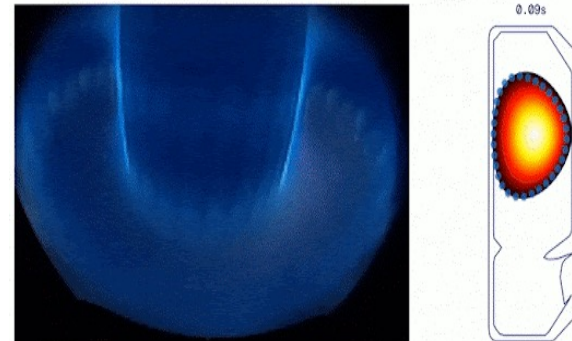
Control parameters →



→ Reward

Plasma physics:

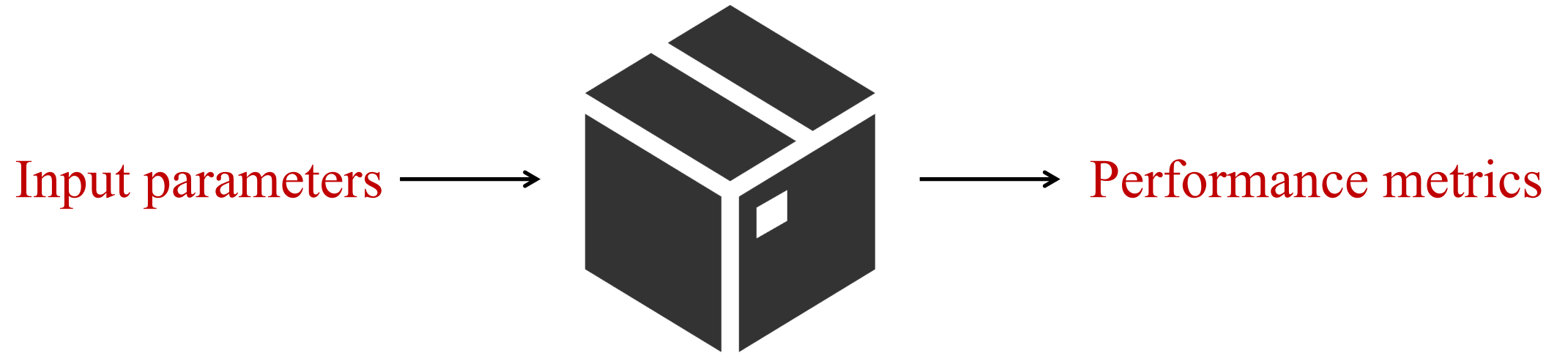
Reactor parameters →



→ Stability

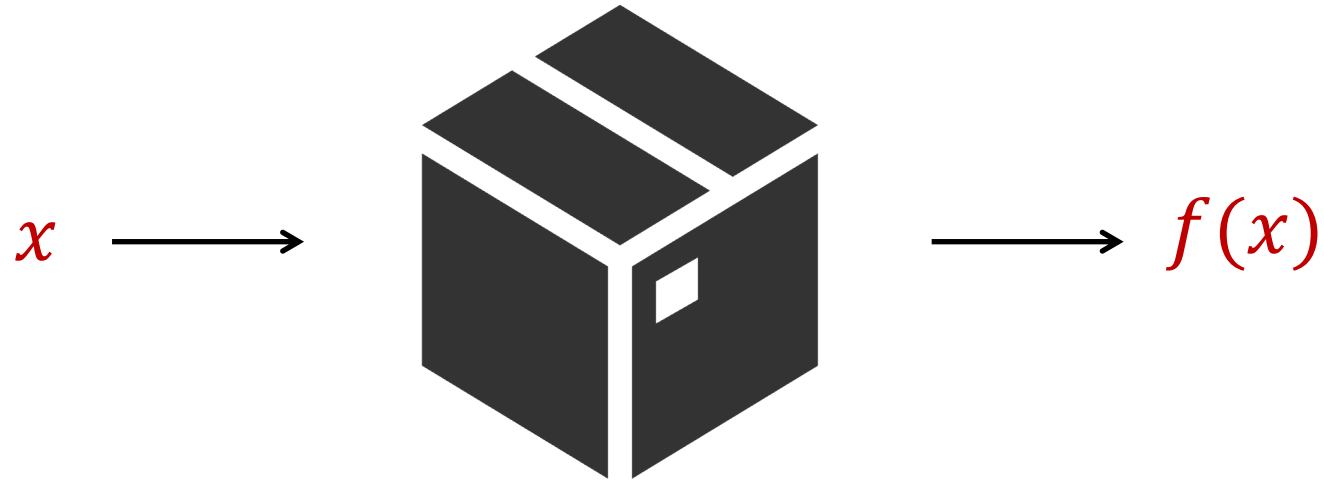
World of Parameter Optimization

Black-box optimization:

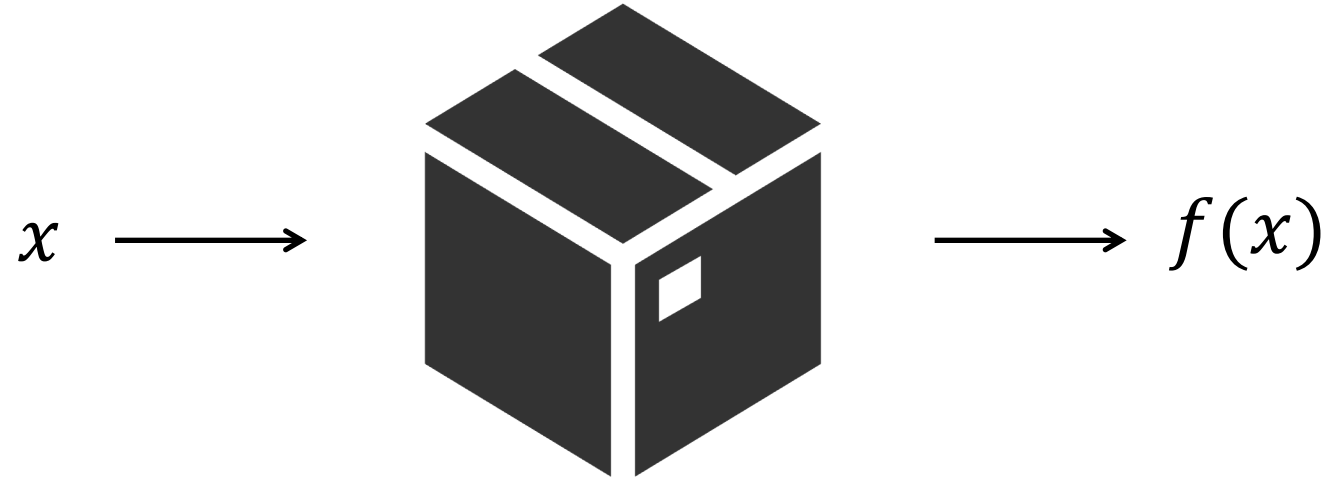


World of Parameter Optimization

Black-box optimization:

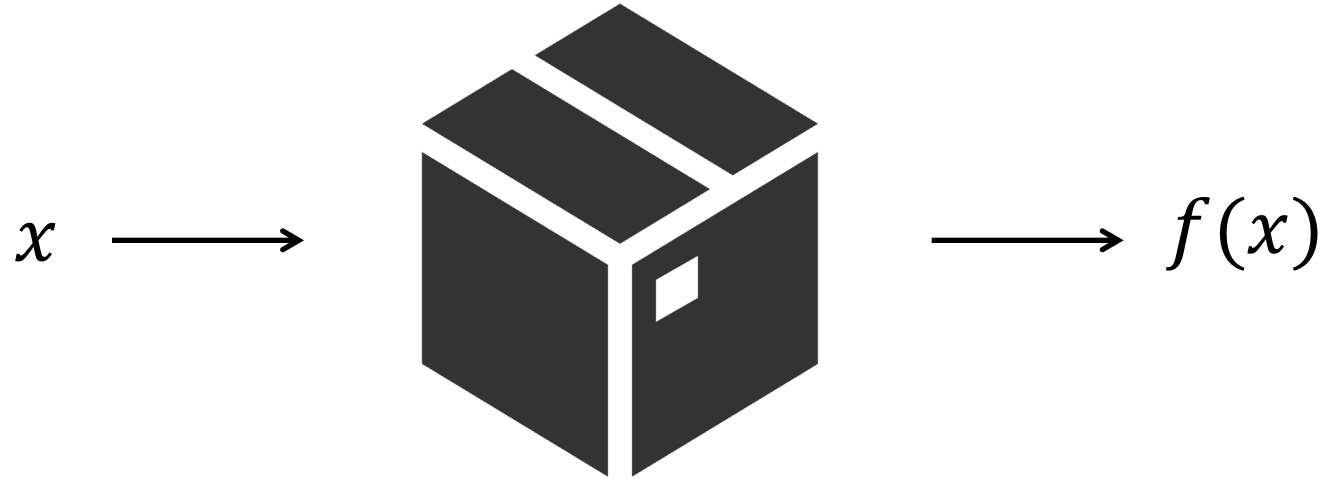


Optimizing Black-box Functions



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

Optimizing Black-box Functions



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

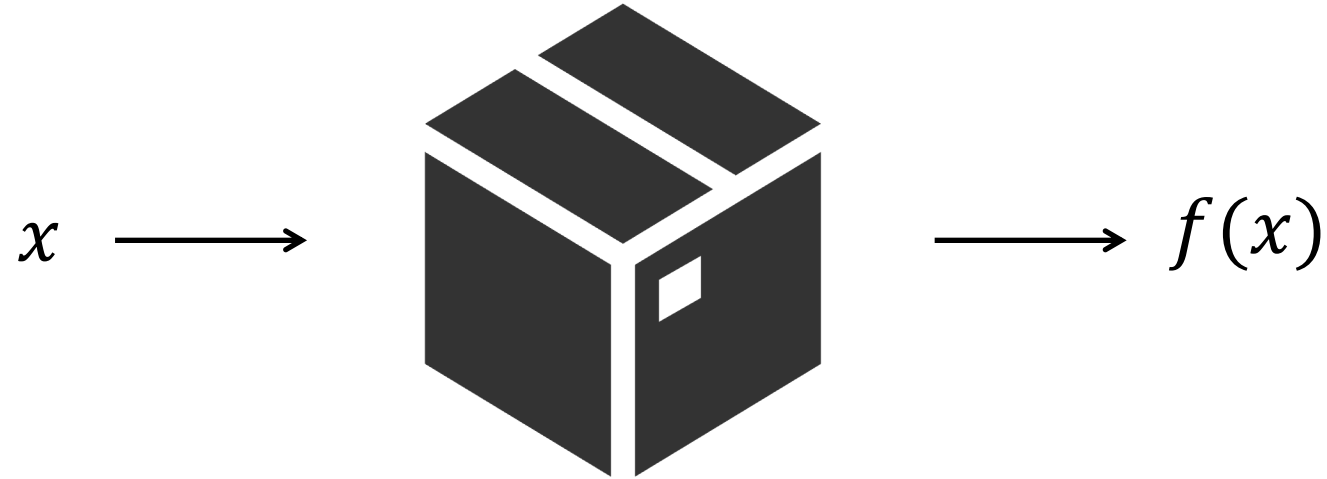
$f \sim \text{Stochastic Process}$

Optimizing Black-box Functions



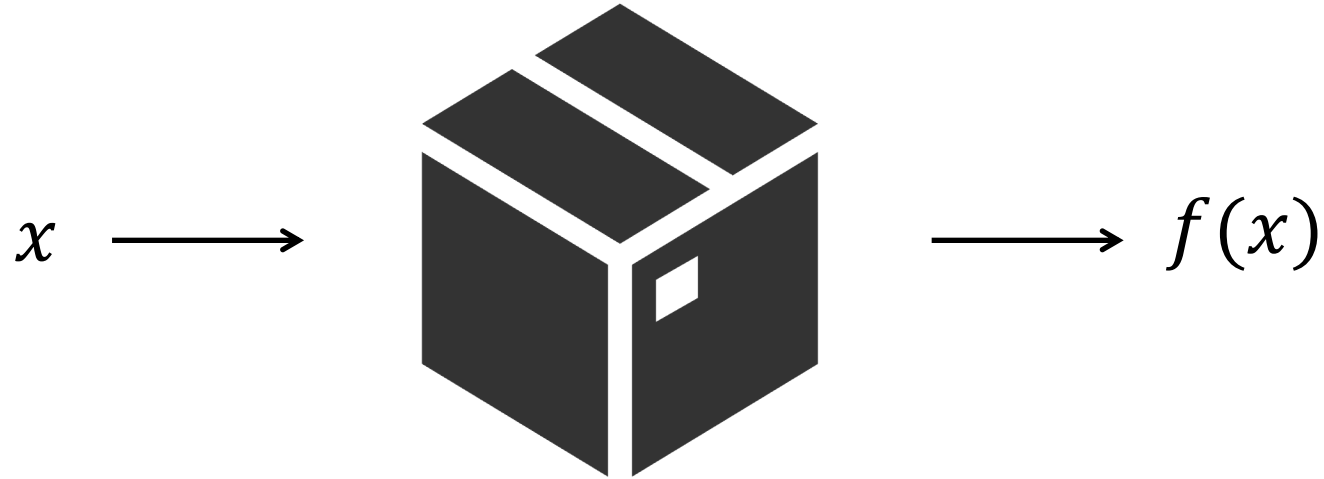
Goal: $\max_{x \in \mathcal{X}} f(x)$

Optimizing Black-box Functions



$$\text{Goal: } \mathbf{\max} \mathbb{E} \max_{t=1,2,\dots,\textcolor{red}{T}} f(x_t)$$

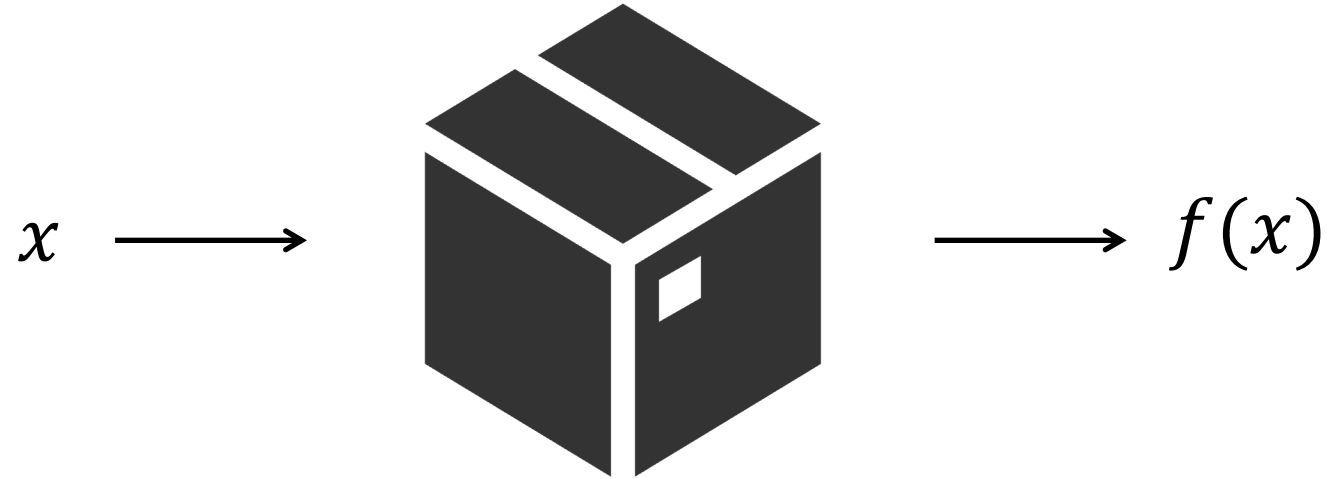
Optimizing Black-box Functions



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

$f \sim \text{Gaussian Process}$

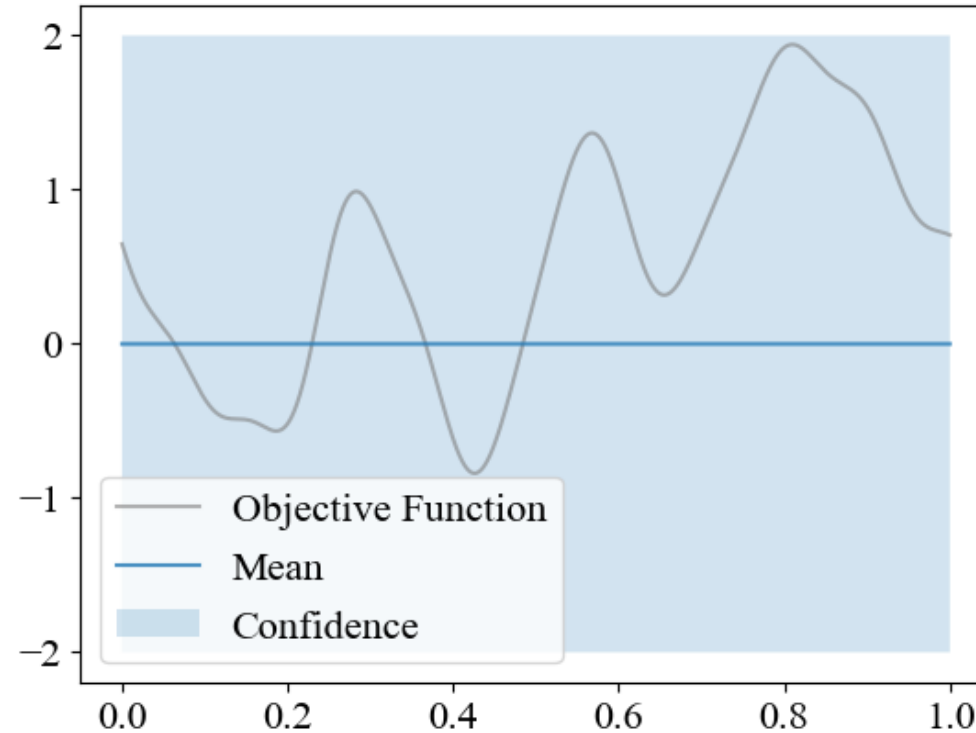
Bayesian Optimization



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

$f \sim \text{Gaussian Process}$

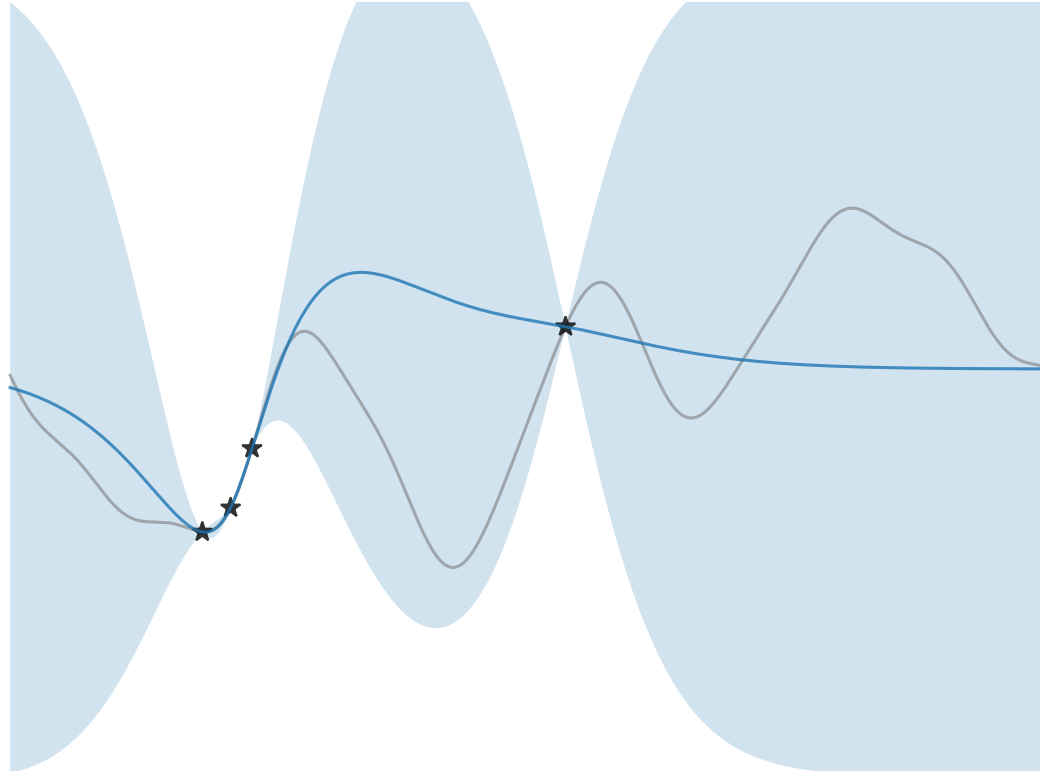
Bayesian Optimization



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

$f \sim \text{Gaussian Process}$

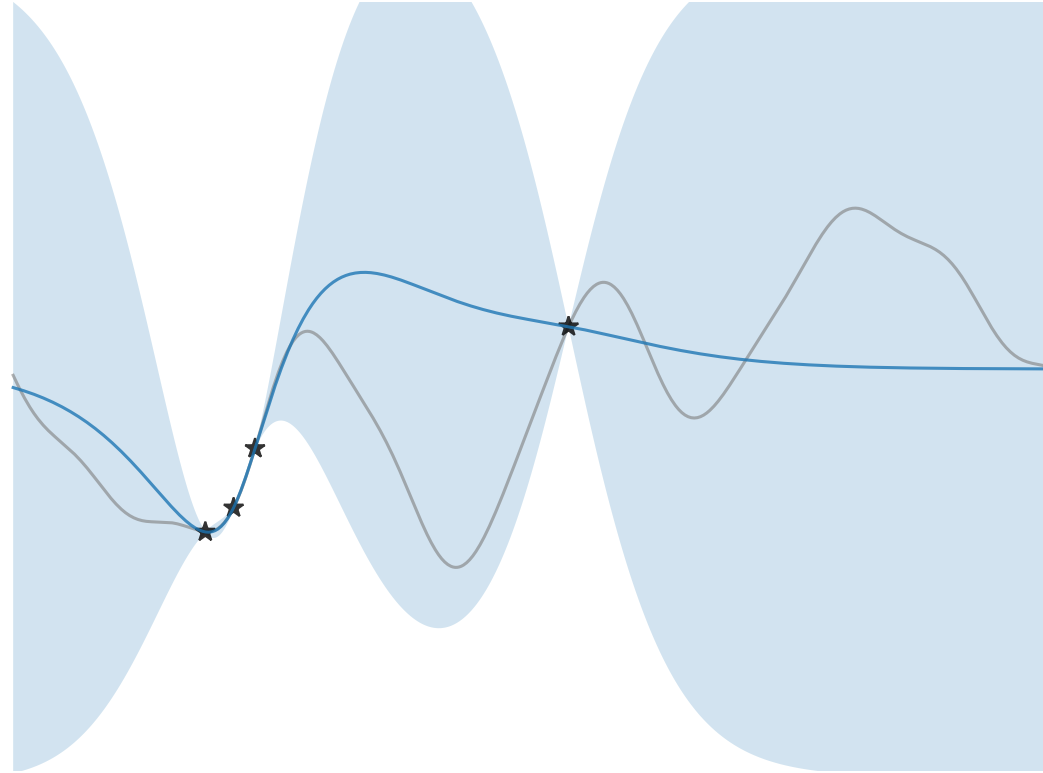
Bayesian Optimization



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

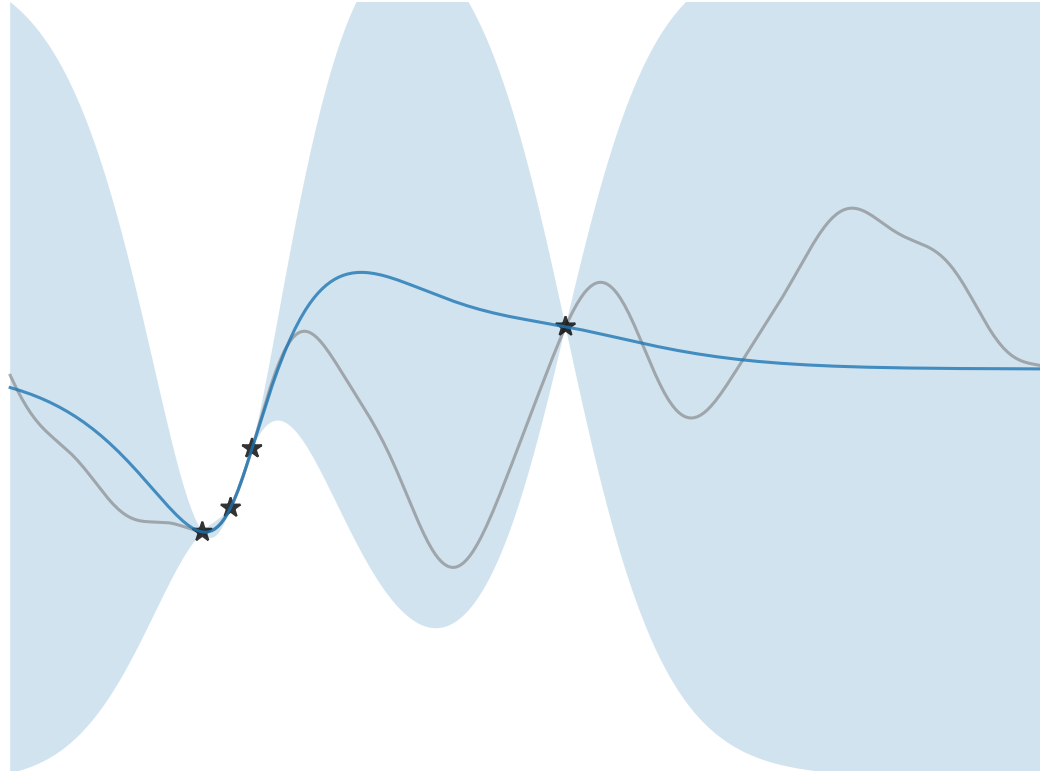
$f \sim \text{Gaussian Process}$

Bayesian Optimization



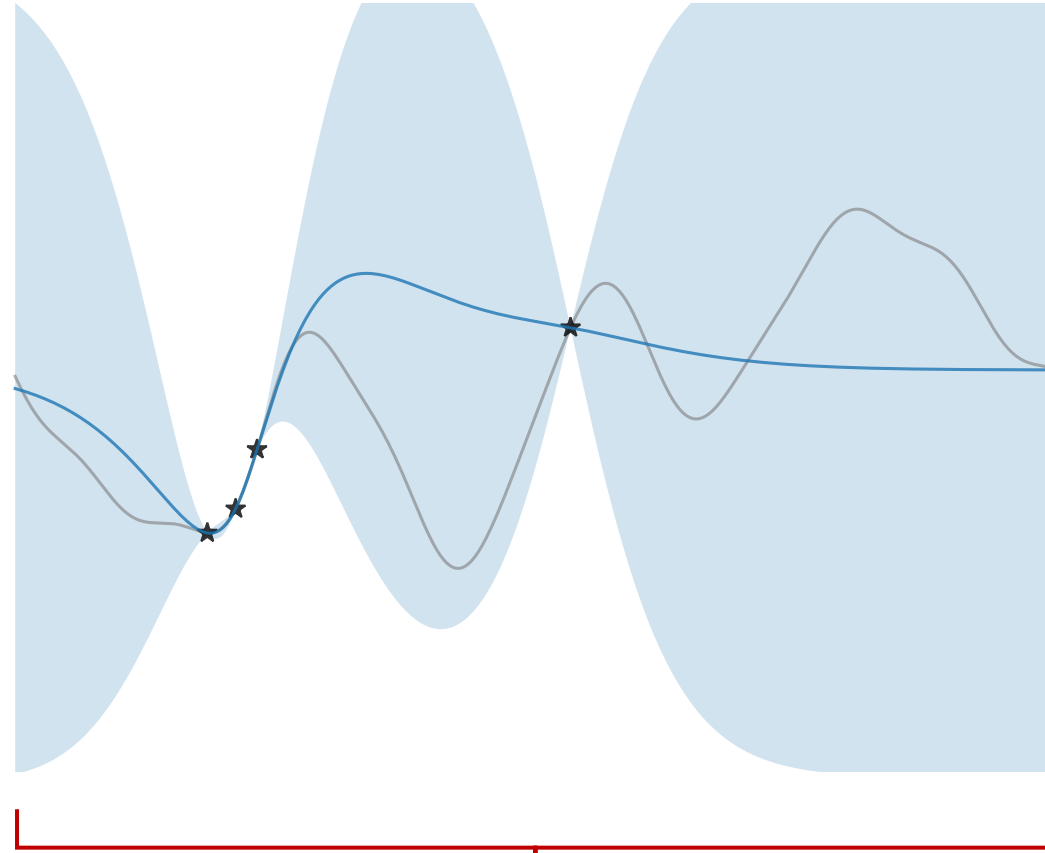
What to evaluate next?

Bayesian Optimization



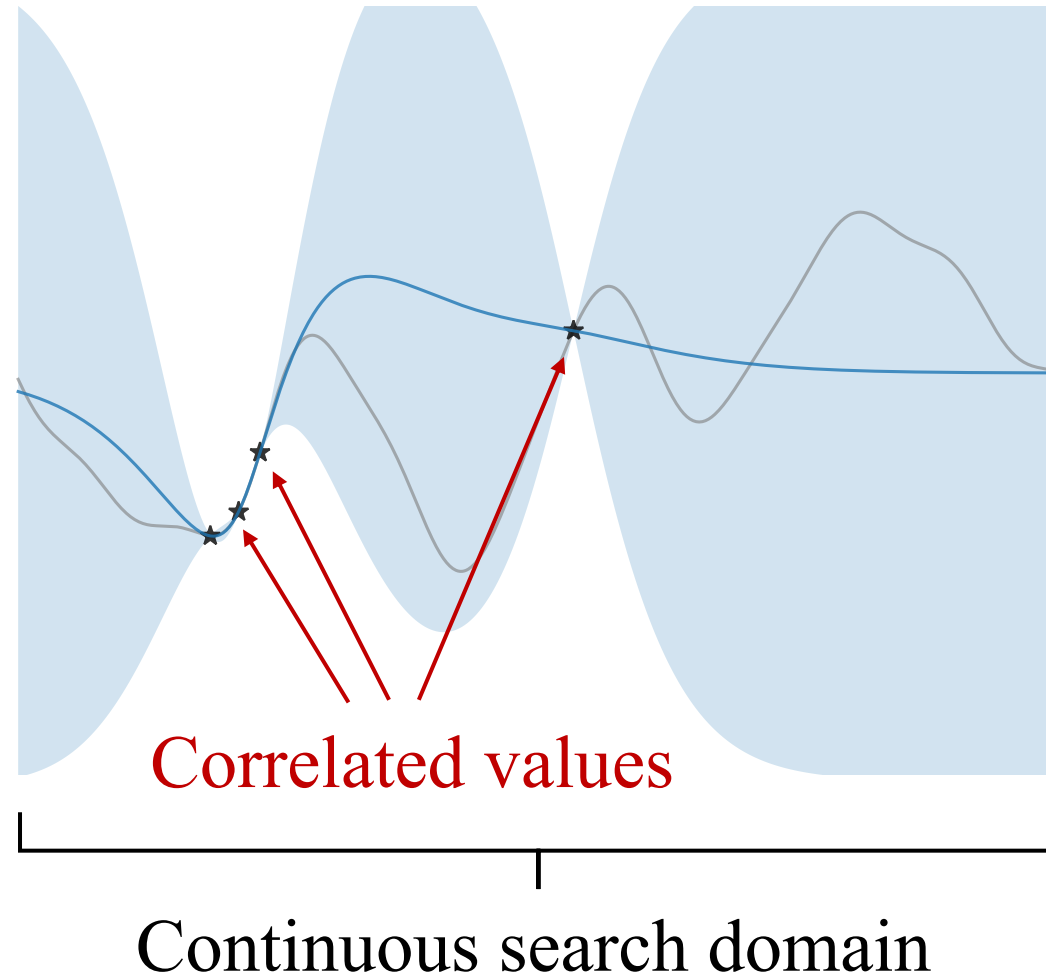
Optimal policy?

Challenges of Bayesian Optimization

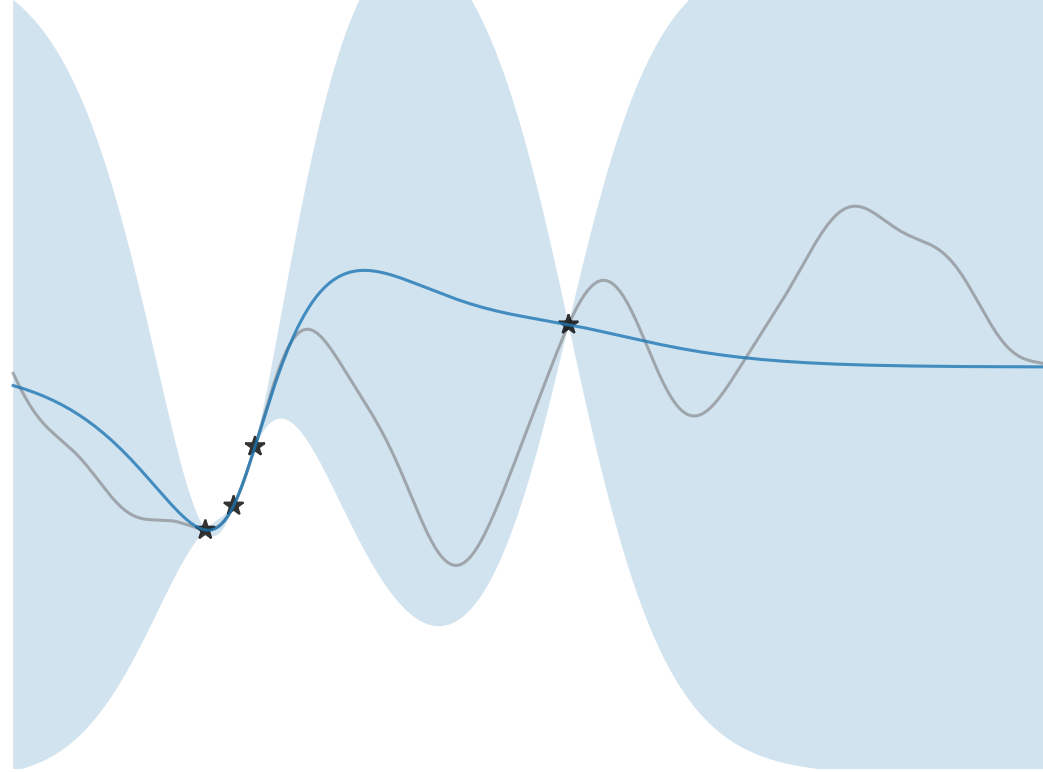


Continuous search domain

Challenges of Bayesian Optimization

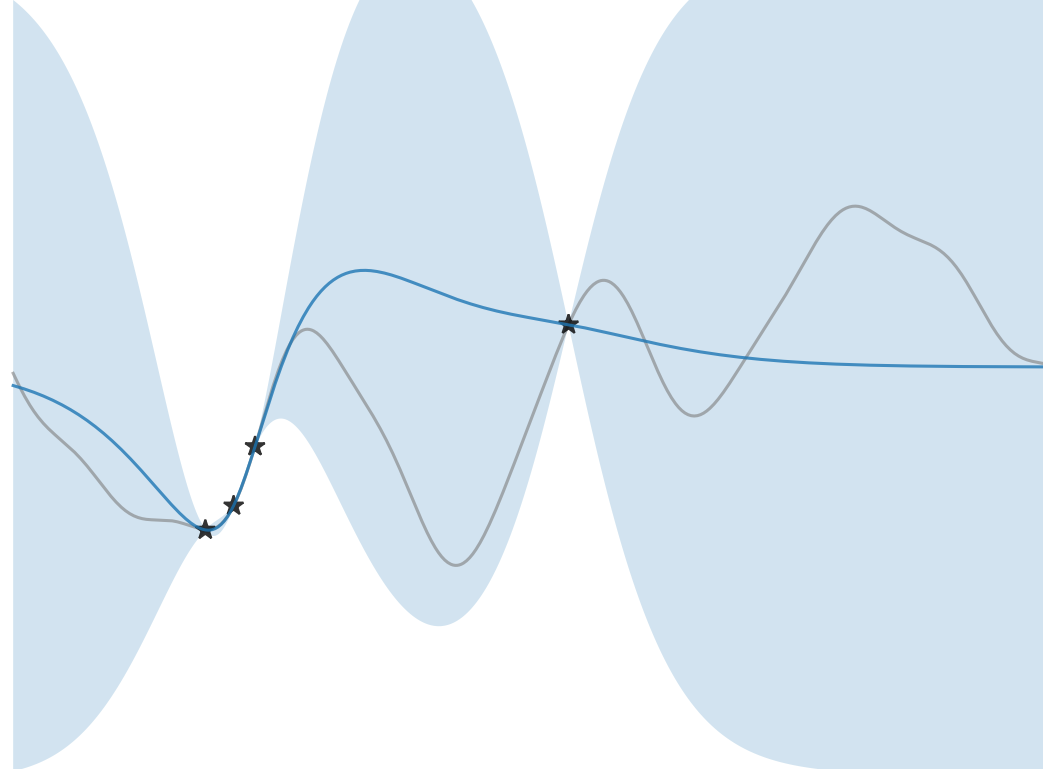


Challenges of Bayesian Optimization



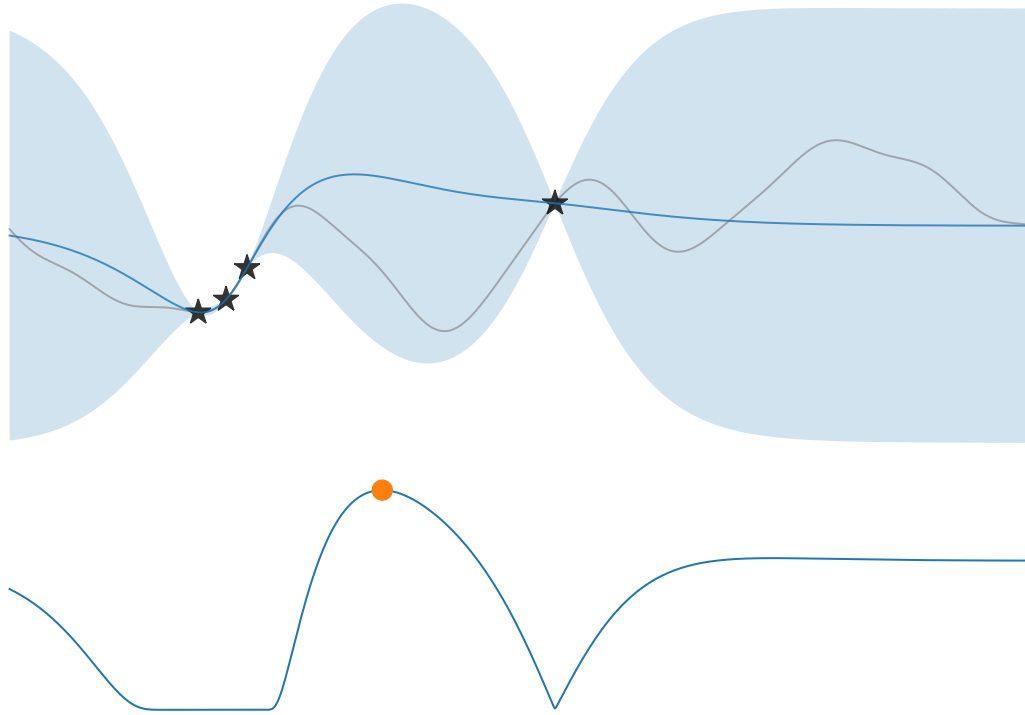
Correlation & continuity \Rightarrow **Intractable MDP**

Challenges of Bayesian Optimization



Intractable MDP \Rightarrow Optimal policy unknown

Popular Policy: Expected Improvement

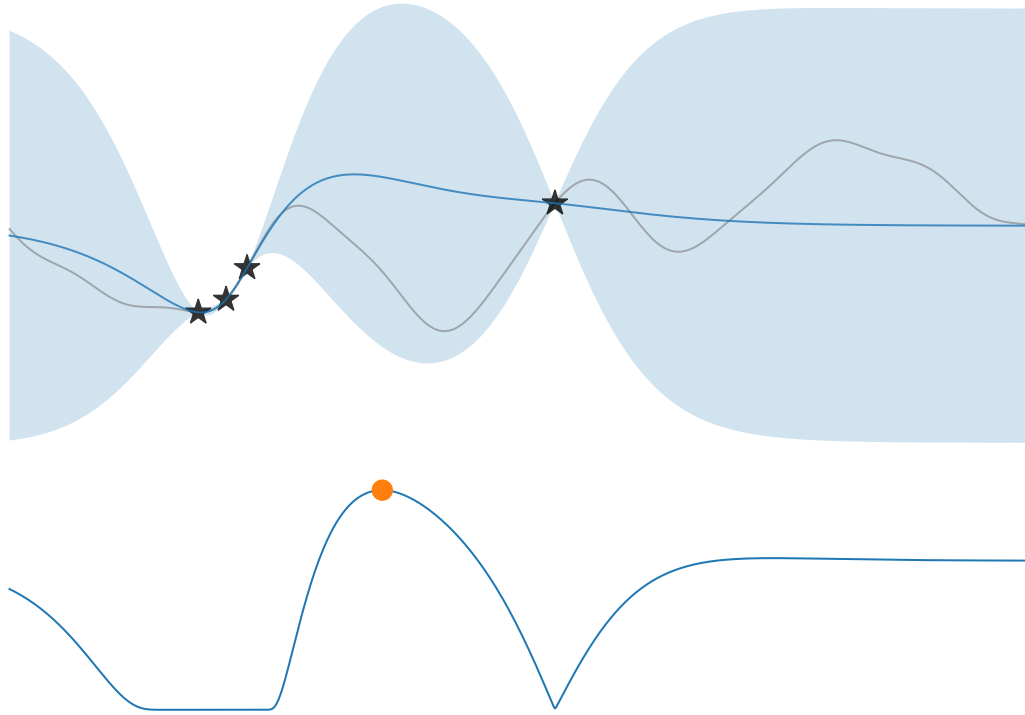


$$\text{EI}(x) = \mathbb{E}[\underbrace{\max(f(x) - y_{\text{best}}, 0)}_{\text{"improvement"}} \mid D]$$

current best observed

data

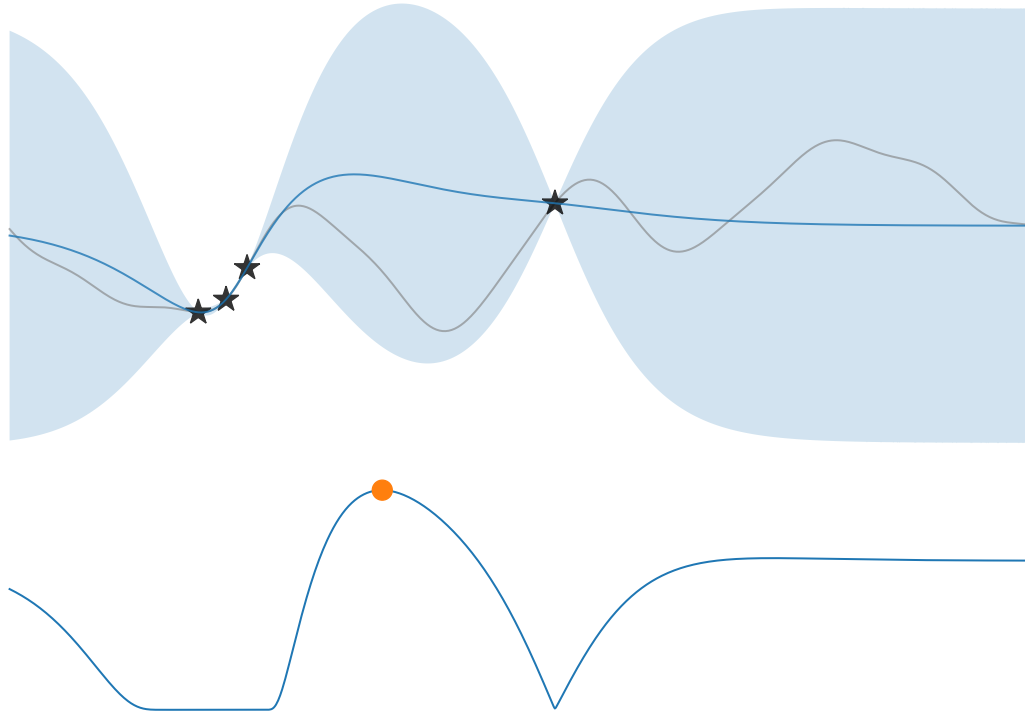
Popular Policy: Expected Improvement



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

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

Popular Policy: Expected Improvement

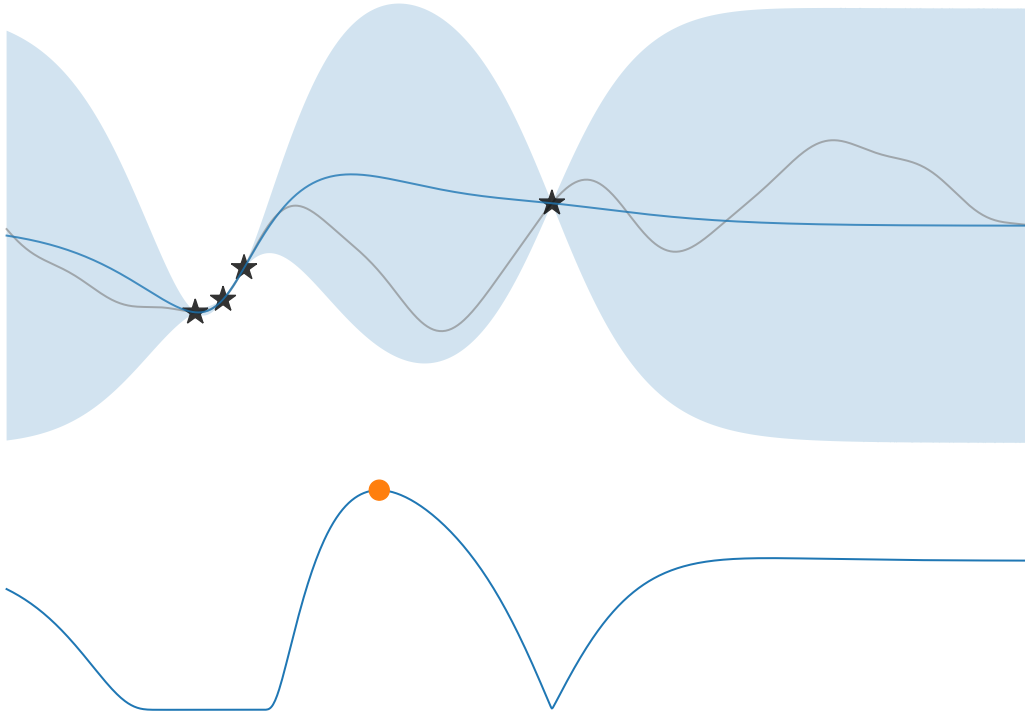


$$\text{EI}(x) = \mathbb{E}[\underbrace{\max(f(x) - y_{\text{best}}, 0)}_{\text{"improvement"}} \mid \underbrace{D}_{\text{data}}]$$

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

One-step approximation to MDP

Popular Policy: Expected Improvement



Other improvement-based policy:

- 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

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

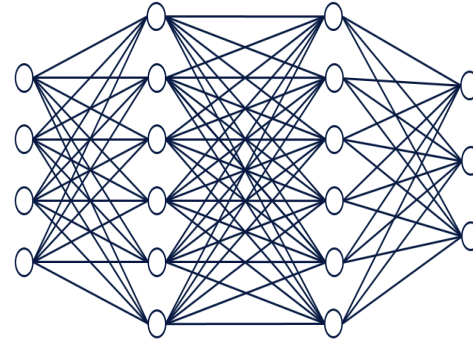
- Improvement-based
- Entropy-based
- Upper Confidence Bound
- Thompson Sampling
- Our work: Gittins Index

Why another approach?

Challenge: Varying Evaluation Costs

Hyperparameter tuning:

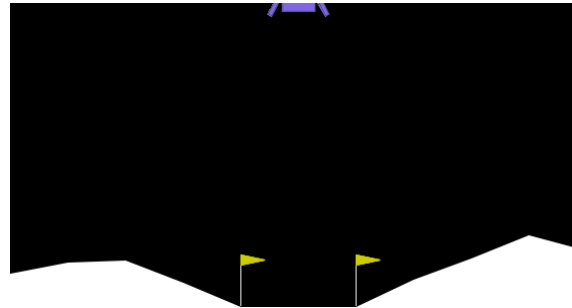
Training parameters →



→ Accuracy

Control optimization:

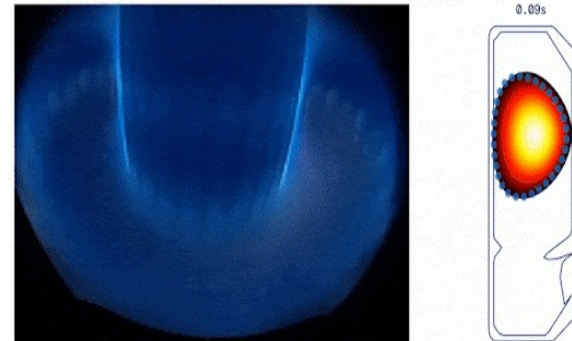
Control parameters →



→ Reward

Plasma physics:

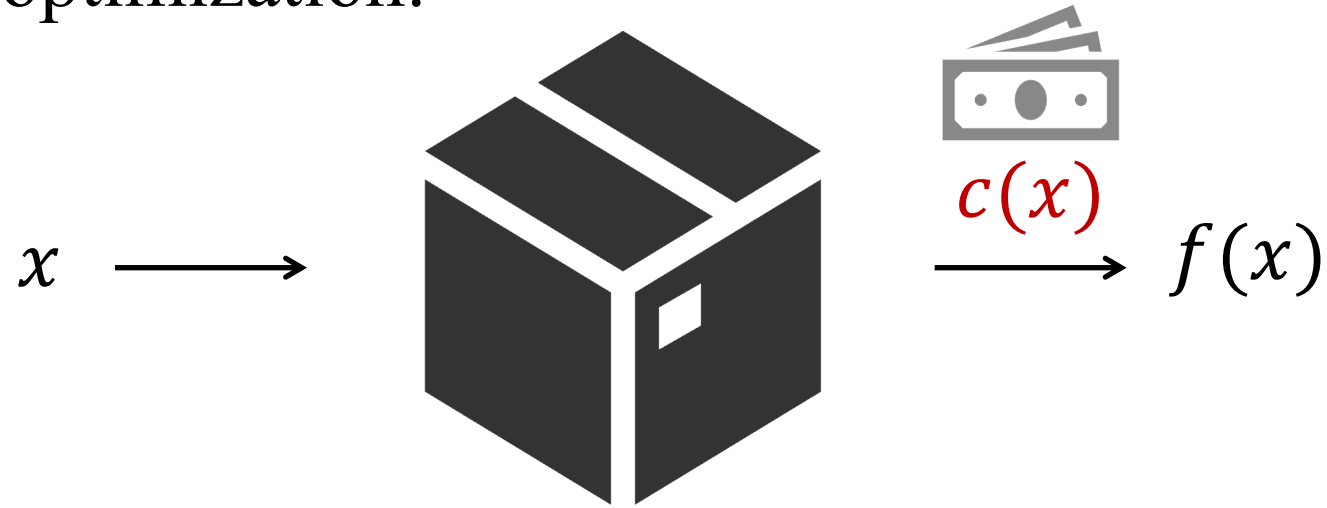
Reactor parameters →



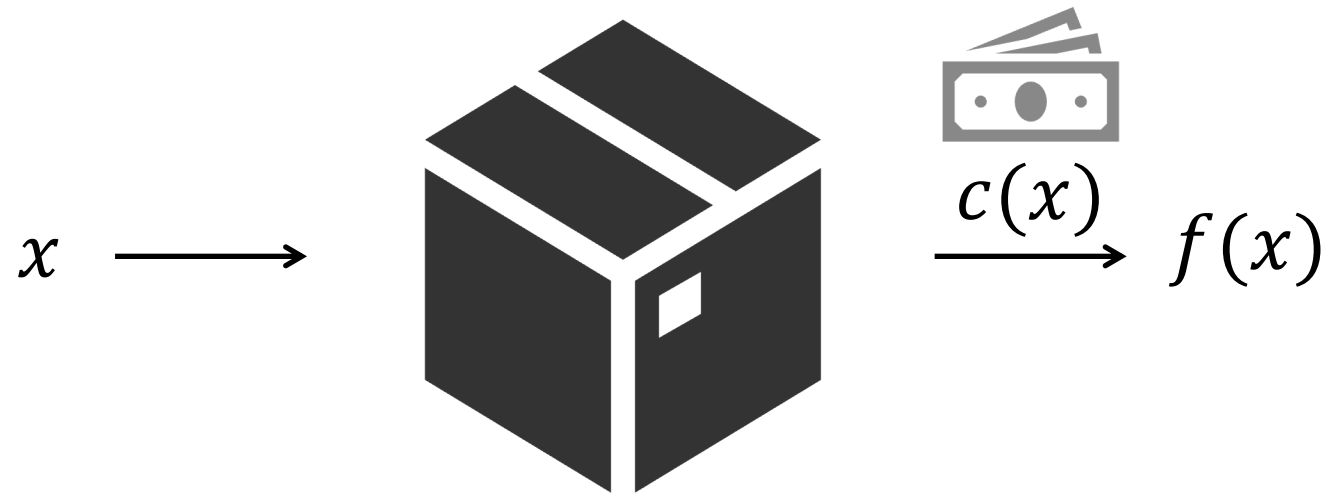
→ Stability

Challenge: Varying Evaluation Costs

Black-box optimization:

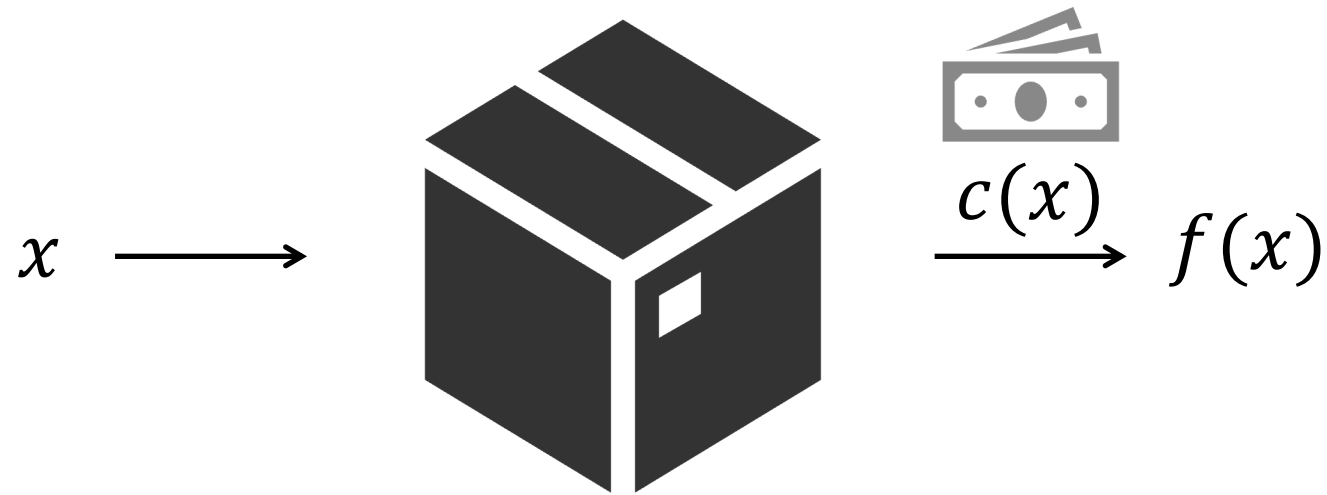


Challenge: Varying Evaluation Costs



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

Cost-aware Bayesian Optimization



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

[Lee, Perrone, Archambeau, Seeger'21]

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

Cost-aware Bayesian Optimization

Uniform costs

Varying costs

Expected improvement

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

Cost-aware Bayesian Optimization

Uniform costs

Expected improvement

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

Varying costs

Expected improvement per cost

$$\max_x \text{EI}_{f|D}(x; y_{\text{best}}) / c(x)$$

Cost-aware Bayesian Optimization

Uniform costs

Expected improvement

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

Varying costs

Expected improvement per cost

$$\max_x \text{El}_{f|D}(x; y_{\text{best}})/c(x)$$

Why divide?

Cost-aware Bayesian Optimization

Uniform costs

Expected improvement

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

Varying costs

Expected improvement per cost

$$\max_x \text{El}_{f|D}(x; y_{\text{best}})/c(x)$$

Why divide?

Our view: lack of a guidance to incorporate costs

Cost-aware Bayesian Optimization

Uniform costs

Expected improvement

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

Varying costs

Expected improvement per cost

$$\max_x \text{EI}_{f|D}(x; y_{\text{best}})/c(x)$$

EI and EIPC policy can be **arbitrarily bad** under varying costs!

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

Cost-aware Bayesian Optimization

Uniform costs

Expected improvement

Multi-step Lookahead EI

Varying costs

Expected improvement per cost

Budgeted Multi-step Lookahead EI

slow

Cost-aware Bayesian Optimization

Uniform costs

Expected improvement

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

Expected improvement

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

Cost-aware Bayesian Optimization

Uniform costs

Expected improvement

Multi-step Lookahead EI

Upper Confidence Bound

Thompson Sampling

⋮

Varying costs

Expected improvement per cost

Budgeted Multi-step Lookahead EI

?

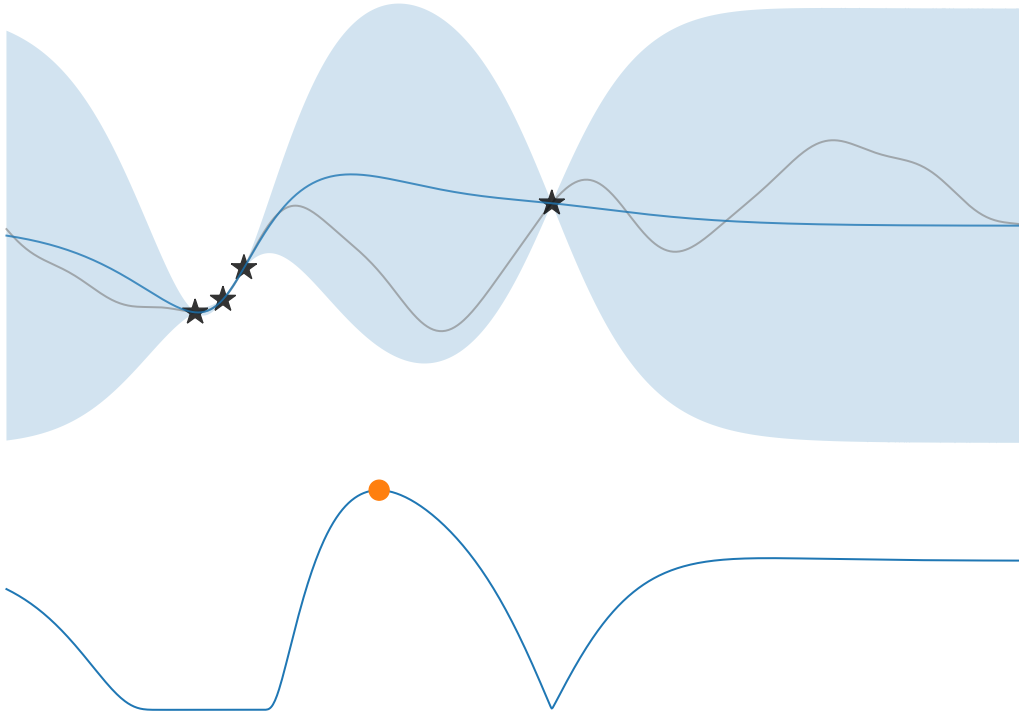
?

⋮

New design principle: Gittins Index

naturally cost-aware

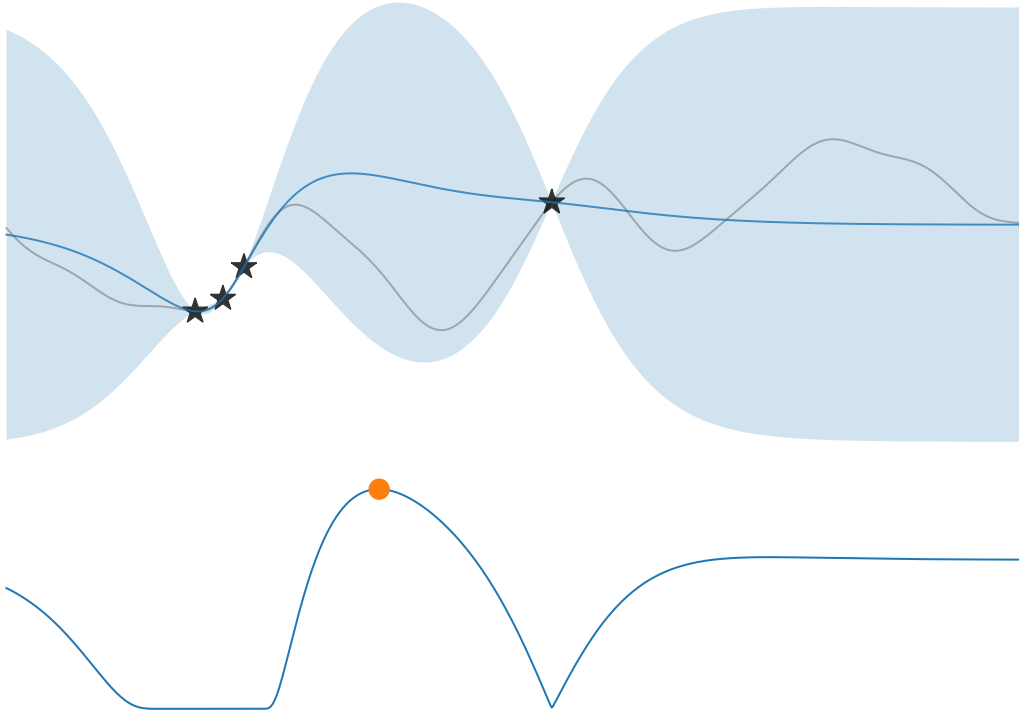
Expected Improvement



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

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

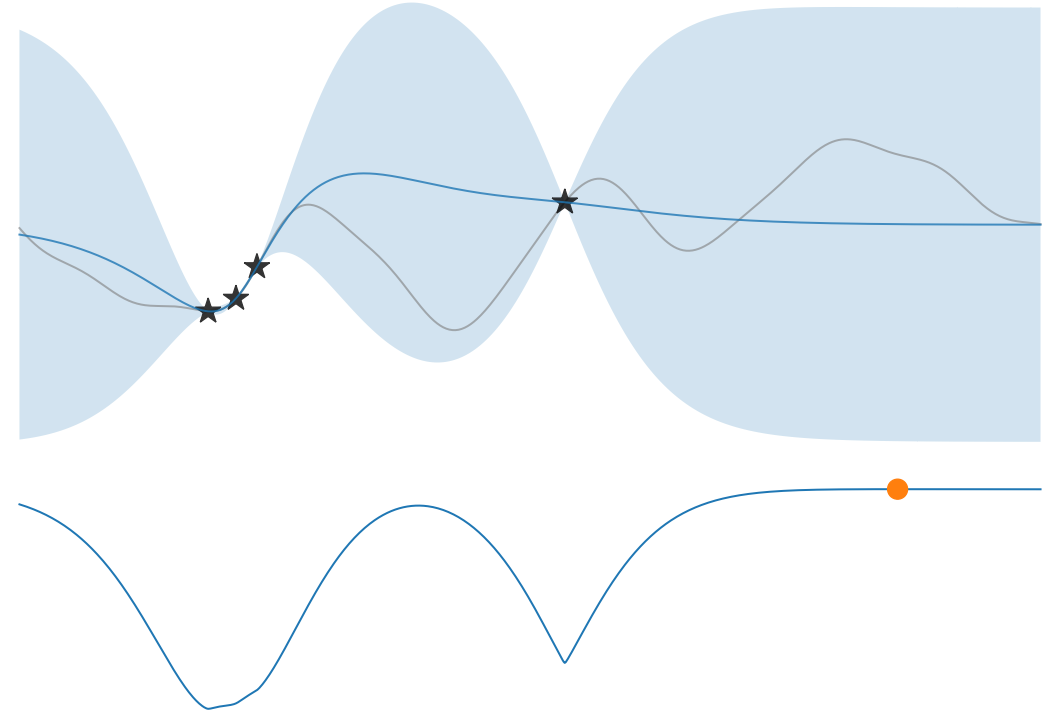
Expected Improvement



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

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

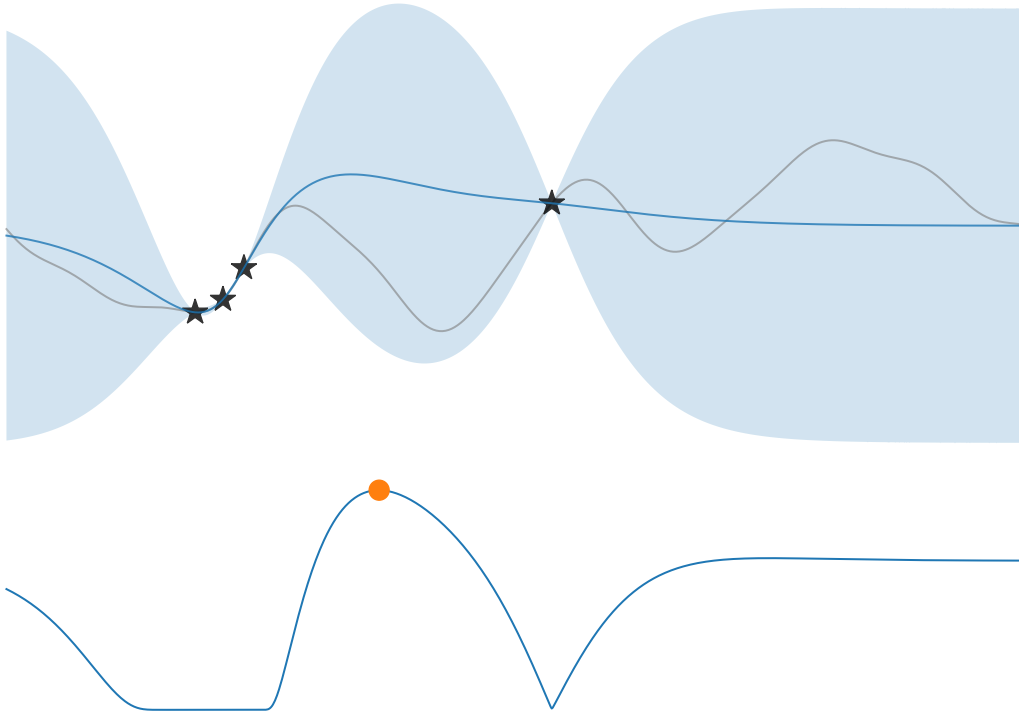
Gittins Index



$$GI_{f|D}(x) = g \text{ s.t. } EI_{f|D}(x; g) = c(x)$$

$$\max_x GI_{f|D}(x)$$

Expected Improvement

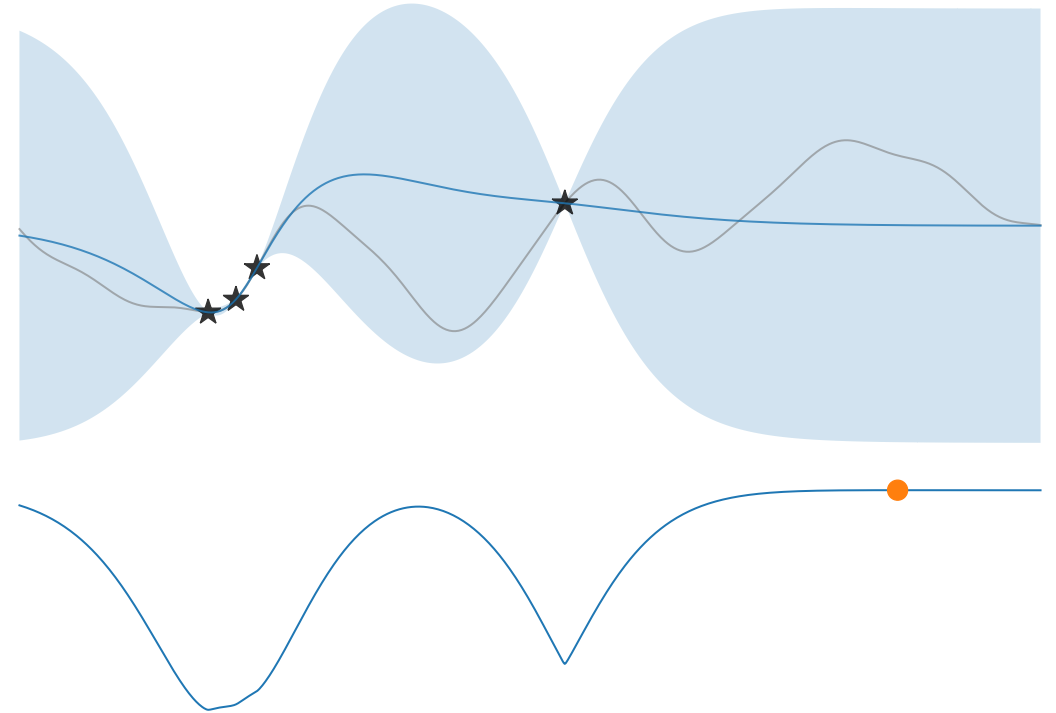


$$\text{EI}(x) = \mathbb{E}[\max(f(x) - y_{\text{best}}, 0) \mid D]$$

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

One-step approximation to MDP

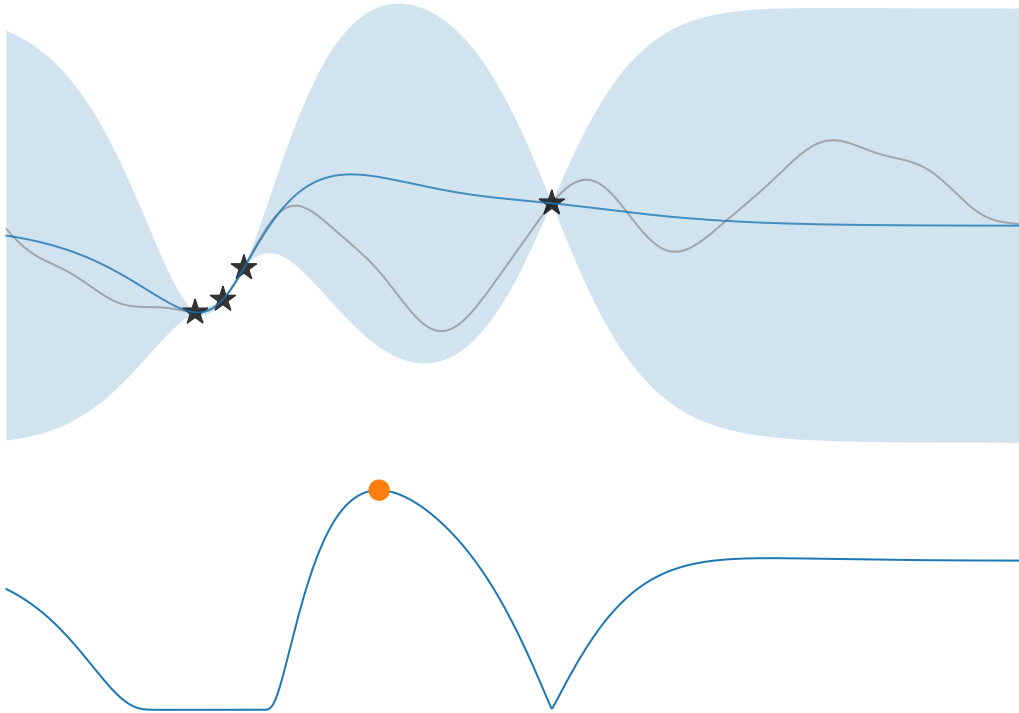
Gittins Index



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

$$\max_x \text{GI}_{f|D}(x)$$

Expected Improvement

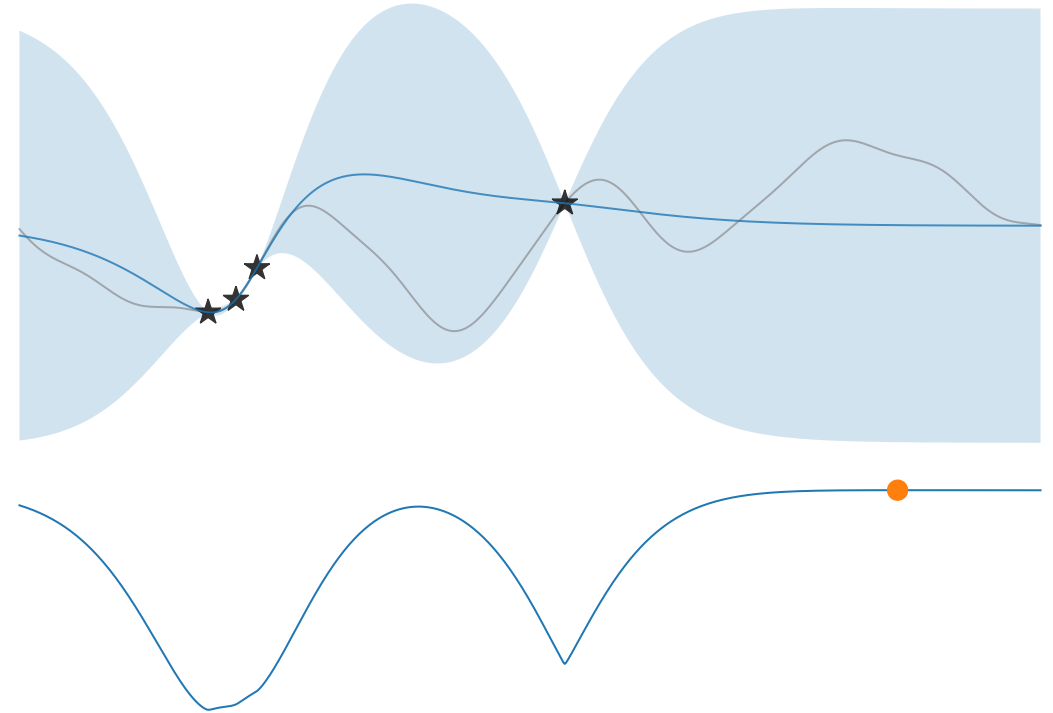


$$\text{EI}(x) = \mathbb{E}[\max(f(x) - y_{\text{best}}, 0) \mid D]$$

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

Temporal simplification to MDP

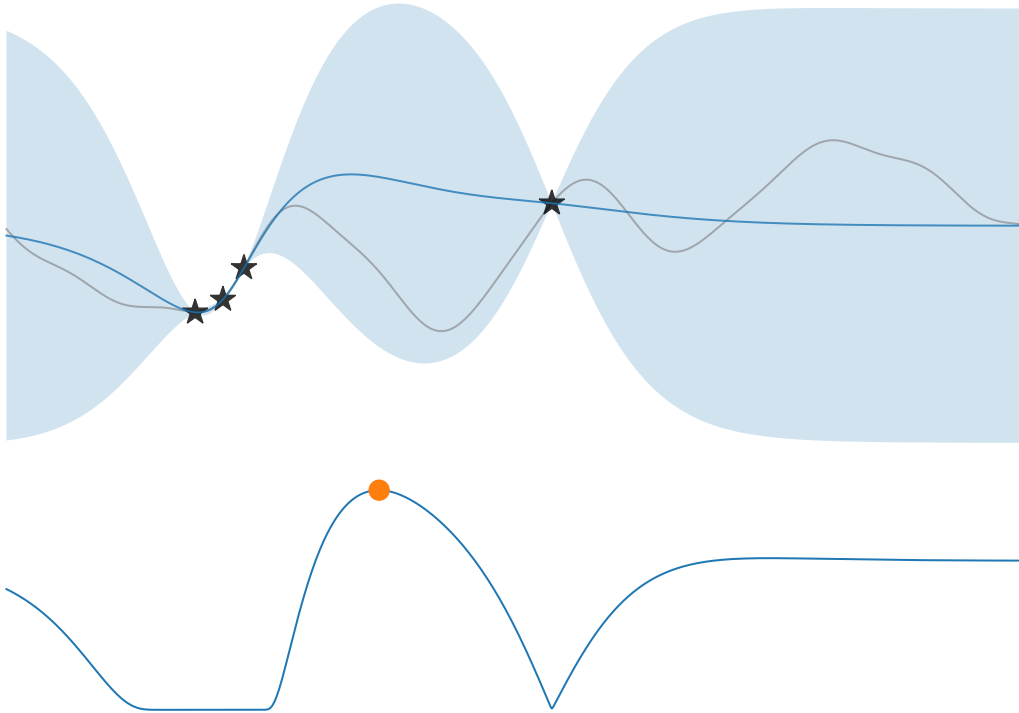
Gittins Index



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

$$\max_x \text{GI}_{f|D}(x)$$

Expected Improvement

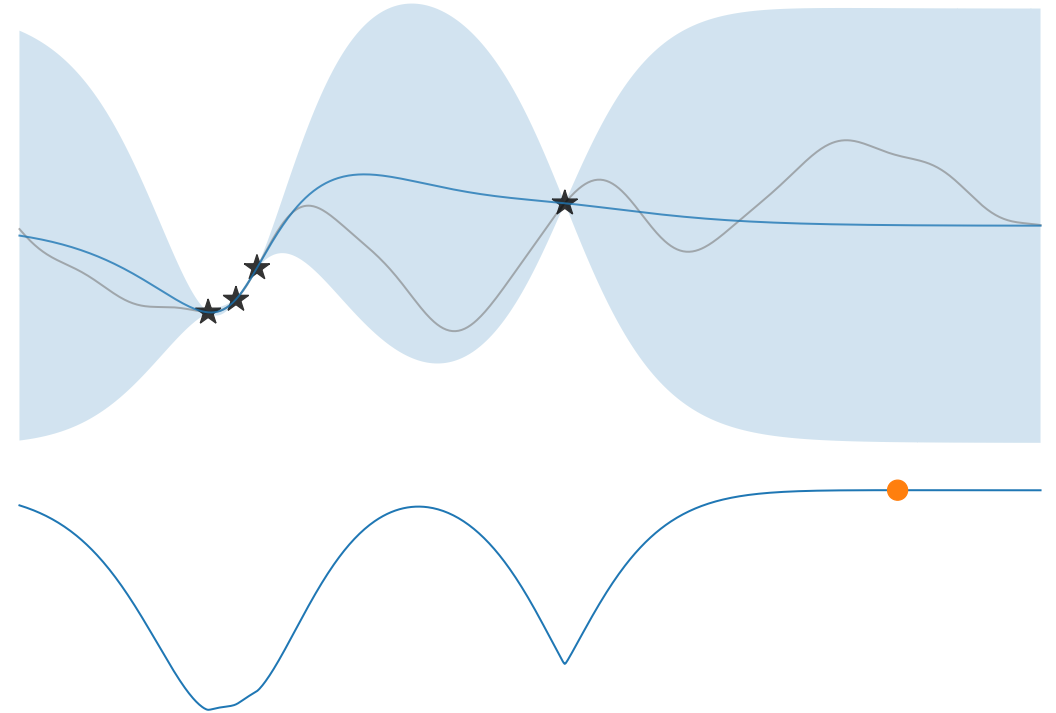


$$\text{EI}(x) = \mathbb{E}[\max(f(x) - y_{\text{best}}, 0) \mid D]$$

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

Temporal simplification to MDP

Gittins Index

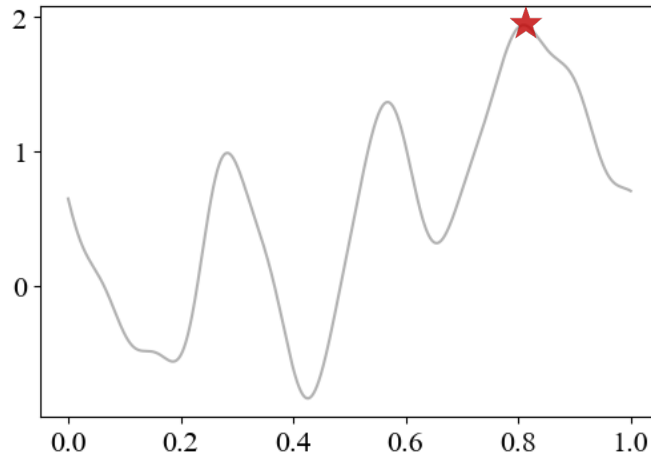


$$\text{GI}_{f|D}(x) = g \text{ s.t. } \text{EI}_{f|D}(x; g) = c(x)$$

$$\max_x \text{GI}_{f|D}(x)$$

Spatial simplification to MDP

Our Approach: Spatial Simplification

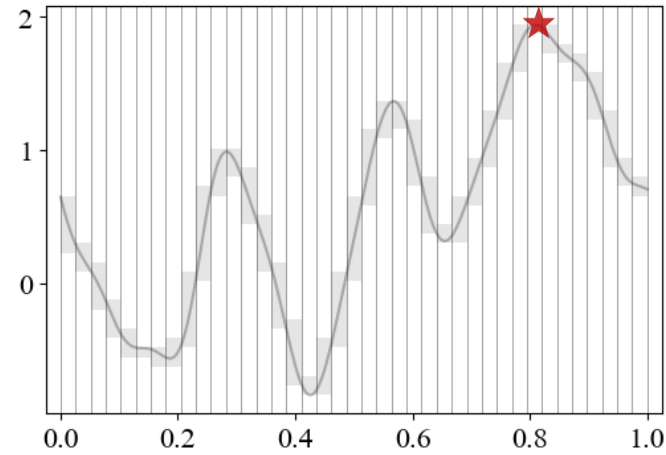
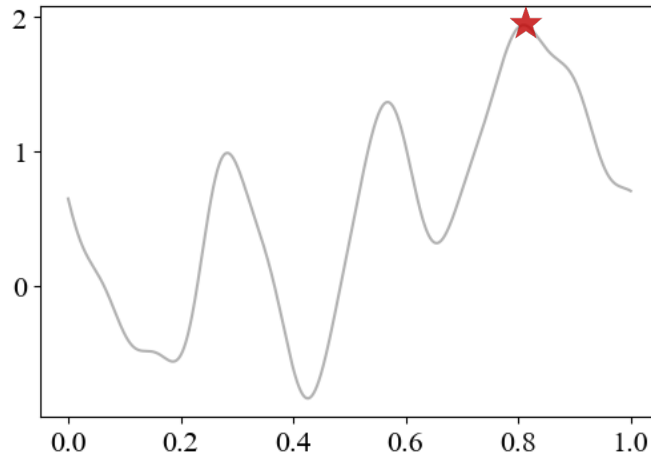


Bayesian Optimization

Continuous

Correlated

Our Approach: Spatial Simplification



Bayesian Optimization

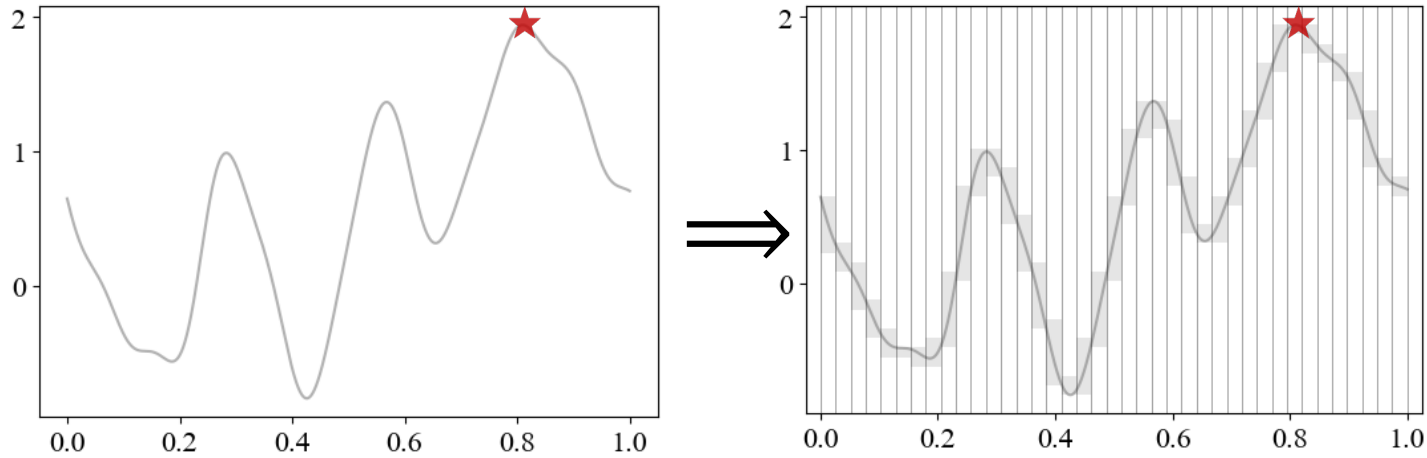
Continuous



Discrete

Correlated

Our Approach: Spatial Simplification



Bayesian Optimization

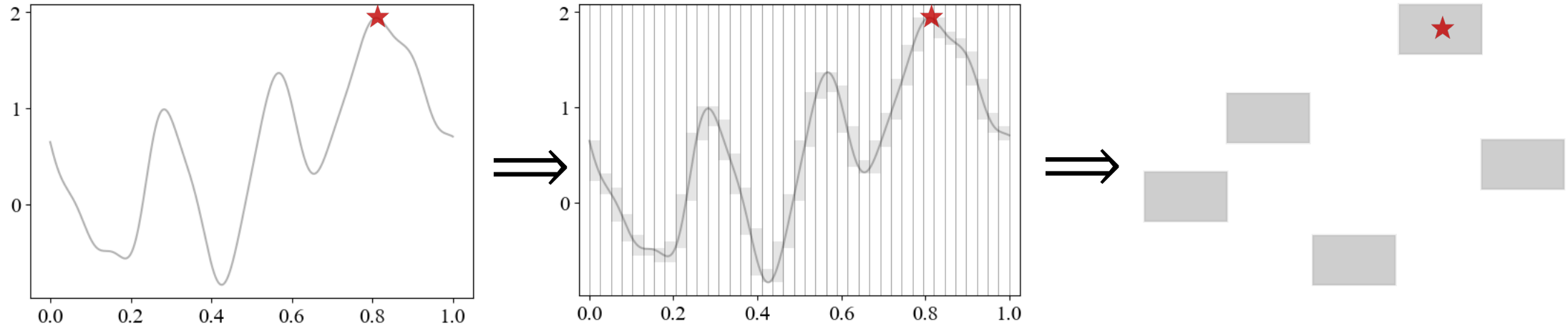
Continuous

\Rightarrow

Discrete

Correlated

Our Approach: Spatial Simplification



Bayesian Optimization

Continuous

⇒

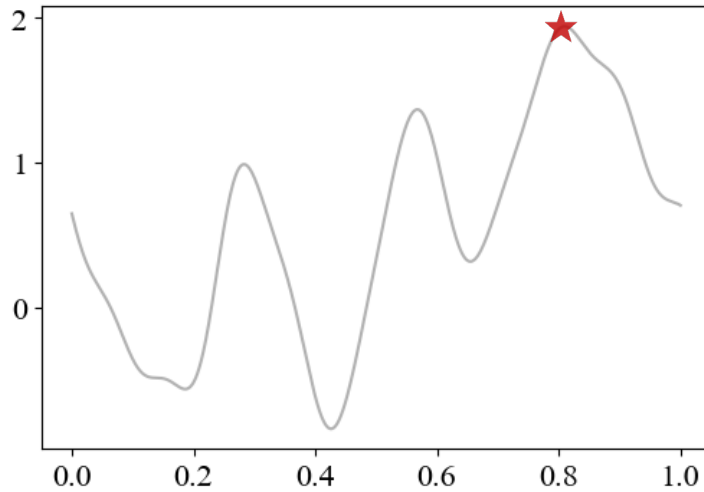
Discrete

Correlated

⇒

Independent

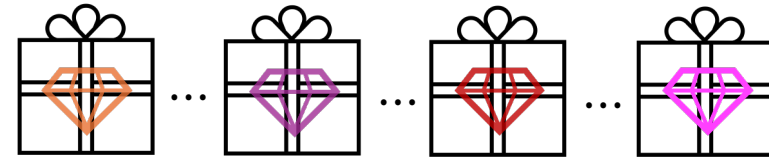
Our Approach: Spatial Simplification



Bayesian Optimization

Continuous

Correlated



Pandora's Box [Weitzman'79]

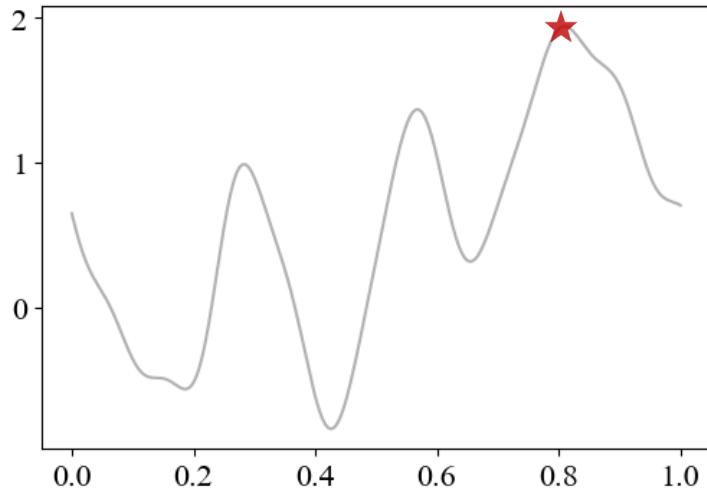


Discrete



Independent

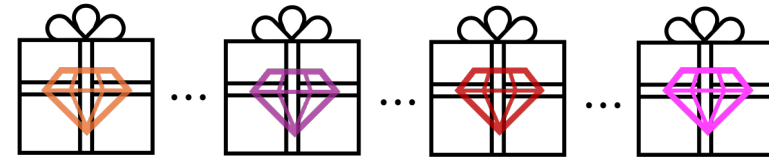
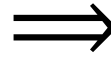
Our Approach: Spatial Simplification



Bayesian Optimization

Continuous

Correlated



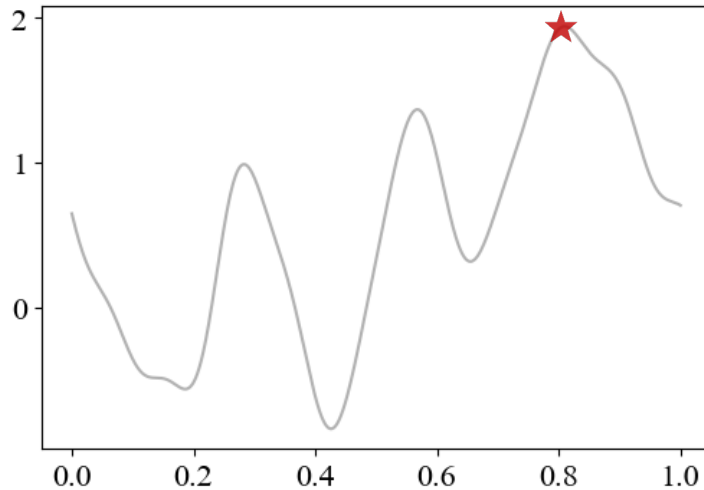
Pandora's Box [Weitzman'79]

Discrete

Independent

Optimal policy: Gittins index

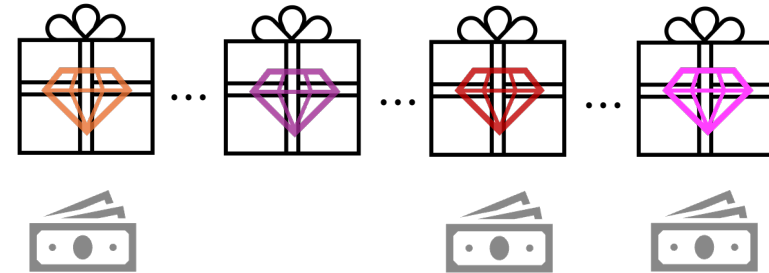
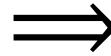
Our Approach: Spatial Simplification



Bayesian Optimization

Continuous

Correlated



Pandora's Box

Discrete

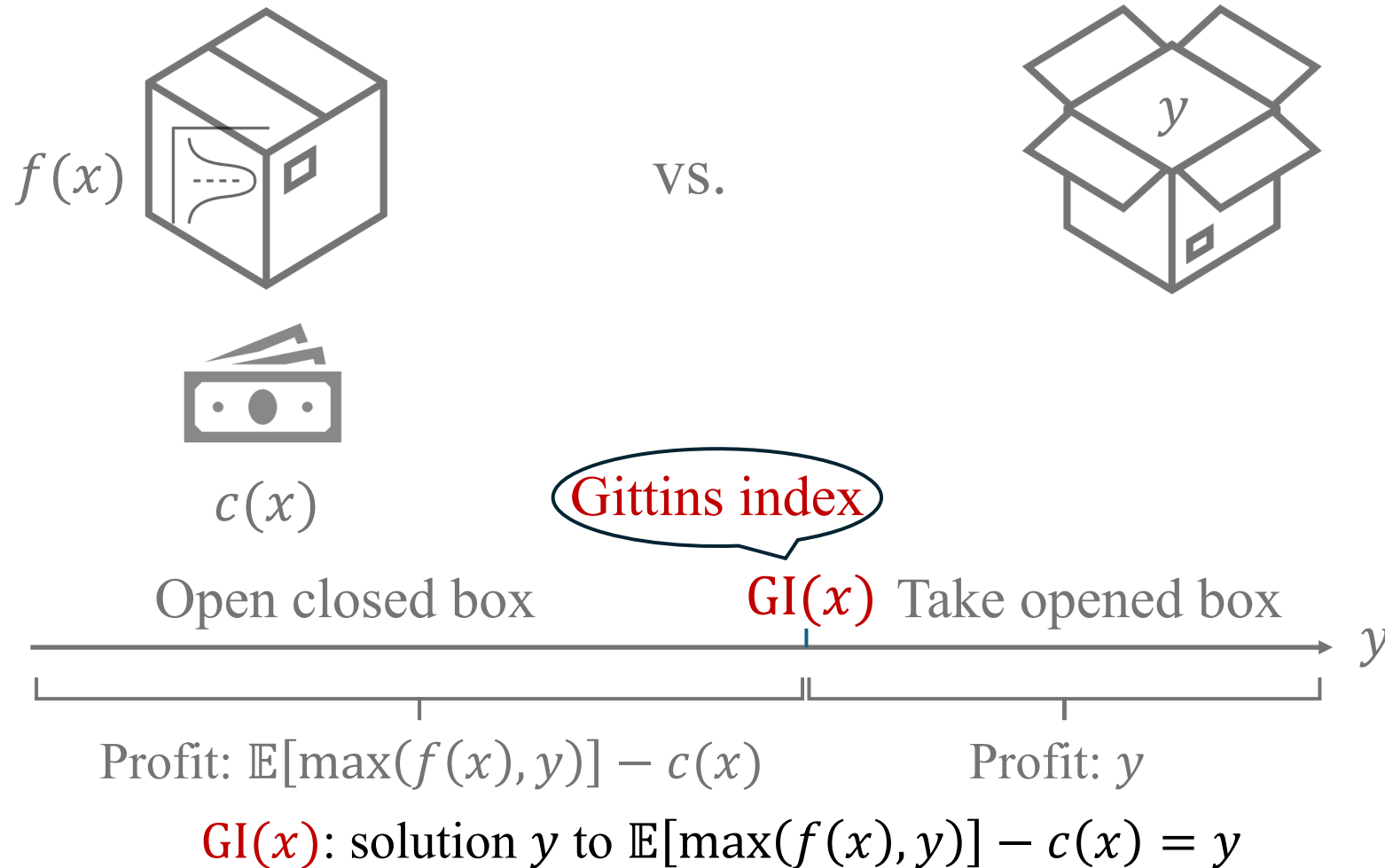
Independent

allow varying costs

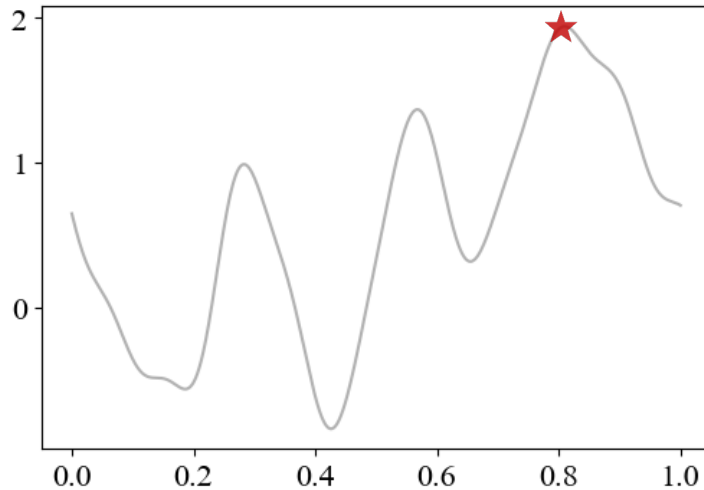
Optimal policy: Gittins index

Intuition Behind Pandora's Box Gittins Index

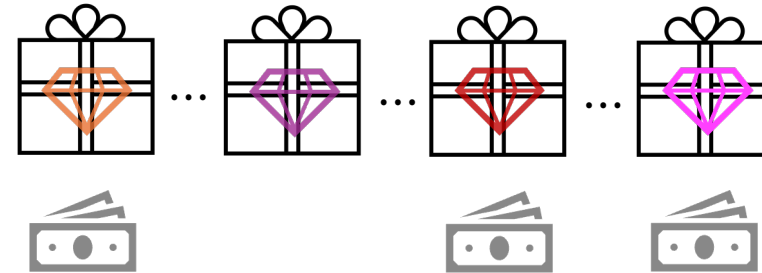
1.5-box problem:



How to translate Gittins index?



Bayesian Optimization



Pandora's Box

Continuous



Discrete

Correlated

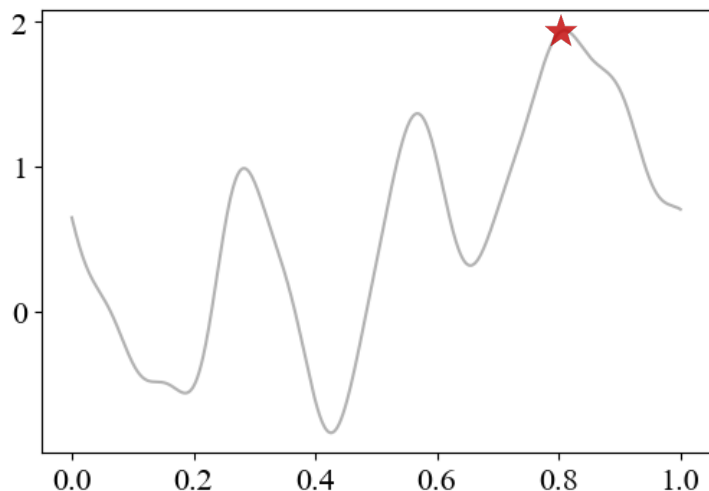


Independent

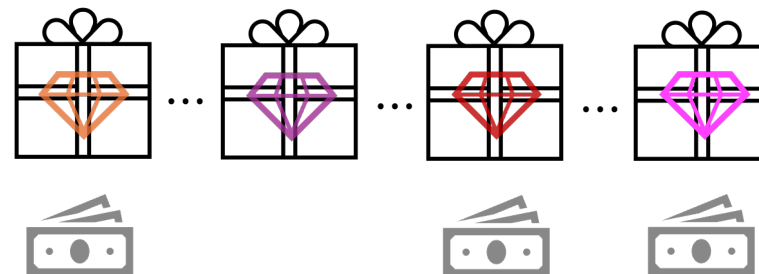
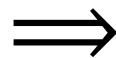
How to translate?

⇐ Optimal policy: Gittins index

How to translate Gittins index?



Bayesian Optimization



Pandora's Box

Continuous



Discrete

Correlated



Independent

incorporate posterior

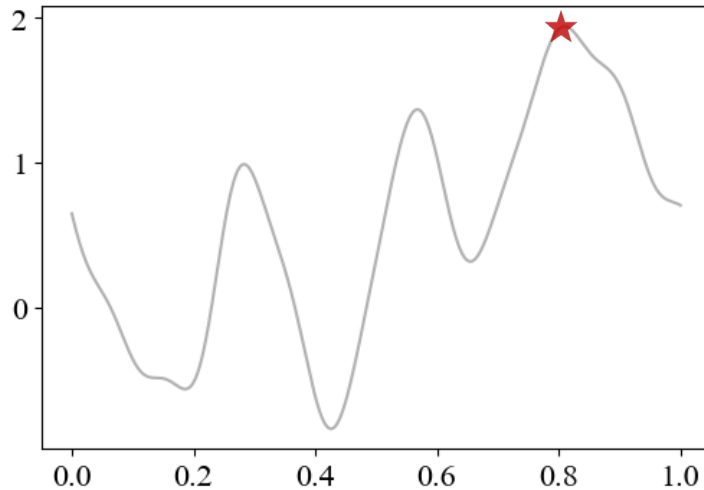
Our policy

$$\max_x \text{GI}_{f|D}(x)$$

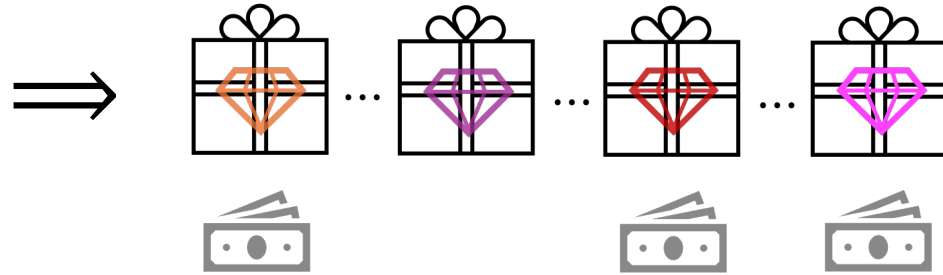
\Leftarrow Optimal policy: Gittins index

$$\max_x \text{GI}(x)$$

Is Gittins good in Bayesian Optimization?



Bayesian Optimization



Pandora's Box [Weitzman'79]

Continuous



Discrete

Correlated



Independent

incorporate posterior

Is Gittins index good?



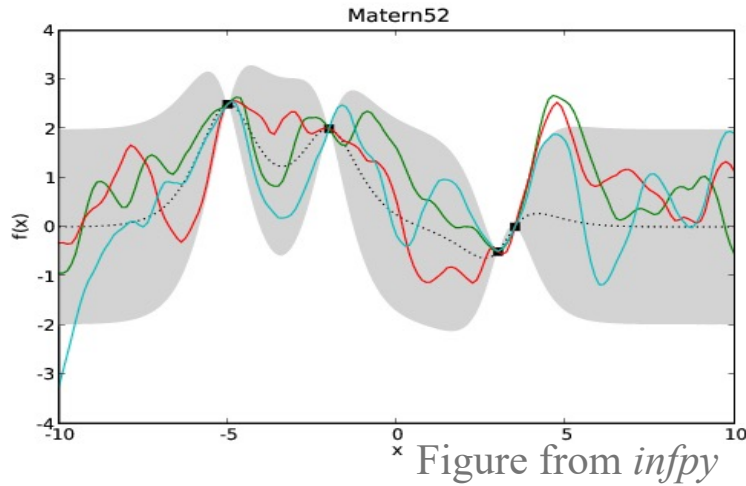
Gittins index is optimal

$$\max_x \text{GI}_{f|D}(x)$$

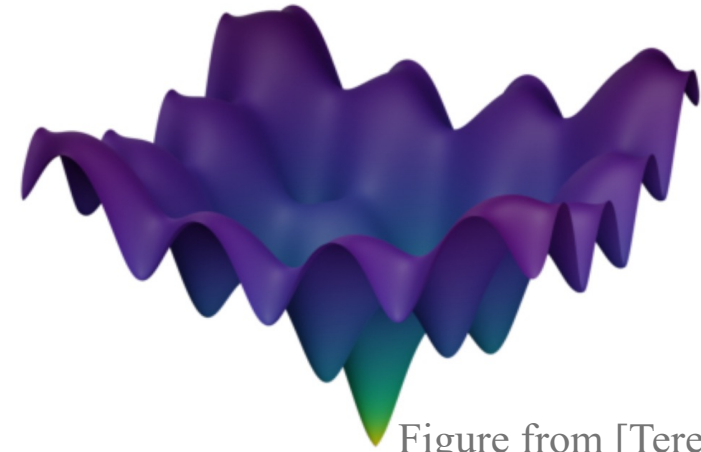
$$\max_x \text{GI}(x)$$

Experiment Setup: Objective Functions

Samples from prior



Ackley function



Pest Control



Lunar Lander

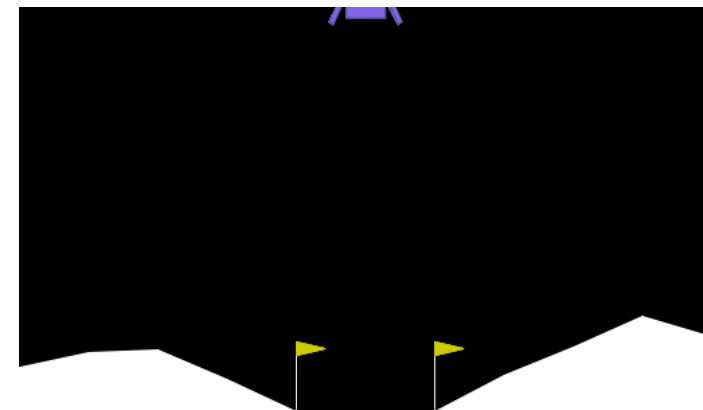
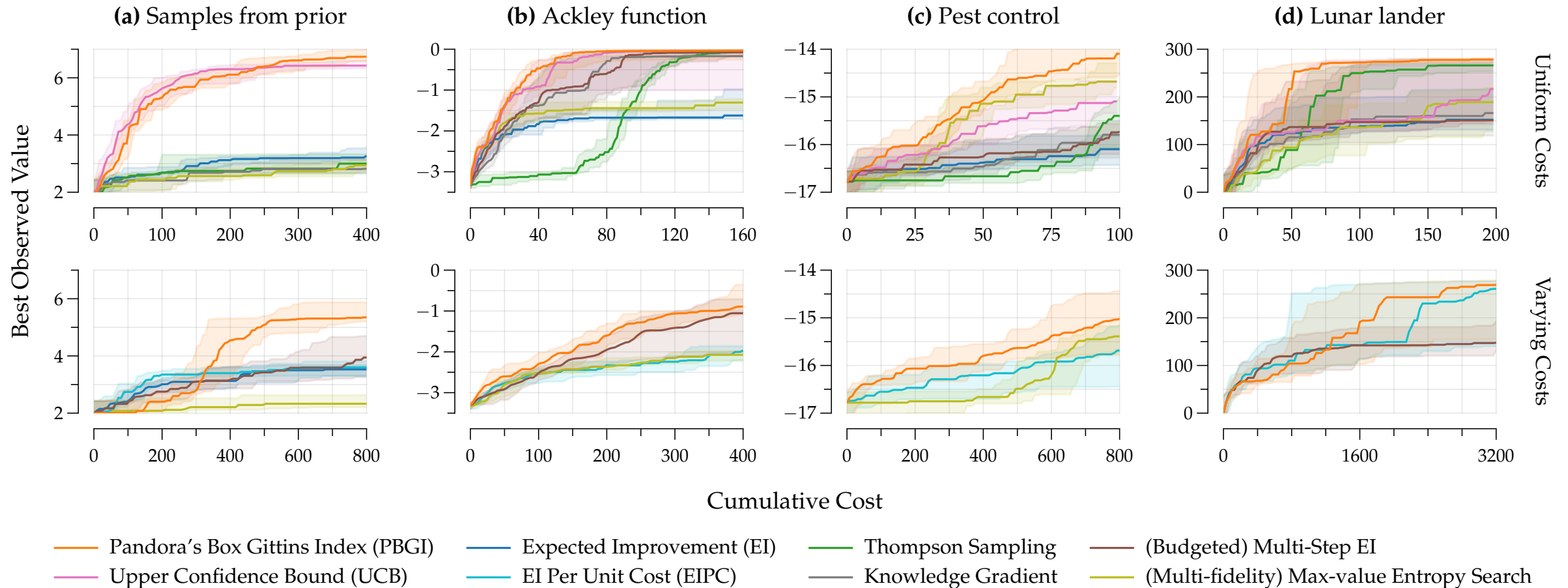


Figure from OpenAI Gym

Experiment Results

Synthetic

Empirical



FAQ

- Easy-to-compute?

FAQ

- Easy-to-compute?
Yes, EI + bisection

FAQ

- Easy-to-compute?
Yes, EI + bisection
- Any theoretical results?

FAQ

- Easy-to-compute?

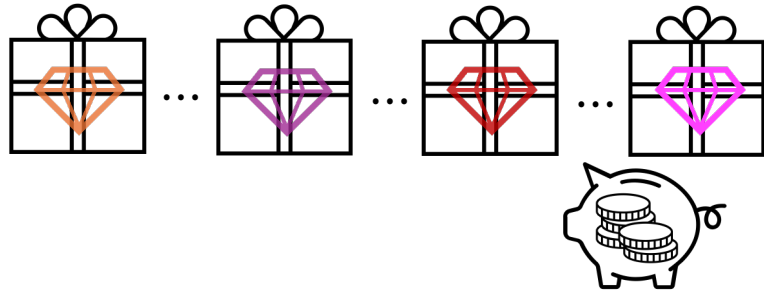
Yes, EI + bisection

- Any theoretical results?

Yes, expected-budget-constrained \cong cost-per-sample

Theoretical Result

max best observed under budget

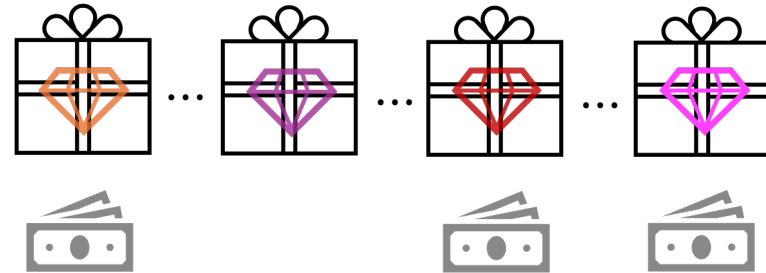


Budgeted Pandora's Box

Expected budget constraint

Optimal policy?

max (best observed – costs)



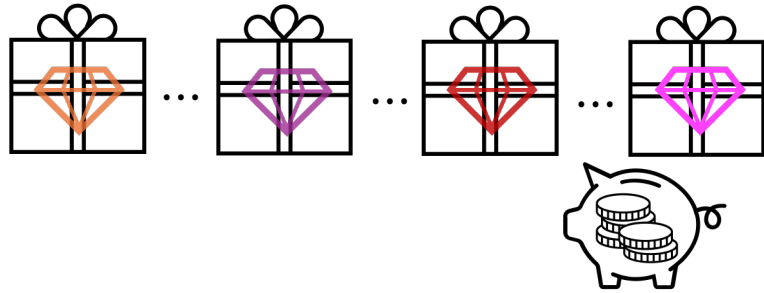
Pandora's Box

Cost per sample

Optimal policy: Gittins index

Theoretical Result

max best observed under budget



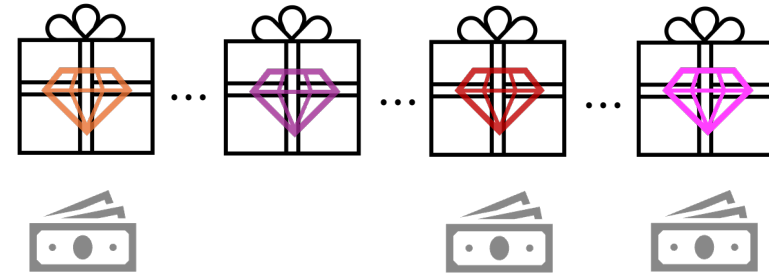
Budgeted Pandora's Box

Expected budget constraint

Optimal policy ✓

extension to [Aminian,
Manshadi, Niazadeh'24]

max (best observed – **scaled** costs)



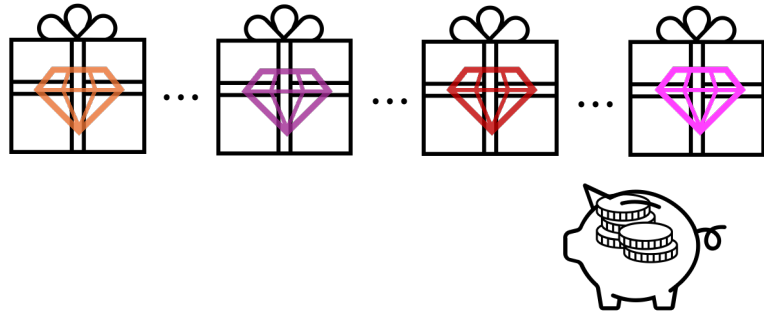
Pandora's Box

Cost per sample

Optimal policy: Gittins index

Theoretical Result

max best observed under budget



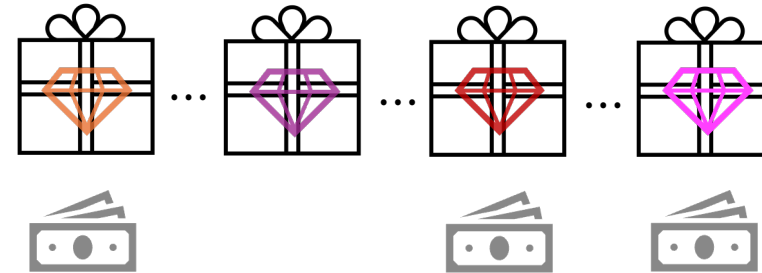
Budgeted Pandora's Box

Expected budget constraint

Optimal policy ✓

extension to [Aminian,
Manshadi, Niazadeh'24]

max (best observed – scaled costs)



Pandora's Box

Cost per sample

Optimal policy: Gittins index

budget-dependent

FAQ

- Easy-to-compute?

Yes, EI + bisection

- Any theoretical results?

Yes, expected-budget-constrained \cong cost-per-sample

- Tuning parameters?

FAQ

- Easy-to-compute?

Yes, EI + bisection

- Any theoretical results?

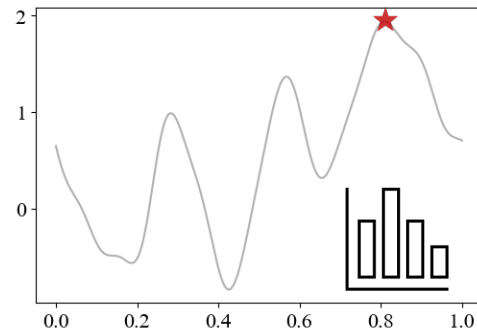
Yes, expected-budget-constrained \cong cost-per-sample

- Tuning parameters?

Yes, control unit conversion

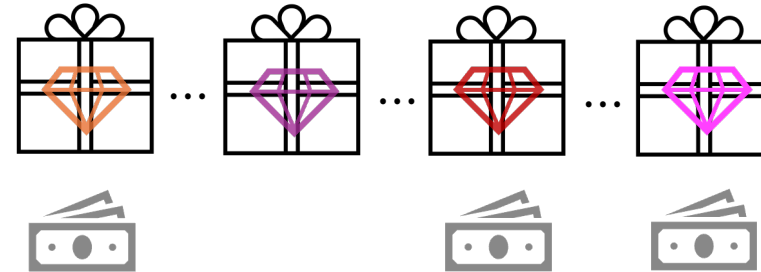
New Design Principle: Gittins Index

Problem



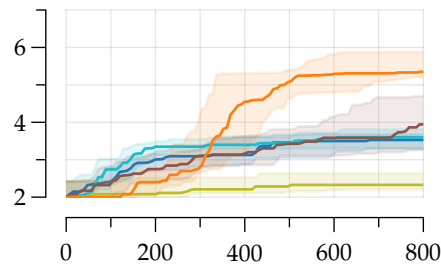
Bayesian optimization
with varying costs

Key idea



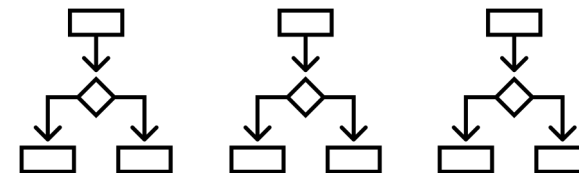
link to Pandora's box
and Gittins index theory

Impact



competitive performance

Future potential



black-box processes
with partial feedback

Check our paper on ArXiv!



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