

Gittins Indices for Bayesian Optimization: Insights from Pandora's Box

Qian Xie (Cornell ORIE)

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

NYC Ops Day 2024

Bayesian Optimization

Goal: optimize expensive-to-evaluate black-box function

∈ decision-making under uncertainty

Bayesian Optimization

Goal: optimize expensive-to-evaluate black-box function

∈ decision-making under uncertainty

Applications:

Hyperparameter tuning

Drug discovery

Control design

Bayesian Optimization

Goal: optimize expensive-to-evaluate **black-box** function

∈ decision-making under uncertainty

Applications:

Hyperparameter tuning

Drug discovery

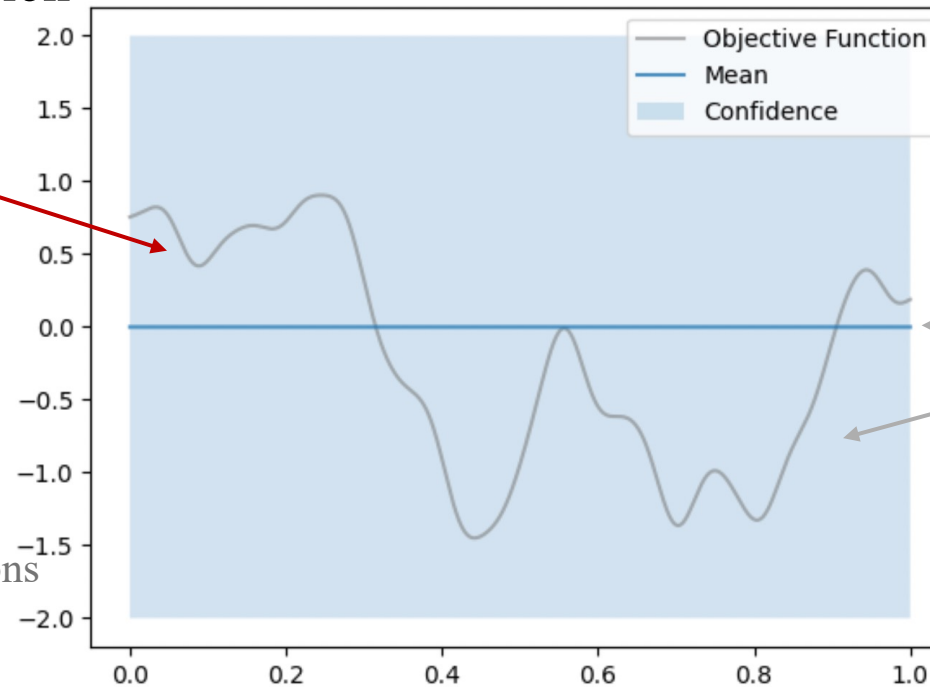
Control design

Bayesian Optimization

Goal: optimize expensive-to-evaluate **black-box** function

An **unknown random** function $f: \mathcal{X} \rightarrow \mathbb{R}$ drawn from a Gaussian process prior

Gaussian process: infinite-dimensional generalization of multivariate normal distributions



Applications:

Hyperparameter tuning

Drug discovery

Control design

x : hyperparameter/configuration

mean: prediction

variance: confidence/uncertainty

Bayesian Optimization

Goal: optimize expensive-to-evaluate black-box function

An unknown random function $f: \mathcal{X} \rightarrow \mathbb{R}$ drawn from a Gaussian process prior

Gaussian process: infinite-dimensional generalization of multivariate normal distributions

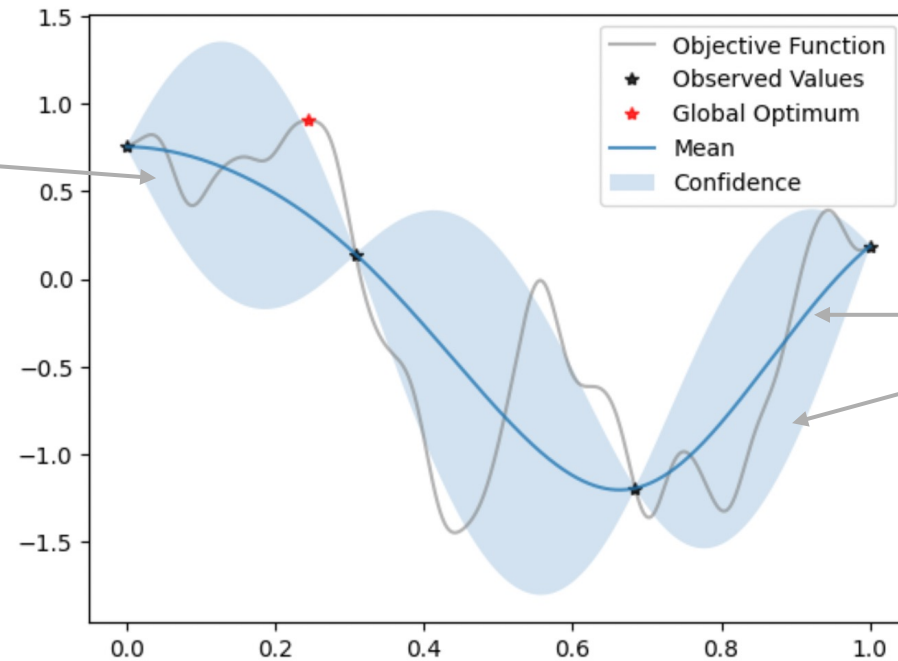
Applications:

Hyperparameter tuning

Drug discovery

Control design

x : hyperparameter/configuration



mean: prediction

variance: confidence/uncertainty

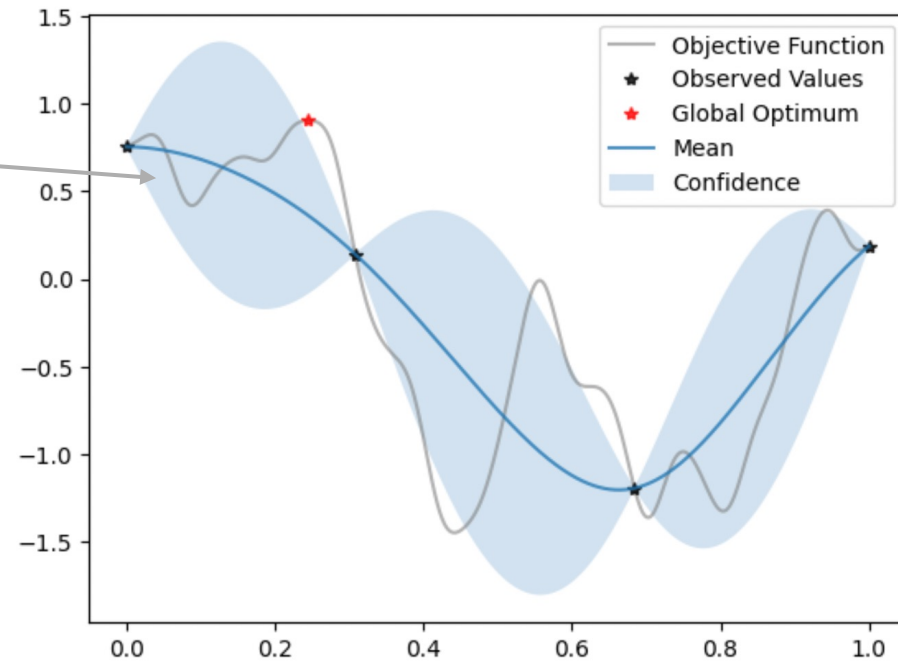
Objective: find global optimum $x^* = \operatorname{argmax}_{x \in \mathcal{X}} f(x)$

Decision: evaluate a set of points

Bayesian Optimization

Goal: optimize **expensive-to-evaluate** black-box function

An unknown random function $f: \mathcal{X} \rightarrow \mathbb{R}$ drawn from a Gaussian process prior



Applications:

Hyperparameter tuning

Drug discovery

Control design

x : hyperparameter/configuration

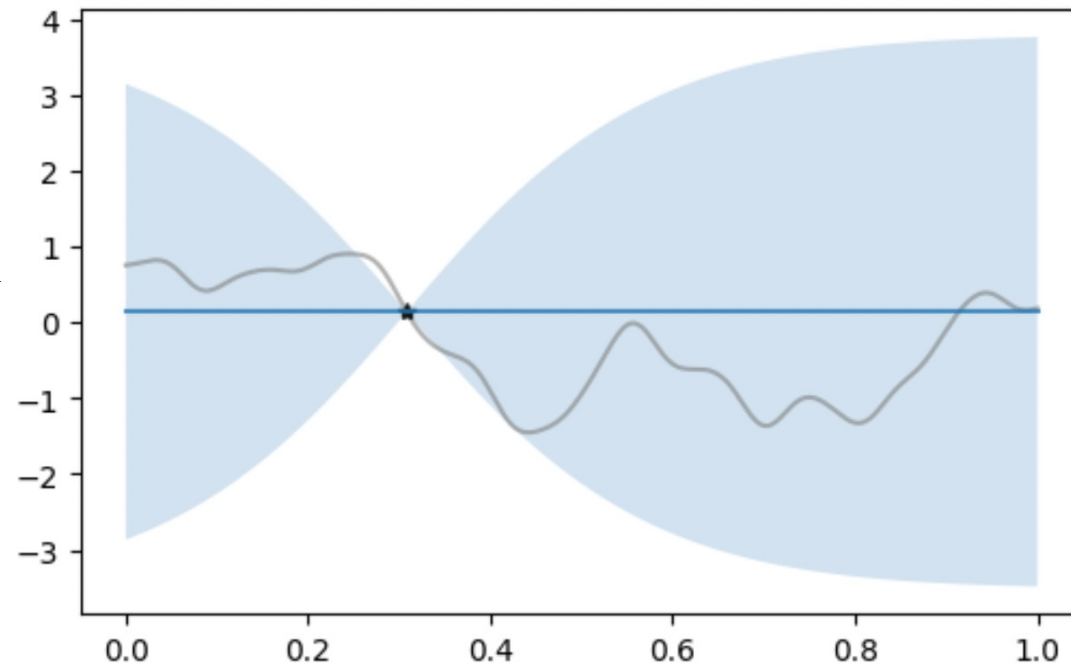
Objective: find global optimum $x^* = \operatorname{argmax}_{x \in \mathcal{X}} f(x)$

Decision: evaluate a set of points

Bayesian Optimization

Goal: optimize **expensive-to-evaluate** black-box function

An unknown random function $f: \mathcal{X} \rightarrow \mathbb{R}$
drawn from a Gaussian process prior



Applications:

Hyperparameter tuning
Drug discovery
Control design

x : hyperparameter/configuration

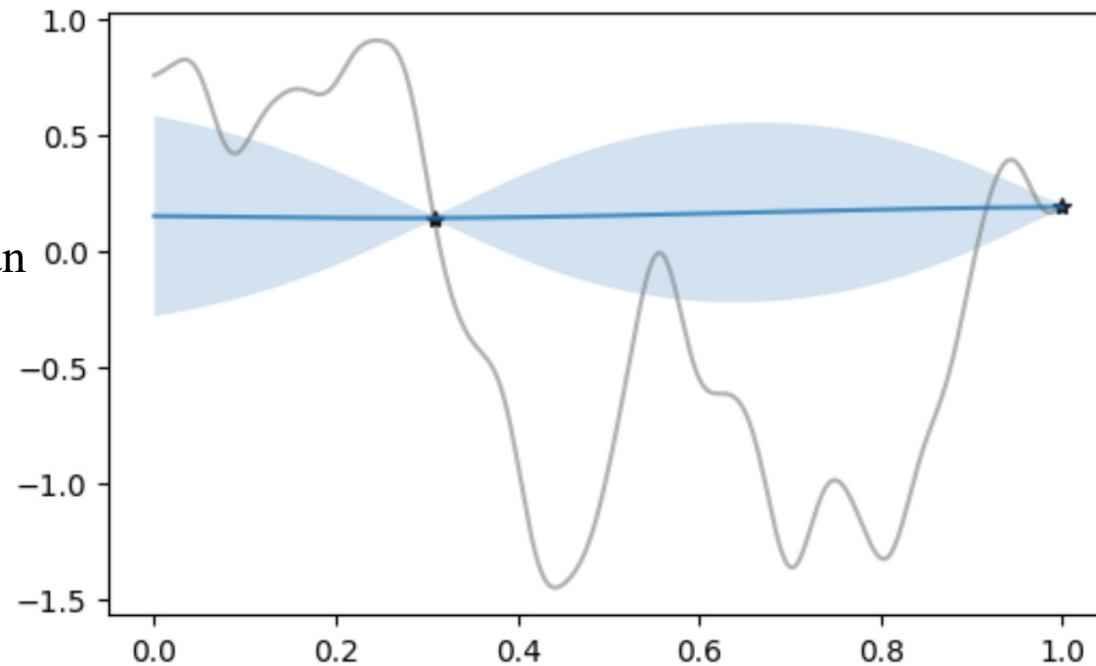
adaptively

Decision: evaluate a set of points

Bayesian Optimization

Goal: optimize **expensive-to-evaluate** black-box function

An unknown random function $f: \mathcal{X} \rightarrow \mathbb{R}$
drawn from a Gaussian process prior



Applications:

Hyperparameter tuning

Drug discovery

Control design

x : hyperparameter/configuration

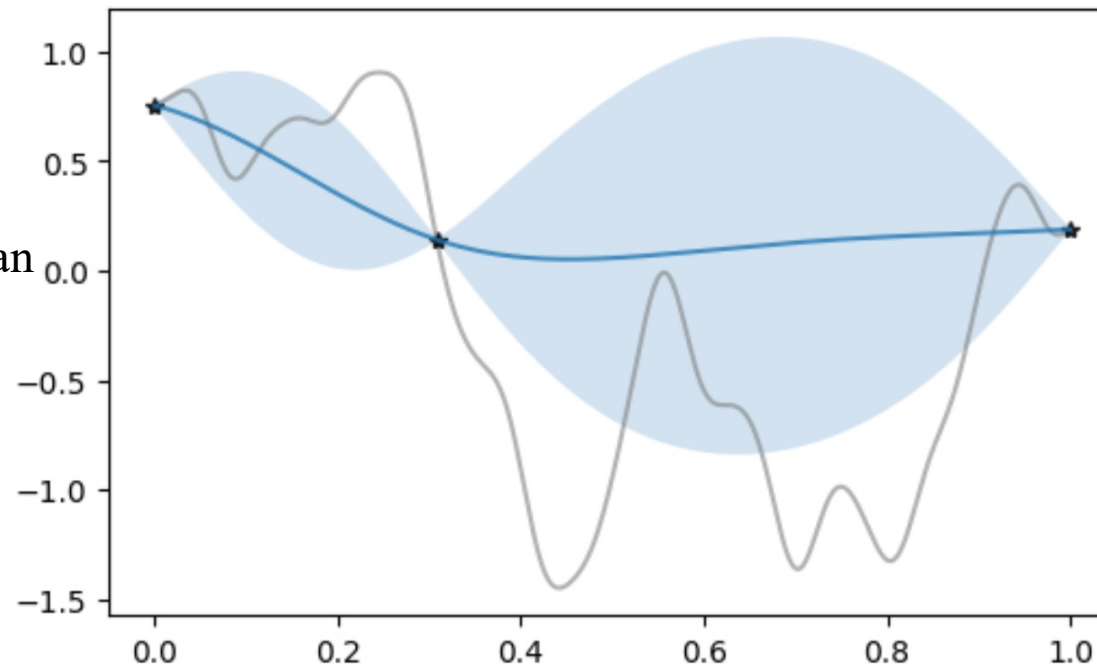
adaptively

Decision: evaluate a set of points

Bayesian Optimization

Goal: optimize **expensive-to-evaluate** black-box function

An unknown random function $f: \mathcal{X} \rightarrow \mathbb{R}$ drawn from a Gaussian process prior



Applications:

Hyperparameter tuning

Drug discovery

Control design

x : hyperparameter/configuration

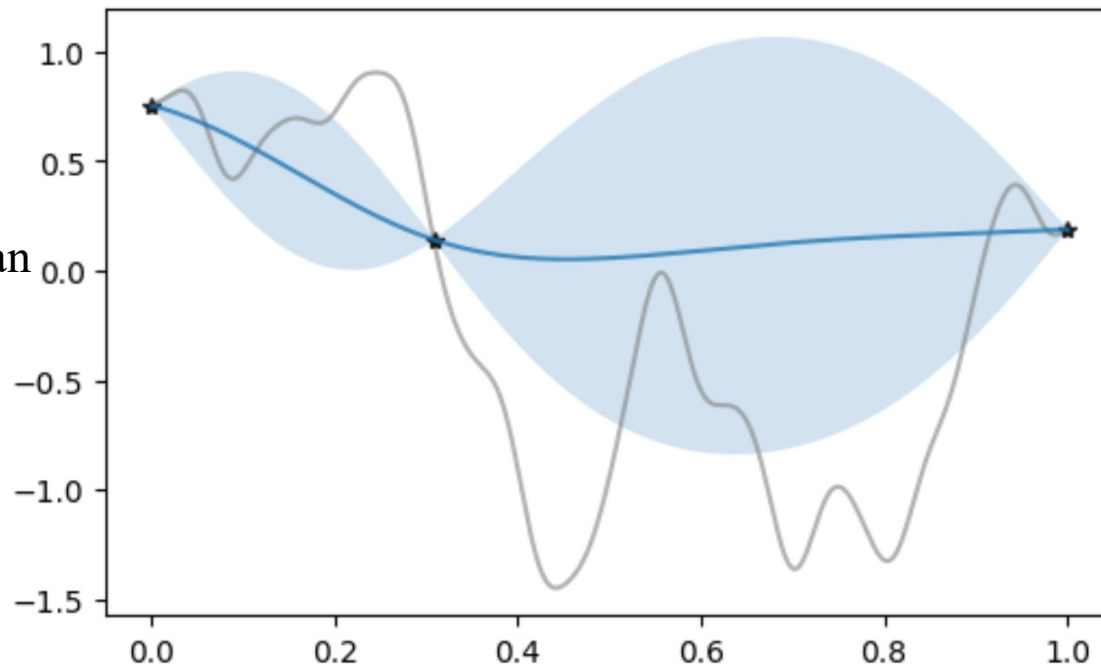
adaptively

Decision: evaluate a set of points

Bayesian Optimization

Goal: optimize **expensive-to-evaluate** black-box function

An unknown random function $f: \mathcal{X} \rightarrow \mathbb{R}$ drawn from a Gaussian process prior



Applications:

Hyperparameter tuning
Drug discovery
Control design

x : hyperparameter/configuration

Decision: **adaptively** evaluate a set of points

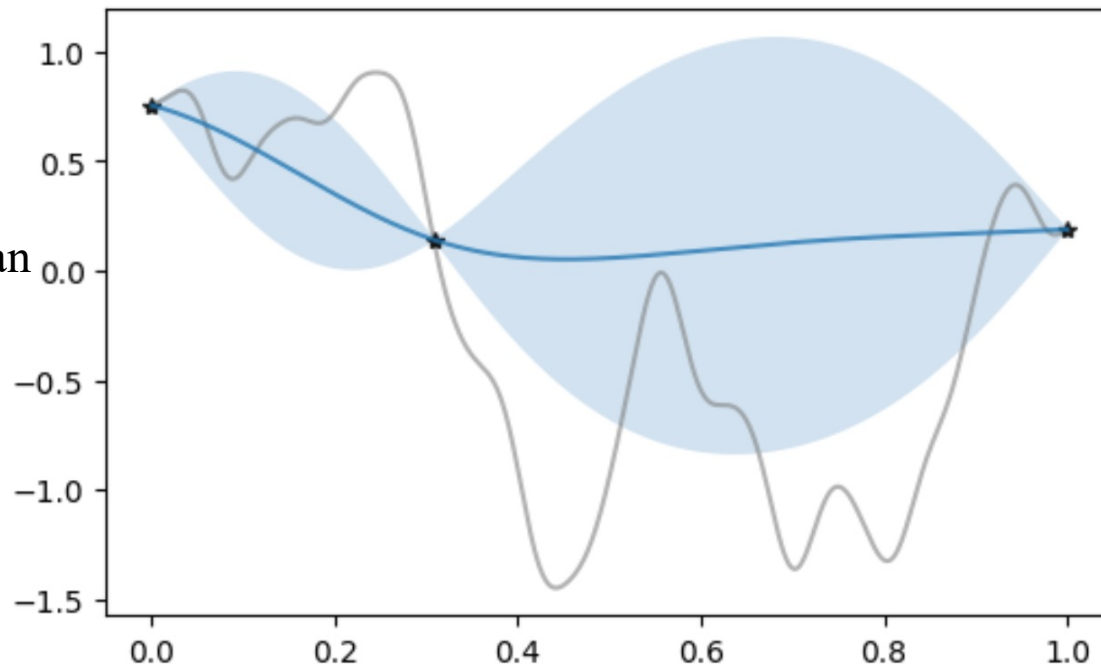
$x_1, x_2, \dots, x_T \in \mathcal{X}$

T : time budget

Bayesian Optimization

Goal: optimize **expensive-to-evaluate** black-box function

An unknown random function $f: \mathcal{X} \rightarrow \mathbb{R}$ drawn from a Gaussian process prior



Applications:

Hyperparameter tuning
Drug discovery
Control design

x : hyperparameter/configuration

Objective: optimize best observed value at time T

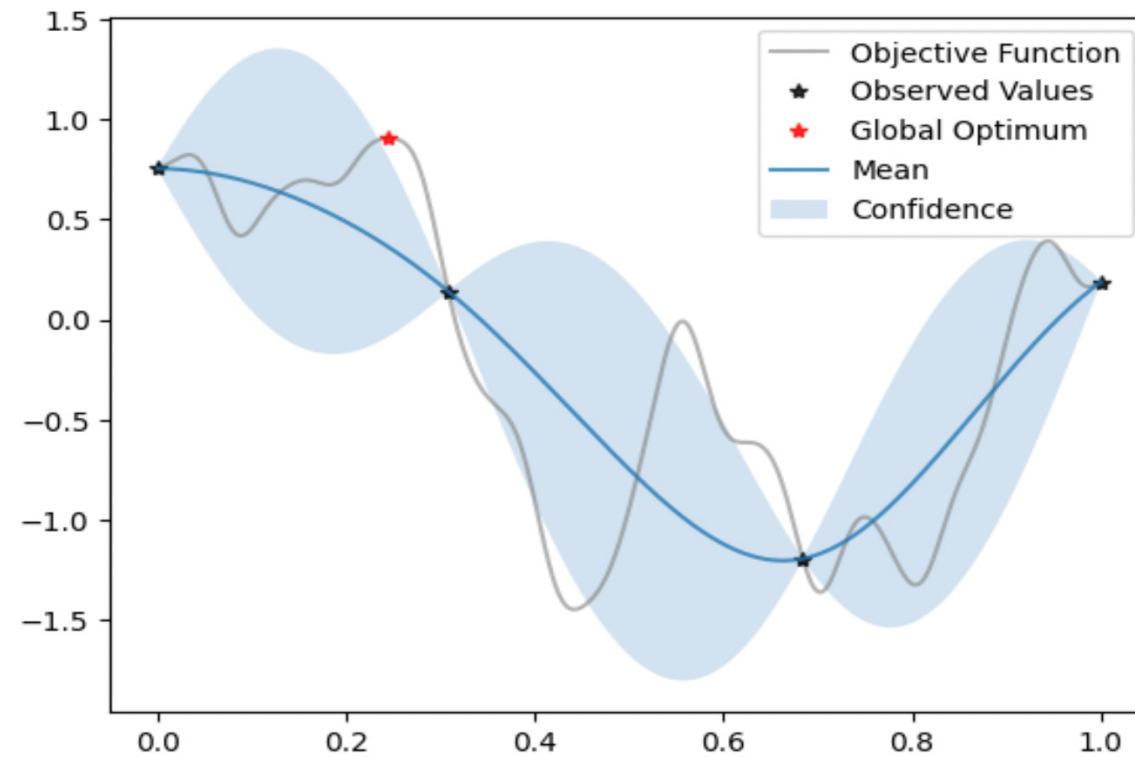
$$\max_{\text{policy}} \mathbb{E} \max_{t=1,2,\dots,T} f(x_t)$$

Decision: **adaptively** evaluate a set of points

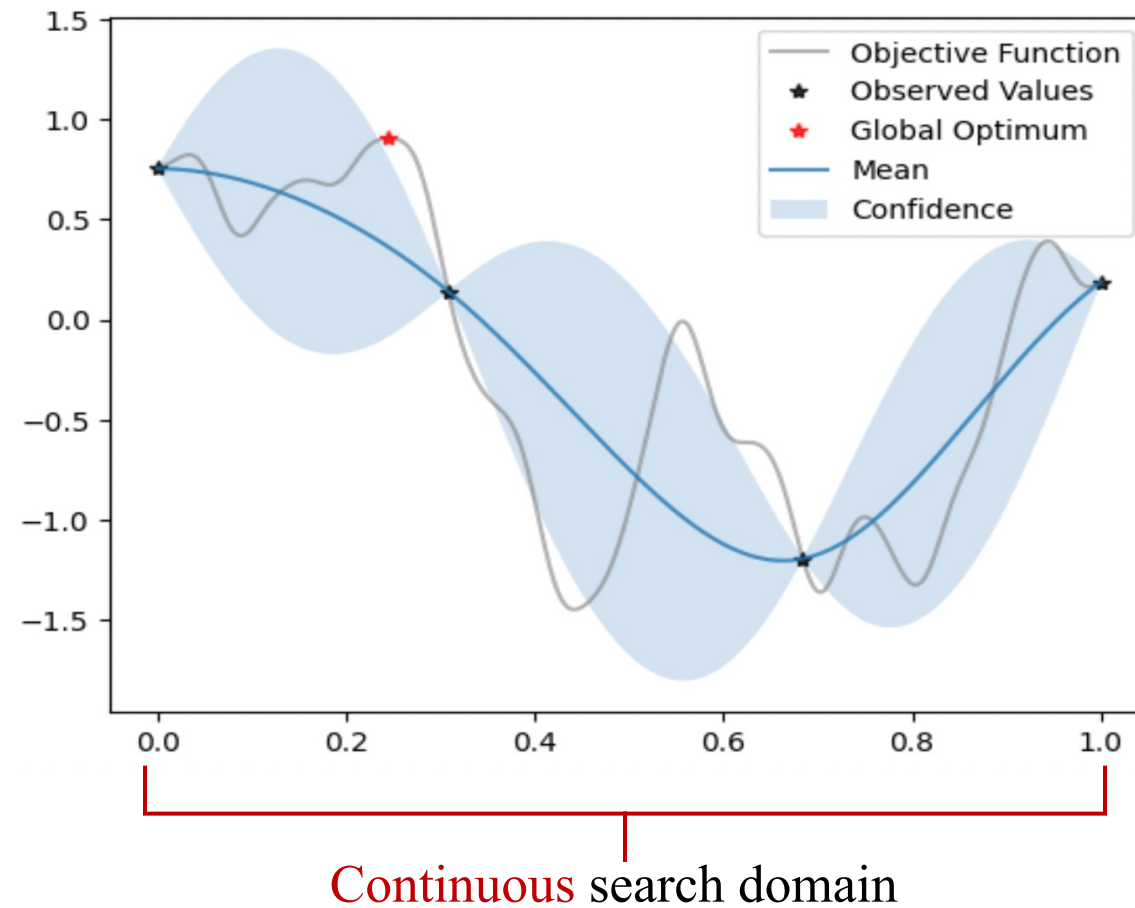
$$x_1, x_2, \dots, x_T \in \mathcal{X}$$

T : time budget

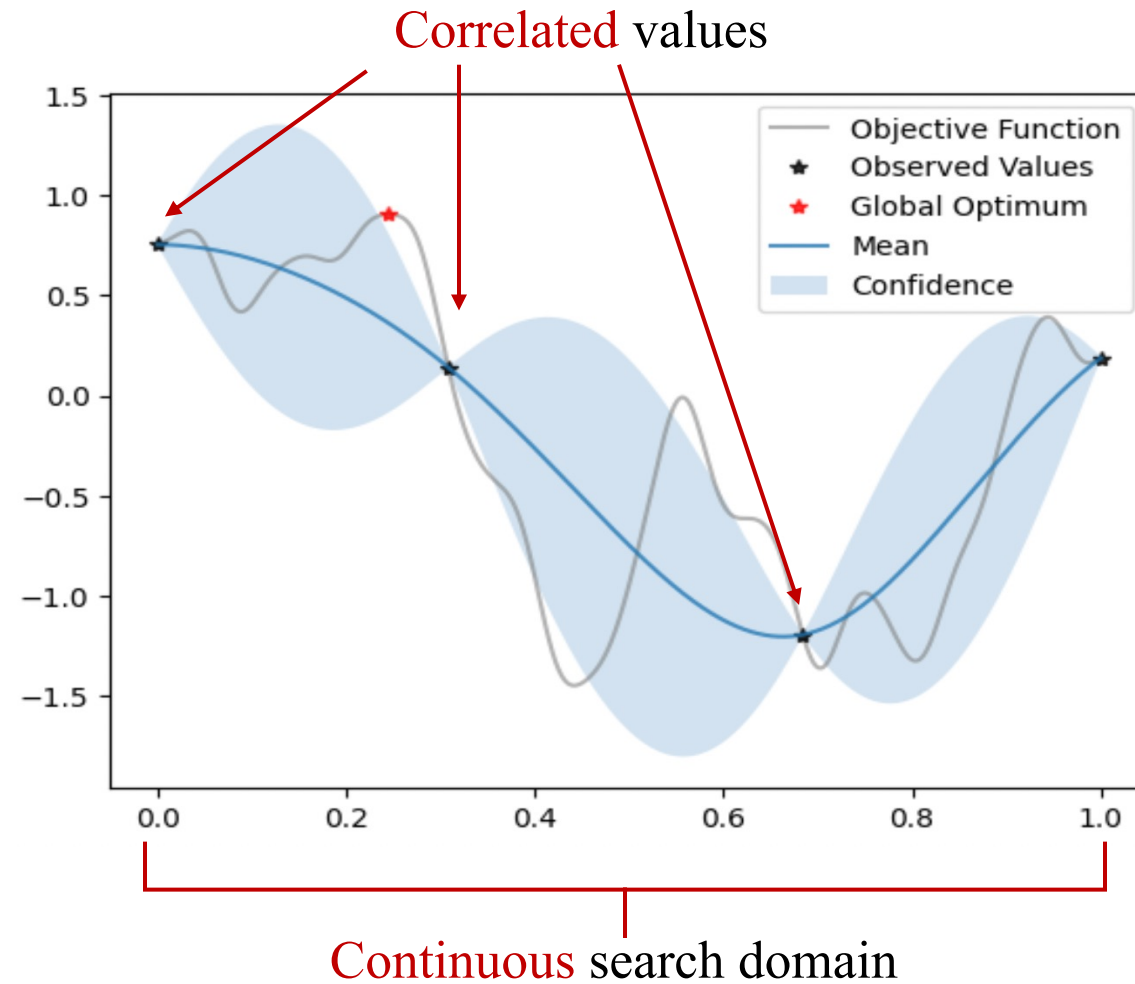
Why is it hard?



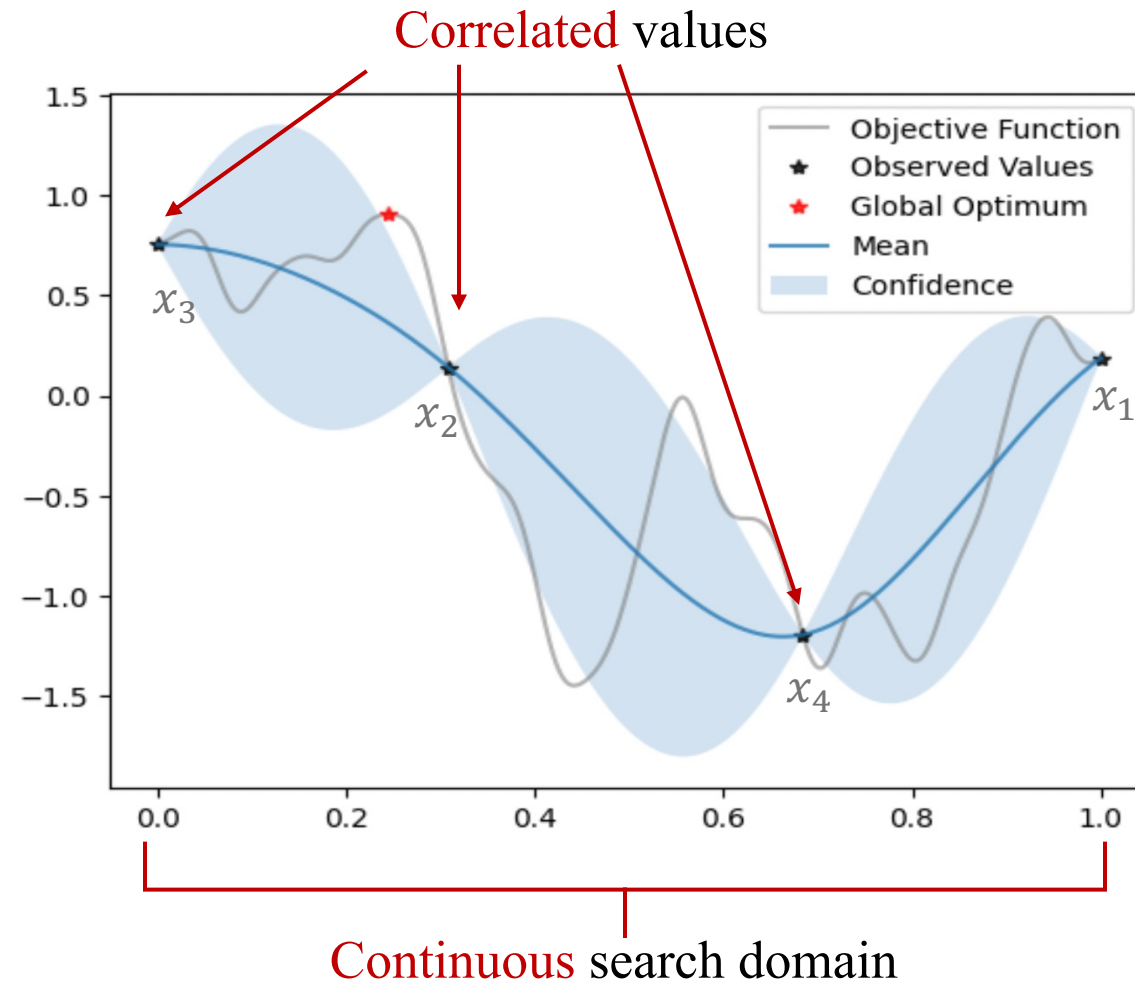
Why is it hard?



Why is it hard?



Why is it hard?



Hard budget **constraint**

~~$t=1$~~



~~$t=2$~~



~~$t=3$~~



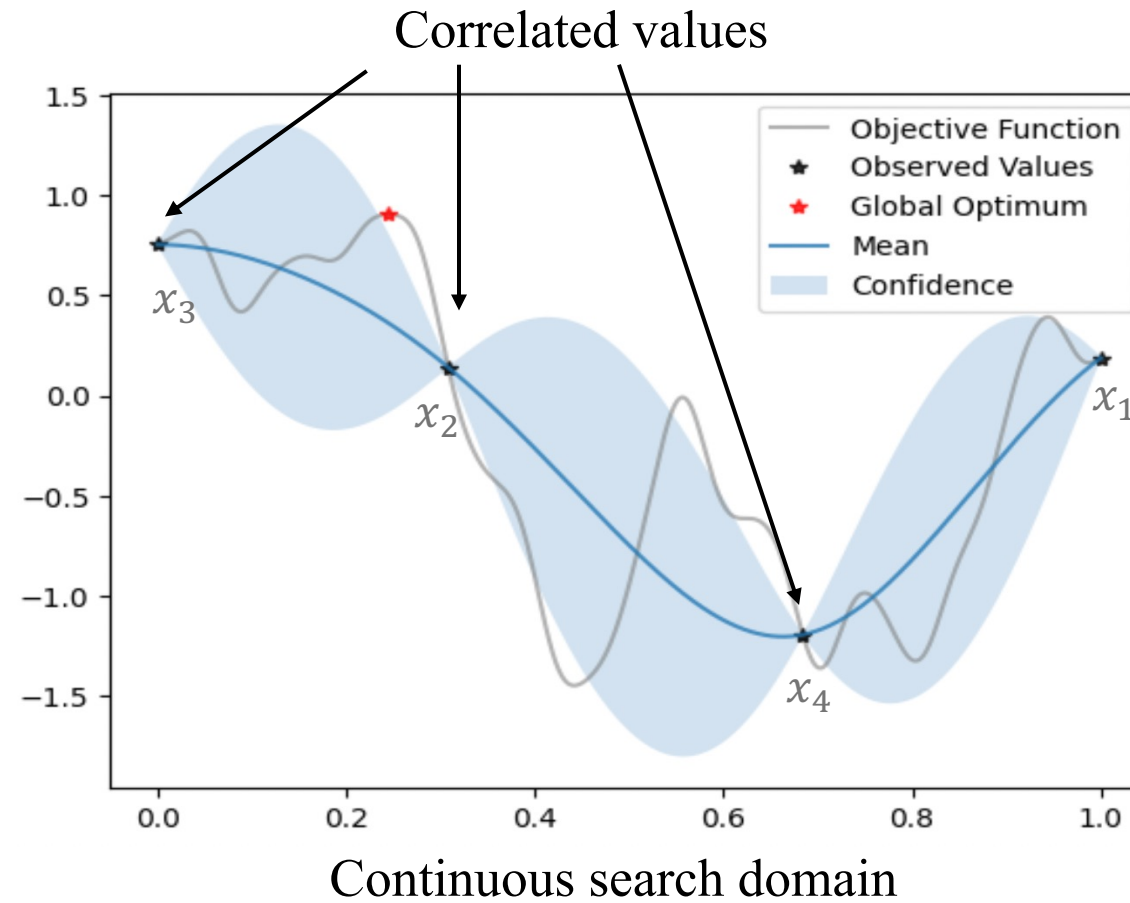
~~$t=4$~~







\vdots

~~$t=T$~~

Why is it hard?

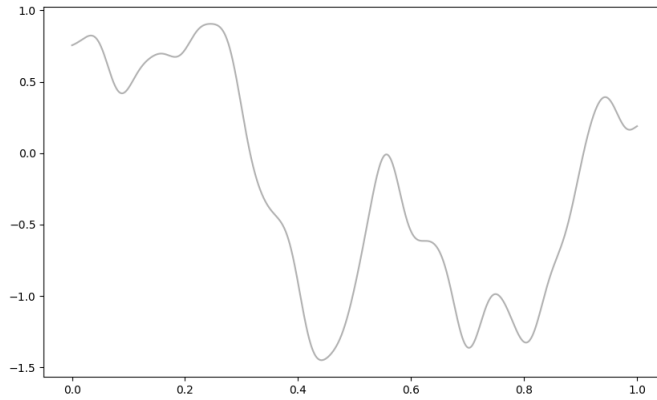


Hard budget constraint

$t=1$ 
 $t=2$ 
 $t=3$ 
 $t=4$ 
 \vdots
 $t=T$

\Rightarrow Optimal policy unknown!

Bayesian Optimization

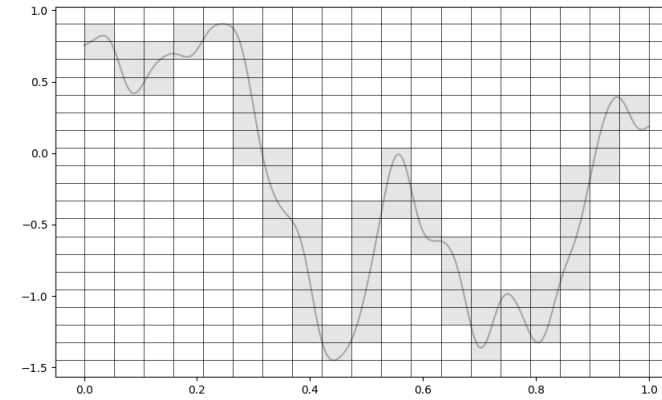
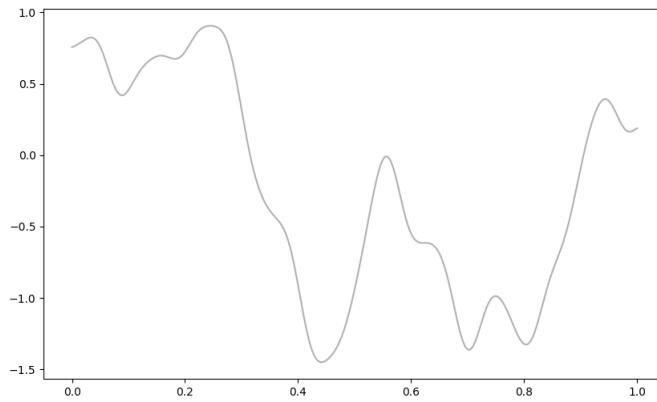


Continuous

Correlated

Hard budget constraint

Bayesian Optimization



Continuous

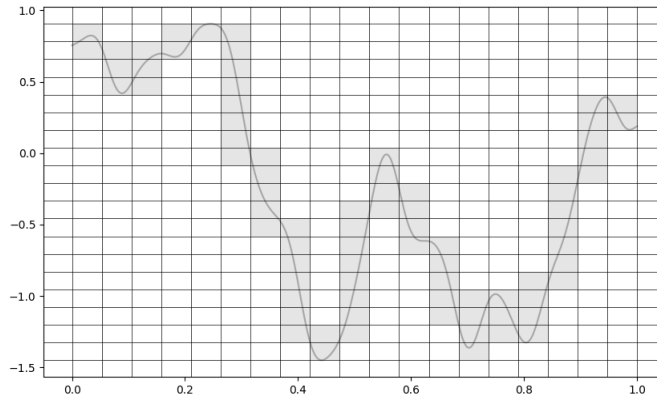


Discrete

Correlated

Hard budget constraint

Bayesian Optimization



Continuous

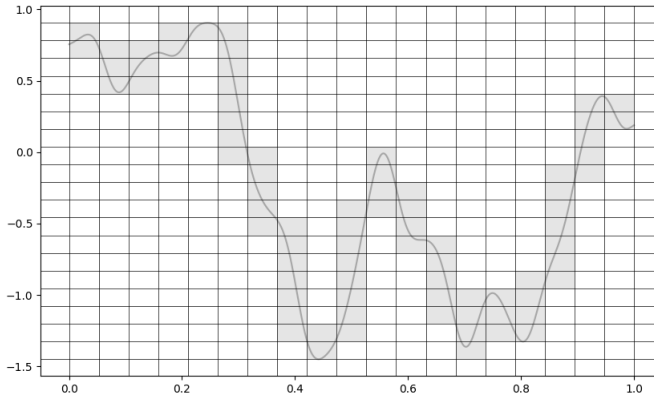


Discrete

Correlated

Hard budget constraint

Bayesian Optimization



Continuous



Discrete

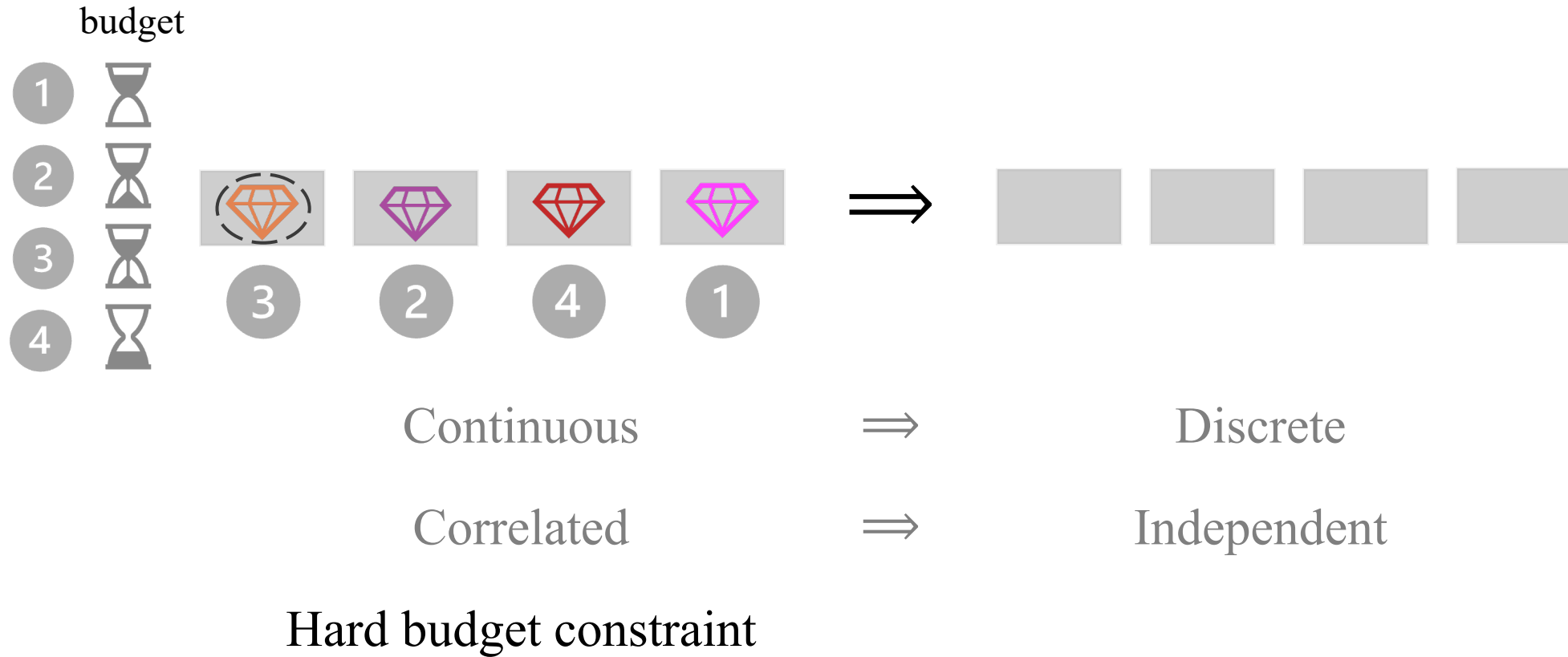
Correlated



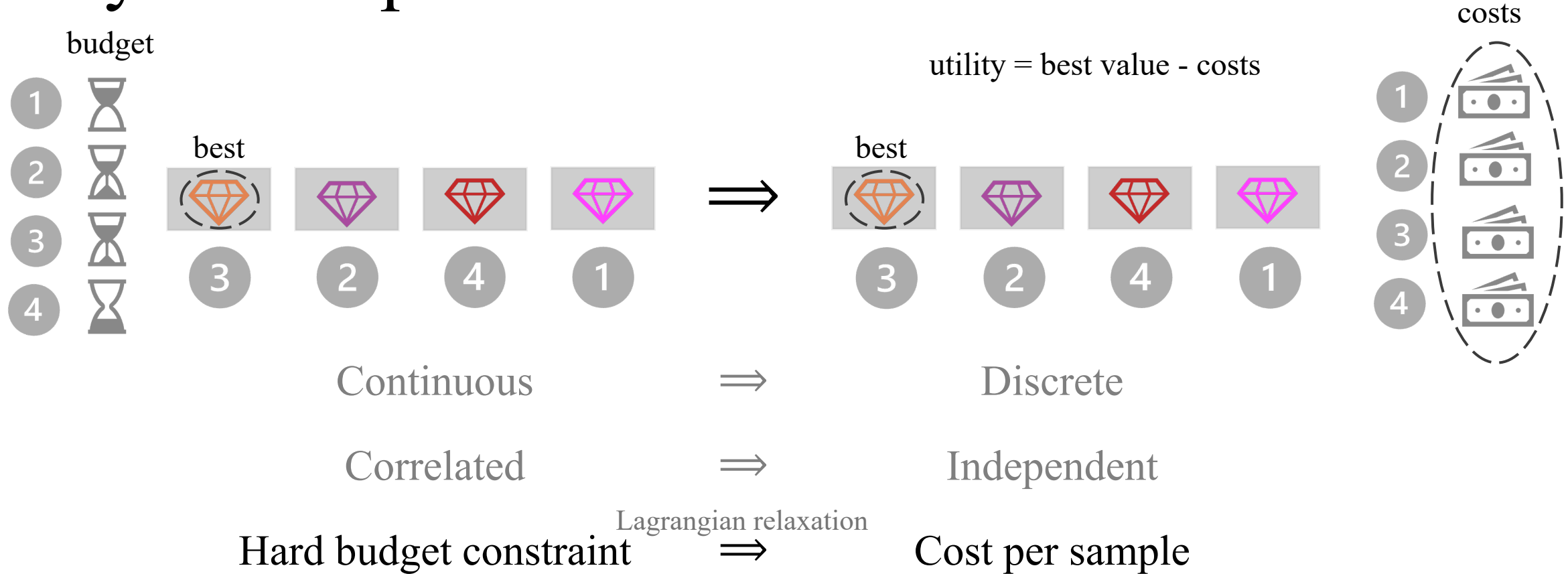
Independent

Hard budget constraint

Bayesian Optimization

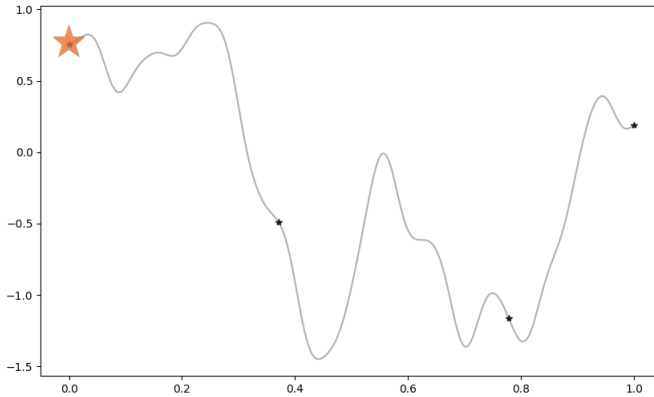


Bayesian Optimization



Bayesian Optimization \Rightarrow Pandora's Box

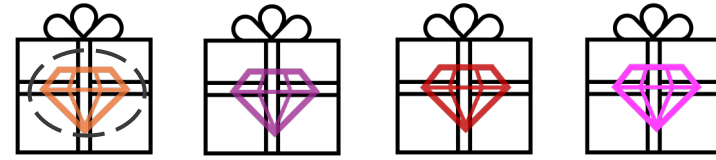
Special case of Markovian/
Bayesian Multi-armed Bandits



Continuous

Correlated

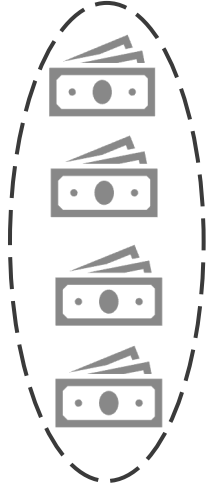
Hard budget constraint



Discrete

Independent

Cost per sample

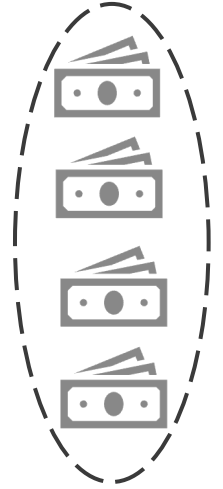
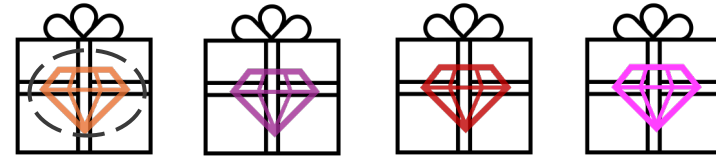


Bayesian Optimization \Rightarrow Pandora's Box

Special case of Markovian/
Bayesian Multi-armed Bandits



Continuous



Discrete



Correlated

Independent

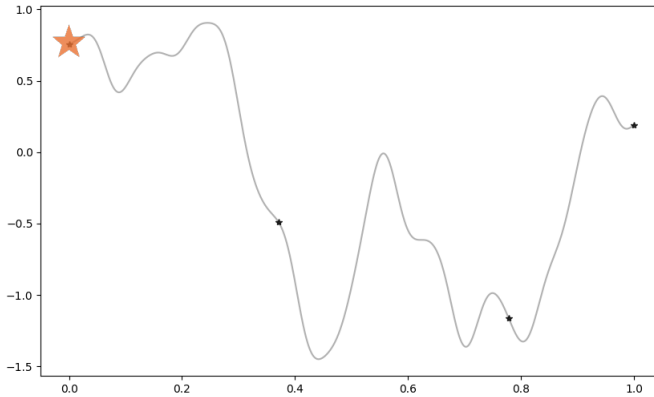


Hard budget constraint

Cost per sample

Optimal policy: Gittins index

Bayesian Optimization \Rightarrow Pandora's Box

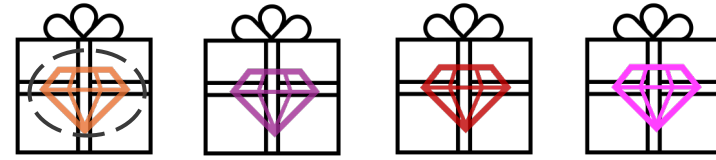


Continuous

Correlated

Hard budget constraint

Is Gittins index good?

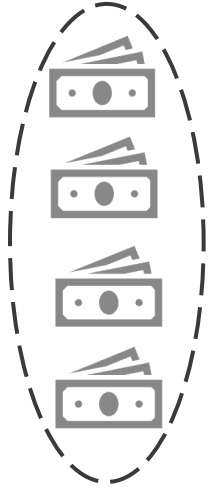


Discrete

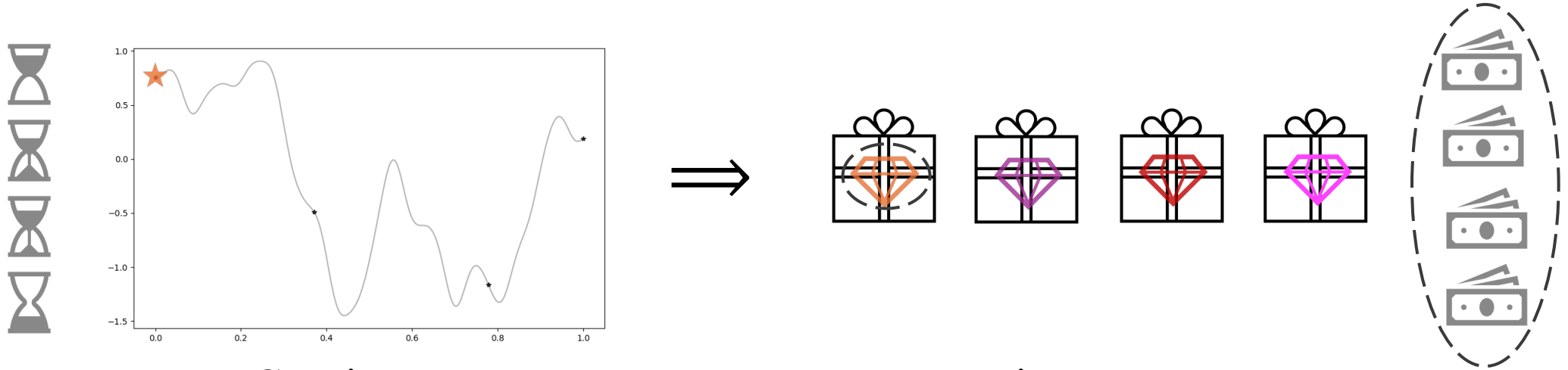
Independent

Cost per sample

Optimal policy: Gittins index



Bayesian Optimization \Rightarrow Pandora's Box



Continuous

Discrete

Correlated

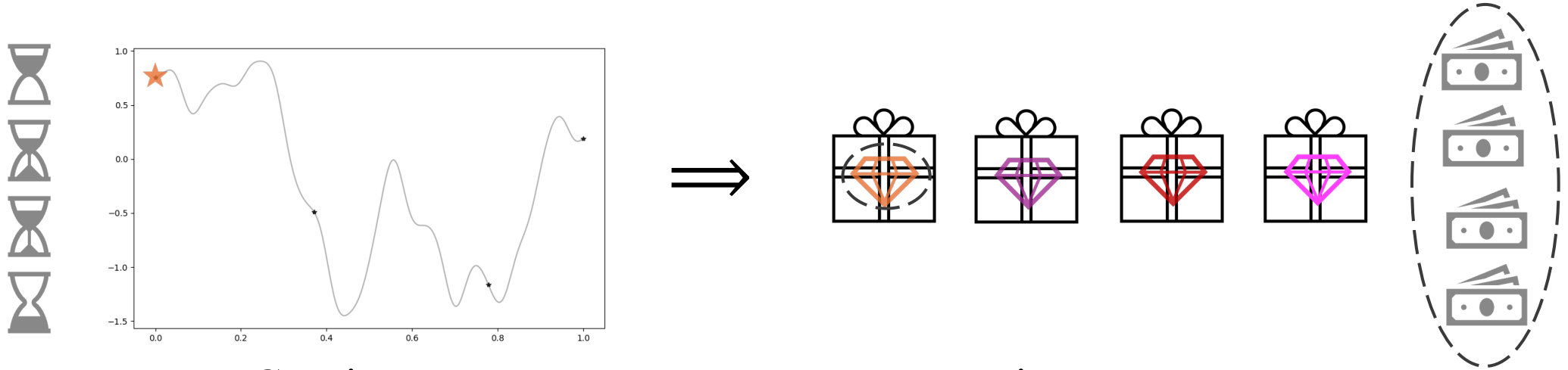
Independent

Hard budget constraint

Cost per sample

Is Gittins index good? $\xRightarrow{\text{How to translate?}}$ Optimal policy: Gittins index

Bayesian Optimization \Rightarrow Pandora's Box



Continuous

\Rightarrow

Discrete

Correlated

\Rightarrow

Independent

Hard budget constraint

\Rightarrow

Cost per sample

Is Gittins index good?

How to translate?

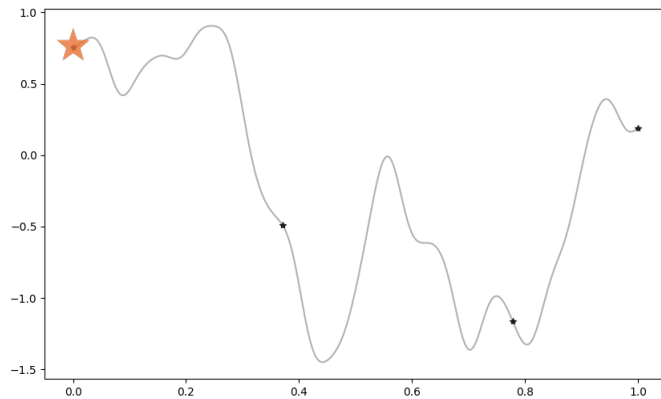
\Leftarrow

Optimal policy: Gittins index

Our contribution!

Our Contributions

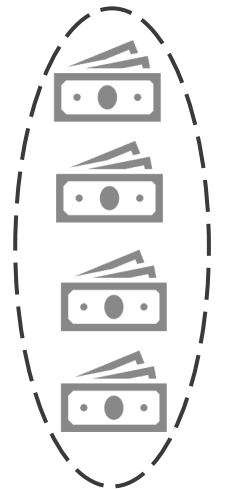
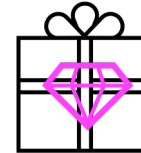
- How to translate?
- Is Pandora's Box Gittins index (PBGI) good?



?

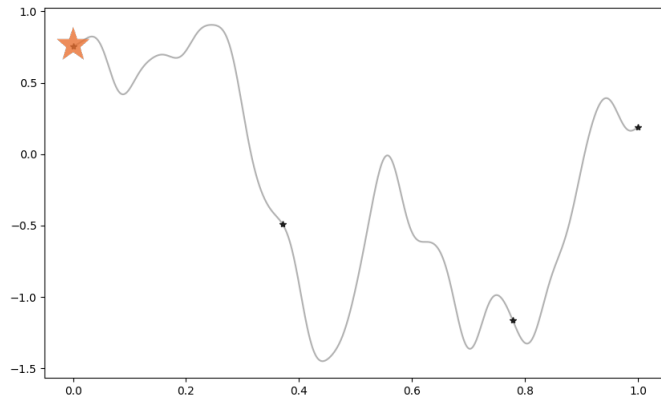


Pandora's Box Gittins index



Our Contributions

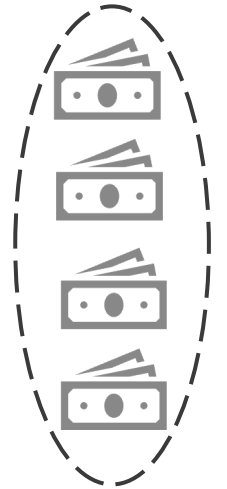
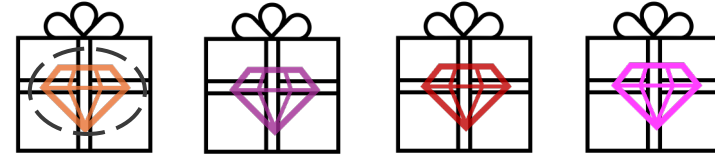
- Develop **Gittins index function** for Bayesian optimization
- Show **performance** against baselines on synthetic & empirical experiments



Our work

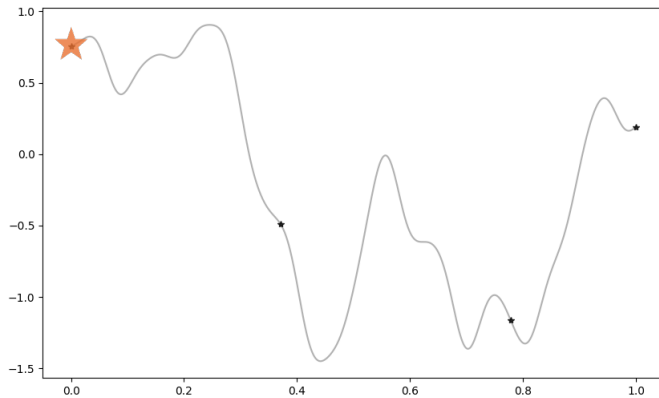


Gittins index



Our Contributions

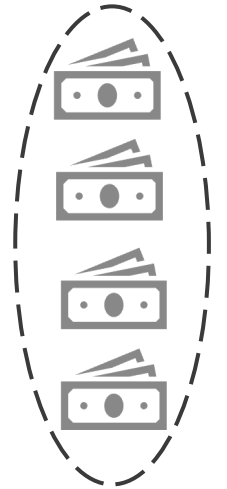
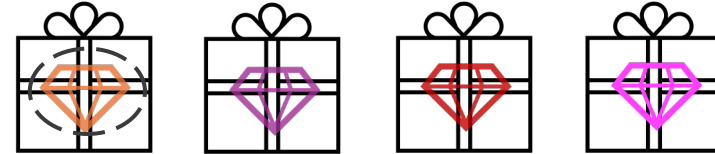
- Develop **Gittins index function** for Bayesian optimization
- Show performance against **baselines** on synthetic & empirical experiments



Our work

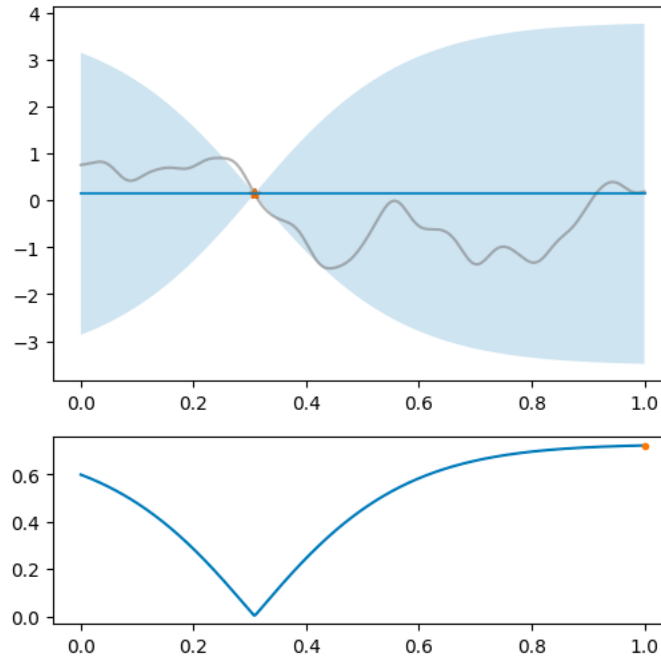


Gittins index



How is our **Gittins index function** different from **baselines**?

Popular One-step Heuristic: EI



mean: prediction
variance: confidence/uncertainty

Trade-off between

- exploitation (high mean) and
- exploration (high uncertainty)

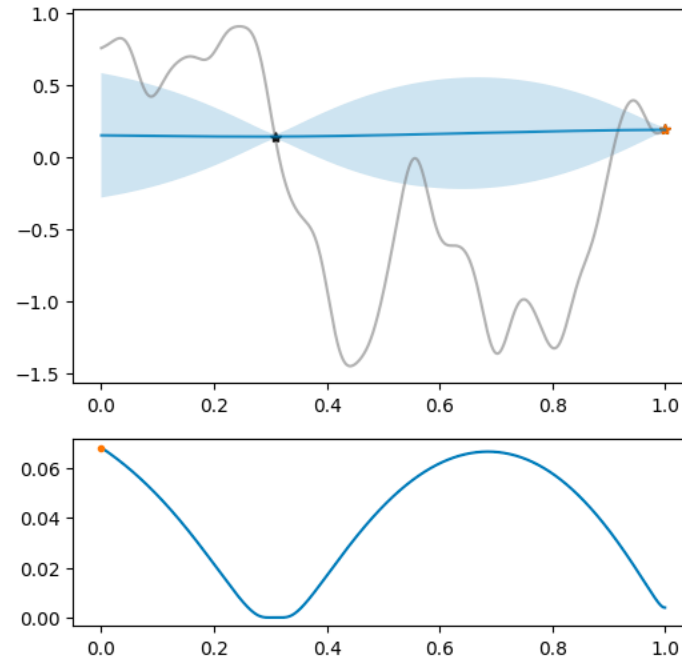
Expected improvement

$$\text{EI}(x; y) = \mathbb{E}[(f(x) - y)^+]$$

y_{best} : current best observed value

EI policy: evaluate $\text{argmax}_x \text{EI}(x; y_{\text{best}})$

Popular One-step Heuristic: EI



mean: prediction
variance: confidence/uncertainty

Trade-off between

- exploitation (high mean) and
- exploration (high uncertainty)

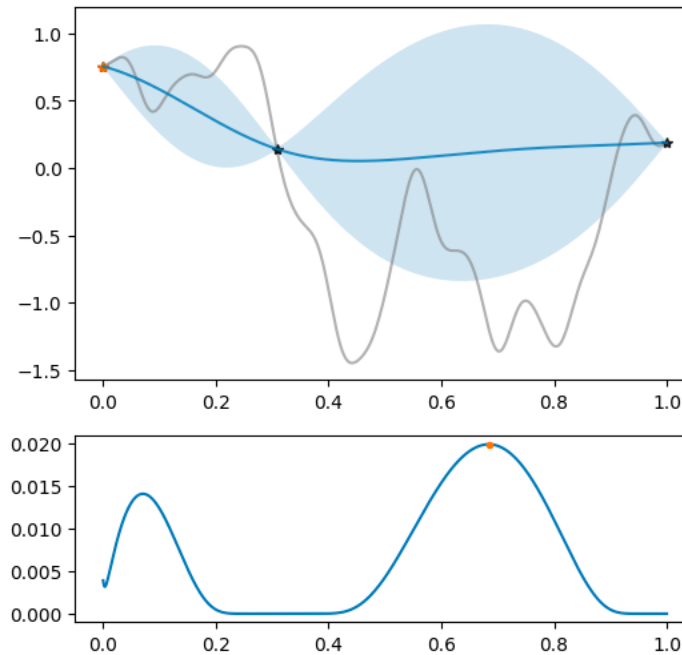
Expected improvement

$$\text{EI}(x; y) = \mathbb{E}[(f(x) - y)^+]$$

y_{best} : current best observed value

EI policy: evaluate $\text{argmax}_x \text{EI}(x; y_{\text{best}})$

Popular One-step Heuristic: EI



mean: prediction
variance: confidence/uncertainty

Trade-off between

- exploitation (high mean) and
- exploration (high uncertainty)

Expected improvement

$$\text{EI}(x; y) = \mathbb{E}[(f(x) - y)^+]$$

y_{best} : current best observed value

EI policy: evaluate $\text{argmax}_x \text{EI}(x; y_{\text{best}})$

Popular One-step Heuristic: EI

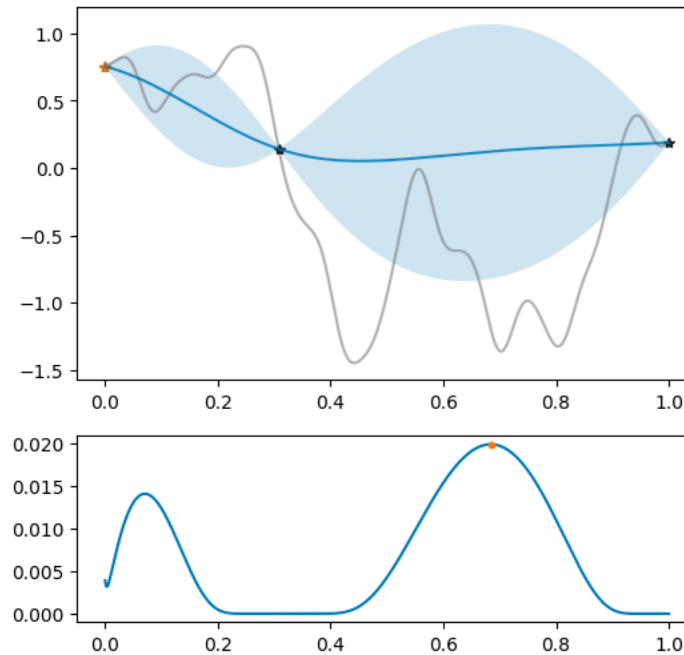
Other heuristics:

simple

- Upper Confidence Bound
- Thompson Sampling (TS)
- Predictive Entropy Search

slow

- Knowledge Gradient
- Multi-step Lookahead EI



mean: prediction

variance: confidence/uncertainty

Trade-off between

- exploitation (high mean) and
- exploration (high uncertainty)

Expected improvement

$$EI(x; y) = \mathbb{E}[(f(x) - y)^+]$$

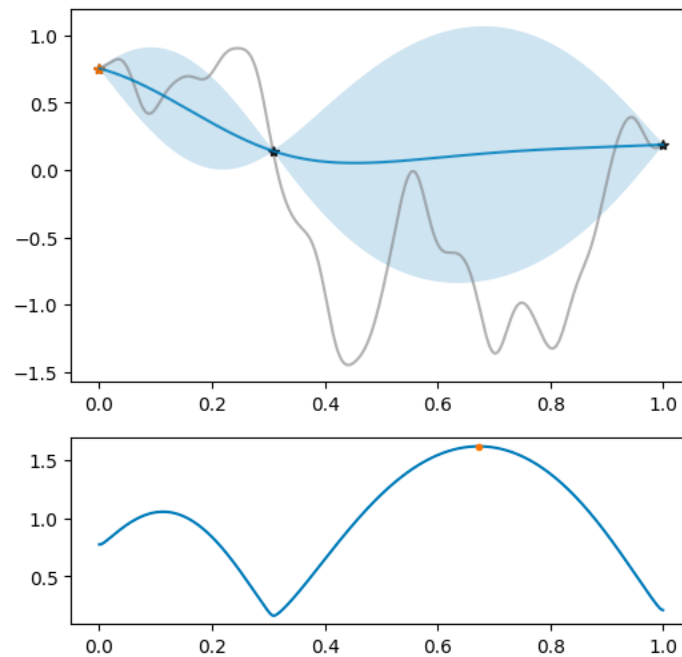
y_{best} : current best observed value

EI policy: evaluate $\operatorname{argmax}_x EI(x; y_{\text{best}})$

New One-step Heuristic: PBGI

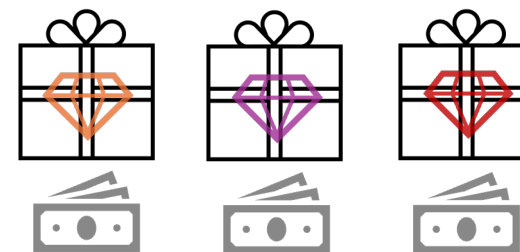
Other heuristics:

- Upper Confidence Bound
- Thompson Sampling (TS)
- Knowledge Gradient
- Predictive Entropy Search
- Multi-step Lookahead EI



Pandora's box Gittins index

Pandora's box



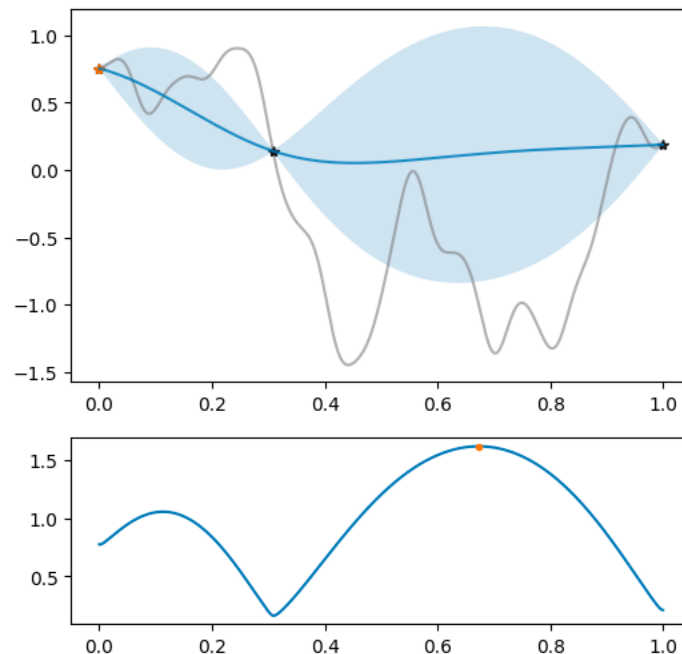
$g(x)$: Gittins index function

PBGI policy: evaluate $\operatorname{argmax}_x g(x)$

New One-step Heuristic: PBGI

Other heuristics:

- Upper Confidence Bound
- Thompson Sampling (TS)
- Knowledge Gradient
- Predictive Entropy Search
- Multi-step Lookahead EI

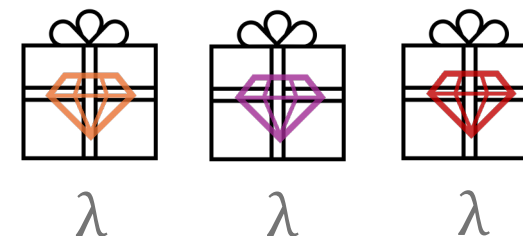


Pandora's box Gittins index

$$\text{EI}(x; y) = \mathbb{E}[(f(x) - y)^+]$$

PBGI policy: evaluate $\arg\max_x g(x)$

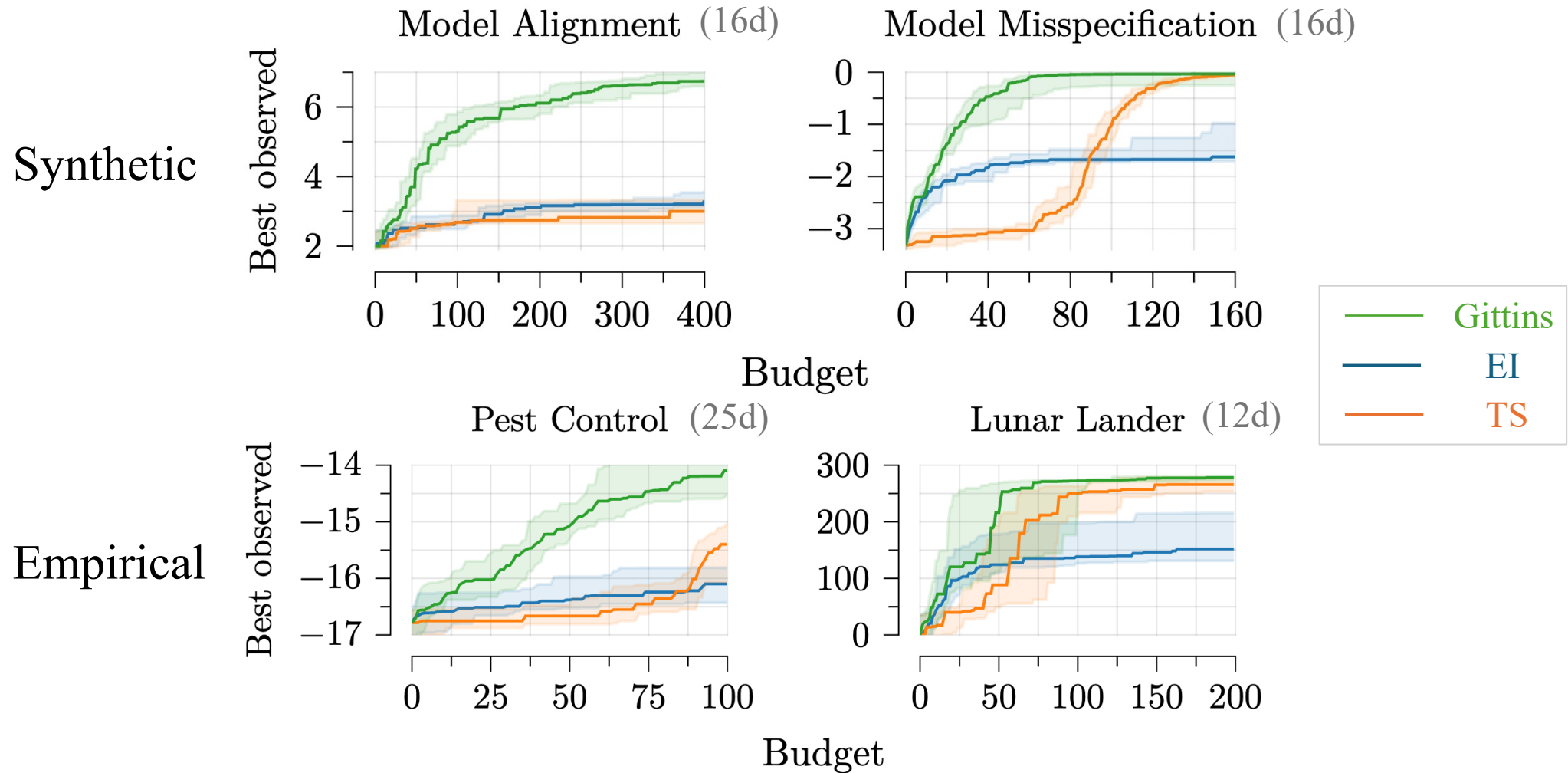
Pandora's box



λ : cost per sample

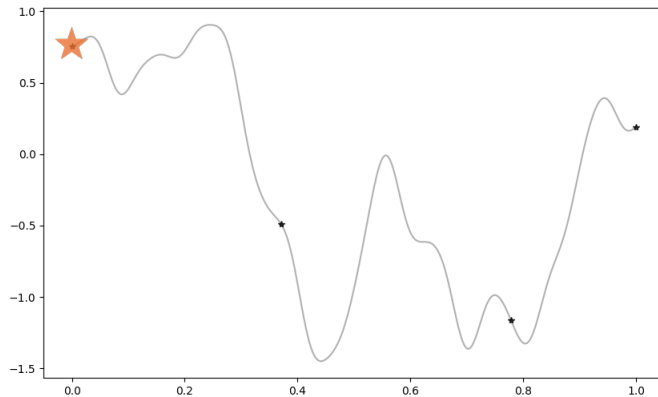
$g(x)$: solution to $\text{EI}(x; g(x)) = \lambda$

Experiment Results: Gittins vs EI vs TS

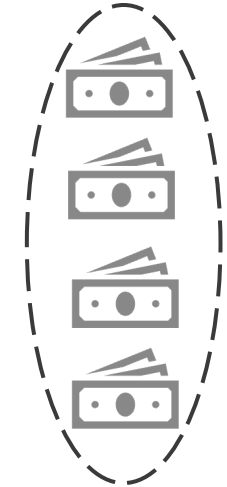
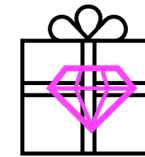
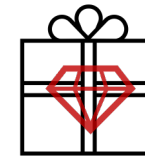


Conclusions

- Propose **easy-to-compute** Gittins index function for Bayesian optimization



Our work

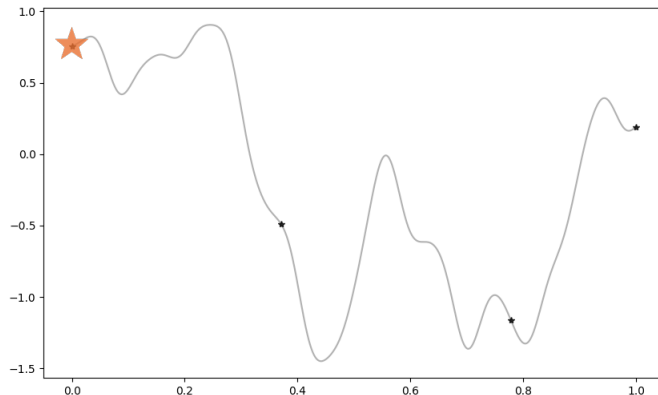


Gittins index

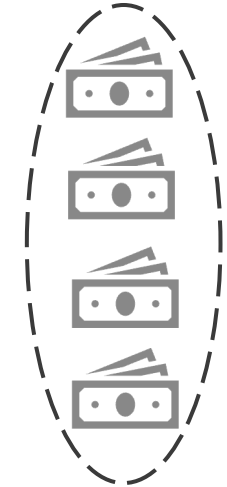
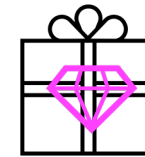
Check our preprint on ArXiv!

Conclusions

- Propose easy-to-compute Gittins index function for Bayesian optimization
- Show **Gittins mostly outperforms baselines** on synthetic & empirical experiments particularly higher dimensions and larger domains!



Our work

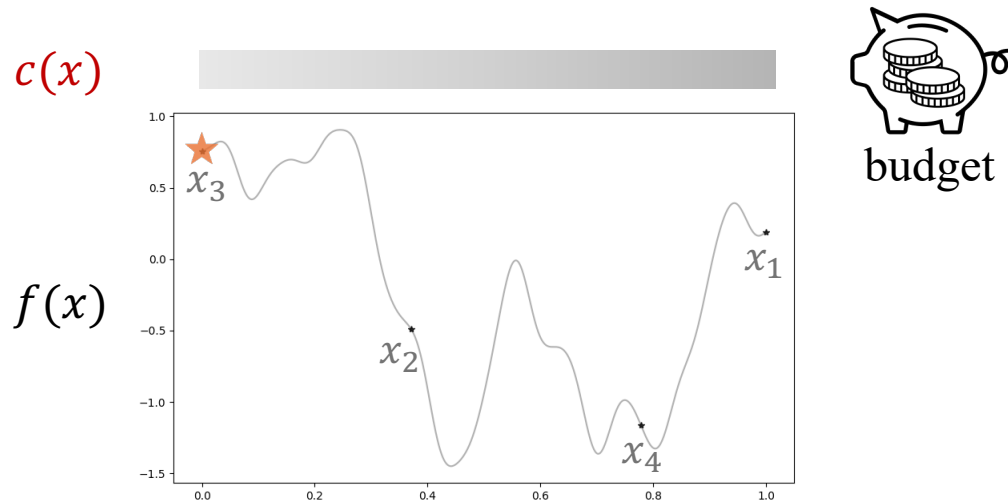


Gittins index

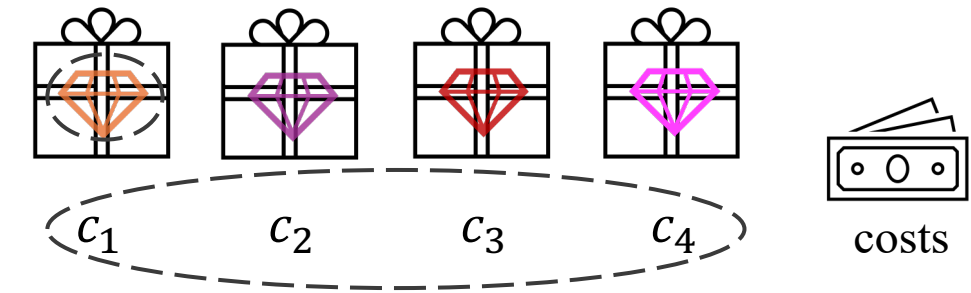
Check our preprint on ArXiv!

Conclusions

- Propose easy-to-compute Gittins index function for Bayesian optimization
- Show Gittins mostly outperforms baselines on synthetic & empirical experiments
- Extend to Bayesian optimization with **heterogeneous evaluation costs**



Our work

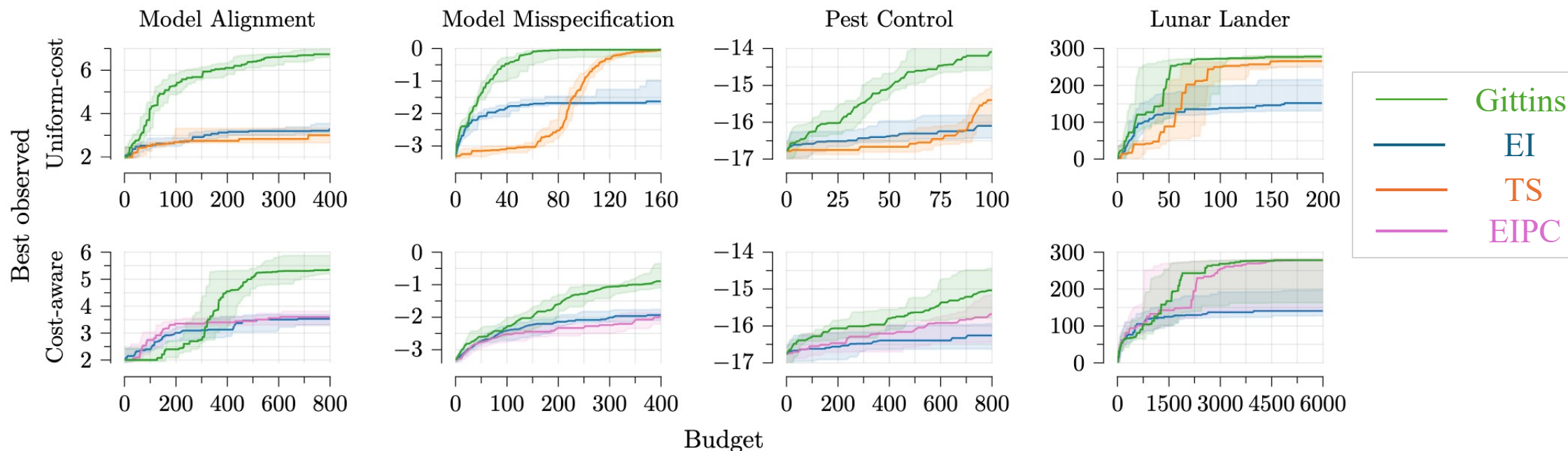


Pandora's Box Gittins index

Check our preprint on ArXiv!

Heterogeneous-cost Experiment Results

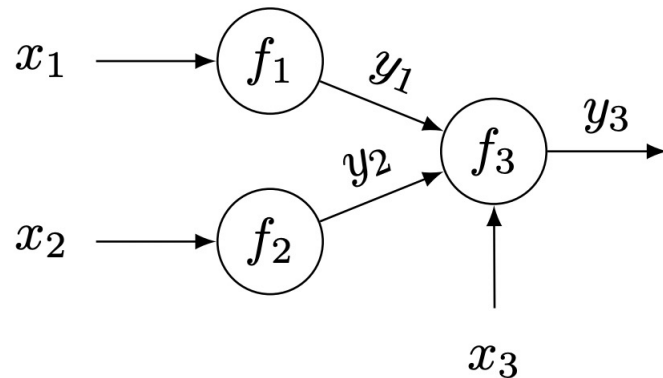
- Show Gittins mostly outperforms baselines on synthetic & empirical experiments
- Extend to Bayesian optimization with **heterogeneous evaluation costs**



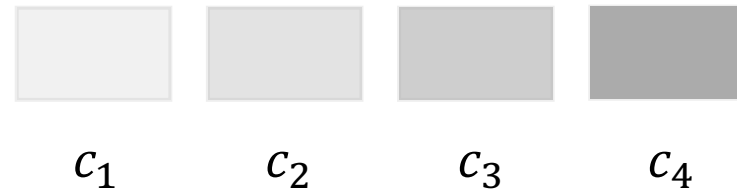
Check our preprint on ArXiv!

Conclusions

- Propose easy-to-compute Gittins index function for Bayesian optimization
- Show Gittins mostly outperforms baselines on synthetic & empirical experiments
- Extend to Bayesian optimization with heterogeneous evaluation costs
- Open door for **exotic BO** (freeze-thaw, multi-fidelity, function network, etc.)



?



Gittins index

Check our preprint on ArXiv!