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

Qian Xie[1], Raul Astudillo[2], Peter Frazier[1], Ziv Scully[1], Alexander Terenin[1]

[1] Cornell University, [2] California Institute of Technology

Introduction to Bayesian optimization

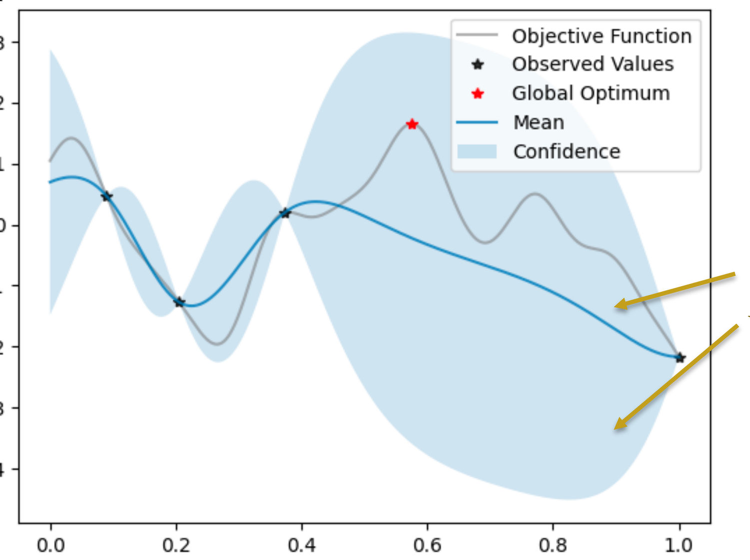
Goal: optimize expensive-to-evaluate black-box function

∈ decision-making under uncertainty

Applications:
Hyperparameter tuning
Drug discovery
Control design

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

Gaussian process: infinite-dimensional generalization of multivariate normal distributions



x: hyperparameter/configuration

mean: prediction
variance: confidence/uncertainty

Trade-off between
- exploitation (high mean) and
- exploration (high uncertainty)

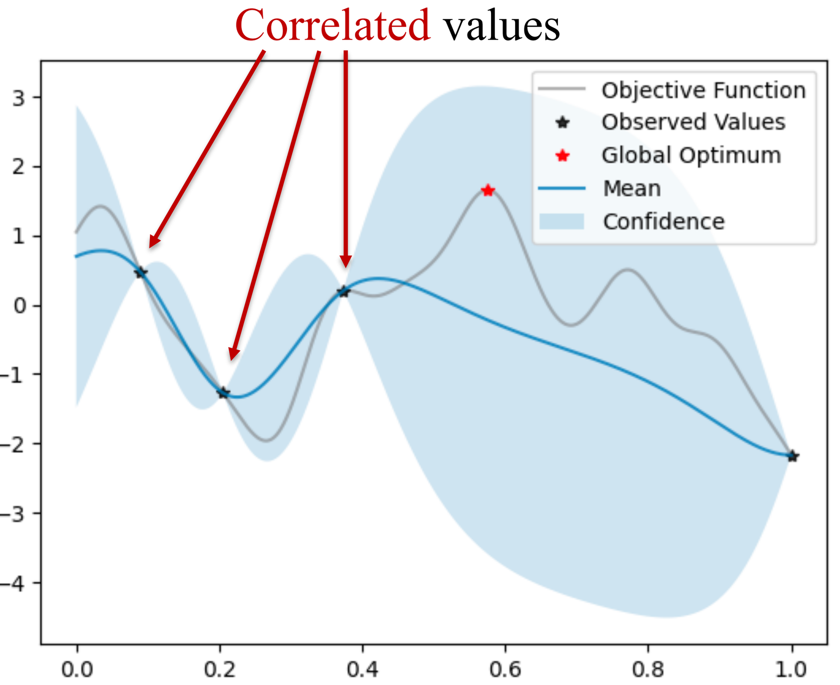
Objective: find global optimum $x^* = \operatorname{argmax}_{x \in \mathcal{X}} f(x)$
⇓
Objective: optimize best observed value at time T

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

Decision: evaluate a set of points
⇓
Decision: adaptively evaluate $x_1, x_2, \dots, x_T \in \mathcal{X}$ given time budget T

Why is Bayesian optimization hard?

Correlated values



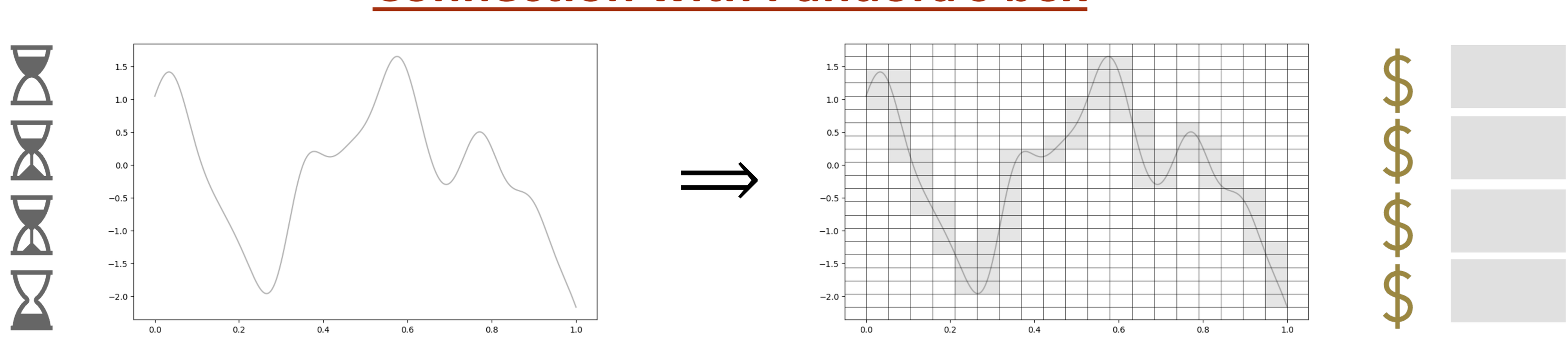
Continuous search domain

Hard budget constraint

~~$t=1$~~
~~$t=2$~~
~~$t=3$~~
~~$t=4$~~
⋮
$t = T$

Optimal policy unknown!

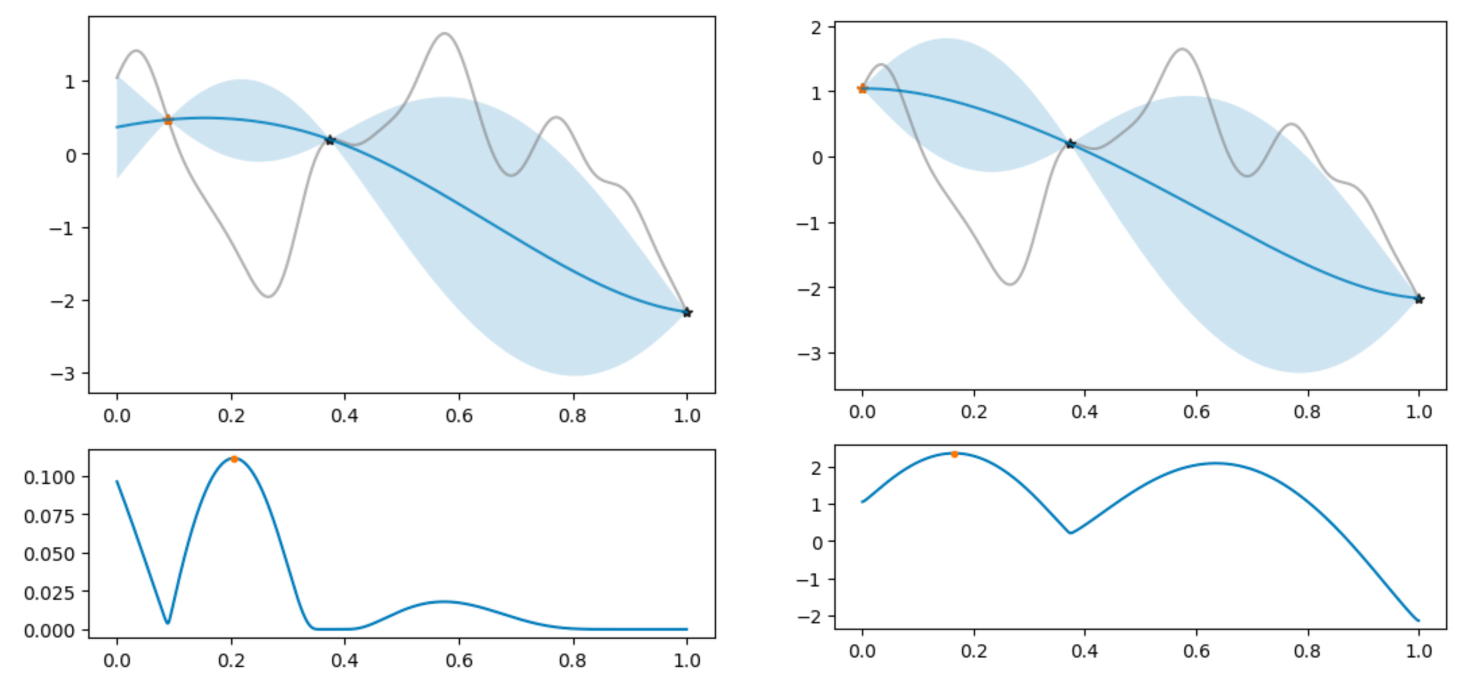
Connection with Pandora's box



Continuous	⟹	Discrete
Correlated	⟹	Independent
Hard budget constraint	⟹ (Lagrangian relaxation)	Cost per sample
Is Gittins index good?	⟸ (How to translate?)	Optimal policy: Gittins index

Special case of Markovian/Bayesian Multi-armed Bandits

Acquisition functions: EI vs Gittins



Expected improvement (EI)

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

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

y_{best}: current best observed value

Pandora's box Gittins index (PBGI)

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

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

λ: cost-per-sample (Lagrangian multiplier)

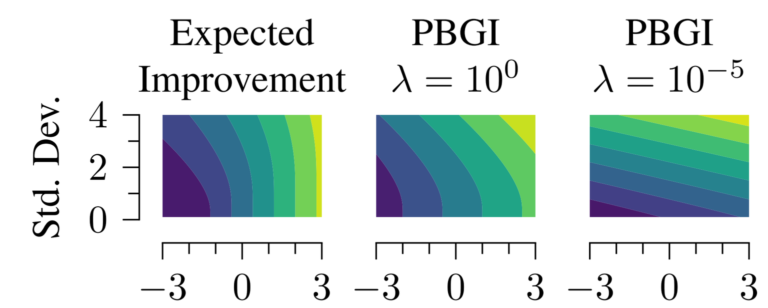
Both are one-step heuristics!

Other acquisition functions:

simple
- Upper Confidence Bound (UCB)
- Thompson Sampling (TS)

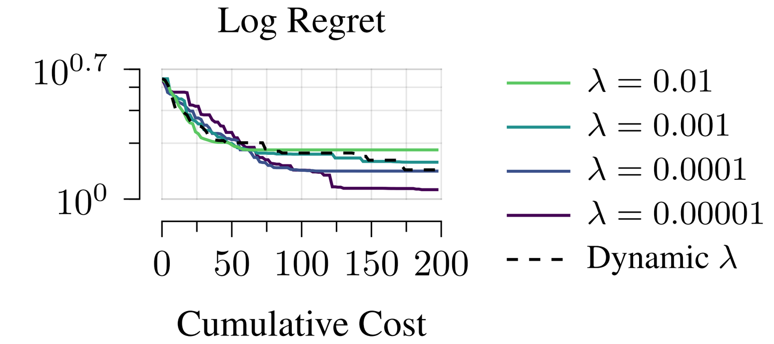
slow
- Predictive Entropy Search
- Knowledge Gradient (KG)
- Multi-step Lookahead EI (MSEI)

Contour Plot



Connection with UCB?

Impact of λ



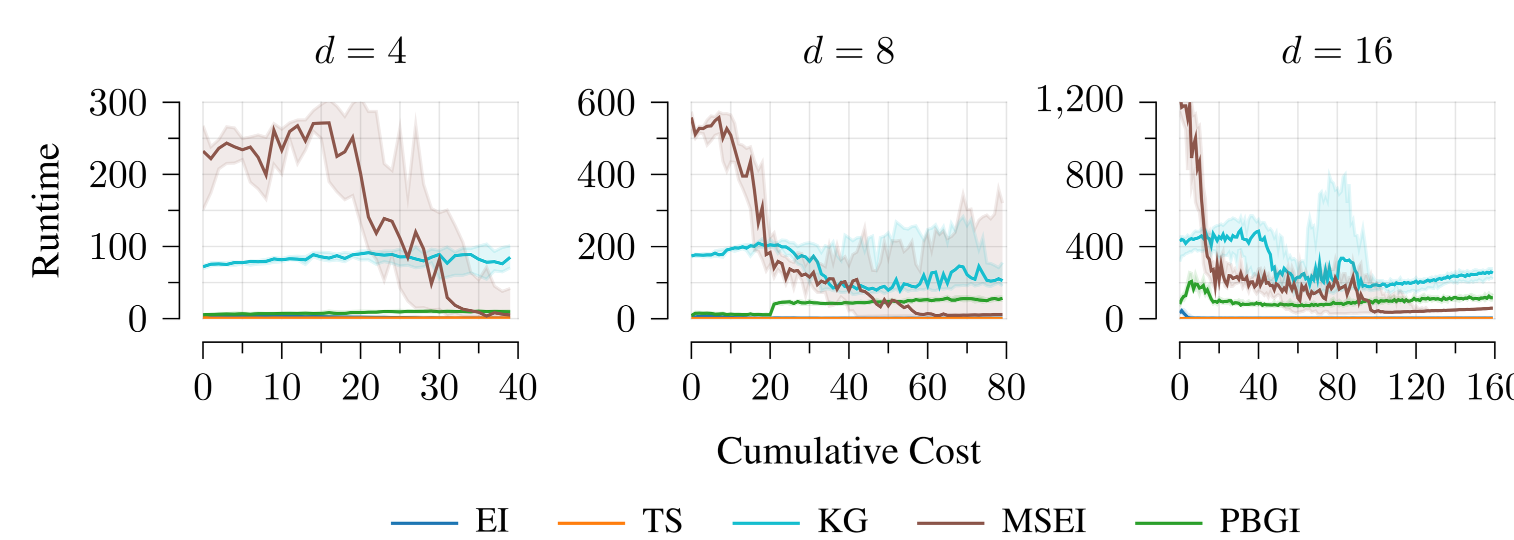
Smaller λ, higher exploration

Extension to heterogeneous costs

- Given cost function $c: \mathcal{X} \to \mathbb{R}^+$ and budget B
- Replace λ with $\lambda c(x)$ to compute $g(x)$ as PBGI

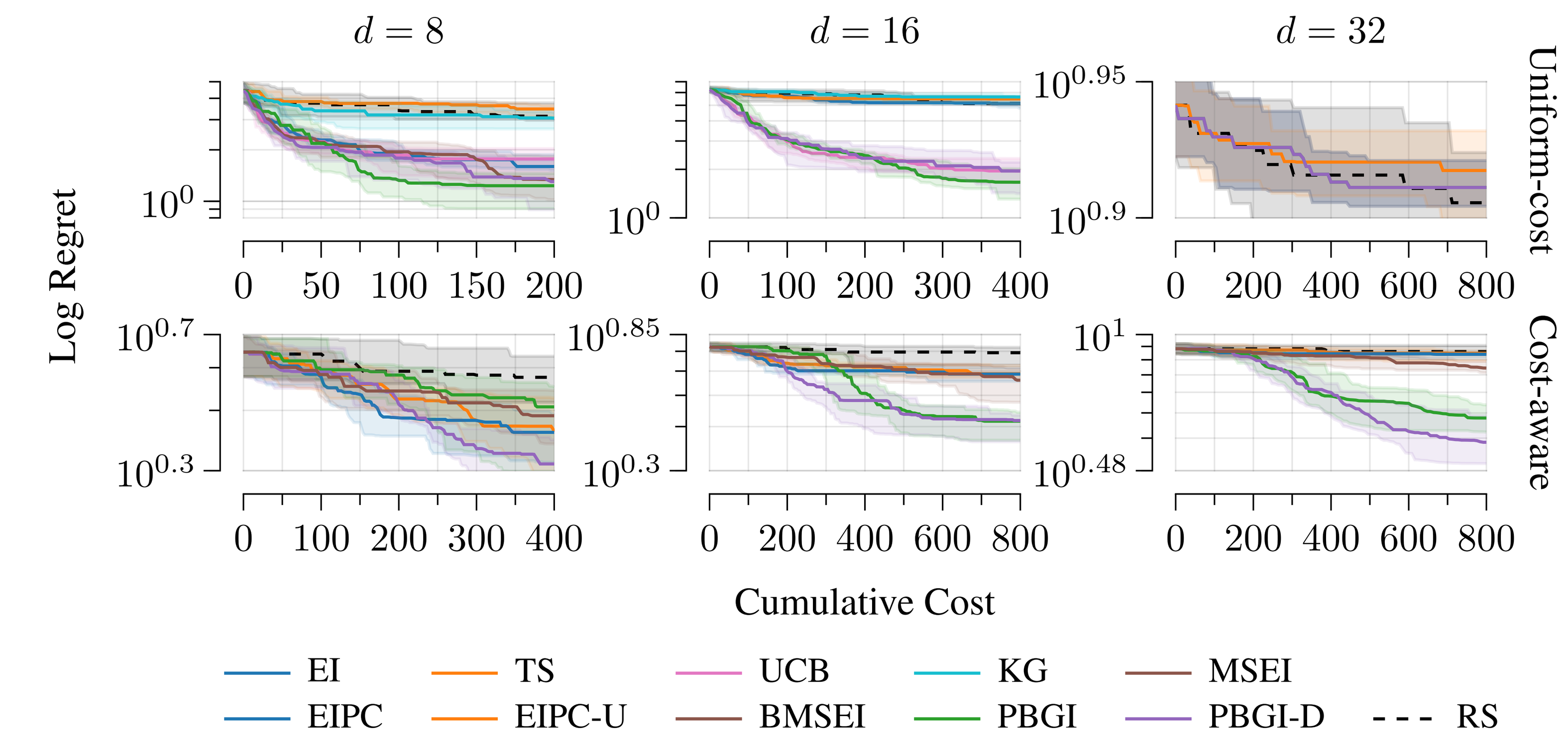
Baselines: arbitrarily bad
- EI Per Unit Cost (EIPC)
- Budgeted MSEI (BMSEI) slow

Experiment: timing

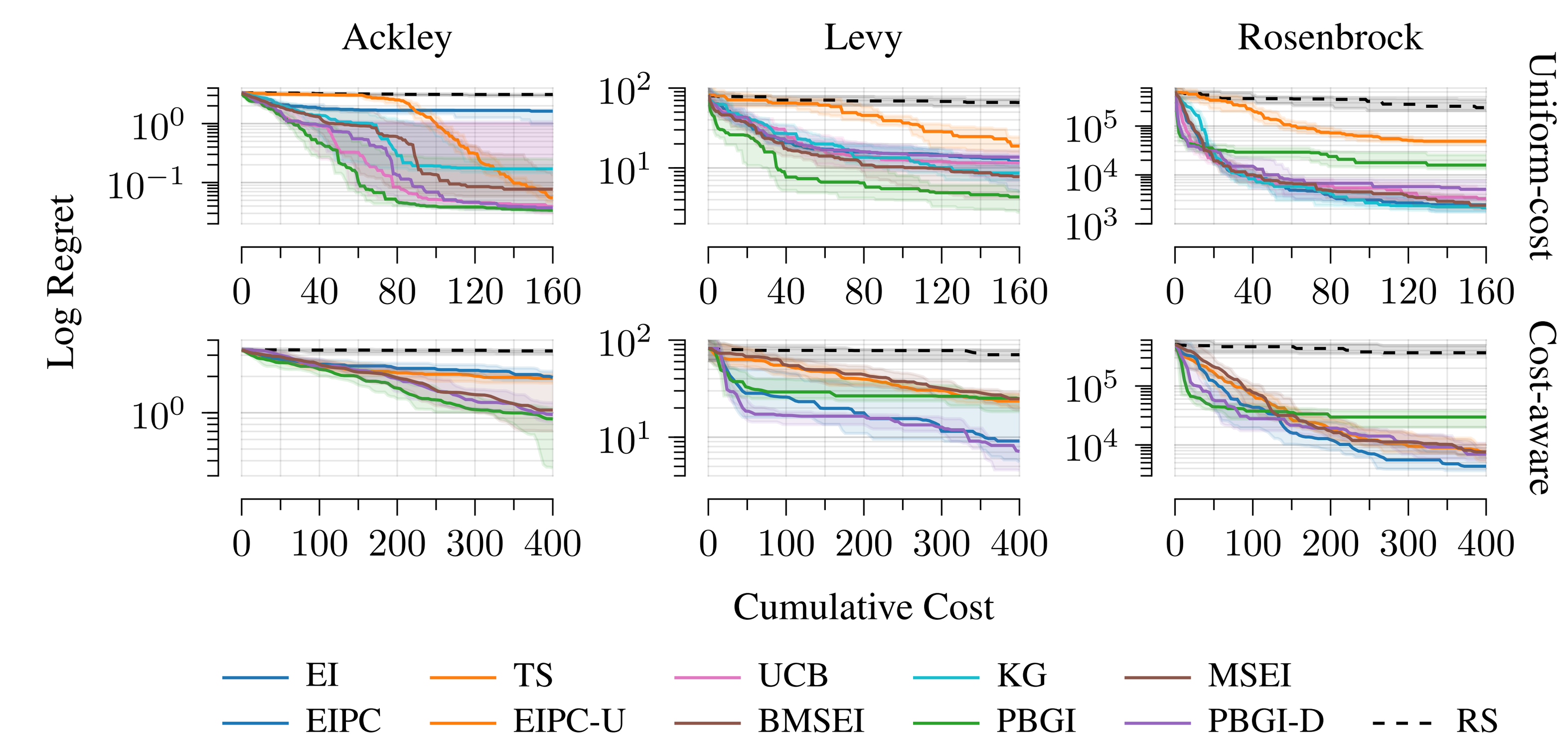


Gittins is easy to compute using bisection method

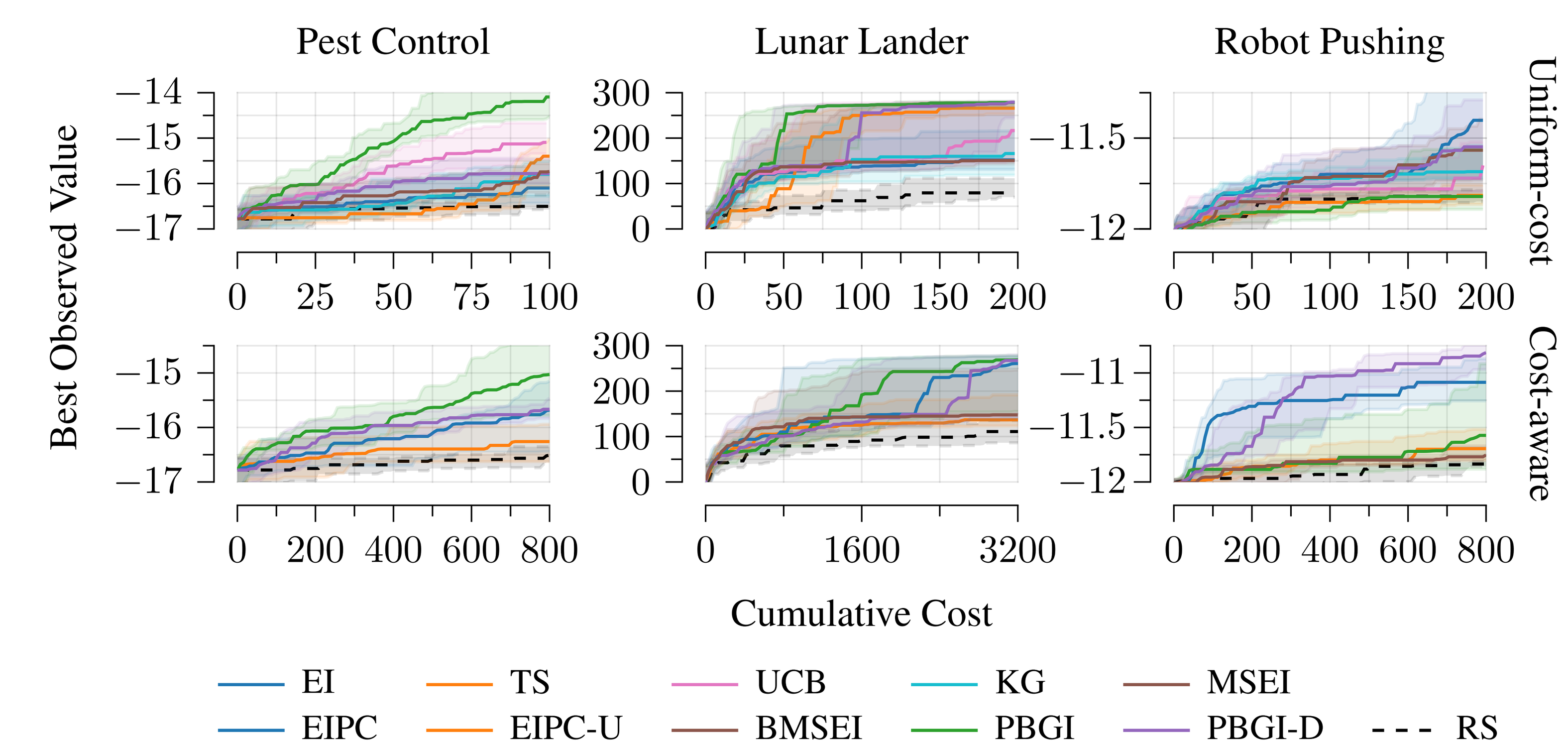
Experiment: Bayesian regret



Experiment: synthetic benchmark



Experiment: empirical



Future work

Extension to exotic BO (freeze-thaw, multi-fidelity, function network, etc.) via Gittins variants (Golf/MDP, optional inspection, etc.)