# Preferential Bayesian Optimization with Hallucination Believer

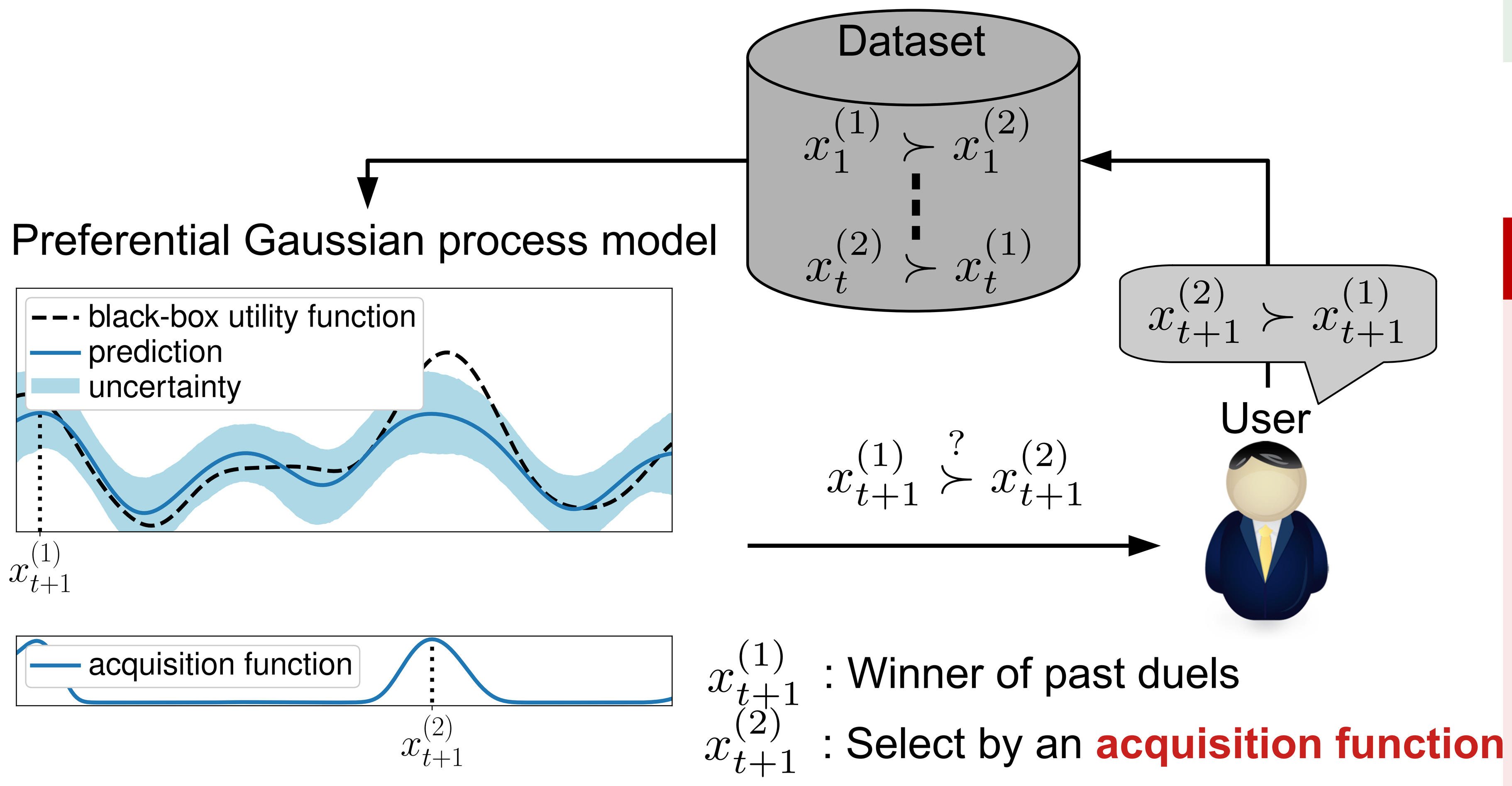
Shion Takeno[1], Masahiro Nomura [2], and Masayuki Karasuyama[1]

[1]Nagoya Institute of Technology, [2]CyberAgent, inc.

## 1. Overview

- Black-box utility function optimization via **dueling feedback**:

$$\operatorname*{argmax}_{\boldsymbol{x}\in\mathcal{X}} f(\boldsymbol{x}), \text{ via } \{\boldsymbol{x}^{(1)} \succ \boldsymbol{x}^{(2)}\} \Leftrightarrow \{f(\boldsymbol{x}^{(1)}) > f(\boldsymbol{x}^{(2)})\}$$




### Applications

1. Recommendation system ($\boldsymbol{x}$ is a recommended item)
2. A/B test ($\boldsymbol{x}$ is a design of webpage)

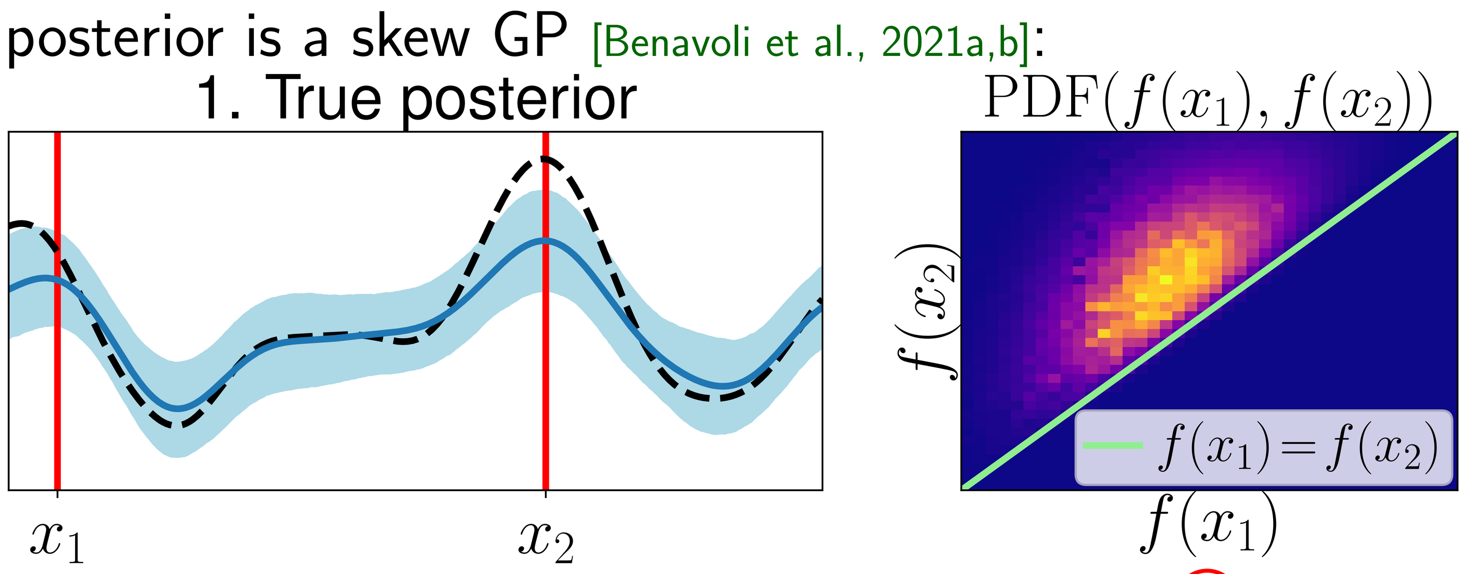
**Difficulty**: Exact posterior is computationally intractable **skew Gaussian process** (skew GP).

### Contributions

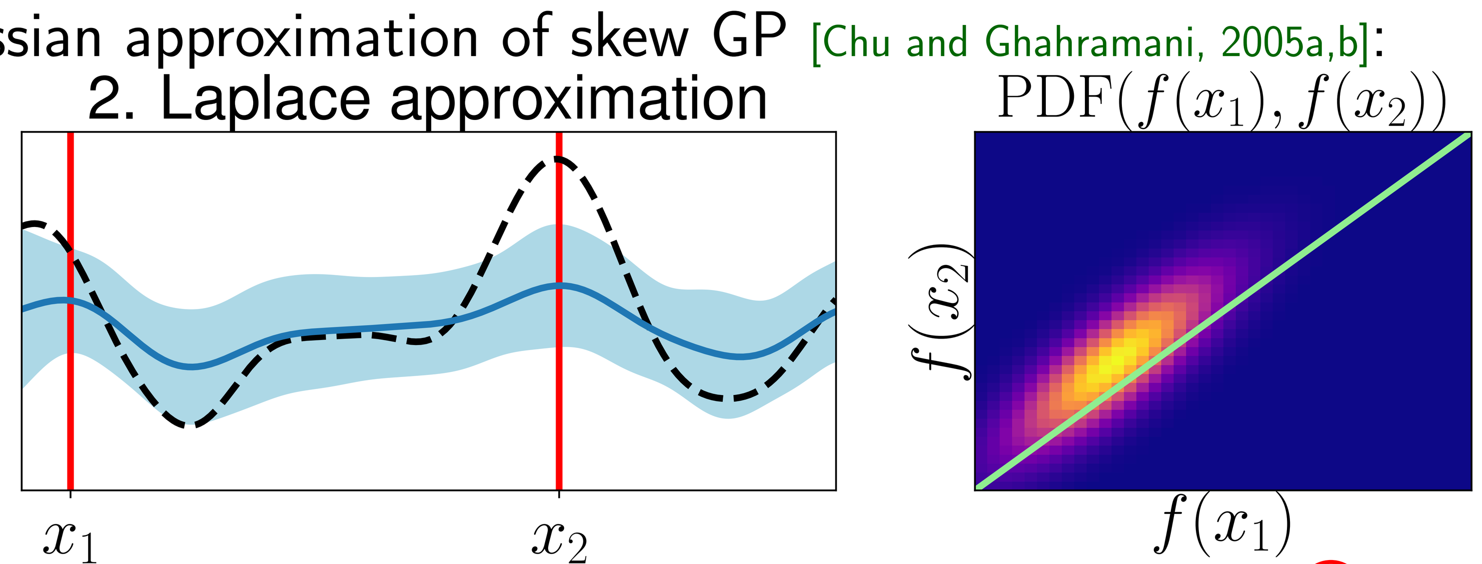
1. We propose easy-to-use preferential Bayesian optimization (BO) method called Hallucination Believer (HB).
   - Conditioning of hallucination avoids a direct use of skew GP.
2. Both complexities of HB are low ☺
   - **Sample complexity**: The number of queries should be small
   - **Computational complexity**: User wait time should be small
3. Arbitrary acquisition function (AF) for BO can be integrated ☺.

## 2. Skew Gaussian process (skew GP)

1. True posterior is a skew GP [Benavoli et al., 2021a,b]:




   - MCMC-based approximation is time-consuming. ☹
2. Gaussian approximation of skew GP [Chu and Ghahramani, 2005a,b]:




   - In right figures, Gauss. approx. shows poor accuracy ☹
     - ⋆ $\Pr\big(f(\boldsymbol{x}_2) < f(\boldsymbol{x}_1)\big) > 0$, which is almost zero in true posterior.

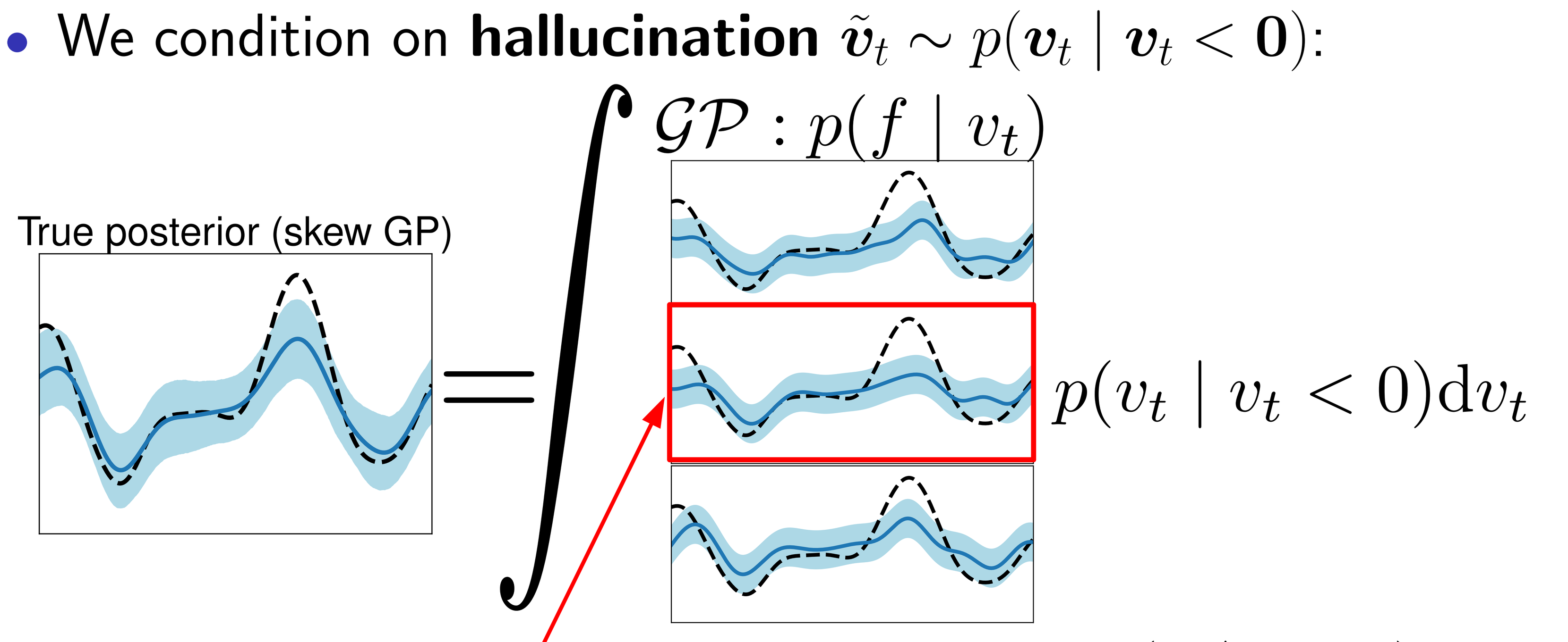
## 3. Key Insight of Hallucination Believer

- Training duels can be rewritten as $\boldsymbol{v}_t < \boldsymbol{0}$:
  - $\boldsymbol{v}_t = \big(f(\boldsymbol{x}_{i,l}) - f(\boldsymbol{x}_{i,w}) + 2\epsilon_i\big)_i \in \mathbb{R}^t$, where $\epsilon_i$ is noise
  - $\boldsymbol{x}_{i,l}$ and $\boldsymbol{x}_{i,w}$ are loser and winner at $i$-th iteration.

### Proposition

*Exact posterior distribution additionally conditioned by $\boldsymbol{v}_t$ is $p(f \mid \boldsymbol{v}_t < \boldsymbol{0}, \boldsymbol{v}_t) = p(f \mid \boldsymbol{v}_t)$, which is a GP.*

- Conditioning on $\boldsymbol{v}_t = \boldsymbol{c}$ by a constant $\boldsymbol{c}$ ignores skewness.
- We condition on **hallucination** $\tilde{\boldsymbol{v}}_t \sim p(\boldsymbol{v}_t \mid \boldsymbol{v}_t < \boldsymbol{0})$:

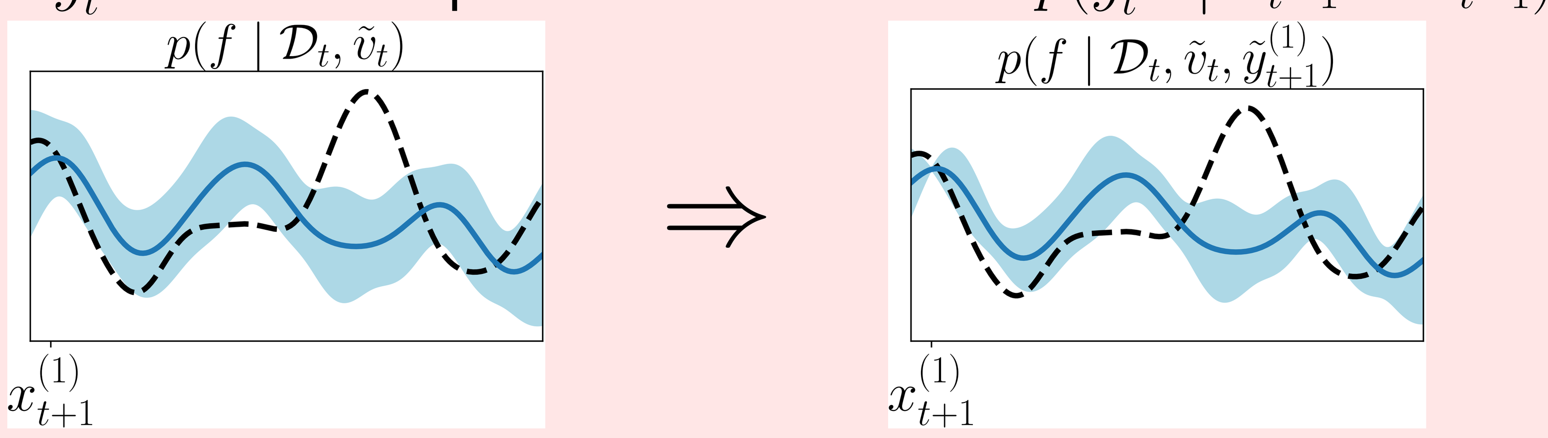



- **Computationally efficient** since only one sample $\tilde{\boldsymbol{v}}_t$ is used
- **Skewness is reflected** via exact posterior $p(\boldsymbol{v}_t \mid \boldsymbol{v}_t < \boldsymbol{0})$

## 4. Algorithm of Hallucination Believer

```
1: for t = 1, ... do
2:     x_t^(1) ← x_{t-1,w}, i.e., winner in past duels
3:     Generate ṽ_{t-1}, ỹ_t^(1) from p(v_{t-1}, y_t^(1) | v_{t-1} < 0)
4:     x_t^(2) ← argmax_{x∈X} α(x) based on the GP f | ṽ_{t-1}, ỹ_t^(1)
5:     Get x_{t,w} and x_{t,l} by duel between x_t^(1) and x_t^(2).
6: end for
```

- $p(\boldsymbol{v}_{t-1} \mid \boldsymbol{v}_{t-1} < \boldsymbol{0})$ is truncated multivariate normal, whose sampling is efficient by Gibbs sampling [Li and Ghosh, 2015]
- Conditioning $y_t^{(1)} := f(\boldsymbol{x}_t^{(1)}) + \epsilon$ aims penalization around $\boldsymbol{x}_t^{(1)}$
  - $y_t^{(1)}$ can be sampled from normal dist. $p(y_t^{(1)} \mid \boldsymbol{v}_{t-1} = \tilde{\boldsymbol{v}}_{t-1})$



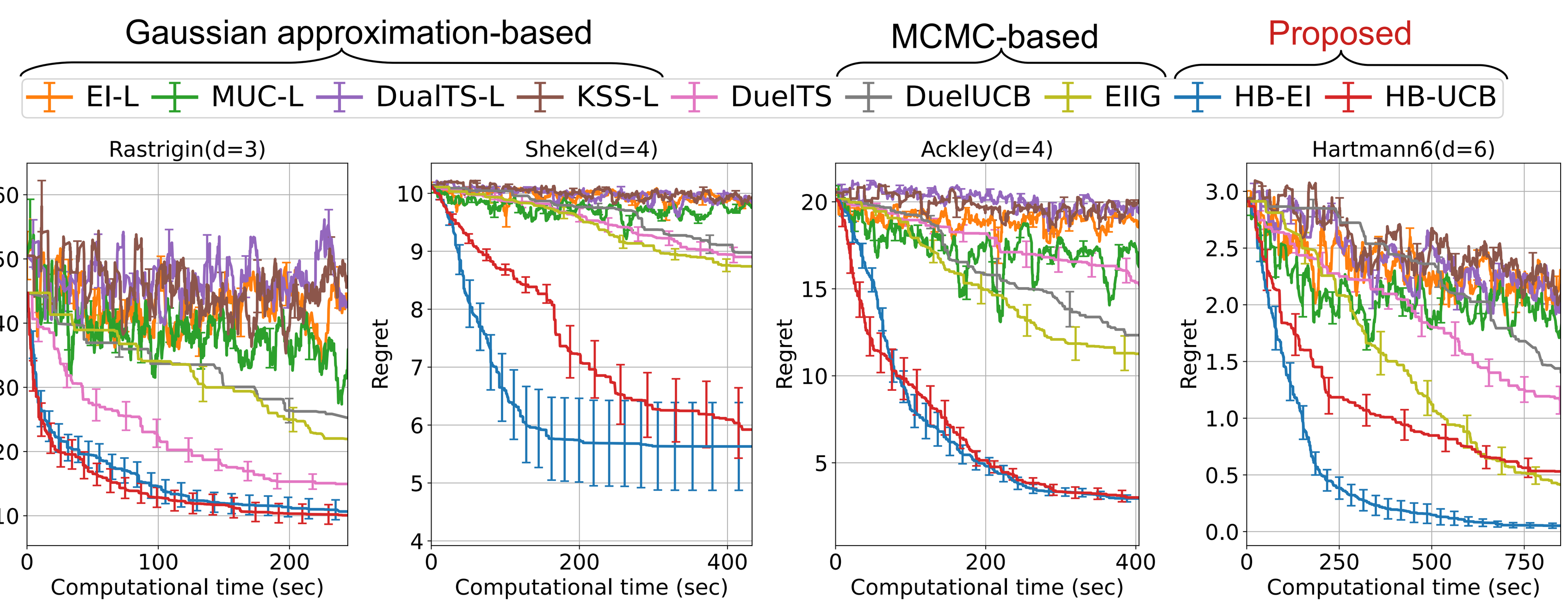

- Since $f \mid \tilde{\boldsymbol{v}}_{t-1}, \tilde{y}_t^{(1)}$ is GP, any AF $\alpha$ can be integrated.

## 5. Experiments

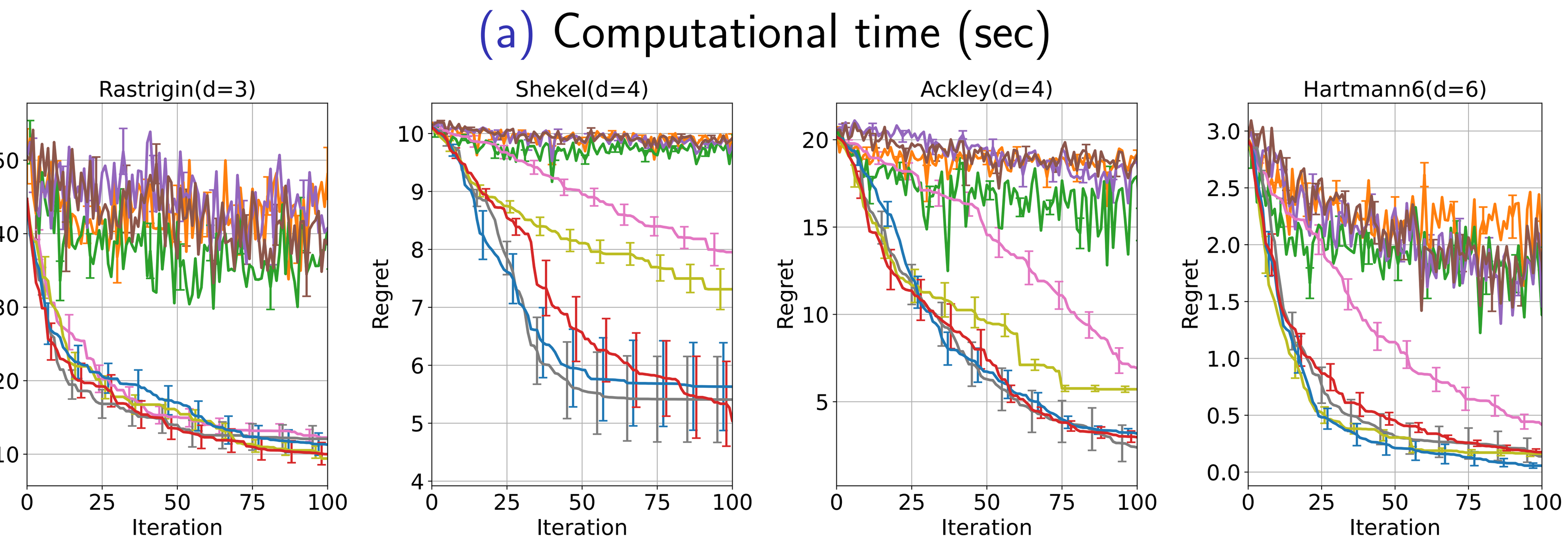
- We report average and standard error of regret:

$$r_t = f_* - f(\hat{\boldsymbol{x}}_t), \text{ where } \hat{\boldsymbol{x}}_t \text{ is a recommendation point.}$$

- **HB with EI (blue) and UCB (red)** were superior.




(a) Computational time (sec)




(b) Iteration