THE UNIVERSITY OF ARIZONA

Noise-Adaptive Confidence Sets for Linear Bandits

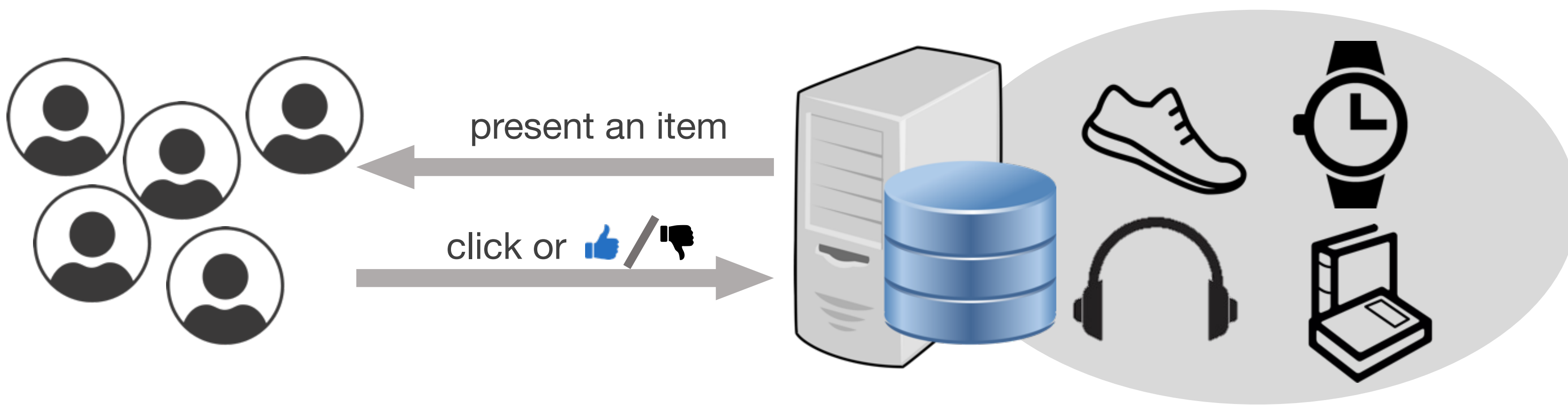
Kwang-Sung Jun (전광성)
Assistant Professor
Department of Computer Science

Joint work with **Jungtaek Kim** (김정택)
University of Pittsburgh

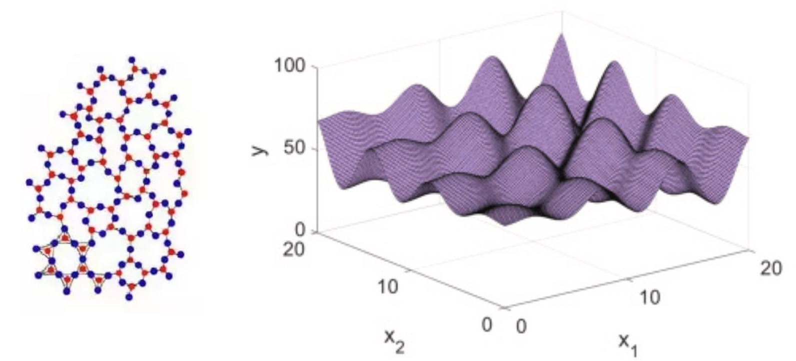
June 20, 2024

Motivating applications

Product recommendation



Materials discovery with Bayesian optimization



<u>Common challenge:</u> **Efficient exploration!**

The contextual bandit problem

For $t = 1, \ldots, T$

- (Optional) Observe a context $c_t \in \mathscr{C}$
- Take an action $a_t \in \mathscr{A}$
- Observe feedback (reward) y_t

	Product recommendation	**Bayesian optimization**
context	user information	N/A
action	item	point/experiment
feedback	click $\in \{0,1\}$	evaluation/measurement
Goal:	maximize $\sum_{t=1}^{T} y_t$	find $a \in \mathscr{A}$ with largest $\mathbb{E} y_t$

Assumption: $y_t = f_t^*(a_t) + \eta_t$ ← σ_*^2-sub-Gaussian noise (zero-mean)

$$f_t^*(a_t) = \langle \theta^*, \phi(a_t, c_t) \rangle$$

(can be extended to kernels)

θ^*: unknown parameter (d-dimensional)

ϕ: known feature map

Theoretical performance measure: Regret

$$\text{Regret}_T = \sum_{t=1}^{T} \max_a f_t^*(a) - f_t^*(a_t)$$

oracle's mean reward　　algorithm's mean reward

<u>Optimal worst-case regret</u>: $\sigma_* d\sqrt{T}$ (Dani et al., 2008)

$$\text{Average regret} = \frac{\text{Regret}_T}{T} \leq \frac{\sigma d}{\sqrt{T}}$$

convergence rate
to the oracle's performance!

For Bayesian optimization,

$$\text{exists } t \in \{1,\ldots,T\} \text{ s.t. } \max_a f^*(a) - f^*(a_t) \leq \frac{\sigma d}{\sqrt{T}}$$

convergence rate
to the maximum!

Key weakness of prior work

Weakness 1: Requires knowledge of σ_* (or its upper bound)

In practice, σ_*^2 is not known $\Rightarrow$ We need to guess it by σ_0^2.

Under-specification: $\sigma_0^2 \le \sigma_*^2 \quad \Rightarrow \quad$ regret $= \Theta(T)$

Over-specification: $\sigma_0^2 \ge \sigma_*^2 \quad \Rightarrow \quad$ regret $\le \sigma_0 d\sqrt{T}$

If $\sigma_* \ll \sigma_0$, then far from $\sigma_* d\sqrt{T}$!

Weakness 2: Assumes the noise level is the same throughout.

In practice, usually not true; i.e., $\sigma_1 \neq \sigma_2 \neq \cdots \neq \sigma_T$.

If $\max_{t=1}^{T} \sigma_t^2 \le \sigma_0^2$, then $\sigma_0 d\sqrt{T} = d\sqrt{\sum_{t=1}^{T} \sigma_0^2}$ $\Rightarrow$ can we attain $d\sqrt{\sum_{t=1}^{T} \sigma_t^2}$?

We made significant progress!

Jun and Kim, "Noise-Adaptive Confidence Sets for Linear Bandits and Application to Bayesian Optimization," ICML'24

Contribution 1: Sub-Gaussian noise

- Novel algorithm **LOSAN** (Linear Optimism with Semi-Adaptivity to Noise)
- σ_*: actual noise level.
- σ_0: specified noise level ($\sigma_0 \geq \sigma_*$).

	regret bound	when $\sigma_* = 0$
OFUL [Abbasi-Yadkori+11]	$\sigma_0\sqrt{d} \cdot \sqrt{dT}$	$\sigma_0\sqrt{d} \cdot \sqrt{dT}$
LOSAN (Ours)	$(\sigma_*\sqrt{d}+\sigma_0) \cdot \sqrt{dT}$	$\sigma_0 \cdot \sqrt{dT}$

if $d = 20$, then 4.5x faster convergence!

LOSAN is the first noise-adaptive algorithm for sub-Gaussian noise!

Contribution 2: Bounded noise

- Novel algorithm **LOFAV** (Linear Optimism with Full Adaptivity to Variance)
- $|\eta_t| \leq R$ for some known R; noise variance at time t is σ_t^2 (unknown)

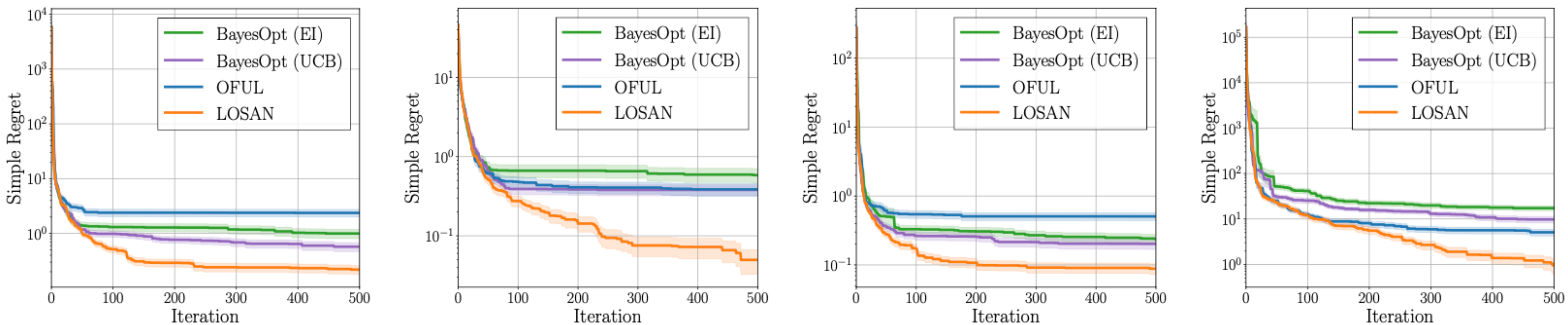
		no additional technical assumption*	uses all samples for learning	time complexity per round
OFUL [Abbasi-Yadkori+11]	$Rd\sqrt{T}$	✔	✔	d^2K
VOFUL [Zhang+21]	$d^{4.5}\sqrt{R^2+\sum_{t=1}^{T}\sigma_t^2}$	✔	✔	e^d
VOFUL2 [KimJ+22]	$d^{1.5}\sqrt{R^2+\sum_{t=1}^{T}\sigma_t^2}$	✔	✔	e^d
SAVE [Zhao+23]	$d\sqrt{R^2+\sum_{t=1}^{T}\sigma_t^2}$ (optimal)	✗	✗	$d^2K\log(T)$
LOFAV (Ours)	$d\sqrt{R^2+\sum_{t=1}^{T}\sigma_t^2}$ (optimal)	✔	✔	$d^2K\log(T)$ (K: number of actions)

LOFAV is the first practical variance-adaptive algorithm!

*i.e., assume that the noise cannot be a function of the chosen action

Numerical results: Sub-Gaussian noise

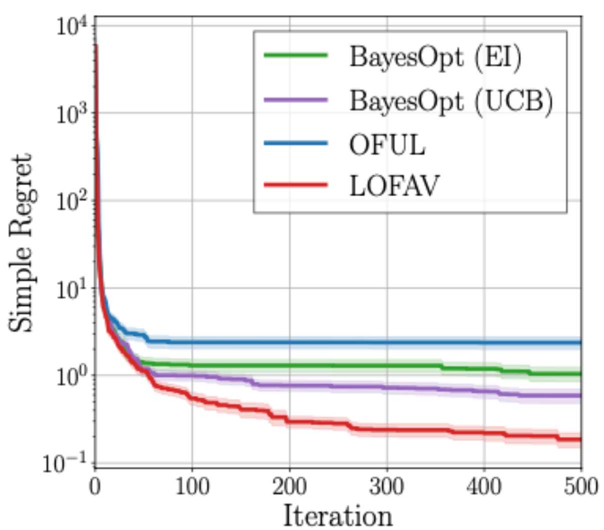
- Optimizing benchmark functions
- Over-specified setting: $\sigma_* = 0.01, \; \sigma_0 = 1$
- Linear model with random Fourier features (d=128) to mock Gaussian kernel.
- BayesOpt (EI/UCB): Bayesian optimization package BayesO



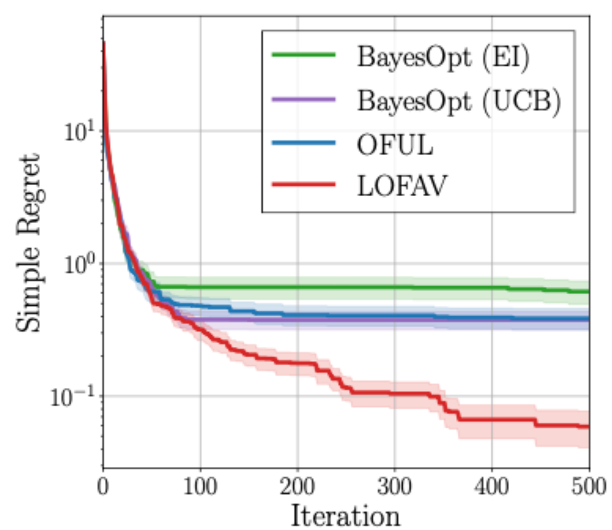
(a) Beale (b) Branin (c) Three-Hump Camel (d) Zakharov 4D

Numerical results: Bounded noise

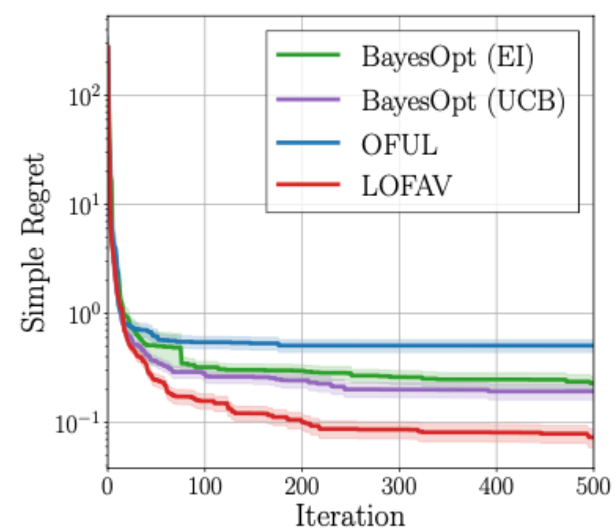
- Optimizing benchmark functions
- Noise bound: $R = 1,$ Noise variance: $\sigma_t^2 = (0.01)^2$
- Linear model with random Fourier features (d=128) to mock Gaussian kernel.
- BayesOpt (EI/UCB): Bayesian optimization package BayesO



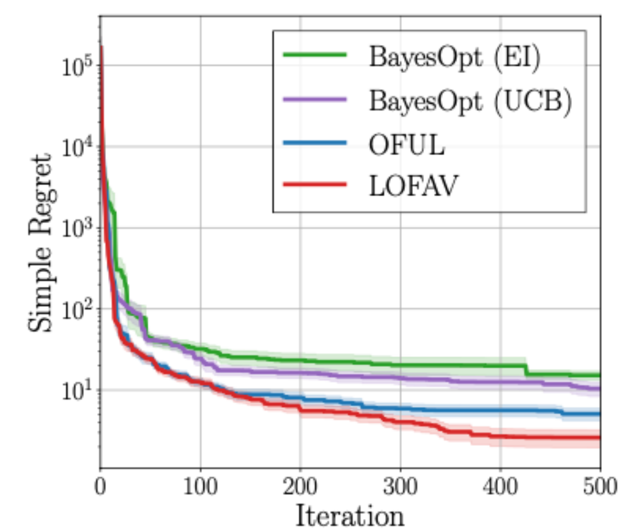
(a) Beale



(b) Branin



(c) Three-Hump Camel



(d) Zakharov 4D

Algorithm: LOSAN (Linear Optimism with Semi-Adaptivity to Noise)

- Optimistic strategy = use upper confidence bound (UCB) [Agrawal'95]
- At time t=1,…,T,
  - Choose action $a_t = \arg\max_{a \in \mathcal{A}} \mathrm{UCB}_t(a)$ — Must be correct with high probability

where $\mathrm{UCB}_t(a) = \langle \phi(a, c_t), \hat{\theta}_{t-1} \rangle + \sqrt{\beta_{t-1}} \|\phi(a, c_t)\|_{V_{t-1}^{-1}}$

- $\hat{\theta}_{t-1}$: Ridge regression
- $\sqrt{\beta_{t-1}}$: Noise factor ← We improved this!
- $\|\phi(a, c_t)\|_{V_{t-1}^{-1}}$: uncertainty of a

$$\|x\|_{A^{-1}} = \sqrt{x^\top A^{-1} x}$$

$$V_{t-1} = \lambda I + \sum_{s=1}^{t-1} \phi(a_s, c_s)\phi(a_s, c_s)^\top$$

Algorithm: LOSAN (Linear Optimism with Semi-Adaptivity to Noise)

- Define $x_t := \phi(a_t, c_t)$
- OFUL: $\beta_t \approx d\sigma_0^2$
- LOSAN: $\beta_t \approx \sigma_0^2 + \sum_{s=1}^{t-1} (x_s^\top \hat{\theta}_{s-1} - y_s)^2 \|x_s\|_{V_s^{-1}}^2$ (by advanced analysis in online learning theory)

If $\hat{\theta}_{s-1} \approx \theta^*$, then $\mathbb{E}[(x_s^\top \theta^* - y_s)^2] \leq \sigma_*^2$

$$\sum_{s=1}^{t-1} (x_s^\top \hat{\theta}_{s-1} - y_s)^2 \|x_s\|_{V_s^{-1}}^2 \lessapprox \sigma_*^2 \sum_{s=1}^{t-1} \|x_s\|_{V_s^{-1}}^2$$

$$\lessapprox \sigma_*^2 d$$ by elliptical potential lemma

$$\lessapprox \sigma_0^2 + d\sigma_*^2$$

- For te

Key technical ingredient for β_t:
"Regret equality" from online learning + martingale concentration

Proof of confidence set

$\hat{\theta}_t$: weighted estimator, $\Sigma_t := \lambda I + \sum_{s=1}^{t} w_s^2 x_s x_s^\top$, $f(\theta) := \frac{1}{2} w_s^2 (x_s^\top \theta - y_s)^2$

Step 1: "Regret equality" from FTRL (Follow The Regularized Leader)

$$\sum_{s=1}^{t} f_s(\hat{\theta}_{s-1}) - f_s(\theta^*) = \frac{\lambda}{2}\|\theta^*\|^2 + \sum_{s=1}^{t} f_s(\hat{\theta}_{s-1})\|w_s x_s\|^2_{\Sigma_s^{-1}} - \underline{\frac{1}{2}\|\hat{\theta}_t - \theta^*\|^2_{\Sigma_t}}$$

usually, throw it away except for [Dekel+10]

$$\iff \frac{1}{2}\|\hat{\theta}_t - \theta^*\|^2_{\Sigma_t} = \frac{\lambda}{2}\|\theta^*\|^2 + \sum_{s=1}^{t} f_s(\hat{\theta}_{s-1})\|w_s x_s\|^2_{\Sigma_s^{-1}} + \underline{\sum_{s=1}^{t} f_s(\theta^*) - f_s(\hat{\theta}_{s-1})}$$

negative (online learning) regret

$\leq \sigma_*^2 \ln(1/\delta)$ // with high probability

Step 2: Bound with known quantities $\leq S^2$ $\leq \sigma_0^2 \ln(1/\delta)$

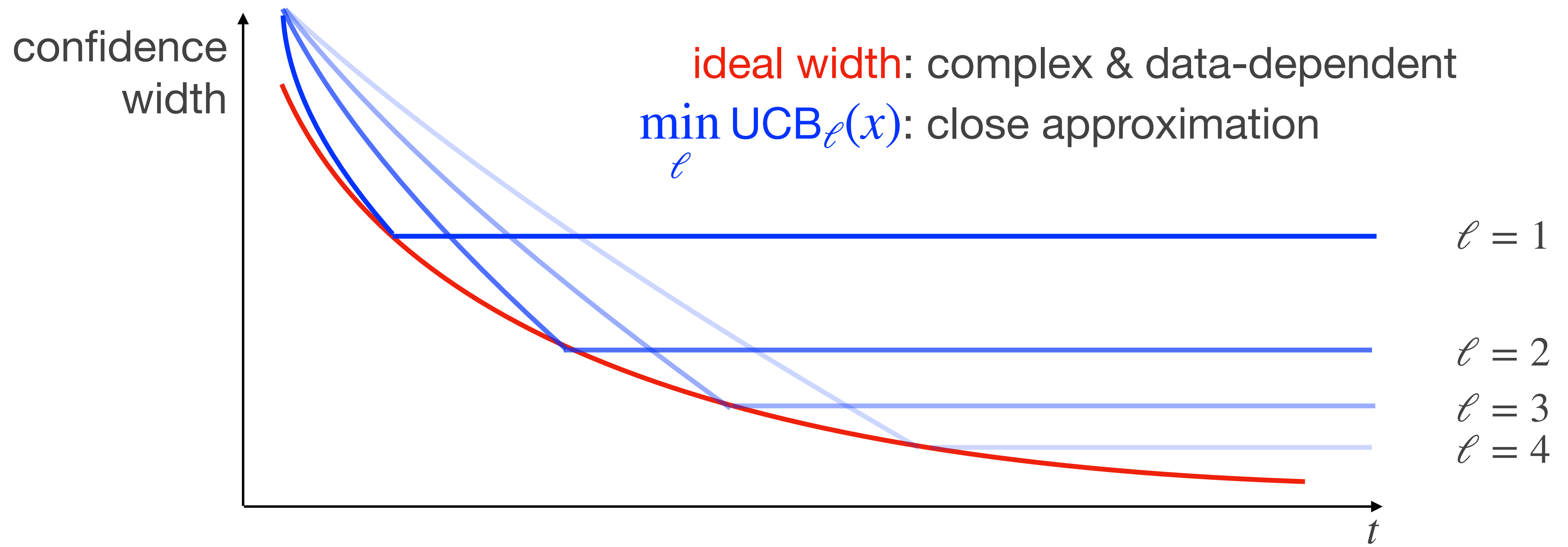
Algorithm: LOFAV (Linear Optimism with Full Adaptivity to Variance)

- Still optimism, but $L = \log_2(T)$ different UCBs

$$\mathrm{UCB}_t(a) = \min_{\ell=1}^{L} \mathrm{UCB}_{t,\ell}(a)$$

- $\mathrm{UCB}_{t,\ell}(a)$: based on weighted ridge regression

$$\hat{\theta}_{t,\ell} = \min_{\theta} \sum_{s=1}^{t} w_{s,\ell}^2 (x_s^\top \theta - y_s)^2 + \lambda_\ell \|\theta\|_2^2 \quad \text{where} \quad w_{s,\ell}^2 = \min\left\{1,\ \frac{2^{-2\ell}}{\|x_s\|_{V_{s-1}^{-1}}^2}\right\}$$



Q&A

Thank you!