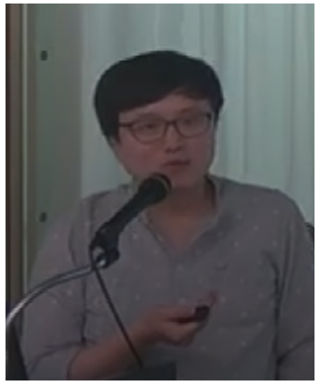
Noise-Adaptive Confidence Sets for Linear Bandits

Kwang-Sung Jun
Assistant Professor

THE UNIVERSITY OF ARIZONA

Joint work with

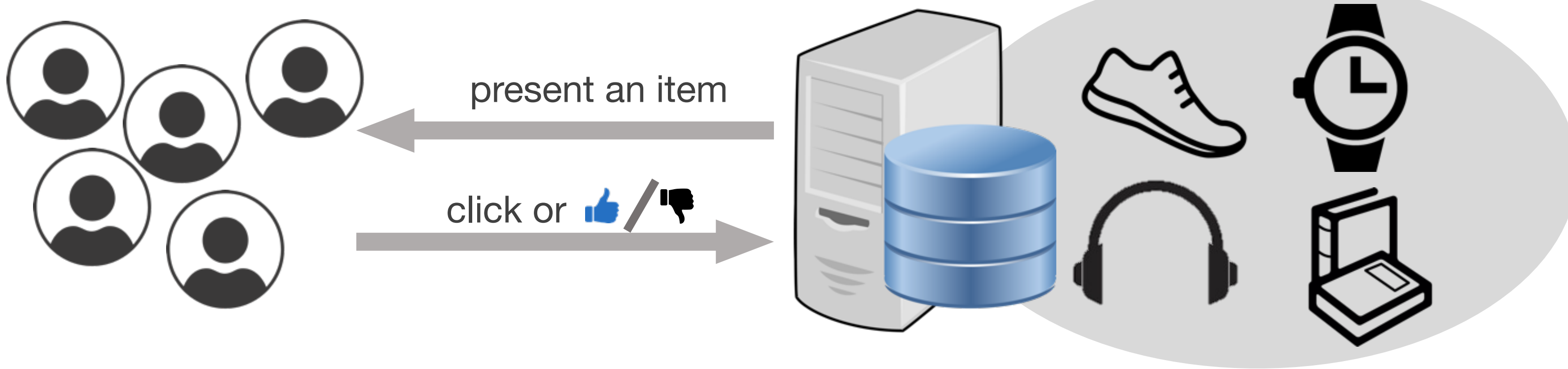
Jungtaek Kim
Postdoc
University of Pittsburgh



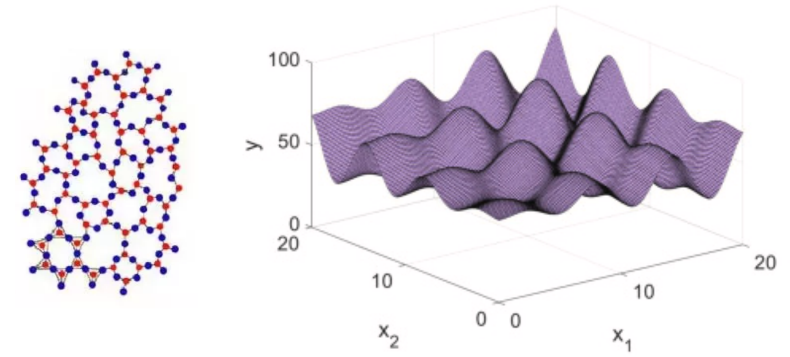
ICML'24

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)

unknown parameter (d-dimensional)

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

Optimal worst-case regret: $\sigma_* d\sqrt{T}$ (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: $\sigma_0^2 \le \sigma_*^2 \;\Rightarrow\;$ regret $= \Theta(T)$

Over-specification: $\sigma_0^2 \ge \sigma_*^2 \;\Rightarrow\;$ 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)
- $|\eta_t| \le R$ for some known R; noise variance at time t is σ_t^2 (unknown)

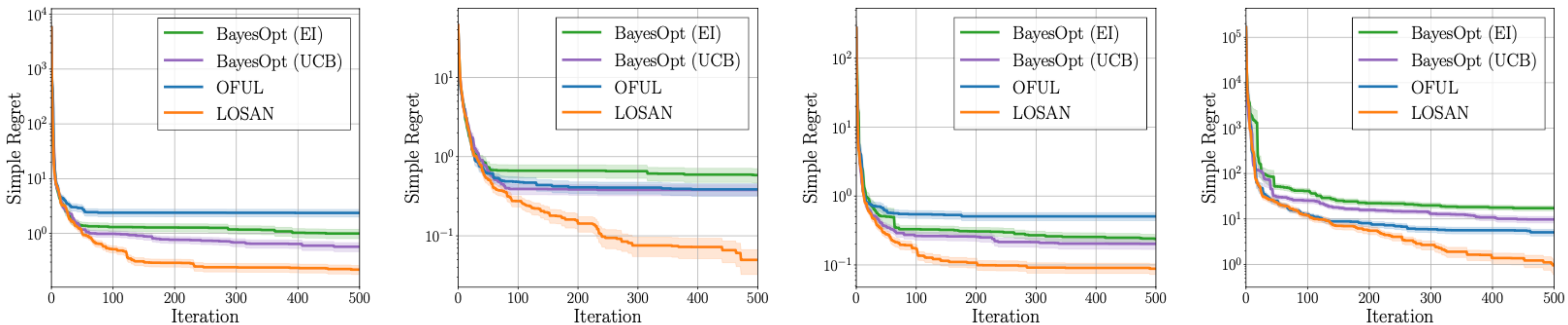
		no additional technical assumption*	uses all samples for learning	time complexity per round
OFUL [Abbasi-Yadkori+11]	$Rd\sqrt{T}$	✔	✔	d^2K
VOFUL [Zhang+21]	$d^{4.5}\sqrt{R^2+\sum_{t=1}^T \sigma_t^2}$	✔	✔	e^d
VOFUL2 [KimJ+22]	$d^{1.5}\sqrt{R^2+\sum_{t=1}^T \sigma_t^2}$	✔	✔	e^d
SAVE [Zhao+23]	$d\sqrt{R^2+\sum_{t=1}^T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

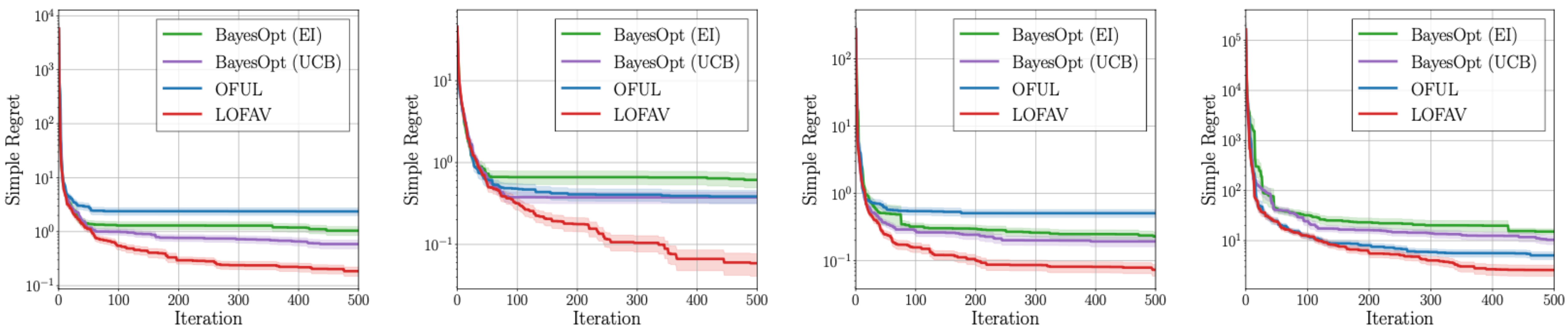
- Optimizing benchmark functions
- Over-specified setting: $\sigma_* = 0.01,\ \sigma_0 = 1$
- Linear model with random Fourier features (d=128) to mock Gaussian kernel.
- BayesOpt (EI/UCB): Bayesian optimization package BayesO



(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 $a_t = \arg\max_{a \in \mathcal{A}} \mathrm{UCB}_t(a)$ where $\mathrm{UCB}_t(a) = \max_{\theta \in \mathscr{C}_{t-1}} \langle \theta, \phi(a, c_t) \rangle$

confidence set: $\mathscr{C}_{t-1} = \left\{ \theta : \frac{1}{2} \|\theta - \underbrace{\hat{\theta}_{t-1}}_{\text{ridge regression}}\|^2_{V_{t-1}} \le \underbrace{\beta_{t-1}}_{\text{radius}} \right\}$

$$\|v\|_A := \sqrt{v^\top A v}$$

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

We improved this!

$\mathrm{UCB}_t(a)$ has a closed form expression!

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

- $x_s := \phi(a_s, c_s)$
- OFUL: $\beta_t \approx d\boldsymbol{\sigma_0^2}$
- LOSAN: $\beta_t \approx \sigma_0^2 + \sum_{s=1}^{t-1} \underline{(x_s^\top \hat{\theta}_{s-1} - y_s)^2} \|x_s\|^2_{V_s^{-1}}$

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

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

$$\lessapprox \sigma_*^2 d \qquad \text{by elliptical potential lemma}$$

$$\lessapprox \sigma_0^2 + d\boldsymbol{\sigma_*^2}$$

- For technical reasons, we turn to **weighted ridge regression** [Zhao+23]

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

Main result

Confidence set: $\mathscr{C}_t = \left\{ \theta : \frac{1}{2}\|\theta - \hat{\theta}_t\|^2_{\Sigma_t} \le \beta_t := \frac{\lambda}{2}S^2 + \sum_{s=1}^{t} \frac{1}{2}(x_s^\top \hat{\theta}_{s-1} - y_s)^2 \|x_s\|^2_{\Sigma_s^{-1}} + \sigma_0^2 \log(1/\delta) \right\}$

$\|\theta^*\| \le S$

$\Sigma_t := \lambda I + \sum_{s=1}^{t} w_s^2 x_s x_s^\top$

Theorem 1. $1 - \delta \le \mathbb{P}(\forall t, \theta^* \in \mathscr{C}_t)$

Theorem 2. $1 - \delta \le \mathbb{P}\left(\forall t, \sum_{s=1}^{t} \frac{1}{2}(x_s^\top \hat{\theta}_{s-1} - y_s)^2 \|x_s\|^2_{\Sigma_s^{-1}} = \tilde{O}(\lambda S^2 + d\sigma_*^2 + \sigma_0^2) \right)$

Theorem 3. LOSAN satisfies $\mathrm{Regret}_T = \tilde{O}\left((\sqrt{d}\sigma_* + \sigma_0)\sqrt{dT} \right)$ with high probability.

Proof of Theorem 1:
"Regret equality" from online learning + martingale concentration

Proof of confidence set

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

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

(proven next slide)

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

Proof of confidence set

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

$$\sum_{s=1}^{t} f_s(\theta^*) - f_s(\hat{\theta}_{s-1}) = \cdots = \sum_{s=1}^{t} r_s \cdot w_s \eta_s - \cancel{\frac{1}{2} \sum_{s=1}^{t} r_s^2}$$

where $r_s = w_s \cdot x_s^\top (\hat{\theta}_{s-1} - \theta^*)$

$$\leq \frac{1}{a} \ln(1/\delta) + \frac{a}{2} \sum_{s=1}^{t} r_s^2 \sigma_*^2$$

w.p. $\geq 1 - \delta$ by (i) Ville's inequality

(ii) $w_s \leq 1$

$$\leq \sigma_*^2 \ln(1/\delta) + \cancel{\frac{1}{2} \sum_{s=1}^{t} r_s^2}$$

by choosing $a = \frac{1}{\sigma_*^2}$

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

Confidence sets via online learning (OL)

	requires $-\frac{1}{2}\|\hat{\theta}_t - \theta^*\|^2_{\Sigma_t}$ in OL regret bound	requires OL regret bound for construction	requires running the OL algorithm
DGS style [DekelGS'10, ZhangYJXZ'16, **ours**]	Y	Y	Y
online-to-confidence-set conversion [Abbasi-YadkoriPS'12, **Jun**BWN'17]	**N**	Y	Y
regret-to-confidence set conversion (for the MLE) [RakhlinS'17, Orabona**J**'24, LeeY**J**'24]	**N**	Y	**N** (Can use computationally intractable OL algorithms)
sequential likelihood ratio [Robbins'72, EmmeneggerMK'23] (This also motivates our confidence set! (a blog article being prepared))	**N**	**N**	Y

(**N** is preferred)

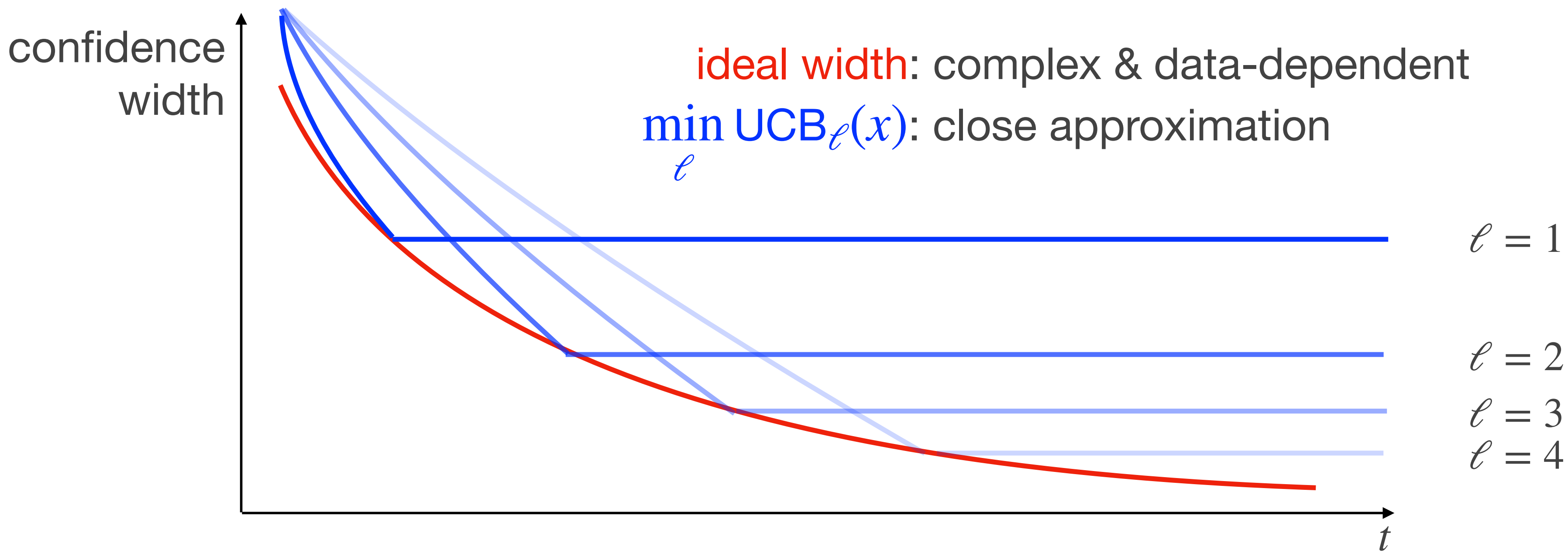
Algorithm: LOFAV (Linear Optimism with Full Adaptivity to Variance)

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

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

- $\mathrm{UCB}_{t,\ell}(a)$: by an ellipsoid centered at the weighted ridge regression

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



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

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

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

- $\mathrm{UCB}_{t,\ell}(a)$: by an ellipsoid centered at the weighted ridge regression

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

$$\beta_{t,\ell} = \tilde{O}\left(2^{-2\ell} S^2 + \sum_{s=1}^{t} f_{s,\ell}(\hat{\theta}_{s-1,\ell}) \|x_s\|^2_{\Sigma^{-1}_{s,\ell}} + 2^{-\ell} \sqrt{\beta_{t-1,\ell}\left(\sum_{s=1}^{t} f_{s,\ell}(\hat{\theta}_{s-1,\ell}) + R^2\right)} + 2^{-\ell} R \sqrt{\beta_{t-1,\ell}}\right)$$

Theorem 4. $1 - \delta \le \mathbb{P}(\forall t, \forall \ell, \theta^* \in \mathscr{C}_{t,\ell})$

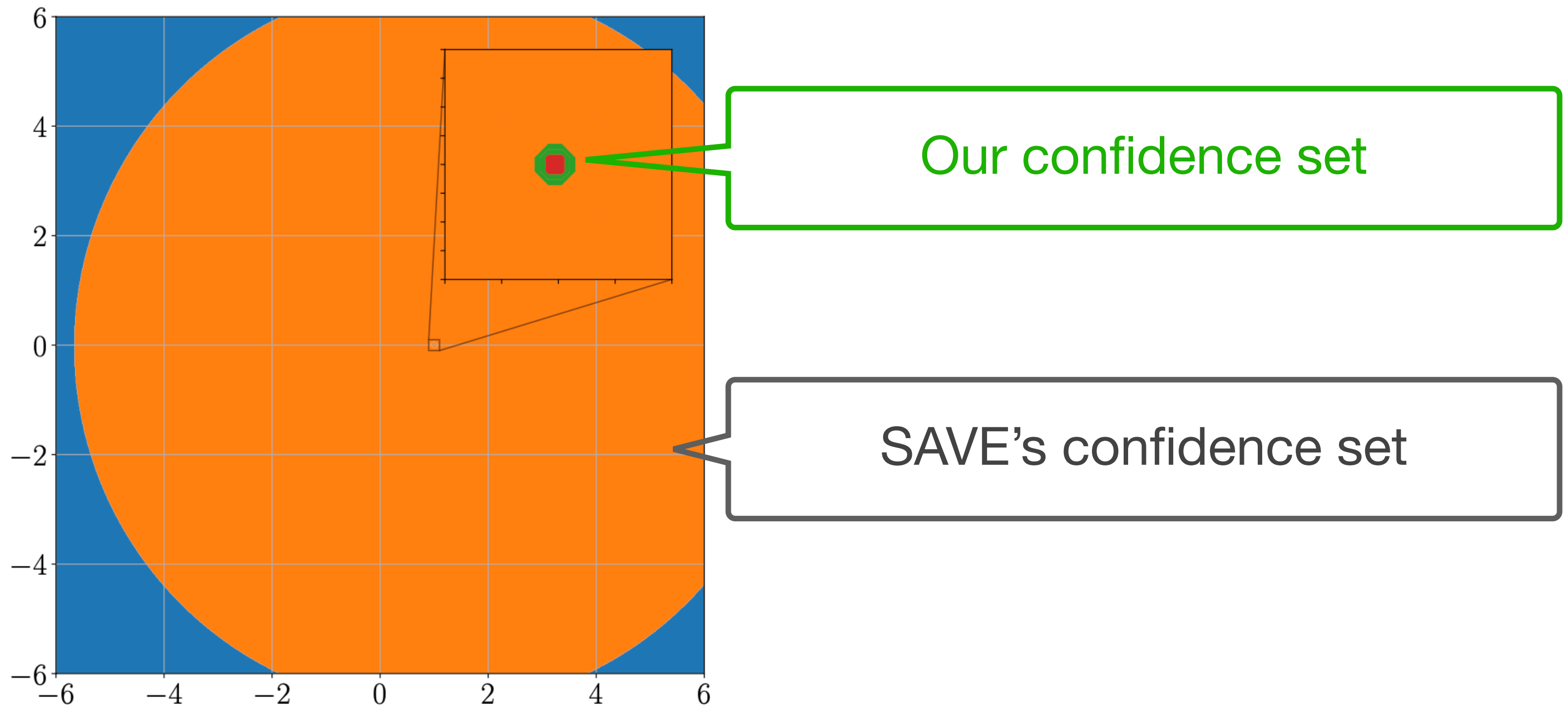
Theorem 5. $1 - \delta \le \mathbb{P}\left(\forall t, \forall \ell, \beta_{t,\ell} = \tilde{O}\left(2^{-2\ell}\left(R^2 + \sum_{s=1}^{t} \sigma_s^2\right)\right)\right)$

Comparison with SAVE [Zhao+23]

Improvement 1: Avoids sample splitting

SAVE is based on SupLinRel [Auer'02] ⇒ Sample splitting kills the performance.

Improvement 2: Tightened confidence set (via regret equality based analysis)



LOFAV regret bound

Theorem 6. LOFAV satisfies $\mathrm{Regret}_T = \tilde{O}\left(d\sqrt{R^2 + \sum_{t=1}^T \sigma_t^2}\right)$ with high probability.

Proof.
$$\mathrm{Regret}_T = \sum_{t=1}^{T} \mathrm{reg}_t \qquad (\beta_T^* = R^2 + \sum_{s=1}^{t} \sigma_s^2)$$

Peeling-based regret analysis [He'21, KimYJ'22]

$$\le \sum_{t=1}^{T}\sum_{\ell=1}^{L} \mathrm{reg}_t \cdot I\left\{2^{-2\ell}\sqrt{\beta_T^*} \le \mathrm{reg}_t \le 2^{-2(\ell-1)}\sqrt{\beta_T^*}\right\} + T \cdot 2^{-L} \qquad (\text{say } \mathrm{reg}_t \le 1)$$

$$\le \sum_{\ell=1}^{L} \sqrt{\beta_T^*} 2^{-2(\ell-1)} \sum_{t=1}^{T} \underline{I\{2^{-2\ell}\sqrt{\beta_T^*} \le \mathrm{reg}_t\}} + T \cdot 2^{-L}$$

$$\Longrightarrow \cdots \Longrightarrow \|w_{t,\ell} x_t\|_{\Sigma_{t-1,\ell}^{-1}} \ge 2^{-\ell}$$

bound $\sum_{t=1}^{T} I\{\|w_{t,\ell} x_t\|_{\Sigma_{t-1,\ell}^{-1}} \ge 2^{-\ell}\} \le O(2^{2\ell} d)$

by elliptical potential 'count' lemma [LattimoreS'20, GalesSJ'22]

Conclusion

- Sub-Gaussian noise: (semi-)adaptivity to the noise level
- Bounded noise: adaptivity to unknown time-varying variance

- Proofs
  - From regret equality to confidence set
  - Peeling based regret analysis + elliptical potential “count” lemma

- Future work
  - How far can we push noise adaptivity to general function class? (second order bound)
  - Algorithms beyond optimism: expected improvement, information directed sampling, etc.