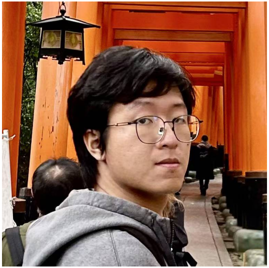
Statistical inference under adaptive sampling with LinUCB

Yuting Wei

Statistics & Data Science, Wharton
University of Pennsylvania

IMSI, Apr 22



Wei Fan, UPenn

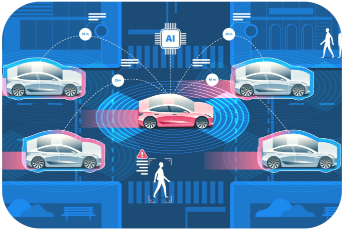
Kevin Tan, UPenn

"*Statistical Inference under Adaptive Sampling with LinUCB,*"

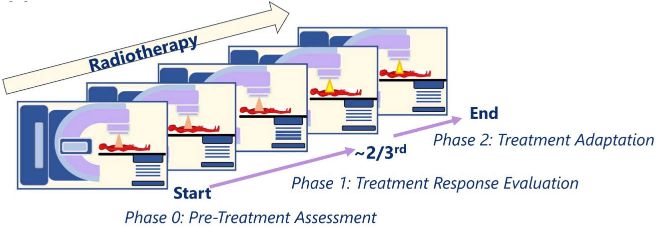
Wei Fan*, Kevin Tan*, Yuting Wei

Adaptively collected data is ubiquitous

Large language models



Autonomous driving cars



Cancer treatment

Regret minimization vs. Statistical inference

Standard Goal for RL: regret minimization

$$\mathrm{Reg}(T) = \sum_{t=1}^{T} \mathsf{Reward}_t(\text{optimal action}) - \sum_{t=1}^{T} \mathsf{Reward}_t(\text{action at } t)$$

Regret minimization vs. Statistical inference

Standard Goal for RL: regret minimization

$$\mathrm{Reg}(T) = \sum_{t=1}^{T} \text{Reward}_t(\text{optimal action}) - \sum_{t=1}^{T} \text{Reward}_t(\text{action at } t)$$

Regret minimization vs. Statistical inference

Alternative Goal: statistical inference/uncertainty quantification

The Washington Post
Democracy Dies in Darkness

Health Health Care Medical Mysteries Science Well+Being

Health & Science

20 percent of patients with serious conditions are first misdiagnosed, study says

April 4, 2017

New York (CNN Business) — It's happened again: A Tesla with its Autopilot feature engaged was involved in a fatal crash.

The crash, which happened March 1 in Florida, left the 50-year-old Tesla driver dead. Authorities revealed new details about the incident on Thursday. The driver turned on Autopilot about 10 seconds before the Model 3 sedan collided with a semi truck, according to the National Transportation Safety Board. The Tesla driver's hands were not detected on the steering wheel for less than eight seconds prior. The truck driver was uninjured.

Regret minimization vs. Statistical inference

Alternative Goal: statistical inference/uncertainty quantification

Challenges:

- data collection method itself introduces complex dependencies
- classical i.i.d. asymptotics fail
- require strong assumptions

Towards a solid inference framework for online collected data...

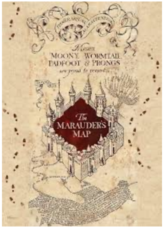
Lai & Wei'82
Abbasi-yadkori, Pal, Szepesvari'11
Deshpande, Mackey, Syrgkanis, Taddy'18
Zhang, Janson, Murphy'20 &'21
Kalvit & Zeevi'21
Deshpande, Javanmard, Mehrabi'23
Khamaru, & Zhang'24
Han, Khamaru, Zhang'24
Srikant'24
Lin, Khamaru, Wainwright'25
Wu, Wei, Rinaldo'25
...

Linear (contextual) bandits

Observe a context x_t

Pick action a_t based on x_t and history

Learner



Environment returns r_t to learner

Environment

At round t:

- Learner is given a context $\boldsymbol{x}_t$, and action set $\mathcal{A}$
- Learner takes action $\boldsymbol{a}_t \in \mathcal{A}$ based on $\boldsymbol{x}_t$ and history $(\boldsymbol{a}_1, r_1, \ldots, \boldsymbol{a}_{t-1}, r_{t-1})$
- Environment returns immediate reward: $r_t = \langle \boldsymbol{\phi}(\boldsymbol{x}_t, \boldsymbol{a}_t), \boldsymbol{\theta}^\star \rangle + \boldsymbol{\epsilon}_t$

Regret minimization vs. Statistical inference

Regret minimization

$$\mathrm{Reg}(T) = \sum_{t=1}^{T} \max_{\boldsymbol{a}_t^\star \in \mathcal{A}} \langle \boldsymbol{\phi}(\boldsymbol{x}_t, \boldsymbol{a}_t^\star), \boldsymbol{\theta}^\star \rangle - \sum_{t=1}^{T} \langle \boldsymbol{\phi}(\boldsymbol{x}_t, \boldsymbol{a}_t), \boldsymbol{\theta}^\star \rangle$$

— *inspired the popular* $\underbrace{\textit{Upper Confidence Bound (UCB) algorithm}}_{\textit{LinUCB for linear bandits}}$

Can UCB algorithm be used not only as a learning algorithm, but also as a vehicle for valid statistical inference?

LinUCB algorithm for linear bandits

LinUCB algorithm

— *Li et al.'10, Abbasi-yadkori et al.'11*

- Compute the UCB score

$$\mathrm{UCB}_t(\boldsymbol{a}) = \underbrace{\langle \boldsymbol{\phi}(\boldsymbol{x}_t, \boldsymbol{a}), \widehat{\boldsymbol{\theta}}_{t-1} \rangle}_{\text{Exploitation}} + \underbrace{\beta \sqrt{\boldsymbol{\phi}(\boldsymbol{x}_t, \boldsymbol{a})^\top \boldsymbol{\Lambda}_{t-1}^{-1} \boldsymbol{\phi}(\boldsymbol{x}_t, \boldsymbol{a})}}_{\text{Exploration bonus}}$$

$$\boldsymbol{\Lambda}_{t-1} = \boldsymbol{I}_d + \sum_{s=1}^{t-1} \boldsymbol{\phi}(\boldsymbol{x}_s, \boldsymbol{a}_s) \boldsymbol{\phi}(\boldsymbol{x}_s, \boldsymbol{a}_s)^\top$$

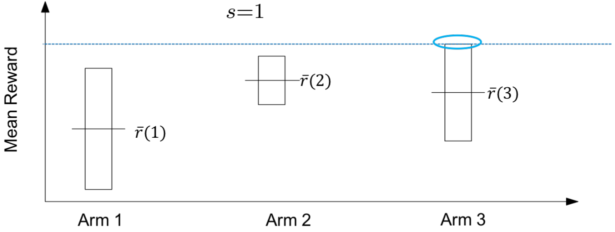
LinUCB algorithm

— *Li et al.'10, Abbasi-yadkori et al.'11*

- Compute the UCB score

$$\mathrm{UCB}_t(\boldsymbol{a}) = \underbrace{\langle \boldsymbol{\phi}(\boldsymbol{x}_t, \boldsymbol{a}), \widehat{\boldsymbol{\theta}}_{t-1} \rangle}_{\text{Exploitation}} + \underbrace{\beta \sqrt{\boldsymbol{\phi}(\boldsymbol{x}_t, \boldsymbol{a})^\top \boldsymbol{\Lambda}_{t-1}^{-1} \boldsymbol{\phi}(\boldsymbol{x}_t, \boldsymbol{a})}}_{\text{Exploration bonus}}$$

$$\boldsymbol{\Lambda}_{t-1} = \boldsymbol{I}_d + \sum_{s=1}^{t-1} \boldsymbol{\phi}(\boldsymbol{x}_s, \boldsymbol{a}_s) \boldsymbol{\phi}(\boldsymbol{x}_s, \boldsymbol{a}_s)^\top$$



LinUCB algorithm

— *Li et al.'10, Abbasi-yadkori et al.'11*

- Compute the UCB score

$$\mathrm{UCB}_t(\boldsymbol{a}) = \underbrace{\langle \boldsymbol{\phi}(\boldsymbol{x}_t, \boldsymbol{a}), \widehat{\boldsymbol{\theta}}_{t-1} \rangle}_{\text{Exploitation}} + \underbrace{\beta \sqrt{\boldsymbol{\phi}(\boldsymbol{x}_t, \boldsymbol{a})^\top \boldsymbol{\Lambda}_{t-1}^{-1} \boldsymbol{\phi}(\boldsymbol{x}_t, \boldsymbol{a})}}_{\text{Exploration bonus}}$$

$$\boldsymbol{\Lambda}_{t-1} = \boldsymbol{I}_d + \sum_{s=1}^{t-1} \boldsymbol{\phi}(\boldsymbol{x}_s, \boldsymbol{a}_s) \boldsymbol{\phi}(\boldsymbol{x}_s, \boldsymbol{a}_s)^\top$$

- Select action to maximize the UCB score

$$\boldsymbol{a}_t = \arg\max_{\boldsymbol{a} \in \mathcal{A}} \mathrm{UCB}_t(\boldsymbol{a})$$

LinUCB algorithm

— *Li et al.'10, Abbasi-yadkori et al.'11*

- Compute the UCB score

$$\mathrm{UCB}_t(\boldsymbol{a}) = \underbrace{\langle \boldsymbol{\phi}(\boldsymbol{x}_t, \boldsymbol{a}), \widehat{\boldsymbol{\theta}}_{t-1} \rangle}_{\text{Exploitation}} + \underbrace{\beta \sqrt{\boldsymbol{\phi}(\boldsymbol{x}_t, \boldsymbol{a})^\top \boldsymbol{\Lambda}_{t-1}^{-1} \boldsymbol{\phi}(\boldsymbol{x}_t, \boldsymbol{a})}}_{\text{Exploration bonus}}$$

$$\boldsymbol{\Lambda}_{t-1} = \boldsymbol{I}_d + \sum_{s=1}^{t-1} \boldsymbol{\phi}(\boldsymbol{x}_s, \boldsymbol{a}_s) \boldsymbol{\phi}(\boldsymbol{x}_s, \boldsymbol{a}_s)^\top$$

- Select action to maximize the UCB score

$$\boldsymbol{a}_t = \arg\max_{\boldsymbol{a} \in \mathcal{A}} \mathrm{UCB}_t(\boldsymbol{a})$$

- Estimator for $\boldsymbol{\theta}^\star$: $\widehat{\boldsymbol{\theta}}_T = \boldsymbol{\Lambda}_T^{-1} \sum_{s=1}^T \boldsymbol{\phi}(\boldsymbol{x}_s, \boldsymbol{a}_s)\, r_s$

LinUCB algorithm

— *Li et al.'10, Abbasi-yadkori et al.'11*

- Compute the UCB score

$$\mathrm{UCB}_t(\boldsymbol{a}) = \underbrace{\langle \boldsymbol{\phi}(\boldsymbol{x}_t, \boldsymbol{a}), \widehat{\boldsymbol{\theta}}_{t-1} \rangle}_{\text{Exploitation}} + \underbrace{\beta \sqrt{\boldsymbol{\phi}(\boldsymbol{x}_t, \boldsymbol{a})^\top \boldsymbol{\Lambda}_{t-1}^{-1} \boldsymbol{\phi}(\boldsymbol{x}_t, \boldsymbol{a})}}_{\text{Exploration bonus}}$$

$$\boldsymbol{\Lambda}_{t-1} = \boldsymbol{I}_d + \sum_{s=1}^{t-1} \boldsymbol{\phi}(\boldsymbol{x}_s, \boldsymbol{a}_s)\boldsymbol{\phi}(\boldsymbol{x}_s, \boldsymbol{a}_s)^\top$$

- Select action to maximize the UCB score

$$\boldsymbol{a}_t = \arg\max_{\boldsymbol{a} \in \mathcal{A}} \mathrm{UCB}_t(\boldsymbol{a})$$

- Estimator for $\boldsymbol{\theta}^\star$: $\widehat{\boldsymbol{\theta}}_T = \boldsymbol{\Lambda}_T^{-1} \sum_{s=1}^{T} \boldsymbol{\phi}(\boldsymbol{x}_s, \boldsymbol{a}_s)\, r_s$

This strategy yields $\widetilde{O}(d\sqrt{T})$ *regret bound (minimax optimal).*

Does $\widehat{\theta}_T$ admit an asymptotic normal distribution?

Stability of design covariance

Stability (Lai and Wei'82)

There exists a sequence of deterministic matrices $\{\boldsymbol{\Sigma}_T\}$ st.

$$\boldsymbol{\Sigma}_T^{-1}\boldsymbol{\Lambda}_T \xrightarrow{\mathsf{p}} \boldsymbol{I}_d$$

- Asymptotic normality holds with stability:

$$\boldsymbol{\Lambda}_T^{1/2}(\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}^\star) \xrightarrow{\mathsf{d}} \mathcal{N}(0, \sigma^2 \boldsymbol{I}_d)$$

- Asymptotic normality fails in the absence of stability

Stability of design covariance

Stability (Lai and Wei'82)

There exists a sequence of deterministic matrices $\{\boldsymbol{\Sigma}_T\}$ st.

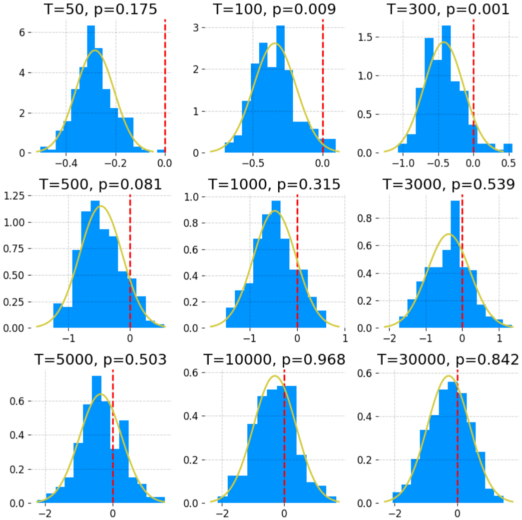
$$\boldsymbol{\Sigma}_T^{-1}\boldsymbol{\Lambda}_T \xrightarrow{\mathsf{p}} \boldsymbol{I}_d$$

- Asymptotic normality holds with stability:

$$\boldsymbol{\Lambda}_T^{1/2}(\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}^\star) \xrightarrow{\mathsf{d}} \mathcal{N}(0, \sigma^2 \boldsymbol{I}_d)$$

- Asymptotic normality fails in the absence of stability

— *stability established for multi-armed bandits*



Asymptotic normality of the LinUCB algorithm. For some random vector u on the unit ball, we plot $\widehat{\sigma}^{-1}\big(\frac{2\beta^2 T}{d+1}\big)^{1/4} u^\top(\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}^\star)$ over 1000 independent trials.

Assumptions

- **Action set:** set of feature maps spans $\mathcal{B}^d$
- **True signal:** $\boldsymbol{\theta}^\star$ lies on the unit sphere, i.e., $\|\boldsymbol{\theta}^\star\|_2 = 1$
- **Stochastic errors:** noise $(\epsilon_t)_{t=1}^T$ is an $(\mathcal{F}_t)$-adapted martingale difference sequence, conditionally mean zero and sub-Gaussian σ

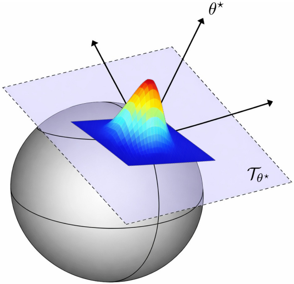
Estimator: $\widehat{\boldsymbol{\theta}}_T = \mathcal{P}(\widehat{\boldsymbol{\theta}}_T)$ projection of the ridge estimator to $\mathcal{B}^d$

Asymptotic normality of LinUCB

Theorem (Fan*, Tan*, Wei'25)

Let $\boldsymbol{U} \in \mathbb{R}^{d\times(d-1)}$ *chosen such that* $\boldsymbol{Q} = (\boldsymbol{\theta}^\star, \boldsymbol{U})$ *is orthogonal matrix* ($\boldsymbol{Q}^\top \boldsymbol{Q} = I$), *then*

$$\left(\frac{2\beta^2 T}{d+1}\right)^{1/4} \boldsymbol{U}^\top(\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}^\star) \to \mathcal{N}(0, \boldsymbol{I}_{d-1}).$$



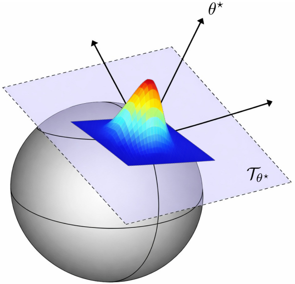
Asymptotic normality of LinUCB

Theorem (Fan*, Tan*, Wei'25)

Let $\boldsymbol{U} \in \mathbb{R}^{d\times(d-1)}$ chosen such that $\boldsymbol{Q} = (\boldsymbol{\theta}^\star, \boldsymbol{U})$ is orthogonal matrix ($\boldsymbol{Q}^\top \boldsymbol{Q} = I$), then

$$\left(\frac{2\beta^2 T}{d+1}\right)^{1/4} \boldsymbol{U}^\top(\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}^\star) \to \mathcal{N}(0, \boldsymbol{I}_{d-1}).$$

- $\left\|\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}^\star\right\|_2 = \widetilde{\Theta}_p\left(T^{-1/4}\right)$



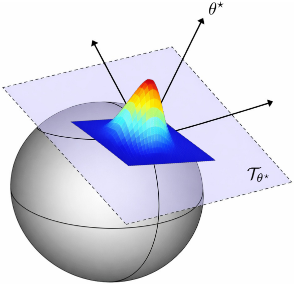
Asymptotic normality of LinUCB

Theorem (Fan*, Tan*, Wei'25)

Let $\boldsymbol{U} \in \mathbb{R}^{d\times(d-1)}$ chosen such that $\boldsymbol{Q} = (\boldsymbol{\theta}^\star, \boldsymbol{U})$ is orthogonal matrix ($\boldsymbol{Q}^\top \boldsymbol{Q} = I$), then

$$\left(\frac{2\beta^2 T}{d+1}\right)^{1/4} \boldsymbol{U}^\top(\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}^\star) \to \mathcal{N}(0, \boldsymbol{I}_{d-1}).$$

- $\left\|\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}^\star\right\|_2 = \widetilde{\Theta}_p\left(T^{-1/4}\right)$
- $\boldsymbol{\Lambda}_t = \boldsymbol{I}_d + \sum_{s=1}^t \boldsymbol{\phi}(\boldsymbol{x}_s, \boldsymbol{a}_s)\boldsymbol{\phi}(\boldsymbol{x}_s, \boldsymbol{a}_s)^\top$



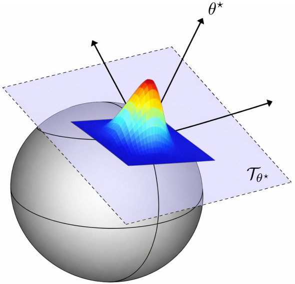
Asymptotic normality of LinUCB

Theorem (Fan*, Tan*, Wei'25)

Let $\boldsymbol{U} \in \mathbb{R}^{d\times(d-1)}$ *chosen such that* $\boldsymbol{Q} = (\boldsymbol{\theta}^\star, \boldsymbol{U})$ *is orthogonal matrix* ($\boldsymbol{Q}^\top \boldsymbol{Q} = I$)*, then*

$$\left(\frac{2\beta^2 T}{d+1}\right)^{1/4} \boldsymbol{U}^\top (\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}^\star) \to \mathcal{N}(0, \boldsymbol{I}_{d-1}).$$

- $\left\|\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}^\star\right\|_2 = \widetilde{\Theta}_p\left(T^{-1/4}\right)$
- $\boldsymbol{\Lambda}_t = \boldsymbol{I}_d + \sum_{s=1}^t \boldsymbol{\phi}(\boldsymbol{x}_s, \boldsymbol{a}_s)\boldsymbol{\phi}(\boldsymbol{x}_s, \boldsymbol{a}_s)^\top$
  - $\lambda_{\mathsf{min}}(\boldsymbol{\Lambda}_T)$ grows like $\Theta(\sqrt{T})$ vs. $\Theta(T)$ in i.i.d. setting



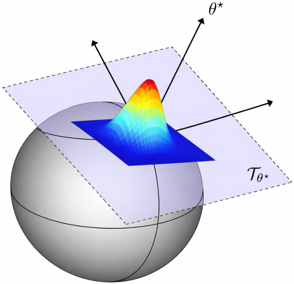
Asymptotic normality of LinUCB

Theorem (Fan*, Tan*, Wei'25)

Let $\boldsymbol{U} \in \mathbb{R}^{d\times(d-1)}$ chosen such that $\boldsymbol{Q} = (\boldsymbol{\theta}^\star, \boldsymbol{U})$ is orthogonal matrix ($\boldsymbol{Q}^\top \boldsymbol{Q} = I$), then

$$\left(\frac{2\beta^2 T}{d+1}\right)^{1/4} \boldsymbol{U}^\top(\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}^\star) \to \mathcal{N}(0, \boldsymbol{I}_{d-1}).$$

- $\left\|\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}^\star\right\|_2 = \widetilde{\Theta}_p\left(T^{-1/4}\right)$
- $\boldsymbol{\Lambda}_t = \boldsymbol{I}_d + \sum_{s=1}^t \boldsymbol{\phi}(\boldsymbol{x}_s, \boldsymbol{a}_s)\boldsymbol{\phi}(\boldsymbol{x}_s, \boldsymbol{a}_s)^\top$
  - $\lambda_{\min}(\boldsymbol{\Lambda}_T)$ grows like $\Theta(\sqrt{T})$ vs. $\Theta(T)$ in i.i.d. setting
  - regret minimizing nature of LinUCB



Confidence set with LinUCB

Corollary (Fan*, Tan*, Wei'25)

For $\beta \gg d^2\sigma\sqrt{d + \log\log T}$, the asymptotic $1-\delta$ confidence set of $\boldsymbol{\theta}^\star$ is given as

$$\mathcal{C}_\delta = \left\{ \boldsymbol{\theta} \in \mathcal{S}^{d-1} : \left\| \widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta} \right\|_2^2 \le \sigma^2 \sqrt{\frac{d+1}{2\beta^2 T}} \cdot \chi^2_{d-1,1-\delta} \right\}$$

Confidence set with LinUCB

Corollary (Fan*, Tan*, Wei'25)

For $\beta \gg d^2\sigma\sqrt{d + \log\log T}$, the asymptotic $1-\delta$ confidence set of $\boldsymbol{\theta}^\star$ is given as

$$\mathcal{C}_\delta = \left\{ \boldsymbol{\theta} \in \mathcal{S}^{d-1} : \left\| \widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta} \right\|_2^2 \le \sigma^2 \sqrt{\frac{d+1}{2\beta^2 T}} \cdot \chi^2_{d-1,1-\delta} \right\}$$

- $\boldsymbol{U}$ does not appear because $\|\boldsymbol{U}^\top(\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}^\star)\|_2 \approx \|\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}^\star\|_2$

Confidence set with LinUCB

Corollary (Fan*, Tan*, Wei'25)

For $\beta \gg d^2\sigma\sqrt{d+\log\log T}$, the asymptotic $1-\delta$ confidence set of $\boldsymbol{\theta}^\star$ is given as

$$\mathcal{C}_\delta = \left\{ \boldsymbol{\theta} \in \mathcal{S}^{d-1} : \left\| \widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta} \right\|_2^2 \le \sigma^2 \sqrt{\frac{d+1}{2\beta^2 T}} \cdot \chi^2_{d-1,1-\delta} \right\}$$

- $\boldsymbol{U}$ does not appear because $\|\boldsymbol{U}^\top(\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}^\star)\|_2 \approx \|\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}^\star\|_2$
- Abbasi-yadkori et al.'11: $\boldsymbol{\theta}^\star$ with prob. $1-\delta$ lies inside

$$C_t = \left\{ \boldsymbol{\theta} \in \mathbb{R}^d \;\middle|\; \|\widehat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}\|_{\boldsymbol{\Lambda}_t} \le \sigma \sqrt{d \log\left(\frac{1+TL^2/\lambda}{\delta}\right)} + \lambda^{1/2} \right\}$$

Confidence set with LinUCB

Corollary (Fan*, Tan*, Wei'25)

For $\beta \gg d^2\sigma\sqrt{d + \log\log T}$, the asymptotic $1-\delta$ confidence set of $\boldsymbol{\theta}^\star$ is given as

$$\mathcal{C}_\delta = \left\{ \boldsymbol{\theta} \in \mathcal{S}^{d-1} : \left\| \widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta} \right\|_2^2 \leq \sigma^2 \sqrt{\frac{d+1}{2\beta^2 T}} \cdot \chi^2_{d-1,1-\delta} \right\}$$

- $\boldsymbol{U}$ does not appear because $\|\boldsymbol{U}^\top(\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}^\star)\|_2 \approx \|\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}^\star\|_2$
- Abbasi-yadkori et al.'11: $\boldsymbol{\theta}^\star$ with prob. $1-\delta$ lies inside

$$C_t = \left\{ \boldsymbol{\theta} \in \mathbb{R}^d \;\middle|\; \|\widehat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}\|_{\boldsymbol{\Lambda}_t} \leq \sigma \sqrt{d \log\left(\frac{1 + TL^2/\lambda}{\delta}\right)} + \lambda^{1/2} \right\}$$

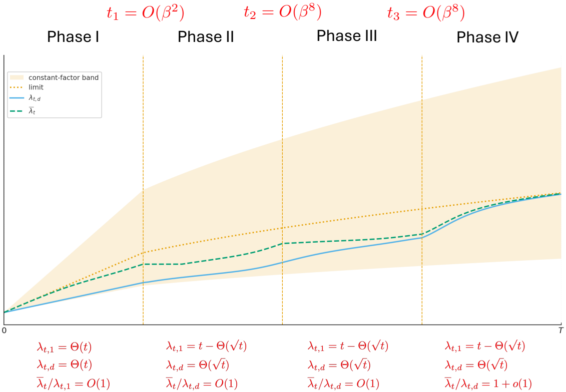
- As $\lambda_{\min}(\boldsymbol{\Lambda}_T) \approx \sqrt{2\beta^2 T/(d+1)}$, this confidence set is $O(\sqrt{\log T})$ times tighter than Abbasi-yadkori et al.'11.

A glimpse of our proof strategy.

Key ingredient: Evolution of Λ_t

Notation: eigenvalues $\lambda_{t,1} \geq \underbrace{\lambda_{t,2} \geq \ldots \geq \lambda_{t,d}}_{\text{mean denoted as } \overline{\lambda}_t}$, eigenvectors $v_{t,1}, \ldots, v_{t,d}$.

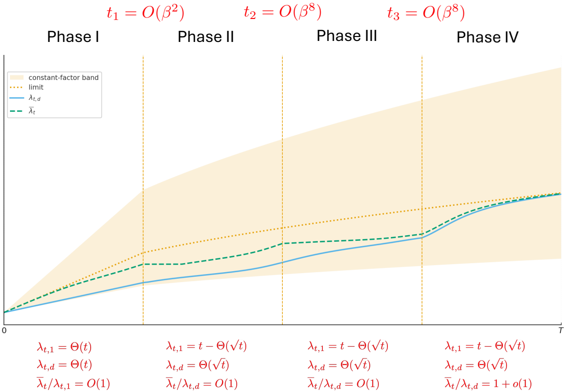
Key ingredient: Evolution of Λ_t



Phase I $(0 \leq t \leq t_1)$: initial exploration on all directions

$$\lambda_{t,d} = \Theta(t), \quad \overline{\lambda}_t/\lambda_{t,d} = O(1), \quad \|\boldsymbol{v}_{t,1} - \boldsymbol{\theta}^\star\|_2 = \Theta(1)$$

Key ingredient: Evolution of Λ_t



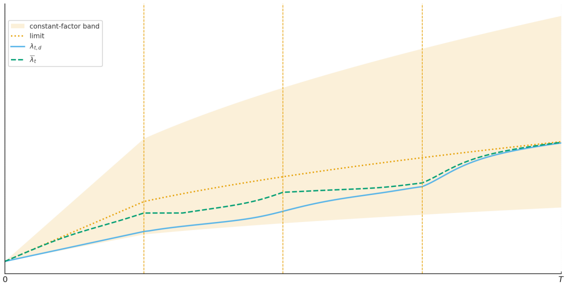
Phase II ($t_1 \le t \le t_2$): $\Theta(\sqrt{t})$ growth of non-leading eigenvalues

$$\lambda_{t,d} = \Theta(\sqrt{t}), \quad \overline{\lambda}_t/\lambda_{t,d} = O(1), \quad \|\boldsymbol{v}_{t,1} - \boldsymbol{\theta}^\star\|_{\boldsymbol{\Lambda}_t} = O(\beta)$$

Key ingredient: Evolution of Λ_t

$t_1 = O(\beta^2)$ $t_2 = O(\beta^8)$ $t_3 = O(\beta^8)$

Phase I Phase II Phase III Phase IV

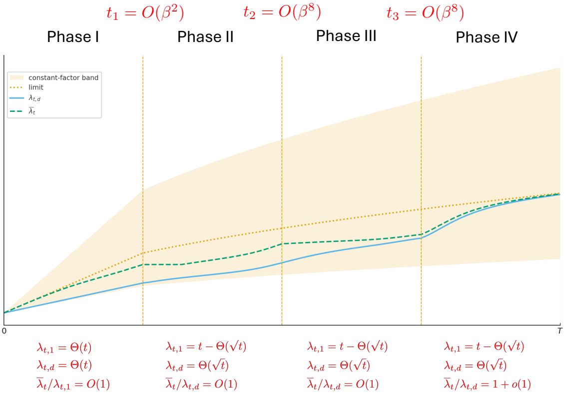


Phase I	Phase II	Phase III	Phase IV
$\lambda_{t,1} = \Theta(t)$	$\lambda_{t,1} = t - \Theta(\sqrt{t})$	$\lambda_{t,1} = t - \Theta(\sqrt{t})$	$\lambda_{t,1} = t - \Theta(\sqrt{t})$
$\lambda_{t,d} = \Theta(t)$	$\lambda_{t,d} = \Theta(\sqrt{t})$	$\lambda_{t,d} = \Theta(\sqrt{t})$	$\lambda_{t,d} = \Theta(\sqrt{t})$
$\overline{\lambda}_t/\lambda_{t,1} = O(1)$	$\overline{\lambda}_t/\lambda_{t,d} = O(1)$	$\overline{\lambda}_t/\lambda_{t,d} = O(1)$	$\overline{\lambda}_t/\lambda_{t,d} = 1 + o(1)$

Phase III ($t_2 \le t \le t_3$): fine-grained concentration of $\boldsymbol{v}_{t,1}$

$$\lambda_{t,d} = \Theta(\sqrt{t}), \quad \overline{\lambda}_t/\lambda_{t,d} = O(1), \quad \|\boldsymbol{v}_{t,1} - \boldsymbol{\theta}^\star\|_{\boldsymbol{\Lambda}_t} = O(\sqrt{\log\log T})$$

Key ingredient: Evolution of Λ_t



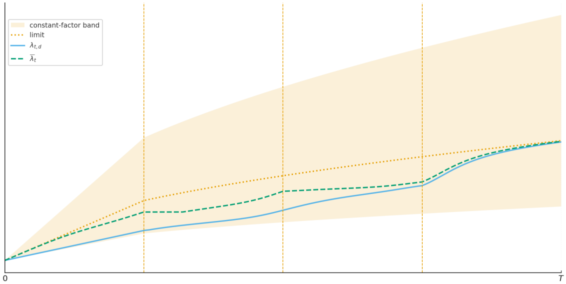
Phase IV ($t_3 \le t \le T$): concentration of non-leading eigenvalues

$$\lambda_{t,d} = \Theta(\sqrt{t}), \quad \overline{\lambda}_t/\lambda_{t,d} = 1 + o(1), \quad \|\boldsymbol{v}_{t,1} - \boldsymbol{\theta}^\star\|_{\boldsymbol{\Lambda}_t} = O(\sqrt{\log\log T})$$

Key ingredient: Evolution of Λ_t

$t_1 = O(\beta^2)$ $t_2 = O(\beta^8)$ $t_3 = O(\beta^8)$

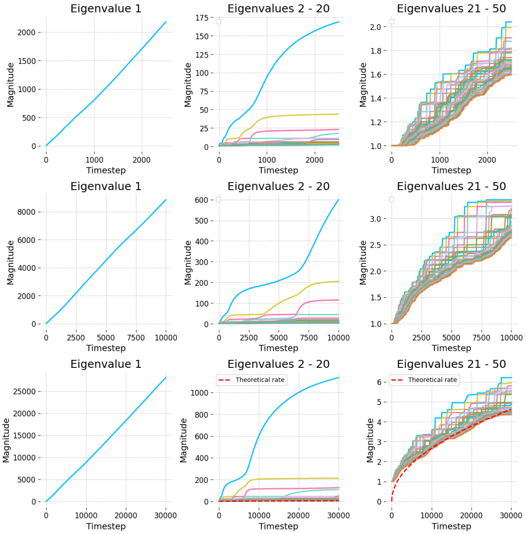
Phase I Phase II Phase III Phase IV



Phase I	Phase II	Phase III	Phase IV
$\lambda_{t,1} = \Theta(t)$	$\lambda_{t,1} = t - \Theta(\sqrt{t})$	$\lambda_{t,1} = t - \Theta(\sqrt{t})$	$\lambda_{t,1} = t - \Theta(\sqrt{t})$
$\lambda_{t,d} = \Theta(t)$	$\lambda_{t,d} = \Theta(\sqrt{t})$	$\lambda_{t,d} = \Theta(\sqrt{t})$	$\lambda_{t,d} = \Theta(\sqrt{t})$
$\overline{\lambda}_t/\lambda_{t,1} = O(1)$	$\overline{\lambda}_t/\lambda_{t,d} = O(1)$	$\overline{\lambda}_t/\lambda_{t,d} = O(1)$	$\overline{\lambda}_t/\lambda_{t,d} = 1 + o(1)$

Terminal covariance ($t = T$): non-leading eigenvalues concentrate

$$\lambda_{T,d},\ \overline{\lambda}_T = (1 + o(1))\sqrt{\frac{2\beta^2 T}{d+1}}, \quad \|\boldsymbol{v}_{t,1} - \boldsymbol{\theta}^\star\|_{\boldsymbol{\Lambda}_t} = O(\sqrt{\log\log T})$$



Rate of growth of the eigenvalues of $\mathbf{\Lambda}_T$.

Step 1: Control of estimation error

Lemma

With probability $1 - (\log T)^{-1}$*, there exists constant* C *such that for any* $1 \le t \le T$*,*

$$\|\widehat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^\star\|_2 \le C\sigma\sqrt{\frac{\log\log T}{\lambda_{t,d}}}.$$

Step 1: Control of estimation error

Lemma

With probability $1 - (\log T)^{-1}$, *there exists constant* C *such that for any* $1 \le t \le T$,

$$\|\widehat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^\star\|_2 \le C\sigma\sqrt{\frac{\log\log T}{\lambda_{t,d}}}.$$

- The influence of noise vanishes when $\beta \gg \sigma\sqrt{\log\log T}$

$$\begin{aligned}\mathrm{UCB}_t(\boldsymbol{a}) &= \langle \boldsymbol{a}, \widehat{\boldsymbol{\theta}}_{t-1}\rangle + \beta\sqrt{\boldsymbol{a}^\top \boldsymbol{\Lambda}_{t-1}^{-1}\boldsymbol{a}} \\ &= \underbrace{\langle \boldsymbol{a}, \boldsymbol{\theta}^\star\rangle}_{\text{deterministic term}} + \underbrace{\langle \boldsymbol{a}, \widehat{\boldsymbol{\theta}}_{t-1} - \boldsymbol{\theta}^\star\rangle}_{\text{controlled by } \|\widehat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^\star\|_2} + \underbrace{\beta\sqrt{\boldsymbol{a}^\top \boldsymbol{\Lambda}_{t-1}^{-1}\boldsymbol{a}}}_{\text{controlled by } O(\beta/\sqrt{\lambda_{t,d}})}\end{aligned}$$

Step 2: Asymptotic behavior

Lemma

The terminal covariance $\boldsymbol{\Lambda}_T$ satisfies

- *Leading eigenvector $\boldsymbol{v}_{T,1}$ concentrates to true signal $\boldsymbol{\theta}^\star$*

$$\|\boldsymbol{v}_{T,1} - \boldsymbol{\theta}^\star\|_2 \lesssim \sqrt{\frac{\log\log T}{\lambda_{T,d}}}$$

- *Non-leading eigenvalues $\lambda_{T,i}$ $(i \geq 2)$ concentrates to a deterministic limit*

$$\lambda_{T,i} = (1 + o(1))\sqrt{\frac{2\beta^2 T}{d+1}}, \quad \forall\, i \geq 2$$

Step 2: Asymptotic behavior

Lemma

The terminal covariance $\boldsymbol{\Lambda}_T$ satisfies

- *Leading eigenvector $\boldsymbol{v}_{T,1}$ concentrates to true signal $\boldsymbol{\theta}^\star$*

$$\|\boldsymbol{v}_{T,1} - \boldsymbol{\theta}^\star\|_2 \lesssim \sqrt{\frac{\log\log T}{\lambda_{T,d}}}$$

- *Non-leading eigenvalues $\lambda_{T,i}$ $(i \geq 2)$ concentrates to a deterministic limit*

$$\lambda_{T,i} = (1 + o(1))\sqrt{\frac{2\beta^2 T}{d+1}}, \quad \forall\, i \geq 2$$

- Controls the precise eigen-structure of terminal covariance

Concluding remarks

- Asymptotic normality of $\widehat{\boldsymbol{\theta}}_T$ when data is adaptively collected with LinUCB

Concluding remarks

- Asymptotic normality of $\widehat{\boldsymbol{\theta}}_T$ when data is adaptively collected with LinUCB
- Asymptotic $(1-\delta)$ Wald-type confidence set for $\boldsymbol{\theta}^{\star}$ with a careful characterization of $\boldsymbol{\Lambda}_t$

Concluding remarks

- From asymptotics to non-asymptotic guarantees

Concluding remarks

- From asymptotics to non-asymptotic guarantees
- From linear bandits to more realistic RL settings

Concluding remarks

- From asymptotics to non-asymptotic guarantees
- From linear bandits to more realistic RL settings
- Trade-off between regret minimization and statistical inference

Concluding remarks

- From asymptotics to non-asymptotic guarantees
- From linear bandits to more realistic RL settings
- Trade-off between regret minimization and statistical inference

Paper:

"*Statistical Inference under Adaptive Sampling with LinUCB*,"

Wei Fan*, Kevin Tan*, Yuting Wei, 2025.

— Thanks for your attention!

Proof overview by phases

- Characterization of $\boldsymbol{a}_t$:

$$\boldsymbol{a}_t = \mathcal{P}\left(\widehat{\boldsymbol{\theta}}_t + \beta \boldsymbol{\Lambda}_t^{-1/2} \boldsymbol{w}_t\right),$$

$$\text{where } \boldsymbol{w}_t = \arg\max_{\|\boldsymbol{w}\|_2=1} \left\|\widehat{\boldsymbol{\theta}}_t + \beta \cdot \boldsymbol{\Lambda}_t^{-1/2} \boldsymbol{w}\right\|_2$$

Proof overview by phases

- Characterization of $\boldsymbol{a}_t$:

$$\boldsymbol{a}_t = \mathcal{P}\left(\widehat{\boldsymbol{\theta}}_t + \beta \boldsymbol{\Lambda}_t^{-1/2} \boldsymbol{w}_t\right),$$

$$\text{where } \boldsymbol{w}_t = \arg\max_{\|\boldsymbol{w}\|_2=1} \left\|\widehat{\boldsymbol{\theta}}_t + \beta \cdot \boldsymbol{\Lambda}_t^{-1/2} \boldsymbol{w}\right\|_2$$

- An equivalent characterization on $\boldsymbol{w}_t$

$$\boldsymbol{w}_t = \arg\max_{\|\boldsymbol{w}\|_2=1} \sum_{i=1}^{d} \left(\nu_{t,i} + \frac{\beta w_i}{\sqrt{\lambda_{t,i}}}\right)^2, \quad \text{where } \sum_{i=1}^{d} \nu_{t,i}^2 = 1$$

Proof of Phase I

Phase I $(0 \le t \le t_1)$: initial exploration on all directions

$$\lambda_{t,d} = \Theta(t), \quad \overline{\lambda}_t / \lambda_{t,d} = O(1), \quad \|\boldsymbol{v}_{t,1} - \boldsymbol{\theta}^\star\|_2 = \Theta(1)$$

- Consider the optimization problem

$$\boldsymbol{w}_t = \arg\max_{\|\boldsymbol{w}\|_2 = 1} \sum_{i=1}^{d} \left(\nu_{t,i} + \frac{\beta w_i}{\sqrt{\lambda_{t,i}}} \right)^2, \quad \text{where } \sum_{i=1}^{d} \nu_{t,i}^2 = 1$$

- When $\beta / \sqrt{\lambda_{t,d}} \gtrsim 1$, bonus term dominates and $\boldsymbol{w}_t$ assigns more mass on directions with small eigenvalues
- All eigenvalues grow with comparable speed

Proof of Phase II

Phase II ($t_1 \le t \le t_2$): $\Theta(\sqrt{t})$ growth of non-leading eigenvalues

$$\lambda_{t,d} = \Theta(\sqrt{t}), \quad \overline{\lambda}_t/\lambda_{t,d} = O(1), \quad \|\boldsymbol{v}_{t,1} - \boldsymbol{\theta}^\star\|_{\boldsymbol{\Lambda}_t} = O(\beta)$$

- Let $\lambda_{t,d} = c_t \beta \sqrt{t}$ and $\underline{c}_t = \min_{t_1 \le s \le t} c_s$
- Consider the optimization problem without the first term

$$\boldsymbol{w}_{t,-1} = \arg \max_{\|\boldsymbol{w}\|_2 = C_0} \sum_{i=2}^{d} \left(\nu_{t,i} + \frac{\beta w_i}{\sqrt{\lambda_{t,i}}} \right)^2, \quad \text{where } \sum_{i=2}^{d} \nu_{t,i}^2 = O\left(\frac{\beta^2}{\lambda_{t,d}} \right)$$

- When $c_t/\underline{c}_t = O(1)$, we have $C_0 \asymp 1.$, indicating all non-leading eigenvalues grow with comparable speed

Proof of Phase III

Phase III $(t_2 \le t \le t_3)$: fine-grained concentration of the leading eigenvector

$$\lambda_{t,d} = \Theta(\sqrt{t}), \quad \overline{\lambda}_t/\lambda_{t,d} = O(1), \quad \|\boldsymbol{v}_{t,1} - \boldsymbol{\theta}^\star\|_{\boldsymbol{\Lambda}_t} = O(\sqrt{\log\log T})$$

- Update on the leading eigenvector $\boldsymbol{v}_{t,1}$

$$\boldsymbol{v}_{t+1,1} = \boldsymbol{v}_{t,1} + \frac{\widehat{\boldsymbol{\theta}}_t - \boldsymbol{v}_{t,1}}{t} + \boldsymbol{\zeta}_t,$$

  where $\boldsymbol{\zeta}_t$ is a high order fluctuation term
- Upper bound on the norm

$$\|\boldsymbol{v}_{t+1,1} - \boldsymbol{\theta}^\star\|_2 \le \left(1 - \frac{1}{t}\right)\|\boldsymbol{v}_{t,1} - \boldsymbol{\theta}^\star\|_2 + \frac{1}{t} \cdot O\left(\sqrt{\frac{\log\log T}{\lambda_{t,d}}}\right)$$

$$+ \text{ high order term}$$

Proof of Phase IV

Phase IV ($t_3 \le t \le T$): concentration of non-leading eigenvalues

$$\lambda_{t,d} = \Theta(\sqrt{t}), \quad \overline{\lambda}_t/\lambda_{t,d} = 1 + o(1), \quad \|\boldsymbol{v}_{t,1} - \boldsymbol{\theta}^\star\|_{\boldsymbol{\Lambda}_t} = O(\sqrt{\log\log T})$$

- Approximating the optimization problem

$$\boldsymbol{w}_t = \arg\max_{\|\boldsymbol{w}\|_2=1} \sum_{i=1}^{d} \left(\nu_{t,i} + \frac{\beta w_i}{\sqrt{\lambda_{t,i}}}\right)^2 \approx \left(1 + \frac{\beta w_1}{\sqrt{\lambda_{t,1}}}\right)^2 + \sum_{i=2}^{d} \frac{\beta^2 w_i^2}{\lambda_{t,i}}$$

- $\boldsymbol{w}_t$ only take mass on the direction of $\lambda_{t,1}$ and $\lambda_{t,d}$
- All non-leading eigenvalues eventually concentrate and grow with the same speed

Asymptotic CLT

- The variance of $\widehat{\boldsymbol{\theta}}_t$ is (asymptotcially) $\widetilde{\boldsymbol{\Lambda}}_T^{-1} = (\boldsymbol{U}^\top \boldsymbol{\Lambda}_T \boldsymbol{U})^{-1}$
- $\{\widetilde{\boldsymbol{\Lambda}}_T\}$ is stable. Let $\widetilde{\boldsymbol{\Sigma}}_T = \sqrt{2\beta^2 T/(d+1)}\boldsymbol{I}_{d-1}$, then

$$\|\widetilde{\boldsymbol{\Sigma}}_T^{-1}\widetilde{\boldsymbol{\Lambda}}_T - \boldsymbol{I}_{d-1}\|_2 \lesssim \left(\frac{\sqrt{\log\log T}}{\beta}\right)^{1/2}$$

- The final result is established with Central Limit Theorem and the stability of $\{\widetilde{\boldsymbol{\Lambda}}_T\}$