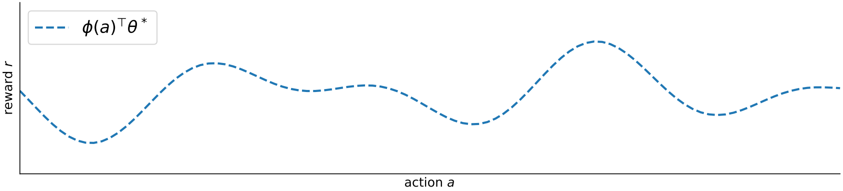
RL THEORY VIRTUAL SEMINARS

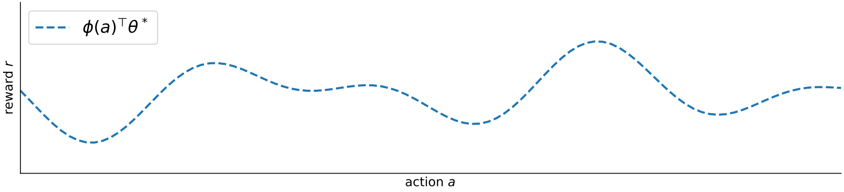
Improved Algorithms for Stochastic Linear Bandits Using Tail Bounds for Martingale Mixtures

Hamish Flynn David Reeb Melih Kandemir Jan Peters

Stochastic Linear Bandits

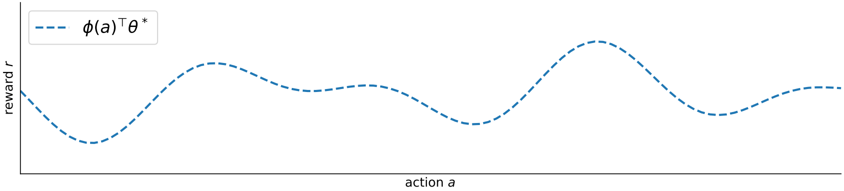


Stochastic Linear Bandits



At round t, query any action $a_t \in \mathcal{A}_t$, receive a noisy reward $r_t = \phi(a_t)^\top \boldsymbol{\theta}^* + \epsilon_t$.

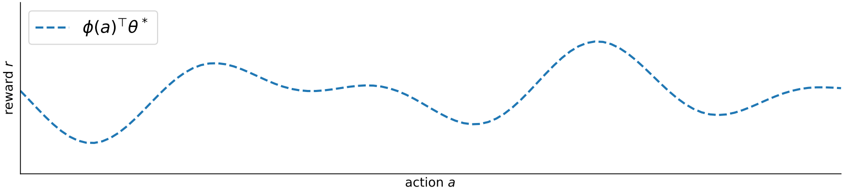
Stochastic Linear Bandits



At round t, query any action $a_t \in \mathcal{A}_t$, receive a noisy reward $r_t = \phi(a_t)^\top \boldsymbol{\theta}^* + \epsilon_t$.

Goal: Minimise cumulative regret.

Stochastic Linear Bandits



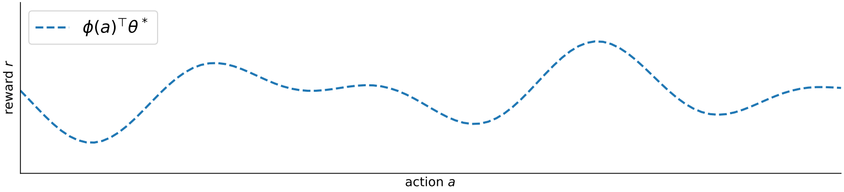
At round t, query any action $a_t \in \mathcal{A}_t$, receive a noisy reward $r_t = \phi(a_t)^\top \boldsymbol{\theta}^* + \epsilon_t$.

Goal: Minimise cumulative regret.

Assumptions: $\epsilon_1, \epsilon_2, \ldots$ are (conditionally) σ-sub-Gaussian and $\|\boldsymbol{\theta}^*\|_2 \leq B$.

$\boldsymbol{\theta}^* \in \mathbb{R}^d$ is unknown, ϕ is known and upper bounds on σ and B are known.

Stochastic Linear Bandits



At round t, query any action $a_t \in \mathcal{A}_t$, receive a noisy reward $r_t = \phi(a_t)^\top \boldsymbol{\theta}^* + \epsilon_t$.

Goal: Minimise cumulative regret.

Assumptions: $\epsilon_1, \epsilon_2, \ldots$ are (conditionally) σ-sub-Gaussian and $\|\boldsymbol{\theta}^*\|_2 \leq B$.

$\boldsymbol{\theta}^* \in \mathbb{R}^d$ is unknown, ϕ is known and upper bounds on σ and B are known.

Examples: Black-box optimisation, recommendation systems, etc.

Confidence Sets/Bounds for Stochastic Linear Bandits

Confidence set: A confidence set Θ_t contains all $\boldsymbol{\theta}$'s that could plausibly be $\boldsymbol{\theta}^*$ given data up to time t.

Confidence Sets/Bounds for Stochastic Linear Bandits

Confidence set: A confidence set Θ_t contains all $\boldsymbol{\theta}$'s that could plausibly be $\boldsymbol{\theta}^*$ given data up to time t.

We want the smallest sequence of confidence sets $\Theta_1, \Theta_2, \ldots$ that satisfies the coverage condition

$$\mathbb{P}_{\substack{a_1, a_2, \ldots \\ r_1, r_2, \ldots}} \left[\forall t \geq 1 : \boldsymbol{\theta}^* \in \Theta_t \right] \geq 1 - \delta.$$

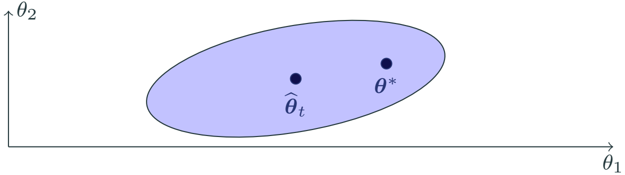
Confidence Sets/Bounds for Stochastic Linear Bandits

Confidence set: A confidence set Θ_t contains all $\boldsymbol{\theta}$'s that could plausibly be $\boldsymbol{\theta}^*$ given data up to time t.

We want the smallest sequence of confidence sets $\Theta_1, \Theta_2, \dots$ that satisfies the coverage condition

$$\mathbb{P}_{\substack{a_1, a_2, \dots \\ r_1, r_2, \dots}} \left[\forall t \geq 1 : \boldsymbol{\theta}^* \in \Theta_t \right] \geq 1 - \delta.$$

Gold Standard (OFUL):[1] Θ_t is an ellipsoid centred at the regularised least squares/Ridge estimate $\widehat{\boldsymbol{\theta}}_t$, with a radius determined using self-normalised concentration and the method of mixtures.



Y. Abbasi-Yadkori et al. (2011) Improved algorithms for linear stochastic bandits. NeurIPS

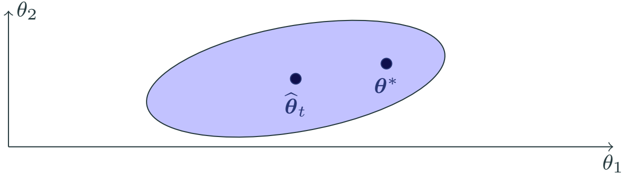
Confidence Sets/Bounds for Stochastic Linear Bandits

Confidence set: A confidence set Θ_t contains all $\boldsymbol{\theta}$'s that could plausibly be $\boldsymbol{\theta}^*$ given data up to time t.

We want the smallest sequence of confidence sets $\Theta_1, \Theta_2, \ldots$ that satisfies the coverage condition

$$\mathbb{P}_{\substack{a_1, a_2, \ldots \\ r_1, r_2, \ldots}} \left[\forall t \geq 1 : \boldsymbol{\theta}^* \in \Theta_t \right] \geq 1 - \delta.$$

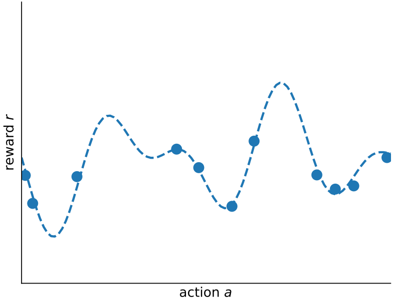
Gold Standard (OFUL):[1] Θ_t is an ellipsoid centred at the regularised least squares/Ridge estimate $\widehat{\boldsymbol{\theta}}_t$, with a radius determined using self-normalised concentration and the method of mixtures.



The corresponding upper confidence bound is $\max_{\boldsymbol{\theta} \in \Theta_t} \{\phi(a)^\top \boldsymbol{\theta}\}$.

1. Y. Abbasi-Yadkori et al. (2011) Improved algorithms for linear stochastic bandits. NeurIPS

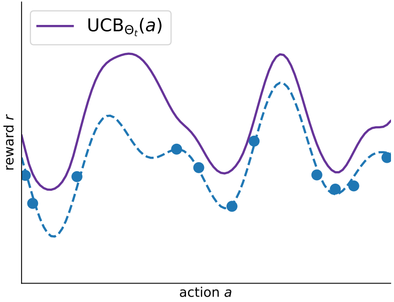
UCB Algorithms for Stochastic Linear Bandits



LinUCB:

For $t = 0, 1, 2, \ldots$

UCB Algorithms for Stochastic Linear Bandits



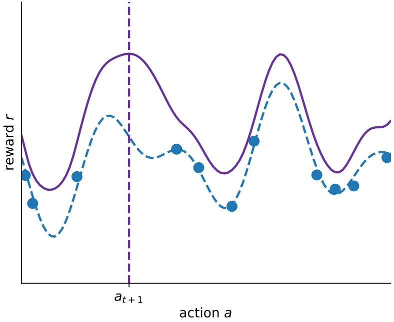
LinUCB:

For $t = 0, 1, 2, \ldots$

- Use $\{(a_k, r_k)\}_{k=1}^{t}$ to construct a confidence set Θ_t and the upper confidence bound

$$\mathrm{UCB}_{\Theta_t}(a) := \max_{\theta \in \Theta_t} \{\phi(a)^\top \boldsymbol{\theta}\}$$

UCB Algorithms for Stochastic Linear Bandits



LinUCB:

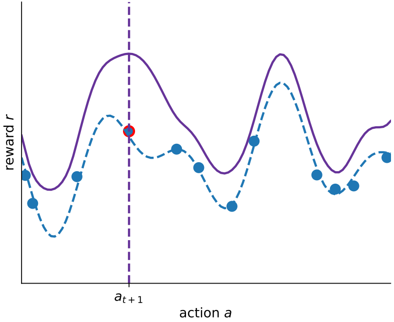
For $t = 0, 1, 2, \ldots$

- Use $\{(a_k, r_k)\}_{k=1}^{t}$ to construct a confidence set Θ_t and the upper confidence bound

$$\mathrm{UCB}_{\Theta_t}(a) := \max_{\theta \in \Theta_t} \{\phi(a)^\top \boldsymbol{\theta}\}$$

- Play $a_{t+1} = \mathrm{argmax}_{a \in \mathcal{A}_{t+1}} \{\mathrm{UCB}_{\Theta_t}(a)\}$

UCB Algorithms for Stochastic Linear Bandits



LinUCB:

For $t = 0, 1, 2, \ldots$

- Use $\{(a_k, r_k)\}_{k=1}^{t}$ to construct a confidence set Θ_t and the upper confidence bound

$$\mathrm{UCB}_{\Theta_t}(a) := \max_{\theta \in \Theta_t} \{\phi(a)^\top \boldsymbol{\theta}\}$$

- Play $a_{t+1} = \mathrm{argmax}_{a \in \mathcal{A}_{t+1}} \{\mathrm{UCB}_{\Theta_t}(a)\}$
- Observe reward $r_{t+1} = \phi(a_{t+1})^\top \boldsymbol{\theta}^* + \epsilon_{t+1}$

This Work

This Work

Question: Is it possible to construct even tighter confidence sets/bounds for linear bandits?

This Work

Question: Is it possible to construct even tighter confidence sets/bounds for linear bandits?

Rest of the Talk:

- Constructing confidence sets for linear bandits
- Computing confidence bounds/solving $\max_{\boldsymbol{\theta} \in \Theta_t} \{\phi(a)^\top \boldsymbol{\theta}\}$
- Regret bounds for LinUCB with our confidence sets
- In what sense is this better than OFUL (and why)?
- Some experimental results
- Open questions

Notation, Etc.

Notation

Notation:

- $\boldsymbol{\theta}^*$ = true parameter vector, $\boldsymbol{\theta}$ = a candidate parameter vector
- $f_1, f_2, \ldots$ = sequence of predictions
- $\mathcal{H}_t = (a_1, r_1, \ldots, a_t, r_t, a_{t+1})$ = history of the bandit problem

Notation

Notation:

- $\boldsymbol{\theta}^*$ = true parameter vector, $\boldsymbol{\theta}$ = a candidate parameter vector
- $f_1, f_2, \ldots$ = sequence of predictions
- $\mathcal{H}_t = (a_1, r_1, \ldots, a_t, r_t, a_{t+1})$ = history of the bandit problem

Predictable sequences:

- I call a sequence of random variables $x_1, x_2, \ldots$ predictable if, given $\mathcal{H}_{t-1}$, x_t is no longer random
- e.g. $a_1, a_2 \ldots.$ are predictable, $f_1, f_2, \ldots$ are predictable, $r_1, r_2, \ldots$ are not predictable

Notation

Notation:

- $\boldsymbol{\theta}^*$ = true parameter vector, $\boldsymbol{\theta}$ = a candidate parameter vector
- $f_1, f_2, \ldots$ = sequence of predictions
- $\mathcal{H}_t = (a_1, r_1, \ldots, a_t, r_t, a_{t+1})$ = history of the bandit problem

Predictable sequences:

- I call a sequence of random variables $x_1, x_2, \ldots$ predictable if, given $\mathcal{H}_{t-1}$, x_t is no longer random
- e.g. $a_1, a_2 \ldots.$ are predictable, $f_1, f_2, \ldots$ are predictable, $r_1, r_2, \ldots$ are not predictable

Matrix/vector notation:

- $\Phi_t = [\phi(a_1), \ldots, \phi(a_t)]^\top \in \mathbb{R}^{t \times d}$ = matrix of first t feature vectors
- $\boldsymbol{r}_t = [r_1, \ldots, r_t]^\top$ = vector of first t rewards
- $\boldsymbol{\epsilon}_t = [\epsilon_1, \ldots, \epsilon_t]^\top$ = vector of first t noise variables
- $\boldsymbol{f}_t = [f_1, \ldots, f_t]^\top$ = vector of first t predictions

Sequences of Predictions

Throughout the talk, $f_1, f_2, \ldots$ is a sequence of predictions for the rewards $r_1, r_2, \ldots,$ where each f_t can depend on the history $\mathcal{H}_{t-1}$.

Sequences of Predictions

Throughout the talk, $f_1, f_2, \ldots$ is a sequence of predictions for the rewards $r_1, r_2, \ldots,$ where each f_t can depend on the history $\mathcal{H}_{t-1}$.

Examples:

- For a fixed $\boldsymbol{\theta}$, we could set $f_t = \phi(a_t)^\top \boldsymbol{\theta}$ for each $t \geq 1$ (or $\boldsymbol{f}_t = \Phi_t \boldsymbol{\theta}$ in matrix notation)

Sequences of Predictions

Throughout the talk, $f_1, f_2, \ldots$ is a sequence of predictions for the rewards $r_1, r_2, \ldots,$ where each f_t can depend on the history $\mathcal{H}_{t-1}$.

Examples:

- For a fixed $\boldsymbol{\theta}$, we could set $f_t = \phi(a_t)^\top \boldsymbol{\theta}$ for each $t \geq 1$ (or $\boldsymbol{f}_t = \Phi_t \boldsymbol{\theta}$ in matrix notation)
- $f_1, f_2, \ldots$ could be a sequence of predictions generated by running an online learning algorithm

Sequences of Predictions

Throughout the talk, $f_1, f_2, \ldots$ is a sequence of predictions for the rewards $r_1, r_2, \ldots$, where each f_t can depend on the history $\mathcal{H}_{t-1}$.

Examples:

- For a fixed $\boldsymbol{\theta}$, we could set $f_t = \phi(a_t)^\top \boldsymbol{\theta}$ for each $t \geq 1$ (or $\boldsymbol{f}_t = \Phi_t \boldsymbol{\theta}$ in matrix notation)
- $f_1, f_2, \ldots$ could be a sequence of predictions generated by running an online learning algorithm
- E.g., $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots$ could be a sequence of parameter estimates, and we could set $f_t = \phi(a_t)^\top \boldsymbol{\theta}_t$

Sequences of Predictions

Throughout the talk, $f_1, f_2, \ldots$ is a sequence of predictions for the rewards $r_1, r_2, \ldots$, where each f_t can depend on the history $\mathcal{H}_{t-1}$.

Examples:

- For a fixed $\boldsymbol{\theta}$, we could set $f_t = \phi(a_t)^\top \boldsymbol{\theta}$ for each $t \geq 1$ (or $\boldsymbol{f}_t = \Phi_t \boldsymbol{\theta}$ in matrix notation)
- $f_1, f_2, \ldots$ could be a sequence of predictions generated by running an online learning algorithm
- E.g., $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots$ could be a sequence of parameter estimates, and we could set $f_t = \phi(a_t)^\top \boldsymbol{\theta}_t$
- We could choose something boring like $f_t \equiv 0$

Sequences of Predictions

Throughout the talk, $f_1, f_2, \ldots$ is a sequence of predictions for the rewards $r_1, r_2, \ldots,$ where each f_t can depend on the history $\mathcal{H}_{t-1}$.

Examples:

- For a fixed $\boldsymbol{\theta}$, we could set $f_t = \phi(a_t)^\top \boldsymbol{\theta}$ for each $t \geq 1$ (or $\boldsymbol{f}_t = \Phi_t \boldsymbol{\theta}$ in matrix notation)
- $f_1, f_2, \ldots$ could be a sequence of predictions generated by running an online learning algorithm
- E.g., $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots$ could be a sequence of parameter estimates, and we could set $f_t = \phi(a_t)^\top \boldsymbol{\theta}_t$
- We could choose something boring like $f_t \equiv 0$

Randomised predictions: Later on in the talk, we will consider distributions over sequences of predictions.

- $\boldsymbol{f}_t$ will be a random draw from P_t, which is distribution on $\mathbb{R}^t$

Sequences of Predictions

Throughout the talk, $f_1, f_2, \ldots$ is a sequence of predictions for the rewards $r_1, r_2, \ldots,$ where each f_t can depend on the history $\mathcal{H}_{t-1}$.

Examples:

- For a fixed $\boldsymbol{\theta}$, we could set $f_t = \phi(a_t)^\top \boldsymbol{\theta}$ for each $t \geq 1$ (or $\boldsymbol{f}_t = \Phi_t \boldsymbol{\theta}$ in matrix notation)
- $f_1, f_2, \ldots$ could be a sequence of predictions generated by running an online learning algorithm
- E.g., $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots$ could be a sequence of parameter estimates, and we could set $f_t = \phi(a_t)^\top \boldsymbol{\theta}_t$
- We could choose something boring like $f_t \equiv 0$

Randomised predictions: Later on in the talk, we will consider distributions over sequences of predictions.

- $\boldsymbol{f}_t$ will be a random draw from P_t, which is distribution on $\mathbb{R}^t$
- Each P_t can depend on the history $\mathcal{H}_{t-1}$

Sequences of Predictions

Throughout the talk, $f_1, f_2, \ldots$ is a sequence of predictions for the rewards $r_1, r_2, \ldots$, where each f_t can depend on the history $\mathcal{H}_{t-1}$.

Examples:

- For a fixed $\boldsymbol{\theta}$, we could set $f_t = \phi(a_t)^\top \boldsymbol{\theta}$ for each $t \geq 1$ (or $\boldsymbol{f}_t = \Phi_t \boldsymbol{\theta}$ in matrix notation)
- $f_1, f_2, \ldots$ could be a sequence of predictions generated by running an online learning algorithm
- E.g., $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots$ could be a sequence of parameter estimates, and we could set $f_t = \phi(a_t)^\top \boldsymbol{\theta}_t$
- We could choose something boring like $f_t \equiv 0$

Randomised predictions: Later on in the talk, we will consider distributions over sequences of predictions.

- $\boldsymbol{f}_t$ will be a random draw from P_t, which is distribution on $\mathbb{R}^t$
- Each P_t can depend on the history $\mathcal{H}_{t-1}$
- If $P_t = \mathcal{N}(\boldsymbol{\mu}_t, \boldsymbol{T}_t)$, $\boldsymbol{\mu}_t$ can still be thought of the first t predictions, and $\boldsymbol{T}_t$ can be thought of as the uncertainty associated with the first t predictions

Sequences of Predictions

Throughout the talk, $f_1, f_2, \ldots$ is a sequence of predictions for the rewards $r_1, r_2, \ldots,$ where each f_t can depend on the history $\mathcal{H}_{t-1}$.

Examples:

- For a fixed $\boldsymbol{\theta}$, we could set $f_t = \phi(a_t)^\top \boldsymbol{\theta}$ for each $t \geq 1$ (or $\boldsymbol{f}_t = \Phi_t \boldsymbol{\theta}$ in matrix notation)
- $f_1, f_2, \ldots$ could be a sequence of predictions generated by running an online learning algorithm
- E.g., $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots$ could be a sequence of parameter estimates, and we could set $f_t = \phi(a_t)^\top \boldsymbol{\theta}_t$
- We could choose something boring like $f_t \equiv 0$

Randomised predictions: Later on in the talk, we will consider distributions over sequences of predictions.

- $\boldsymbol{f}_t$ will be a random draw from P_t, which is distribution on $\mathbb{R}^t$
- Each P_t can depend on the history $\mathcal{H}_{t-1}$
- If $P_t = \mathcal{N}(\boldsymbol{\mu}_t, \boldsymbol{T}_t)$, $\boldsymbol{\mu}_t$ can still be thought of the first t predictions, and $\boldsymbol{T}_t$ can be thought of as the uncertainty associated with the first t predictions
- E.g. suppose $\boldsymbol{\theta} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I})$ and define P_t to be the induced distribution on $\Phi_t \boldsymbol{\theta}$, so $P_t = \mathcal{N}(\mathbf{0}, \Phi_t \Phi_t^\top)$

Constructing Confidence Sets

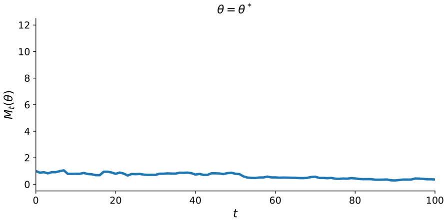
General Plan (Part 1)

Step 1: Construct a collection of non-negative random processes $M_t(\boldsymbol{f}_t, \boldsymbol{\theta})$ such that:

General Plan (Part 1)

Step 1: Construct a collection of non-negative random processes $M_t(\boldsymbol{f}_t, \boldsymbol{\theta})$ such that:

$$\mathbb{E}[M_t(\boldsymbol{f}_t, \boldsymbol{\theta}^*)|\mathcal{H}_{t-1}] \leq M_{t-1}(\boldsymbol{f}_{t-1}, \boldsymbol{\theta}^*)$$

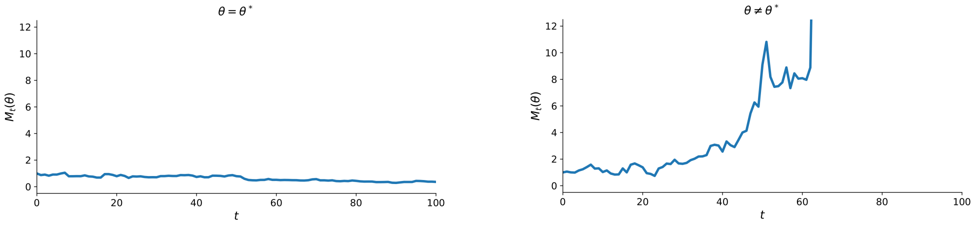


General Plan (Part 1)

Step 1: Construct a collection of non-negative random processes $M_t(\boldsymbol{f}_t, \boldsymbol{\theta})$ such that:

$\mathbb{E}[M_t(\boldsymbol{f}_t, \boldsymbol{\theta}^*)|\mathcal{H}_{t-1}] \leq M_{t-1}(\boldsymbol{f}_{t-1}, \boldsymbol{\theta}^*)$

If $\boldsymbol{\theta} \neq \boldsymbol{\theta}^*$, $M_t(\boldsymbol{f}_t, \boldsymbol{\theta})$ blows up

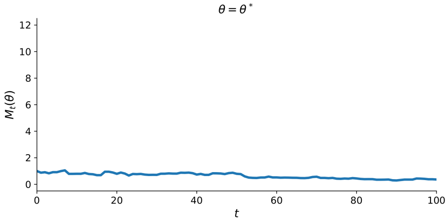


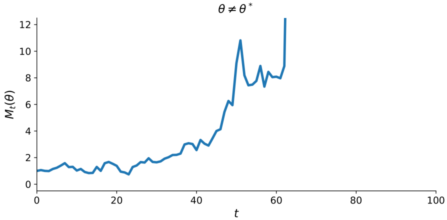
General Plan (Part 1)

Step 1: Construct a collection of non-negative random processes $M_t(\boldsymbol{f}_t, \boldsymbol{\theta})$ such that:

$\mathbb{E}[M_t(\boldsymbol{f}_t, \boldsymbol{\theta}^*)|\mathcal{H}_{t-1}] \leq M_{t-1}(\boldsymbol{f}_{t-1}, \boldsymbol{\theta}^*)$

If $\boldsymbol{\theta} \neq \boldsymbol{\theta}^*$, $M_t(\boldsymbol{f}_t, \boldsymbol{\theta})$ blows up





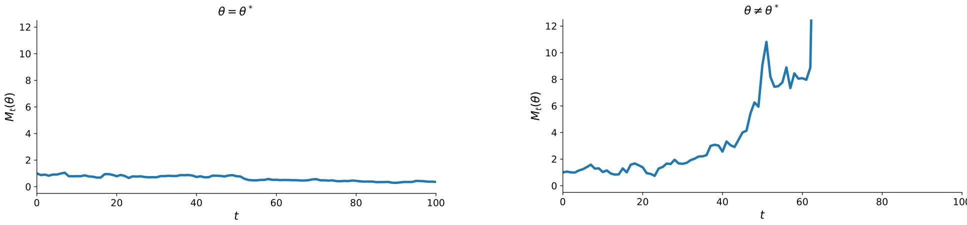
We want to maximise $M_t(\boldsymbol{f}_t, \boldsymbol{\theta})$ w.r.t. $\boldsymbol{f}_t$, but the maximiser is not a predictable sequence.

General Plan (Part 1)

Step 1: Construct a collection of non-negative random processes $M_t(\boldsymbol{f}_t, \boldsymbol{\theta})$ such that:

$\mathbb{E}[M_t(\boldsymbol{f}_t, \boldsymbol{\theta}^*)|\mathcal{H}_{t-1}] \leq M_{t-1}(\boldsymbol{f}_{t-1}, \boldsymbol{\theta}^*)$

If $\boldsymbol{\theta} \neq \boldsymbol{\theta}^*$, $M_t(\boldsymbol{f}_t, \boldsymbol{\theta})$ blows up



We want to maximise $M_t(\boldsymbol{f}_t, \boldsymbol{\theta})$ w.r.t. $\boldsymbol{f}_t$, but the maximiser is not a predictable sequence.

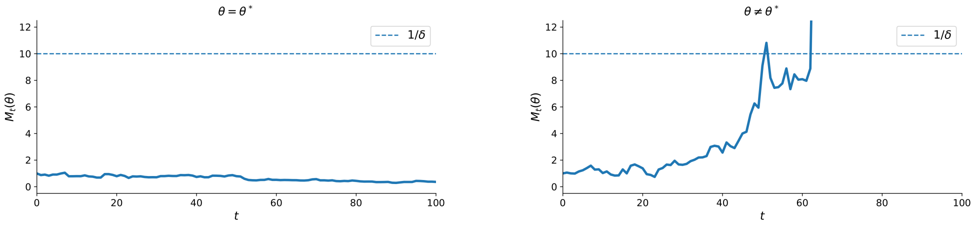
Step 2: Laplace's method/pseudo-maximisation/method of mixtures

$$\mathbb{E}_{\boldsymbol{f}_t \sim P_t}[M_t(\boldsymbol{f}_t, \boldsymbol{\theta})] \approx \max_{\boldsymbol{f}_t} M_t(\boldsymbol{f}_t, \boldsymbol{\theta}).$$

General Plan (Part 2)

Step 3: Use Ville's inequality to determine a threshold level

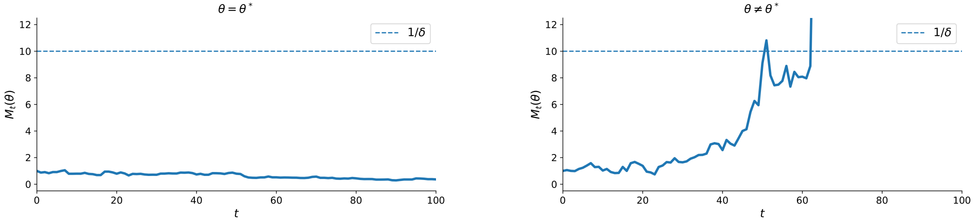
$$\mathbb{P}\left(\forall t \geq 1 : \mathbb{E}_{\boldsymbol{f}_t \sim P_t}\left[M_t(\boldsymbol{f}_t, \boldsymbol{\theta}^*)\right] \leq 1/\delta\right) \geq 1 - \delta.$$



General Plan (Part 2)

Step 3: Use Ville's inequality to determine a threshold level

$$\mathbb{P}\left(\forall t \geq 1 : \mathbb{E}_{\boldsymbol{f}_t \sim P_t}\left[M_t(\boldsymbol{f}_t, \boldsymbol{\theta}^*)\right] \leq 1/\delta\right) \geq 1 - \delta.$$



Step 4: We define our confidence sets as

$$\Theta_t := \left\{\boldsymbol{\theta} \in \mathbb{R}^d : \mathbb{E}_{\boldsymbol{f}_t \sim P_t}\left[M_t(\boldsymbol{f}_t, \boldsymbol{\theta})\right] \leq 1/\delta\right\} \cap \left\{\boldsymbol{\theta} \in \mathbb{R}^d : \|\boldsymbol{\theta}\|_2 \leq B\right\}.$$

What Do We Want From $M_t(\boldsymbol{f}_t, \boldsymbol{\theta}^*)$?

We want to construct a collection of supermartingales $M_t(\boldsymbol{f}_t, \boldsymbol{\theta}^*)$ such that:

- $M_t(\boldsymbol{f}_t, \boldsymbol{\theta}^*)$ is always non-negative
- $\mathbb{E}_{\boldsymbol{f}_t \sim P_t}\left[M_t(\boldsymbol{f}_t, \boldsymbol{\theta}^*)\right]$ has a closed-form expression whenever P_t is Gaussian
- $\mathbb{E}_{\boldsymbol{f}_t \sim P_t}\left[M_t(\boldsymbol{f}_t, \boldsymbol{\theta}^*)\right] \leq 1/\delta$ is a convex constraint for $\boldsymbol{\theta}^*$

What Do We Want From $M_t(\boldsymbol{f}_t, \boldsymbol{\theta}^*)$?

We want to construct a collection of supermartingales $M_t(\boldsymbol{f}_t, \boldsymbol{\theta}^*)$ such that:

- $M_t(\boldsymbol{f}_t, \boldsymbol{\theta}^*)$ is always non-negative
- $\mathbb{E}_{\boldsymbol{f}_t \sim P_t}\left[M_t(\boldsymbol{f}_t, \boldsymbol{\theta}^*)\right]$ has a closed-form expression whenever P_t is Gaussian
- $\mathbb{E}_{\boldsymbol{f}_t \sim P_t}\left[M_t(\boldsymbol{f}_t, \boldsymbol{\theta}^*)\right] \leq 1/\delta$ is a convex constraint for $\boldsymbol{\theta}^*$

Look for $M_t(\boldsymbol{f}_t, \boldsymbol{\theta}^*)$ in the form

$$M_t(\boldsymbol{f}_t, \boldsymbol{\theta}^*) = \exp\left(\sum_{k=1}^{t} \mathrm{quad}(f_k, \phi(a_k)^\top \boldsymbol{\theta}^*)\right) = \prod_{k=1}^{t} \exp\Big(\mathrm{quad}(f_k, \phi(a_t)^\top \boldsymbol{\theta}^*)\Big)$$

What Do We Want From $M_t(\boldsymbol{f}_t, \boldsymbol{\theta}^*)$?

We want to construct a collection of supermartingales $M_t(\boldsymbol{f}_t, \boldsymbol{\theta}^*)$ such that:

- $M_t(\boldsymbol{f}_t, \boldsymbol{\theta}^*)$ is always non-negative
- $\mathbb{E}_{\boldsymbol{f}_t \sim P_t}\left[M_t(\boldsymbol{f}_t, \boldsymbol{\theta}^*)\right]$ has a closed-form expression whenever P_t is Gaussian
- $\mathbb{E}_{\boldsymbol{f}_t \sim P_t}\left[M_t(\boldsymbol{f}_t, \boldsymbol{\theta}^*)\right] \leq 1/\delta$ is a convex constraint for $\boldsymbol{\theta}^*$

Look for $M_t(\boldsymbol{f}_t, \boldsymbol{\theta}^*)$ in the form

$$M_t(\boldsymbol{f}_t, \boldsymbol{\theta}^*) = \exp\left(\sum_{k=1}^{t} \mathrm{quad}(f_k, \phi(a_k)^\top \boldsymbol{\theta}^*)\right) = \prod_{k=1}^{t} \exp\Big(\mathrm{quad}(f_k, \phi(a_t)^\top \boldsymbol{\theta}^*)\Big)$$

This is a supermartingale if, for all $t \geq 1$,

$$\mathbb{E}\left[\exp\Big(\mathrm{quad}(f_t, \phi(a_t)^\top \boldsymbol{\theta}^*)\Big) \,|\mathcal{H}_{t-1}\right] \leq 1.$$

Choosing The Quadratic Bits

Since $\epsilon_1, \epsilon_2, \ldots$ are (conditionally) σ-sub-Gaussian, we know that for any predictable $\lambda_1, \lambda_2, \ldots,$

$$\mathbb{E}\left[\exp(\lambda_t(f_t - \phi(a_t)^\top \boldsymbol{\theta}^*)\epsilon_t) | \mathcal{H}_{t-1}\right] \leq \exp\left(\frac{\sigma^2 \lambda_t^2 (f_t - \phi(a_t)^\top \boldsymbol{\theta}^*)^2}{2}\right).$$

Choosing The Quadratic Bits

Since $\epsilon_1, \epsilon_2, \ldots$ are (conditionally) σ-sub-Gaussian, we know that for any predictable $\lambda_1, \lambda_2, \ldots,$

$$\mathbb{E}\left[\exp(\lambda_t(f_t - \phi(a_t)^\top \boldsymbol{\theta}^*)\epsilon_t) | \mathcal{H}_{t-1}\right] \leq \exp\left(\frac{\sigma^2 \lambda_t^2 (f_t - \phi(a_t)^\top \boldsymbol{\theta}^*)^2}{2}\right).$$

Therefore, we know that

$$\mathbb{E}\left[\exp\left(\lambda_t(f_t - \phi(a_t)^\top \boldsymbol{\theta}^*)\epsilon_t - \frac{\sigma^2 \lambda_t^2 (f_t - \phi(a_t)^\top \boldsymbol{\theta}^*)^2}{2}\right) | \mathcal{H}_{t-1}\right] \leq 1. \tag{1}$$

Choosing The Quadratic Bits

Since $\epsilon_1, \epsilon_2, \ldots$ are (conditionally) σ-sub-Gaussian, we know that for any predictable $\lambda_1, \lambda_2, \ldots,$

$$\mathbb{E}\left[\exp(\lambda_t(f_t - \phi(a_t)^\top \boldsymbol{\theta}^*)\epsilon_t)|\mathcal{H}_{t-1}\right] \leq \exp\left(\frac{\sigma^2\lambda_t^2(f_t - \phi(a_t)^\top \boldsymbol{\theta}^*)^2}{2}\right).$$

Therefore, we know that

$$\mathbb{E}\left[\exp\left(\lambda_t(f_t - \phi(a_t)^\top \boldsymbol{\theta}^*)\epsilon_t - \frac{\sigma^2\lambda_t^2(f_t - \phi(a_t)^\top \boldsymbol{\theta}^*)^2}{2}\right)|\mathcal{H}_{t-1}\right] \leq 1. \tag{1}$$

Using $\epsilon_t = r_t - \phi(a_t)^\top \boldsymbol{\theta}^*$, the $\exp(\cdots)$ term in (1) can be re-written as

$$\exp\left(\frac{\lambda_t}{2}(\phi(a_t)^\top \boldsymbol{\theta}^* - r_t)^2 - \frac{\lambda_t}{2}(f_t - r_t)^2 + \frac{1}{2}(\lambda_t - \sigma^2\lambda_t^2)(f_t - \phi(a_t)^\top \boldsymbol{\theta}^*)^2\right).$$

Choosing The Quadratic Bits

Since $\epsilon_1, \epsilon_2, \dots$ are (conditionally) σ-sub-Gaussian, we know that for any predictable $\lambda_1, \lambda_2, \dots,$

$$\mathbb{E}\left[\exp(\lambda_t(f_t - \phi(a_t)^\top \boldsymbol{\theta}^*)\epsilon_t)|\mathcal{H}_{t-1}\right] \leq \exp\left(\frac{\sigma^2\lambda_t^2(f_t - \phi(a_t)^\top \boldsymbol{\theta}^*)^2}{2}\right).$$

Therefore, we know that

$$\mathbb{E}\left[\exp\left(\lambda_t(f_t - \phi(a_t)^\top \boldsymbol{\theta}^*)\epsilon_t - \frac{\sigma^2\lambda_t^2(f_t - \phi(a_t)^\top \boldsymbol{\theta}^*)^2}{2}\right)|\mathcal{H}_{t-1}\right] \leq 1. \tag{1}$$

Using $\epsilon_t = r_t - \phi(a_t)^\top \boldsymbol{\theta}^*$, the $\exp(\cdots)$ term in (1) can be re-written as

$$\exp\left(\frac{\lambda_t}{2}(\phi(a_t)^\top \boldsymbol{\theta}^* - r_t)^2 - \frac{\lambda_t}{2}(f_t - r_t)^2 + \frac{1}{2}(\lambda_t - \sigma^2\lambda_t^2)(f_t - \phi(a_t)^\top \boldsymbol{\theta}^*)^2\right).$$

Setting $\lambda_t \equiv 1/\sigma^2$, this becomes

$$\exp\left(\frac{1}{2\sigma^2}(\phi(a_t)^\top \boldsymbol{\theta}^* - r_t)^2 - \frac{1}{2\sigma^2}(f_t - r_t)^2\right).$$

Putting Everything Together

Multiplying the $\exp(\text{quad}(\cdots))$ terms together, we obtain

$$M_t(\boldsymbol{f}_t, \boldsymbol{\theta}^*) = \prod_{k=1}^{t} \exp\left(\frac{1}{2\sigma^2}(\phi(a_t)^\top \boldsymbol{\theta}^* - r_t)^2 - \frac{1}{2\sigma^2}(f_t - r_t)^2\right)$$
$$= \exp\left(\frac{1}{2\sigma^2}\left\|\Phi_t \boldsymbol{\theta}^* - \boldsymbol{r}_t\right\|_2^2 - \frac{1}{2\sigma^2}\left\|\boldsymbol{f}_t - \boldsymbol{r}_t\right\|_2^2\right).$$

Putting Everything Together

Multiplying the $\exp(\text{quad}(\cdots))$ terms together, we obtain

$$M_t(\boldsymbol{f}_t, \boldsymbol{\theta}^*) = \prod_{k=1}^{t} \exp\left(\frac{1}{2\sigma^2}(\phi(a_t)^\top \boldsymbol{\theta}^* - r_t)^2 - \frac{1}{2\sigma^2}(f_t - r_t)^2\right)$$
$$= \exp\left(\frac{1}{2\sigma^2}\|\Phi_t \boldsymbol{\theta}^* - \boldsymbol{r}_t\|_2^2 - \frac{1}{2\sigma^2}\|\boldsymbol{f}_t - \boldsymbol{r}_t\|_2^2\right).$$

Closed-form integration. $M_t(\boldsymbol{f}_t, \boldsymbol{\theta}^*)$ is an unnormalised Gaussian density function (with mean $\boldsymbol{r}_t$ and covariance $\sigma^2 \boldsymbol{I}$), so we can use known tricks for integrating products of Gaussian densities.

Putting Everything Together

Multiplying the $\exp(\text{quad}(\cdots))$ terms together, we obtain

$$M_t(\boldsymbol{f}_t, \boldsymbol{\theta}^*) = \prod_{k=1}^{t} \exp\left(\frac{1}{2\sigma^2}(\phi(a_t)^\top \boldsymbol{\theta}^* - r_t)^2 - \frac{1}{2\sigma^2}(f_t - r_t)^2\right)$$
$$= \exp\left(\frac{1}{2\sigma^2}\left\|\Phi_t \boldsymbol{\theta}^* - \boldsymbol{r}_t\right\|_2^2 - \frac{1}{2\sigma^2}\left\|\boldsymbol{f}_t - \boldsymbol{r}_t\right\|_2^2\right).$$

Closed-form integration. $M_t(\boldsymbol{f}_t, \boldsymbol{\theta}^*)$ is an unnormalised Gaussian density function (with mean $\boldsymbol{r}_t$ and covariance $\sigma^2 \boldsymbol{I}$), so we can use known tricks for integrating products of Gaussian densities.

Convex constraint. $M_t(\boldsymbol{f}_t, \boldsymbol{\theta}^*)$ is the composition of $\exp(\cdot)$ and a convex function of $\boldsymbol{\theta}^*$, which means $\mathbb{E}_{\boldsymbol{f}_t \sim P_t}\left[M_t(\boldsymbol{f}_t, \boldsymbol{\theta}^*)\right] \leq 1/\delta$ is a convex constraint for $\boldsymbol{\theta}^*$.

Putting Everything Together

Multiplying the $\exp(\text{quad}(\cdots))$ terms together, we obtain

$$M_t(\boldsymbol{f}_t, \boldsymbol{\theta}^*) = \prod_{k=1}^{t} \exp\left(\frac{1}{2\sigma^2}(\phi(a_t)^\top \boldsymbol{\theta}^* - r_t)^2 - \frac{1}{2\sigma^2}(f_t - r_t)^2\right)$$
$$= \exp\left(\frac{1}{2\sigma^2}\left\|\Phi_t \boldsymbol{\theta}^* - \boldsymbol{r}_t\right\|_2^2 - \frac{1}{2\sigma^2}\left\|\boldsymbol{f}_t - \boldsymbol{r}_t\right\|_2^2\right).$$

Closed-form integration. $M_t(\boldsymbol{f}_t, \boldsymbol{\theta}^*)$ is an unnormalised Gaussian density function (with mean $\boldsymbol{r}_t$ and covariance $\sigma^2 \boldsymbol{I}$), so we can use known tricks for integrating products of Gaussian densities.

Convex constraint. $M_t(\boldsymbol{f}_t, \boldsymbol{\theta}^*)$ is the composition of $\exp(\cdot)$ and a convex function of $\boldsymbol{\theta}^*$, which means $\mathbb{E}_{\boldsymbol{f}_t \sim P_t}\left[M_t(\boldsymbol{f}_t, \boldsymbol{\theta}^*)\right] \leq 1/\delta$ is a convex constraint for $\boldsymbol{\theta}^*$.

Blowing up when $\boldsymbol{\theta} \neq \boldsymbol{\theta}^*$. If $f_1, f_2, \ldots$ predicts the rewards better than $\phi(a_1)^\top \boldsymbol{\theta}, \phi(a_2)^\top \boldsymbol{\theta}, \ldots$, then $M_t(\boldsymbol{f}_t, \boldsymbol{\theta})$ will grow exponentially with t (in expectation).

Motivation for Mixing

We would like to maximise $M_t(\boldsymbol{f}_t, \boldsymbol{\theta}^*)$ w.r.t. $\boldsymbol{f}_t$, but the maximiser is not a predictable sequence.

$$\underset{\boldsymbol{f}_t \in \mathbb{R}^t}{\operatorname{argmax}} \left\{ \exp\left(\frac{1}{2\sigma^2} \left\| \Phi_t \boldsymbol{\theta}^* - \boldsymbol{r}_t \right\|_2^2 - \frac{1}{2\sigma^2} \left\| \boldsymbol{f}_t - \boldsymbol{r}_t \right\|_2^2 \right) \right\} = \boldsymbol{r}_t.$$

Motivation for Mixing

We would like to maximise $M_t(\boldsymbol{f}_t, \boldsymbol{\theta}^*)$ w.r.t. $\boldsymbol{f}_t$, but the maximiser is not a predictable sequence.

$$\underset{\boldsymbol{f}_t \in \mathbb{R}^t}{\operatorname{argmax}} \left\{ \exp\left(\frac{1}{2\sigma^2} \|\Phi_t \boldsymbol{\theta}^* - \boldsymbol{r}_t\|_2^2 - \frac{1}{2\sigma^2} \|\boldsymbol{f}_t - \boldsymbol{r}_t\|_2^2 \right) \right\} = \boldsymbol{r}_t.$$

For a function $g(x)$ with a minimiser x^*, Laplace's asymptotic formula tells us that,

$$\int_{-\infty}^{\infty} \exp(-\lambda g(x)) \mathrm{d}x \approx \int_{-\infty}^{\infty} \exp\left(-\lambda g(x^*) - \frac{\lambda}{2} g''(x^*)(x - x^*)^2 \right) \mathrm{d}x$$

$$= \exp(-\lambda g(x^*)) \sqrt{\frac{2\pi}{g''(x^*)}} = \max_x \{\exp(-\lambda g(x))\} \sqrt{\frac{2\pi}{g''(x^*)}}$$

Motivation for Mixing

We would like to maximise $M_t(\boldsymbol{f}_t, \boldsymbol{\theta}^*)$ w.r.t. $\boldsymbol{f}_t$, but the maximiser is not a predictable sequence.

$$\underset{\boldsymbol{f}_t \in \mathbb{R}^t}{\operatorname{argmax}} \left\{ \exp\left(\frac{1}{2\sigma^2} \|\Phi_t \boldsymbol{\theta}^* - \boldsymbol{r}_t\|_2^2 - \frac{1}{2\sigma^2} \|\boldsymbol{f}_t - \boldsymbol{r}_t\|_2^2 \right) \right\} = \boldsymbol{r}_t.$$

For a function $g(x)$ with a minimiser x^*, Laplace's asymptotic formula tells us that,

$$\begin{aligned} \int_{-\infty}^{\infty} \exp(-\lambda g(x)) \mathrm{d}x &\approx \int_{-\infty}^{\infty} \exp\left(-\lambda g(x^*) - \frac{\lambda}{2} g''(x^*)(x - x^*)^2 \right) \mathrm{d}x \\ &= \exp(-\lambda g(x^*)) \sqrt{\frac{2\pi}{g''(x^*)}} = \max_x \{\exp(-\lambda g(x))\} \sqrt{\frac{2\pi}{g''(x^*)}} \end{aligned}$$

This suggests that we can perform "pseudo-maximisation" w.r.t. $\boldsymbol{f}_t$ via integration w.r.t. a (probability) measure, i.e.

$$\mathbb{E}_{\boldsymbol{f}_t \sim P_t} \left[\exp\left(\frac{1}{2\sigma^2} \|\Phi_t \boldsymbol{\theta}^* - \boldsymbol{r}_t\|_2^2 - \frac{1}{2\sigma^2} \|\boldsymbol{f}_t - \boldsymbol{r}_t\|_2^2 \right) \right] \approx \max_{\boldsymbol{f}_t} \left\{ \exp\left(\frac{1}{2\sigma^2} \|\Phi_t \boldsymbol{\theta}^* - \boldsymbol{r}_t\|_2^2 - \frac{1}{2\sigma^2} \|\boldsymbol{f}_t - \boldsymbol{r}_t\|_2^2 \right) \right\}$$

Which Mixture Distributions Are Allowed?

We can choose any sequence of mixture distributions as long as $\mathbb{E}_{\boldsymbol{f}_t \sim P_t}\left[M_t(\boldsymbol{f}_t, \boldsymbol{\theta}^*)\right]$ is a supermartingale, i.e.

$$\mathbb{E}\left[\mathbb{E}_{\boldsymbol{f}_t \sim P_t}\left[M_t(\boldsymbol{f}_t, \boldsymbol{\theta}^*)\right] | \mathcal{H}_{t-1}\right] \leq \mathbb{E}_{\boldsymbol{f}_{t-1} \sim P_{t-1}}\left[M_{t-1}(\boldsymbol{f}_{t-1}, \boldsymbol{\theta}^*)\right].$$

Which Mixture Distributions Are Allowed?

We can choose any sequence of mixture distributions as long as $\mathbb{E}_{\boldsymbol{f}_t \sim P_t}\left[M_t(\boldsymbol{f}_t, \boldsymbol{\theta}^*)\right]$ is a supermartingale, i.e.

$$\mathbb{E}\left[\mathbb{E}_{\boldsymbol{f}_t \sim P_t}\left[M_t(\boldsymbol{f}_t, \boldsymbol{\theta}^*)\right] | \mathcal{H}_{t-1}\right] \leq \mathbb{E}_{\boldsymbol{f}_{t-1} \sim P_{t-1}}\left[M_{t-1}(\boldsymbol{f}_{t-1}, \boldsymbol{\theta}^*)\right].$$

Suppose that $P_1, P_2, \ldots$ satisfies

1. P_t depends on only the history $\mathcal{H}_{t-1}$ (and not the future $r_t, a_{t+1}, r_{t+1}, \ldots$)
2. $P_t(\boldsymbol{f}_t) = p_t(f_t | \boldsymbol{f}_{t-1}) P_{t-1}(\boldsymbol{f}_{t-1})$

Which Mixture Distributions Are Allowed?

We can choose any sequence of mixture distributions as long as $\mathbb{E}_{\boldsymbol{f}_t \sim P_t}\left[M_t(\boldsymbol{f}_t, \boldsymbol{\theta}^*)\right]$ is a supermartingale, i.e.

$$\mathbb{E}\left[\mathbb{E}_{\boldsymbol{f}_t \sim P_t}\left[M_t(\boldsymbol{f}_t, \boldsymbol{\theta}^*)\right] | \mathcal{H}_{t-1}\right] \leq \mathbb{E}_{\boldsymbol{f}_{t-1} \sim P_{t-1}}\left[M_{t-1}(\boldsymbol{f}_{t-1}, \boldsymbol{\theta}^*)\right].$$

Suppose that $P_1, P_2, \ldots$ satisfies

1. P_t depends on only the history $\mathcal{H}_{t-1}$ (and not the future $r_t, a_{t+1}, r_{t+1}, \ldots$)
2. $P_t(\boldsymbol{f}_t) = p_t(f_t | \boldsymbol{f}_{t-1}) P_{t-1}(\boldsymbol{f}_{t-1})$

In this case, we have

$$\begin{aligned}
\mathbb{E}\left[\mathbb{E}_{\boldsymbol{f}_t \sim P_t}\left[M_t(\boldsymbol{f}_t, \boldsymbol{\theta}^*)\right] | \mathcal{H}_{t-1}\right] &= \mathbb{E}_{\boldsymbol{f}_t \sim P_t}\left[\mathbb{E}\left[M_t(\boldsymbol{f}_t, \boldsymbol{\theta}^*) | \mathcal{H}_{t-1}\right]\right] && \text{(1.)} \\
&\leq \mathbb{E}_{\boldsymbol{f}_t \sim P_t}\left[M_{t-1}(\boldsymbol{f}_{t-1}, \boldsymbol{\theta}^*)\right] && (M_t \text{ is a supermartingale}) \\
&= \mathbb{E}_{\boldsymbol{f}_{t-1} \sim P_{t-1}}\left[M_{t-1}(\boldsymbol{f}_{t-1}, \boldsymbol{\theta}^*)\right] && \text{(2.)}
\end{aligned}$$

Gaussian Mixture Distributions

A sequence $P_1 = \mathcal{N}(\boldsymbol{\mu}_1, \boldsymbol{T}_1), P_2 = \mathcal{N}(\boldsymbol{\mu}_2, \boldsymbol{T}_2), \ldots$ satisfies 1. and 2. if

$$\boldsymbol{\mu}_t = \left[\begin{array}{c} | \\ \boldsymbol{\mu}_{t-1} \\ | \\ \hline \mu_t \end{array}\right], \quad \boldsymbol{T}_t = \left[\begin{array}{ccc|c} & & & T_1 \\ & \boldsymbol{T}_{t-1} & & \vdots \\ & & & T_{t-1} \\ \hline T_1 & \cdots & T_{t-1} & T_t \end{array}\right],$$

Each new element μ_t and row/column $T_1, \ldots, T_t$ can depend on the history $\mathcal{H}_{t-1}$.

Gaussian Mixture Distributions

A sequence $P_1 = \mathcal{N}(\boldsymbol{\mu}_1, \boldsymbol{T}_1), P_2 = \mathcal{N}(\boldsymbol{\mu}_2, \boldsymbol{T}_2), \ldots$ satisfies 1. and 2. if

$$\boldsymbol{\mu}_t = \left[\begin{array}{c} | \\ \boldsymbol{\mu}_{t-1} \\ | \\ \hline \mu_t \end{array}\right], \quad \boldsymbol{T}_t = \left[\begin{array}{ccc|c} & & & T_1 \\ & \boldsymbol{T}_{t-1} & & \vdots \\ & & & T_{t-1} \\ \hline T_1 & \cdots & T_{t-1} & T_t \end{array}\right],$$

Each new element μ_t and row/column $T_1, \ldots, T_t$ can depend on the history $\mathcal{H}_{t-1}$.

With $P_t = \mathcal{N}(\boldsymbol{\mu}_t, \boldsymbol{T}_t)$, the martingale mixture is

$$\mathbb{E}_{\boldsymbol{f}_t \sim \mathcal{N}(\boldsymbol{\mu}_t, \boldsymbol{T}_t)} \left[M_t(\boldsymbol{f}_t, \boldsymbol{\theta}^*) \right] = \frac{1}{\sqrt{\det(\boldsymbol{I} + \boldsymbol{T}_t / \sigma^2)}} \exp\left(\frac{1}{2\sigma^2} \|\Phi_t \boldsymbol{\theta}^* - \boldsymbol{r}_t\|_2^2 - \frac{1}{2\sigma^2} \|\boldsymbol{\mu}_t - \boldsymbol{r}_t\|^2_{(\boldsymbol{I} + \boldsymbol{T}_t/\sigma^2)^{-1}} \right).$$

Martingale Mixture Tail Bound

The constraint $\mathbb{E}_{\boldsymbol{f}_t \sim \mathcal{N}(\boldsymbol{\mu}_t, \boldsymbol{T}_t)}\left[M_t(\boldsymbol{f}_t, \boldsymbol{\theta}^*)\right] \leq 1/\delta$ can be rearranged into

$$\|\Phi_t \boldsymbol{\theta}^* - \boldsymbol{r}_t\|_2^2 \leq (\boldsymbol{\mu}_t - \boldsymbol{r}_t)^\top \left(\boldsymbol{I} + \frac{\boldsymbol{T}_t}{\sigma^2}\right)^{-1} (\boldsymbol{\mu}_t - \boldsymbol{r}_t) + \sigma^2 \ln\left(\det\left(\boldsymbol{I} + \frac{\boldsymbol{T}_t}{\sigma^2}\right)\right) + 2\sigma^2 \ln(1/\delta).$$

Martingale Mixture Tail Bound

The constraint $\mathbb{E}_{\boldsymbol{f}_t \sim \mathcal{N}(\boldsymbol{\mu}_t, \boldsymbol{T}_t)} [M_t(\boldsymbol{f}_t, \boldsymbol{\theta}^*)] \leq 1/\delta$ can be rearranged into

$$\|\Phi_t \boldsymbol{\theta}^* - \boldsymbol{r}_t\|_2^2 \leq (\boldsymbol{\mu}_t - \boldsymbol{r}_t)^\top \left(\boldsymbol{I} + \frac{\boldsymbol{T}_t}{\sigma^2}\right)^{-1} (\boldsymbol{\mu}_t - \boldsymbol{r}_t) + \sigma^2 \ln\left(\det\left(\boldsymbol{I} + \frac{\boldsymbol{T}_t}{\sigma^2}\right)\right) + 2\sigma^2 \ln(1/\delta).$$

Standard mixture distributions: $P_t = \mathcal{N}(\mathbf{0}, c\Phi_t \Phi_t^\top)$

$$\|\Phi_t \boldsymbol{\theta}^* - \boldsymbol{r}_t\|_2^2 \leq \boldsymbol{r}_t^\top \left(\boldsymbol{I} + \frac{c\Phi_t \Phi_t^\top}{\sigma^2}\right)^{-1} \boldsymbol{r}_t + \sigma^2 \ln\left(\det\left(\boldsymbol{I} + \frac{c\Phi_t \Phi_t^\top}{\sigma^2}\right)\right) + 2\sigma^2 \ln(1/\delta) =: R^2_{\mathrm{MM},t}.$$

Martingale Mixture Tail Bound

The constraint $\mathbb{E}_{\boldsymbol{f}_t \sim \mathcal{N}(\boldsymbol{\mu}_t, \boldsymbol{T}_t)}[M_t(\boldsymbol{f}_t, \boldsymbol{\theta}^*)] \leq 1/\delta$ can be rearranged into

$$\|\Phi_t \boldsymbol{\theta}^* - \boldsymbol{r}_t\|_2^2 \leq (\boldsymbol{\mu}_t - \boldsymbol{r}_t)^\top \left(\boldsymbol{I} + \frac{\boldsymbol{T}_t}{\sigma^2}\right)^{-1} (\boldsymbol{\mu}_t - \boldsymbol{r}_t) + \sigma^2 \ln\left(\det\left(\boldsymbol{I} + \frac{\boldsymbol{T}_t}{\sigma^2}\right)\right) + 2\sigma^2 \ln(1/\delta).$$

Standard mixture distributions: $P_t = \mathcal{N}(\mathbf{0}, c\Phi_t\Phi_t^\top)$

$$\|\Phi_t \boldsymbol{\theta}^* - \boldsymbol{r}_t\|_2^2 \leq \boldsymbol{r}_t^\top \left(\boldsymbol{I} + \frac{c\Phi_t\Phi_t^\top}{\sigma^2}\right)^{-1} \boldsymbol{r}_t + \sigma^2 \ln\left(\det\left(\boldsymbol{I} + \frac{c\Phi_t\Phi_t^\top}{\sigma^2}\right)\right) + 2\sigma^2 \ln(1/\delta) =: R_{\mathrm{MM},t}^2.$$

On the one hand...

- $\mathcal{N}(\mathbf{0}, c\Phi_t\Phi_t^\top)$ is good enough to give us tighter confidence sets/bounds
- $c\Phi_t\Phi_t^\top$ is rank d, so $R_{\mathrm{MM},t}^2$ can be computed relatively cheaply

Martingale Mixture Tail Bound

The constraint $\mathbb{E}_{\boldsymbol{f}_t \sim \mathcal{N}(\boldsymbol{\mu}_t, \boldsymbol{T}_t)} [M_t(\boldsymbol{f}_t, \boldsymbol{\theta}^*)] \leq 1/\delta$ can be rearranged into

$$\|\Phi_t \boldsymbol{\theta}^* - \boldsymbol{r}_t\|_2^2 \leq (\boldsymbol{\mu}_t - \boldsymbol{r}_t)^\top \left(\boldsymbol{I} + \frac{\boldsymbol{T}_t}{\sigma^2} \right)^{-1} (\boldsymbol{\mu}_t - \boldsymbol{r}_t) + \sigma^2 \ln \left(\det \left(\boldsymbol{I} + \frac{\boldsymbol{T}_t}{\sigma^2} \right) \right) + 2\sigma^2 \ln(1/\delta).$$

Standard mixture distributions: $P_t = \mathcal{N}(\mathbf{0}, c\Phi_t \Phi_t^\top)$

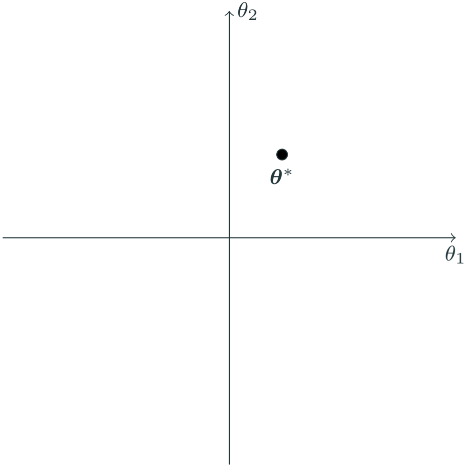
$$\|\Phi_t \boldsymbol{\theta}^* - \boldsymbol{r}_t\|_2^2 \leq \boldsymbol{r}_t^\top \left(\boldsymbol{I} + \frac{c\Phi_t \Phi_t^\top}{\sigma^2} \right)^{-1} \boldsymbol{r}_t + \sigma^2 \ln \left(\det \left(\boldsymbol{I} + \frac{c\Phi_t \Phi_t^\top}{\sigma^2} \right) \right) + 2\sigma^2 \ln(1/\delta) =: R_{\mathrm{MM},t}^2.$$

On the one hand...

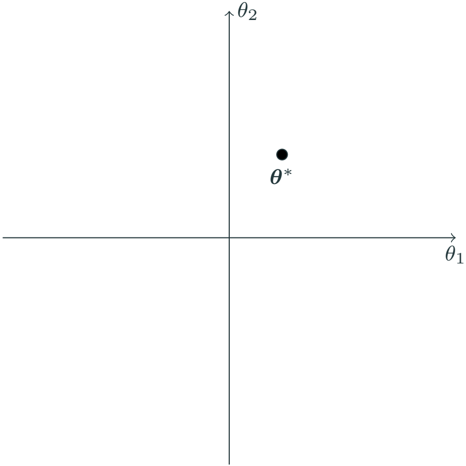
- $\mathcal{N}(\mathbf{0}, c\Phi_t \Phi_t^\top)$ is good enough to give us tighter confidence sets/bounds
- $c\Phi_t \Phi_t^\top$ is rank d, so $R_{\mathrm{MM},t}^2$ can be computed relatively cheaply

On the other hand, $\boldsymbol{\mu}_t = \mathbf{0}$ seems bit a silly.

Confidence Sets For Linear Bandits



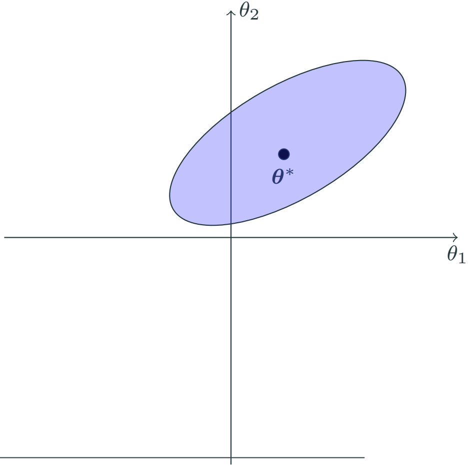
Confidence Sets For Linear Bandits



Using our martingale tail bound, we have

$$\|\Phi_t \boldsymbol{\theta}^* - \boldsymbol{r}_t\|_2 \leq R_{\mathrm{MM},t},$$

Confidence Sets For Linear Bandits



Using our martingale tail bound, we have

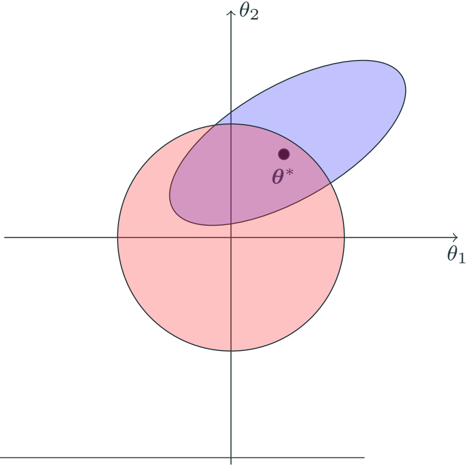
$$\|\Phi_t \boldsymbol{\theta}^* - \boldsymbol{r}_t\|_2 \leq R_{\mathrm{MM},t},$$

This means that $\boldsymbol{\theta}^*$ lies within the set*

$$\{\boldsymbol{\theta} \in \mathbb{R}^d : \|\Phi_t \boldsymbol{\theta} - \boldsymbol{r}_t\|_2 \leq R_{\mathrm{MM},t}\}.$$

* this set can be re-written as $\{\boldsymbol{\theta} \in \mathbb{R}^d : \|\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_t\|_{\Phi_t^\top \Phi_t} \leq \tilde{R}_t\}$, where $\widehat{\boldsymbol{\theta}}_t = \Phi_t^\dagger \boldsymbol{r}_t$ and $\tilde{R}_t$ is some other radius quantity

Confidence Sets For Linear Bandits



Using our martingale tail bound, we have

$$\|\Phi_t \boldsymbol{\theta}^* - \boldsymbol{r}_t\|_2 \leq R_{\mathrm{MM},t},$$

This means that $\boldsymbol{\theta}^*$ lies within the set*

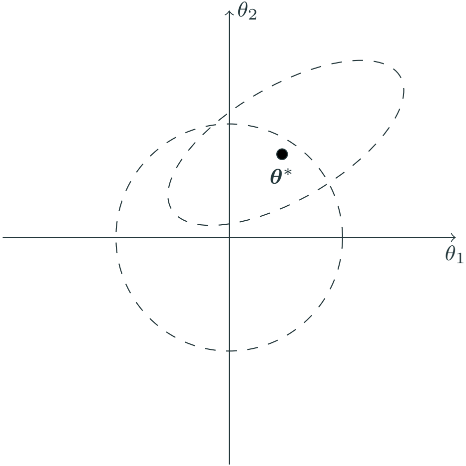
$$\{\boldsymbol{\theta} \in \mathbb{R}^d : \|\Phi_t \boldsymbol{\theta} - \boldsymbol{r}_t\|_2 \leq R_{\mathrm{MM},t}\}.$$

Incorporating the smoothness assumption, we obtain

$$\Theta_t = \{\boldsymbol{\theta} \in \mathbb{R}^d : \|\Phi_t \boldsymbol{\theta} - \boldsymbol{r}_t\|_2 \leq R_{\mathrm{MM},t}, \|\boldsymbol{\theta}\|_2 \leq B\}.$$

* this set can be re-written as $\{\boldsymbol{\theta} \in \mathbb{R}^d : \|\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_t\|_{\Phi_t^\top \Phi_t} \leq \tilde{R}_t\}$, where $\widehat{\boldsymbol{\theta}}_t = \Phi_t^\dagger \boldsymbol{r}_t$ and $\tilde{R}_t$ is some other radius quantity

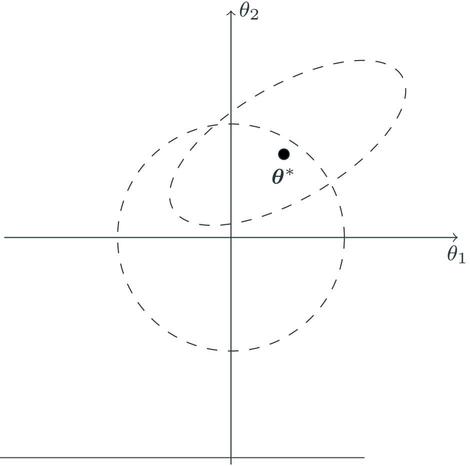
Confidence Sets For Linear Bandits (Single Ellipsoid)



By taking a weighted (by $\alpha > 0$) sum, we obtain a single quadratic constraint for $\boldsymbol{\theta}^*$

$$\|\Phi_t \boldsymbol{\theta}^* - \boldsymbol{r}_t\|_2^2 + \alpha \|\boldsymbol{\theta}^*\|_2^2 \leq R_{\mathrm{MM},t}^2 + \alpha B^2,$$

Confidence Sets For Linear Bandits (Single Ellipsoid)



By taking a weighted (by $\alpha > 0$) sum, we obtain a single quadratic constraint for $\boldsymbol{\theta}^*$

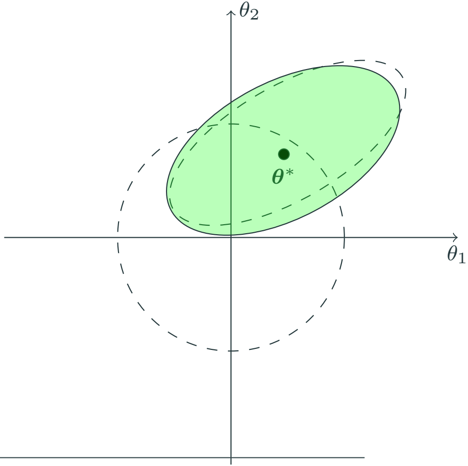
$$\|\Phi_t \boldsymbol{\theta}^* - \boldsymbol{r}_t\|_2^2 + \alpha \|\boldsymbol{\theta}^*\|_2^2 \leq R_{\mathrm{MM},t}^2 + \alpha B^2,$$

By completing the square on the LHS, this constraint can be re-written as*

$$\|\boldsymbol{\theta}^* - \widehat{\boldsymbol{\theta}}_{\alpha,t}\|_{(\Phi_t^\top \Phi_t + \alpha \boldsymbol{I})} \leq R_{\mathrm{AMM},t}$$

* where $\widehat{\boldsymbol{\theta}}_{\alpha,t} = (\Phi_t^\top \Phi_t + \alpha \boldsymbol{I})^{-1} \Phi_t^\top \boldsymbol{r}_t$, $\quad R_{\mathrm{AMM,t}}^2 = R_{\mathrm{MM,t}}^2 + \alpha B^2 - \boldsymbol{r}_t^\top \boldsymbol{r}_t + \boldsymbol{r}_t^\top \Phi_t \left(\Phi_t^\top \Phi_t + \alpha \boldsymbol{I}\right)^{-1} \Phi_t^\top \boldsymbol{r}_t$

Confidence Sets For Linear Bandits (Single Ellipsoid)



By taking a weighted (by $\alpha > 0$) sum, we obtain a single quadratic constraint for $\boldsymbol{\theta}^*$

$$\|\Phi_t \boldsymbol{\theta}^* - \boldsymbol{r}_t\|_2^2 + \alpha \|\boldsymbol{\theta}^*\|_2^2 \leq R_{\mathrm{MM},t}^2 + \alpha B^2,$$

By completing the square on the LHS, this constraint can be re-written as*

$$\|\boldsymbol{\theta}^* - \widehat{\boldsymbol{\theta}}_{\alpha,t}\|_{(\Phi_t^\top \Phi_t + \alpha \boldsymbol{I})} \leq R_{\mathrm{AMM},t}$$

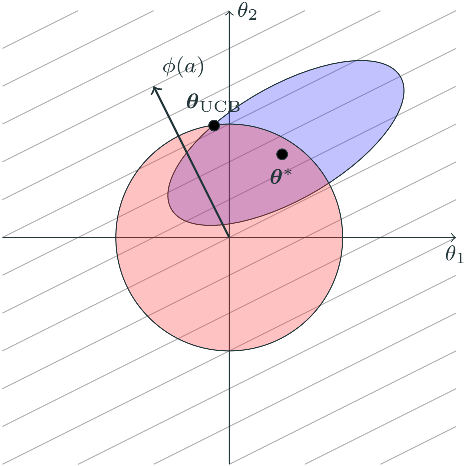
This means that $\boldsymbol{\theta}^*$ lies within the ellipsoid

$$\Theta_t^\alpha = \{\boldsymbol{\theta} \in \mathbb{R}^d : \|\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_{\alpha,t}\|_{(\Phi_t^\top \Phi_t + \alpha \boldsymbol{I})} \leq R_{\mathrm{AMM},t}\}.$$

* where $\widehat{\boldsymbol{\theta}}_{\alpha,t} = (\Phi_t^\top \Phi_t + \alpha \boldsymbol{I})^{-1} \Phi_t^\top \boldsymbol{r}_t,\qquad R_{\mathrm{AMM,t}}^2 = R_{\mathrm{MM,t}}^2 + \alpha B^2 - \boldsymbol{r}_t^\top \boldsymbol{r}_t + \boldsymbol{r}_t^\top \Phi_t \left(\Phi_t^\top \Phi_t + \alpha \boldsymbol{I}\right)^{-1} \Phi_t^\top \boldsymbol{r}_t$

Computing and Maximising Confidence Bounds

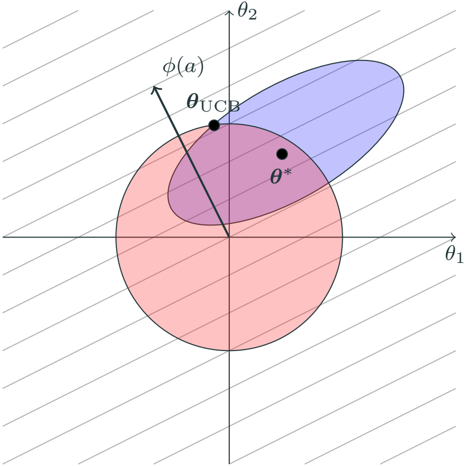
Convex Martingale Mixture UCB



The UCB for our double ellipsoid confidence set is

$$\begin{aligned} \mathrm{UCB}_{\Theta_t}(a) &= \max_{\boldsymbol{\theta} \in \mathbb{R}^d} \ \phi(a)^\top \boldsymbol{\theta} \\ &\quad\ \text{s.t. } \ \|\Phi_t \boldsymbol{\theta} - \boldsymbol{r}_t\|_2 \leq R_{\mathrm{MM},t} \\ &\quad\ \text{and } \ \|\boldsymbol{\theta}\|_2 \leq B \\ &= \phi(a)^\top \boldsymbol{\theta}_{\mathrm{UCB}}. \end{aligned}$$

Convex Martingale Mixture UCB

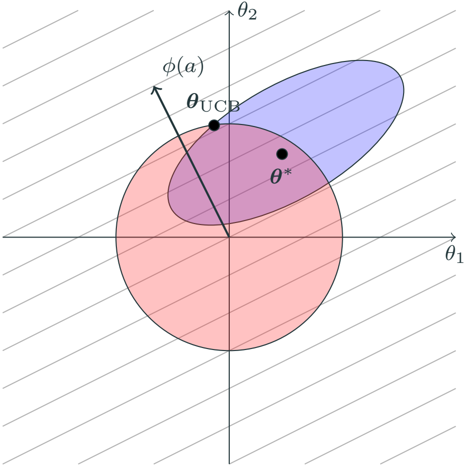


The UCB for our double ellipsoid confidence set is

$$\begin{aligned}\mathrm{UCB}_{\Theta_t}(a) &= \max_{\boldsymbol{\theta}\in\mathbb{R}^d} \ \phi(a)^\top \boldsymbol{\theta} \\ &\quad\ \text{s.t. } \ \|\Phi_t \boldsymbol{\theta} - \boldsymbol{r}_t\|_2 \le R_{\mathrm{MM},t} \\ &\quad\ \text{and } \ \|\boldsymbol{\theta}\|_2 \le B \\ &= \phi(a)^\top \boldsymbol{\theta}_{\mathrm{UCB}}.\end{aligned}$$

$\mathrm{UCB}_{\Theta_t}(a)$ can be computed in $\mathcal{O}(d^3)$ time complexity via interior point methods.

Convex Martingale Mixture UCB



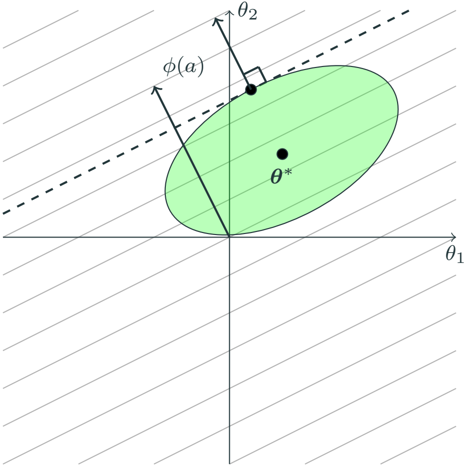
The UCB for our double ellipsoid confidence set is

$$\begin{aligned}\mathrm{UCB}_{\Theta_t}(a) &= \max_{\boldsymbol{\theta}\in\mathbb{R}^d} \ \phi(a)^\top\boldsymbol{\theta} \\ &\quad\ \text{s.t.}\ \ \|\Phi_t\boldsymbol{\theta} - \boldsymbol{r}_t\|_2 \le R_{\mathrm{MM},t} \\ &\quad\ \text{and}\ \ \|\boldsymbol{\theta}\|_2 \le B \\ &= \phi(a)^\top\boldsymbol{\theta}_{\mathrm{UCB}}.\end{aligned}$$

$\mathrm{UCB}_{\Theta_t}(a)$ can be computed in $\mathcal{O}(d^3)$ time complexity via interior point methods.

We call LinUCB with these confidence sets/bounds Convex Martingale Mixture (CMM-)UCB.

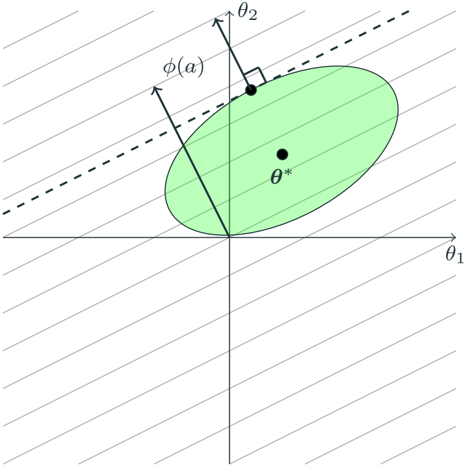
Analytic Martingale Mixture UCB



The UCB for our single ellipsoid confidence set is

$$\begin{aligned} \mathrm{UCB}_{\Theta_t^\alpha}(a) \;=\; & \max_{\boldsymbol{\theta} \in \mathbb{R}^d} \;\; \phi(a)^\top \boldsymbol{\theta} \\ & \text{s.t. } \|\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_{\alpha,t}\|^2_{(\Phi_t^\top \Phi_t + \alpha \boldsymbol{I})} \leq R^2_{\mathrm{AMM},t}. \end{aligned}$$

Analytic Martingale Mixture UCB



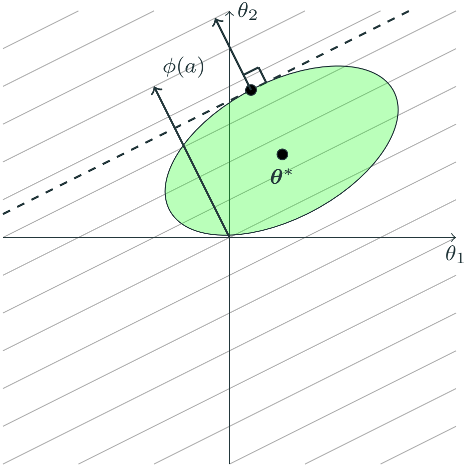
The UCB for our single ellipsoid confidence set is

$$\mathrm{UCB}_{\Theta_t^\alpha}(a) = \max_{\boldsymbol{\theta}\in\mathbb{R}^d} \phi(a)^\top \boldsymbol{\theta}$$
$$\text{s.t. } \|\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_{\alpha,t}\|^2_{(\Phi_t^\top \Phi_t + \alpha \boldsymbol{I})} \leq R^2_{\mathrm{AMM},t}.$$

This time, there is a closed-form solution.

$$\mathrm{UCB}_{\Theta_t^\alpha}(a) = \phi(a)^\top \widehat{\boldsymbol{\theta}}_{\alpha,t} + R_{\mathrm{AMM},t}\|\phi(a)\|_{(\Phi_t^\top \Phi_t + \alpha \boldsymbol{I})^{-1}}.$$

Analytic Martingale Mixture UCB



The UCB for our single ellipsoid confidence set is

$$\mathrm{UCB}_{\Theta_t^\alpha}(a) = \max_{\boldsymbol{\theta} \in \mathbb{R}^d} \phi(a)^\top \boldsymbol{\theta}$$
$$\text{s.t. } \|\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_{\alpha,t}\|^2_{(\Phi_t^\top \Phi_t + \alpha \boldsymbol{I})} \leq R^2_{\mathrm{AMM},t}.$$

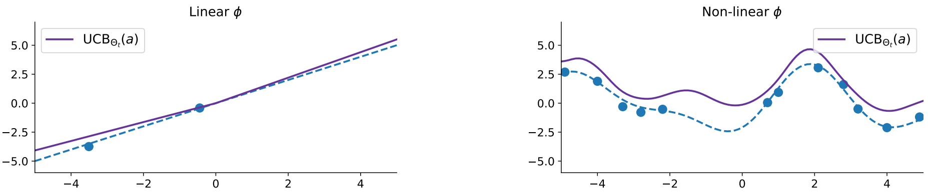
This time, there is a closed-form solution.

$$\mathrm{UCB}_{\Theta_t^\alpha}(a) = \phi(a)^\top \widehat{\boldsymbol{\theta}}_{\alpha,t} + R_{\mathrm{AMM},t} \|\phi(a)\|_{(\Phi_t^\top \Phi_t + \alpha \boldsymbol{I})^{-1}}.$$

We call LinUCB with these confidence sets/bounds Analytic Martingale Mixture (AMM-)UCB.

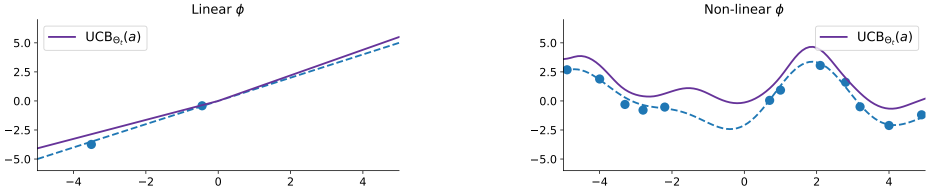
Confidence Bound Maximisation

To run LinUCB with our confidence sets, we need to maximise $\mathrm{UCB}_{\Theta_t}(a) = \max_{\boldsymbol{\theta} \in \Theta_t} \{\phi(a)^\top \boldsymbol{\theta}\}$ w.r.t. a



Confidence Bound Maximisation

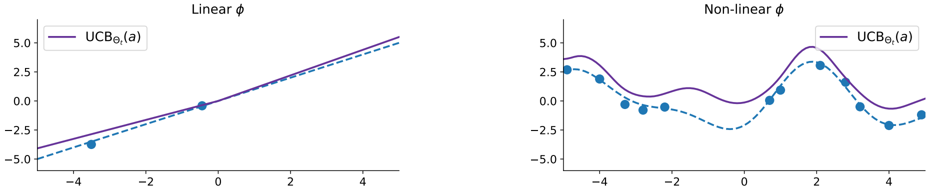
To run LinUCB with our confidence sets, we need to maximise $\mathrm{UCB}_{\Theta_t}(a) = \max_{\boldsymbol{\theta} \in \Theta_t} \{\phi(a)^\top \boldsymbol{\theta}\}$ w.r.t. a



For continuous action sets, we approximately maximise $\mathrm{UCB}_{\Theta_t}(a)$ w.r.t a using gradient-based methods.

Confidence Bound Maximisation

To run LinUCB with our confidence sets, we need to maximise $\mathrm{UCB}_{\Theta_t}(a) = \max_{\boldsymbol{\theta} \in \Theta_t} \{\phi(a)^\top \boldsymbol{\theta}\}$ w.r.t. a



For continuous action sets, we approximately maximise $\mathrm{UCB}_{\Theta_t}(a)$ w.r.t a using gradient-based methods.

For CMM-UCB, $\mathrm{UCB}_{\Theta_t}(a)$ and $\nabla_a \mathrm{UCB}_{\Theta_t}(a)$ can be computed numerically using differentiable convex optimisation (surprisingly easy with cvxpylayers[2]).

[2] A. Agrawal et al. (2019) Differentiable convex optimization layers. NeurIPS

Regret Bounds

Bounding the Radius

The full expression for the (squared) AMM radius is

$$R^2_{\mathrm{AMM,t}} = \boldsymbol{r}_t \left(\boldsymbol{I} + \frac{c\Phi_t\Phi_t^\top}{\sigma^2} \right)^{-1} \boldsymbol{r}_t + \sigma^2 \ln\left(\det\left(\boldsymbol{I} + \frac{c\Phi_t\Phi_t^\top}{\sigma^2} \right) \right) + 2\sigma^2 \ln(1/\delta)$$
$$+ \alpha B^2 - \boldsymbol{r}_t^\top \boldsymbol{r}_t + \boldsymbol{r}_t^\top \Phi_t \left(\Phi_t^\top \Phi_t + \alpha \boldsymbol{I} \right)^{-1} \Phi_t^\top \boldsymbol{r}_t .$$

Bounding the Radius

The full expression for the (squared) AMM radius is

$$R^2_{\mathrm{AMM,t}} = \boldsymbol{r}_t \left(\boldsymbol{I} + \frac{c\Phi_t\Phi_t^\top}{\sigma^2}\right)^{-1} \boldsymbol{r}_t + \sigma^2 \ln\left(\det\left(\boldsymbol{I} + \frac{c\Phi_t\Phi_t^\top}{\sigma^2}\right)\right) + 2\sigma^2 \ln(1/\delta)$$
$$+ \alpha B^2 - \boldsymbol{r}_t^\top \boldsymbol{r}_t + \boldsymbol{r}_t^\top \Phi_t \left(\Phi_t^\top \Phi_t + \alpha \boldsymbol{I}\right)^{-1} \Phi_t^\top \boldsymbol{r}_t.$$

Using the Matrix Inversion Lemma, we have

$$\boldsymbol{r}_t \left(\boldsymbol{I} + \frac{c\Phi_t\Phi_t^\top}{\sigma^2}\right)^{-1} \boldsymbol{r}_t = \boldsymbol{r}_t^\top \boldsymbol{r}_t - \boldsymbol{r}_t^\top \Phi_t \left(\Phi_t^\top \Phi_t + \frac{\sigma^2}{c}\boldsymbol{I}\right)^{-1} \Phi_t^\top \boldsymbol{r}_t.$$

Bounding the Radius

The full expression for the (squared) AMM radius is

$$R^2_{\mathrm{AMM,t}} = \boldsymbol{r}_t \left(\boldsymbol{I} + \frac{c\Phi_t\Phi_t^\top}{\sigma^2} \right)^{-1} \boldsymbol{r}_t + \sigma^2 \ln \left(\det \left(\boldsymbol{I} + \frac{c\Phi_t\Phi_t^\top}{\sigma^2} \right) \right) + 2\sigma^2 \ln(1/\delta)$$
$$+ \alpha B^2 - \boldsymbol{r}_t^\top \boldsymbol{r}_t + \boldsymbol{r}_t^\top \Phi_t \left(\Phi_t^\top \Phi_t + \alpha \boldsymbol{I} \right)^{-1} \Phi_t^\top \boldsymbol{r}_t.$$

Using the Matrix Inversion Lemma, we have

$$\boldsymbol{r}_t \left(\boldsymbol{I} + \frac{c\Phi_t\Phi_t^\top}{\sigma^2} \right)^{-1} \boldsymbol{r}_t = \boldsymbol{r}_t^\top \boldsymbol{r}_t - \boldsymbol{r}_t^\top \Phi_t \left(\Phi_t^\top \Phi_t + \frac{\sigma^2}{c} \boldsymbol{I} \right)^{-1} \Phi_t^\top \boldsymbol{r}_t.$$

If we set $\alpha = \sigma^2/c$, the quadratic terms cancel, and we can use the determinant-trace inequality*

$$R^2_{\mathrm{AMM,t}} = \sigma^2 \left(\ln \left(\det \left(\boldsymbol{I} + \frac{c\Phi_t\Phi_t^\top}{\sigma^2} \right) \right) + \frac{B^2}{c} + 2\ln(1/\delta) \right) \le \sigma^2 \left(d \ln \left(1 + \frac{ctL^2}{\sigma^2 d} \right) + \frac{B^2}{c} + 2\ln(1/\delta) \right)$$

* Assuming $\|\phi(a)\|_2 \le L$.

OFUL Analysis

Step 1. Use optimism to bound the cumulative regret by the confidence bound widths.

$$\sum_{t=1}^{T} \phi(a^*)^\top \boldsymbol{\theta}^* - \phi(a_t)^\top \boldsymbol{\theta}^* \leq \sum_{t=1}^{T} \mathrm{UCB}_{\Theta_{t-1}}(a^*) - \mathrm{LCB}_{\Theta_{t-1}}(a_t) \leq \sum_{t=1}^{T} \mathrm{UCB}_{\Theta_{t-1}}(a_t) - \mathrm{LCB}_{\Theta_{t-1}}(a_t).$$

OFUL Analysis

Step 1. Use optimism to bound the cumulative regret by the confidence bound widths.

$$\sum_{t=1}^{T} \phi(a^*)^\top \boldsymbol{\theta}^* - \phi(a_t)^\top \boldsymbol{\theta}^* \leq \sum_{t=1}^{T} \mathrm{UCB}_{\Theta_{t-1}}(a^*) - \mathrm{LCB}_{\Theta_{t-1}}(a_t) \leq \sum_{t=1}^{T} \mathrm{UCB}_{\Theta_{t-1}}(a_t) - \mathrm{LCB}_{\Theta_{t-1}}(a_t).$$

Step 2. For both CMM-UCB and AMM-UCB, we have

$$\sum_{t=1}^{T} \mathrm{UCB}_{\Theta_{t-1}}(a_t) - \mathrm{LCB}_{\Theta_{t-1}}(a_t) \leq \sum_{t=1}^{T} 2R_{\mathrm{AMM},t-1} \|\phi(a_t)\|_{(\Phi_{t-1}^\top \Phi_{t-1} + \alpha \boldsymbol{I})^{-1}}.$$

OFUL Analysis

Step 1. Use optimism to bound the cumulative regret by the confidence bound widths.

$$\sum_{t=1}^{T} \phi(a^*)^\top \boldsymbol{\theta}^* - \phi(a_t)^\top \boldsymbol{\theta}^* \le \sum_{t=1}^{T} \mathrm{UCB}_{\Theta_{t-1}}(a^*) - \mathrm{LCB}_{\Theta_{t-1}}(a_t) \le \sum_{t=1}^{T} \mathrm{UCB}_{\Theta_{t-1}}(a_t) - \mathrm{LCB}_{\Theta_{t-1}}(a_t).$$

Step 2. For both CMM-UCB and AMM-UCB, we have

$$\sum_{t=1}^{T} \mathrm{UCB}_{\Theta_{t-1}}(a_t) - \mathrm{LCB}_{\Theta_{t-1}}(a_t) \le \sum_{t=1}^{T} 2R_{\mathrm{AMM},t-1} \|\phi(a_t)\|_{(\Phi_{t-1}^\top \Phi_{t-1} + \alpha \boldsymbol{I})^{-1}}.$$

Step 3. Separately upper bound $R_{\mathrm{AMM},T-1}$ and $\sum_{t=1}^{T} \|\phi(a_t)\|_{(\Phi_{t-1}^\top \Phi_{t-1} + \alpha \boldsymbol{I})^{-1}}$, to obtain

$$\sum_{t=1}^{T} \phi(a^*)^\top \boldsymbol{\theta}^* - \phi(a_t)^\top \boldsymbol{\theta}^* \le \mathcal{O}(d\sqrt{T} \ln(T)).$$

Comparison With OFUL

How Does AMM-UCB Compare To OFUL?

We derive and use a bound on the norm of the noise vector

$$\|\boldsymbol{\epsilon}_t\|_2 = \|\Phi_t \boldsymbol{\theta}^* - \boldsymbol{r}_t\|_2 \leq R_{\mathrm{MM},t}.$$

How Does AMM-UCB Compare To OFUL?

We derive and use a bound on the norm of the noise vector

$$\|\boldsymbol{\epsilon}_t\|_2 = \|\Phi_t \boldsymbol{\theta}^* - \boldsymbol{r}_t\|_2 \leq R_{\mathrm{MM},t}.$$

For $c = \sigma^2/\alpha$ and any $\alpha > 0$, this leads to the inequality

$$\|\boldsymbol{\theta}^* - \widehat{\boldsymbol{\theta}}_{\alpha,t}\|_{(\Phi_t^\top \Phi_t + \alpha \boldsymbol{I})} \leq R_{\mathrm{AMM},t} = \sqrt{\sigma^2 \ln\left(\det\left(\frac{1}{\alpha}\Phi_t^\top \Phi_t + \boldsymbol{I}\right)\right) + \alpha B^2 + 2\sigma^2 \ln(1/\delta)}.$$

How Does AMM-UCB Compare To OFUL?

We derive and use a bound on the norm of the noise vector

$$\|\boldsymbol{\epsilon}_t\|_2 = \|\Phi_t \boldsymbol{\theta}^* - \boldsymbol{r}_t\|_2 \leq R_{\mathrm{MM},t}.$$

For $c = \sigma^2/\alpha$ and any $\alpha > 0$, this leads to the inequality

$$\|\boldsymbol{\theta}^* - \widehat{\boldsymbol{\theta}}_{\alpha,t}\|_{(\Phi_t^\top \Phi_t + \alpha \boldsymbol{I})} \leq R_{\mathrm{AMM},t} = \sqrt{\sigma^2 \ln\left(\det\left(\frac{1}{\alpha}\Phi_t^\top \Phi_t + \boldsymbol{I}\right)\right) + \alpha B^2 + 2\sigma^2 \ln(1/\delta)}.$$

OFUL uses a bound on the (weighted) norm of the projection of the noise vector

$$\|\Phi_t^\top \boldsymbol{\epsilon}_t\|_{(\Phi_t^\top \Phi_t + \alpha \boldsymbol{I})^{-1}} \leq \sigma \sqrt{\ln\left(\det\left(\frac{1}{\alpha}\Phi_t^\top \Phi_t + \boldsymbol{I}\right)\right) + 2\ln(1/\delta)}.$$

How Does AMM-UCB Compare To OFUL?

We derive and use a bound on the norm of the noise vector

$$\|\boldsymbol{\epsilon}_t\|_2 = \|\Phi_t \boldsymbol{\theta}^* - \boldsymbol{r}_t\|_2 \leq R_{\mathrm{MM},t}.$$

For $c = \sigma^2/\alpha$ and any $\alpha > 0$, this leads to the inequality

$$\|\boldsymbol{\theta}^* - \widehat{\boldsymbol{\theta}}_{\alpha,t}\|_{(\Phi_t^\top \Phi_t + \alpha \boldsymbol{I})} \leq R_{\mathrm{AMM},t} = \sqrt{\sigma^2 \ln\left(\det\left(\frac{1}{\alpha}\Phi_t^\top \Phi_t + \boldsymbol{I}\right)\right) + \alpha B^2 + 2\sigma^2 \ln(1/\delta)}.$$

OFUL uses a bound on the (weighted) norm of the projection of the noise vector

$$\|\Phi_t^\top \boldsymbol{\epsilon}_t\|_{(\Phi_t^\top \Phi_t + \alpha \boldsymbol{I})^{-1}} \leq \sigma \sqrt{\ln\left(\det\left(\frac{1}{\alpha}\Phi_t^\top \Phi_t + \boldsymbol{I}\right)\right) + 2\ln(1/\delta)}.$$

This leads to a similar, but looser (due to $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$) inequality

$$\|\boldsymbol{\theta}^* - \widehat{\boldsymbol{\theta}}_{\alpha,t}\|_{(\Phi_t^\top \Phi_t + \alpha \boldsymbol{I})} \leq \sigma \sqrt{\ln\left(\det\left(\frac{1}{\alpha}\Phi_t^\top \Phi_t + \boldsymbol{I}\right)\right) + 2\ln(1/\delta)} + \sqrt{\alpha} B.$$

Why Is AMM-UCB Better Than OFUL?

Bounds on $\|\Phi_t \boldsymbol{\theta}^* - \boldsymbol{r}_t\|_2$ and $\|\boldsymbol{\theta}^*\|_2$ fit together better than bounds on $\|\Phi_t^\top \boldsymbol{\epsilon}_t\|_{(\Phi_t^\top \Phi_t + \alpha \boldsymbol{I})^{-1}}$ and $\|\boldsymbol{\theta}^*\|_2$.

Why Is AMM-UCB Better Than OFUL?

Bounds on $\|\Phi_t\boldsymbol{\theta}^* - \boldsymbol{r}_t\|_2$ and $\|\boldsymbol{\theta}^*\|_2$ fit together better than bounds on $\|\Phi_t^\top \boldsymbol{\epsilon}_t\|_{(\Phi_t^\top\Phi_t+\alpha\boldsymbol{I})^{-1}}$ and $\|\boldsymbol{\theta}^*\|_2$.

OFUL: Using the definition of $\widehat{\boldsymbol{\theta}}_{\alpha,t}$, and then the triangle inequality,

$$\begin{aligned}
\|\boldsymbol{\theta}^* - \widehat{\boldsymbol{\theta}}_{\alpha,t}\|_{(\Phi_t^\top\Phi_t+\alpha\boldsymbol{I})} &= \|\Phi_t^\top\boldsymbol{\epsilon}_t + \alpha\boldsymbol{\theta}^*\|_{(\Phi_t^\top\Phi_t+\alpha\boldsymbol{I})^{-1}} \\
&\leq \|\Phi_t^\top\boldsymbol{\epsilon}_t\|_{(\Phi_t^\top\Phi_t+\alpha\boldsymbol{I})^{-1}} + \alpha\|\boldsymbol{\theta}^*\|_{(\Phi_t^\top\Phi_t+\alpha\boldsymbol{I})^{-1}} \\
&\leq \sigma\sqrt{\ln\left(\det\left(\frac{1}{\alpha}\Phi_t^\top\Phi_t + \boldsymbol{I}\right)\right) + 2\ln(1/\delta)} + \sqrt{\alpha}B
\end{aligned}$$

Why Is AMM-UCB Better Than OFUL?

Bounds on $\|\Phi_t \boldsymbol{\theta}^* - \boldsymbol{r}_t\|_2$ and $\|\boldsymbol{\theta}^*\|_2$ fit together better than bounds on $\|\Phi_t^\top \boldsymbol{\epsilon}_t\|_{(\Phi_t^\top \Phi_t + \alpha \boldsymbol{I})^{-1}}$ and $\|\boldsymbol{\theta}^*\|_2$.

OFUL: Using the definition of $\widehat{\boldsymbol{\theta}}_{\alpha,t}$, and then the triangle inequality,

$$\begin{aligned}
\|\boldsymbol{\theta}^* - \widehat{\boldsymbol{\theta}}_{\alpha,t}\|_{(\Phi_t^\top \Phi_t + \alpha \boldsymbol{I})} &= \|\Phi_t^\top \boldsymbol{\epsilon}_t + \alpha \boldsymbol{\theta}^*\|_{(\Phi_t^\top \Phi_t + \alpha \boldsymbol{I})^{-1}} \\
&\leq \|\Phi_t^\top \boldsymbol{\epsilon}_t\|_{(\Phi_t^\top \Phi_t + \alpha \boldsymbol{I})^{-1}} + \alpha \|\boldsymbol{\theta}^*\|_{(\Phi_t^\top \Phi_t + \alpha \boldsymbol{I})^{-1}} \\
&\leq \sigma \sqrt{\ln\left(\det\left(\frac{1}{\alpha} \Phi_t^\top \Phi_t + \boldsymbol{I}\right)\right) + 2\ln(1/\delta)} + \sqrt{\alpha} B
\end{aligned}$$

The triangle inquality step causes the $\ln \det$ and αB^2 terms to appear under separate square roots.

Why Is AMM-UCB Better Than OFUL?

Bounds on $\|\Phi_t \boldsymbol{\theta}^* - \boldsymbol{r}_t\|_2$ and $\|\boldsymbol{\theta}^*\|_2$ fit together better than bounds on $\|\Phi_t^\top \boldsymbol{\epsilon}_t\|_{(\Phi_t^\top \Phi_t + \alpha \boldsymbol{I})^{-1}}$ and $\|\boldsymbol{\theta}^*\|_2$.

OFUL: Using the definition of $\widehat{\boldsymbol{\theta}}_{\alpha,t}$, and then the triangle inequality,

$$\begin{aligned}
\|\boldsymbol{\theta}^* - \widehat{\boldsymbol{\theta}}_{\alpha,t}\|_{(\Phi_t^\top \Phi_t + \alpha \boldsymbol{I})} &= \|\Phi_t^\top \boldsymbol{\epsilon}_t + \alpha \boldsymbol{\theta}^*\|_{(\Phi_t^\top \Phi_t + \alpha \boldsymbol{I})^{-1}} \\
&\leq \|\Phi_t^\top \boldsymbol{\epsilon}_t\|_{(\Phi_t^\top \Phi_t + \alpha \boldsymbol{I})^{-1}} + \alpha \|\boldsymbol{\theta}^*\|_{(\Phi_t^\top \Phi_t + \alpha \boldsymbol{I})^{-1}} \\
&\leq \sigma \sqrt{\ln\left(\det\left(\frac{1}{\alpha}\Phi_t^\top \Phi_t + \boldsymbol{I}\right)\right) + 2\ln(1/\delta)} + \sqrt{\alpha} B
\end{aligned}$$

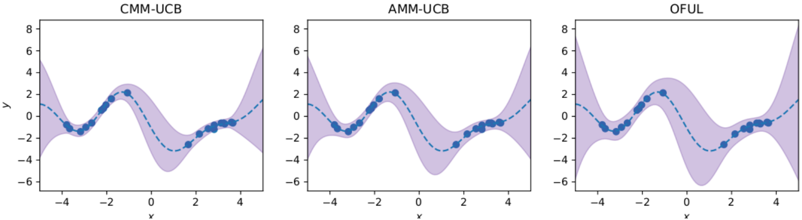
The triangle inquality step causes the $\ln \det$ and αB^2 terms to appear under separate square roots.

Ours: We combine our constraints by completing the square on the LHS of

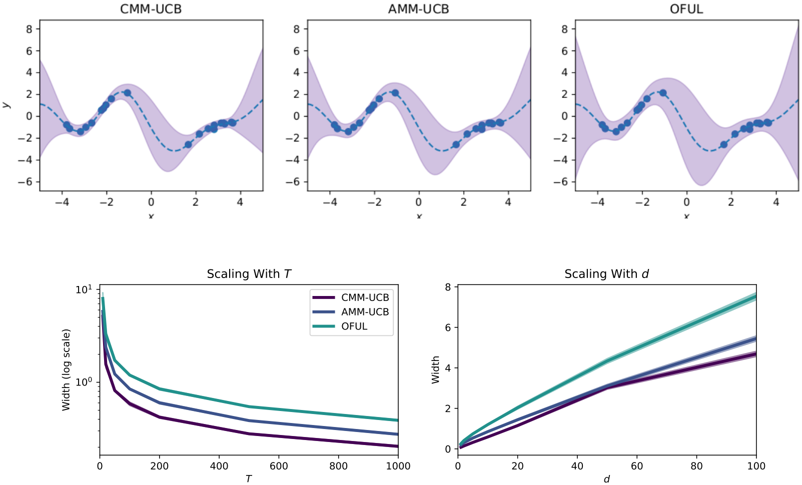
$$\|\Phi_t \boldsymbol{\theta}^* - \boldsymbol{r}_t\|_2^2 + \alpha \|\boldsymbol{\theta}^*\|_2^2 \leq R_{\mathrm{MM},t}^2 + \alpha B^2$$

Some Experimental Results

Confidence Bound Comparison

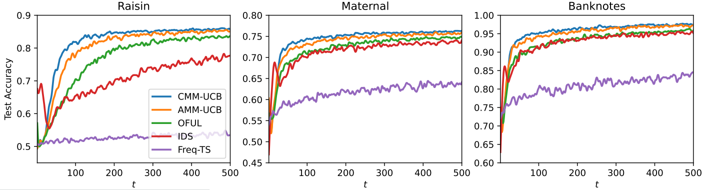


Confidence Bound Comparison



Hyperparameter Tuning

	Raisin		Maternal		Banknotes	
	Mean Acc	Max Acc	Mean Acc	Max Acc	Mean Acc	Max Acc
CMM-UCB (Ours)	**0.818** ± 0.018	**0.893** ± 0.019	**0.744** ± 0.020	**0.829** ± 0.023	**0.954** ± 0.005	**1.000** ± 0.000
AMM-UCB (Ours)	0.800 ± 0.017	0.892 ± 0.020	0.736 ± 0.020	**0.829** ± 0.023	0.948 ± 0.005	**1.000** ± 0.000
OFUL	0.764 ± 0.019	0.891 ± 0.019	0.722 ± 0.019	0.827 ± 0.022	0.929 ± 0.006	**1.000** ± 0.000
IDS[3]	0.706 ± 0.048	0.891 ± 0.020	0.714 ± 0.019	0.827 ± 0.024	0.926 ± 0.007	**1.000** ± 0.000
Freq-TS[4]	0.527 ± 0.022	0.884 ± 0.019	0.616 ± 0.018	0.823 ± 0.022	0.808 ± 0.012	**1.000** ± 0.000



[3] J. Kirschner and A. Krause. (2018) Information directed sampling and bandits with heteroscedastic noise, COLT

[4] S. Agrawal and N. Goyal. (2013) Thompson sampling for contextual bandits with linear payoffs, ICML

Open Questions

When Do More Adaptive Mixture Distributions Help?

The means $\boldsymbol{\mu}_t$ and covariances $\boldsymbol{T}_t$ of the standard mixture distributions can be written in the form

$$\boldsymbol{\mu}_t = \begin{bmatrix} m(a_1) \\ m(a_2) \\ \vdots \\ m(a_t) \end{bmatrix}, \qquad \boldsymbol{T}_t = \begin{bmatrix} k(a_1, a_1) & k(a_1, a_2) & \cdots & k(a_1, a_t) \\ k(a_2, a_1) & k(a_2, a_2) & \cdots & k(a_2, a_t) \\ \vdots & \vdots & \ddots & \vdots \\ k(a_t, a_1) & k(a_t, a_2) & \cdots & k(a_t, a_t) \end{bmatrix},$$

where $m(a) = 0$ and $k(a, a') = \phi(a)^\top \phi(a')$.

When Do More Adaptive Mixture Distributions Help?

The means $\boldsymbol{\mu}_t$ and covariances $\boldsymbol{T}_t$ of the standard mixture distributions can be written in the form

$$\boldsymbol{\mu}_t = \begin{bmatrix} m(a_1) \\ m(a_2) \\ \vdots \\ m(a_t) \end{bmatrix}, \qquad \boldsymbol{T}_t = \begin{bmatrix} k(a_1, a_1) & k(a_1, a_2) & \cdots & k(a_1, a_t) \\ k(a_2, a_1) & k(a_2, a_2) & \cdots & k(a_2, a_t) \\ \vdots & \vdots & \ddots & \vdots \\ k(a_t, a_1) & k(a_t, a_2) & \cdots & k(a_t, a_t) \end{bmatrix},$$

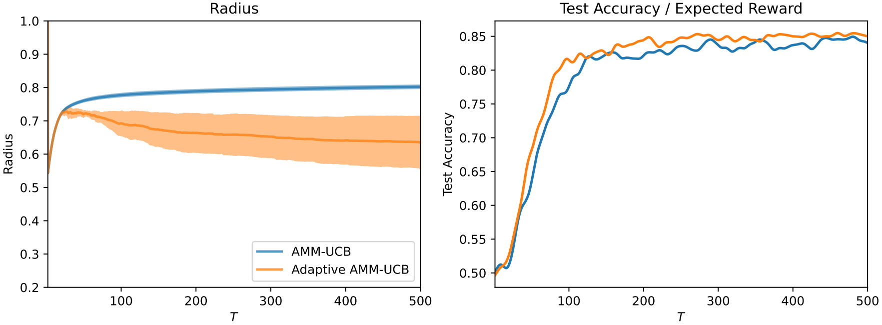
where $m(a) = 0$ and $k(a, a') = \phi(a)^\top \phi(a')$.

We also tried out

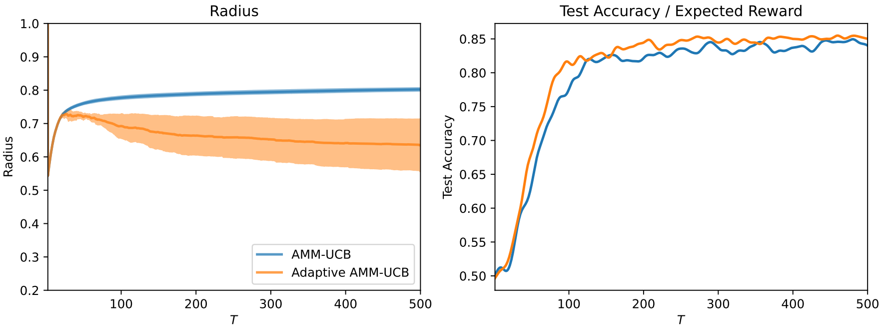
$$\boldsymbol{\mu}_t = \begin{bmatrix} m_0(a_1) \\ m_1(a_2) \\ \vdots \\ m_{t-1}(a_t) \end{bmatrix}, \qquad \boldsymbol{T}_t = \begin{bmatrix} k_0(a_1, a_1) & k_1(a_1, a_2) & \cdots & k_{t-1}(a_1, a_t) \\ k_1(a_2, a_1) & k_1(a_2, a_2) & \cdots & k_{t-1}(a_2, a_t) \\ \vdots & \vdots & \ddots & \vdots \\ k_{t-1}(a_t, a_1) & k_{t-1}(a_t, a_2) & \cdots & k_{t-1}(a_t, a_t) \end{bmatrix},$$

where $m_t(a) = \boldsymbol{k}_t(a)^\top \left(\boldsymbol{K}_t + \beta \boldsymbol{I}\right)^{-1} \boldsymbol{r}_t$ and $k_t(a, a') = k(a, a') - \boldsymbol{k}_t(a)^\top \left(\boldsymbol{K}_t + \beta \boldsymbol{I}\right)^{-1} \boldsymbol{k}_t(a')$.

When Do More Adaptive Mixture Distributions Help?



When Do More Adaptive Mixture Distributions Help?



Adaptive mixture distributions don't always help this much though.

Thank you for listening!