CRITEO

AI Lab

Jointly Efficient and Optimal Algorithms for Logistic Bandits

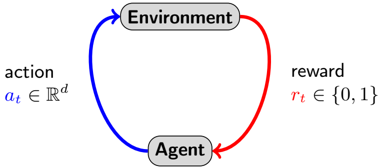
ML Big Weeks

Marc Abeille, Louis Faury

OPEN TOGETHER IMPACTFUL

The Learning Problem

- Repeated game with **structured** and **binary** feedback.



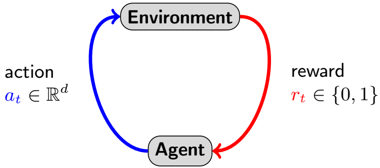
unknown

where $r_t \sim$ Bernoulli $\left(\mu({a_t}^\top \theta_\star)\right)$

logistic function
$\mu(z) = (1 + e^{-z})^{-1}$

The Learning Problem

- Repeated game with **structured** and **binary** feedback.



unknown

where $r_t \sim$ Bernoulli $\left(\mu({a_t}^\top \theta_\star)\right)$

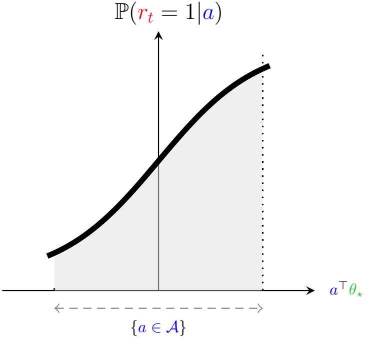
logistic function
$\mu(z) = (1 + e^{-z})^{-1}$

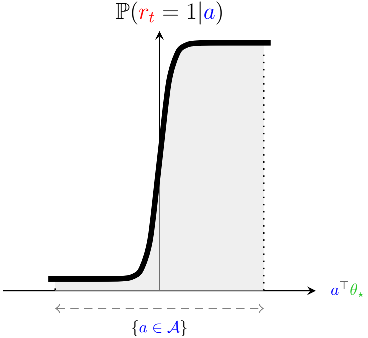
- **Goal.** Maximize $\sum_{t=1}^{T} \mu({a_t}^\top \theta_\star)$ = performance over time.

$\theta_\star$ is unknown! $\Rightarrow$ exploration-exploitation dilemma.

Reward Model: Closer Look

Different regimes





$\forall a \in \mathcal{A}$, $\mathbb{P}(r_t = 1|a) \simeq 0.5$

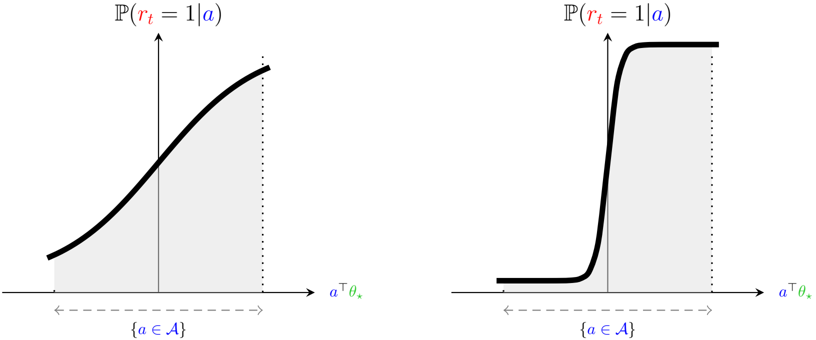
✔ SOTA!

$\exists a \in \mathcal{A}$, $\mathbb{P}(r_t = 1|a) \approx 0$

✘ SOTA!

Reward Model: Closer Look

Different regimes



$\forall a \in \mathcal{A},\ \mathbb{P}(r_t = 1|a) \simeq 0.5$

✔ SOTA!

$\exists a \in \mathcal{A},\ \mathbb{P}(r_t = 1|a) \approx 0$

✘ SOTA!

Key Quantity

$$\boldsymbol{\kappa} \simeq \max_{a\in\mathcal{A}} 1/\mathbb{P}(r_t = 1|a) \qquad \leftarrow \text{ typically } \; 10^3 \; !$$

Objective

- The performance metric is the regret:

$$\text{Regret}(T) = T \max_{a \in \mathcal{A}} \mu(a^\top \theta_\star) - \sum_{t=1}^{T} \mu(a_t{}^\top \theta_\star) \ .$$

Objective

- The performance metric is the regret:

$$\text{Regret}(T) = T \max_{a \in \mathcal{A}} \mu(a^\top \theta_\star) - \sum_{t=1}^{T} \mu(a_t{}^\top \theta_\star) \, .$$

- **Challenges.** Obtain optimal regret with efficient algorithms.

Lower-bound *[Abeille et al. AISTATS21]*			
Algorithm	**Regret Bound**	**Minimax**	**Efficient**
`GLM-UCB` *[Filippi et al. NIPS10.]*			
`OFULog-r` *[Faury et al. ICML20]*			
`OFU-ECOLog` (submitted)			

Objective

- The performance metric is the regret:

$$\text{Regret}(T) = T \max_{a\in\mathcal{A}} \mu(a^\top \theta_\star) - \sum_{t=1}^{T} \mu(a_t{}^\top \theta_\star) \ .$$

- **Challenges.** Obtain optimal regret with efficient algorithms.

Lower-bound *[Abeille et al. AISTATS21]*			
Algorithm	**Regret Bound**	**Minimax**	**Efficient**
`GLM-UCB` *[Filippi et al. NIPS10.]*	$\mathcal{O}(\boldsymbol{\kappa} d\sqrt{T})$	✘	✘
`OFULog-r` *[Faury et al. ICML20]*			
`OFU-ECOLog` (submitted)			

Objective

- The performance metric is the regret:

$$\text{Regret}(T) = T \max_{a \in \mathcal{A}} \mu(a^\top \theta_\star) - \sum_{t=1}^{T} \mu({a_t}^\top \theta_\star) \ .$$

- **Challenges.** Obtain optimal regret with efficient algorithms.

Lower-bound *[Abeille et al. AISTATS21]*			
Algorithm	**Regret Bound**	**Minimax**	**Efficient**
GLM-UCB *[Filippi et al. NIPS10.]*	$\mathcal{O}(\boldsymbol{\kappa} d\sqrt{T})$	✘	✘
OFULog-r *[Faury et al. ICML20]*	$\mathcal{O}(d\sqrt{T/\boldsymbol{\kappa}})$	✔	✘
OFU-ECOLog (submitted)			

Objective

- The performance metric is the regret:

$$\text{Regret}(T) = T \max_{a \in \mathcal{A}} \mu(a^\top \theta_\star) - \sum_{t=1}^{T} \mu(a_t{}^\top \theta_\star) \ .$$

- **Challenges.** Obtain optimal regret with efficient algorithms.

Lower-bound *[Abeille et al. AISTATS21]*	$\Omega(d\sqrt{T/\kappa})$		
Algorithm	**Regret Bound**	**Minimax**	**Efficient**
`GLM-UCB` *[Filippi et al. NIPS10.]*	$\mathcal{O}(\kappa d\sqrt{T})$	✘	✘
`OFULog-r` *[Faury et al. ICML20]*	$\mathcal{O}(d\sqrt{T/\kappa})$	✔	✘
`OFU-ECOLog` (submitted)			

Objective

- The performance metric is the regret:

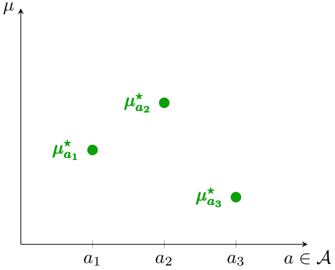
$$\text{Regret}(T) = T \max_{a\in\mathcal{A}} \mu(a^\top \theta_\star) - \sum_{t=1}^{T} \mu({a_t}^\top \theta_\star) \ .$$

- **Challenges.** Obtain optimal regret with efficient algorithms.

Lower-bound *[Abeille et al. AISTATS21]*	$\Omega(d\sqrt{T/\kappa})$		
Algorithm	**Regret Bound**	**Minimax**	**Efficient**
`GLM-UCB` *[Filippi et al. NIPS10.]*	$\mathcal{O}(\kappa d\sqrt{T})$	✘	✘
`OFULog-r` *[Faury et al. ICML20]*	$\mathcal{O}(d\sqrt{T/\kappa})$	✔	✘
`OFU-ECOLog` (submitted)	$\mathcal{O}(d\sqrt{T/\kappa})$	✔	✔

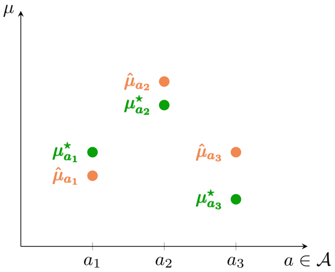
The exploration-exploitation dilemma

High-level idea.



The exploration-exploitation dilemma

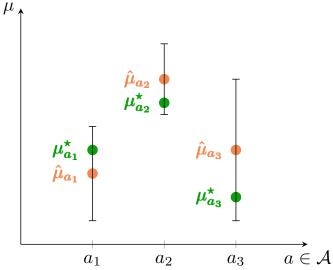
High-level idea.



- **Exploit.** Predict rewards

The exploration-exploitation dilemma

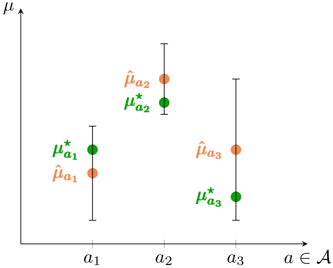
High-level idea.



- **Exploit.** Predict rewards
- **Explore.** Quantify uncertainty

The exploration-exploitation dilemma

High-level idea.



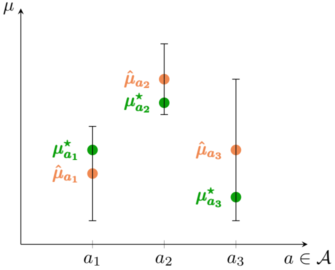
- **Exploit.** Predict rewards
- **Explore.** Quantify uncertainty

Leveraging the structure.

$$\mu^\star_a = \mu(a^\top \theta_\star)$$

The exploration-exploitation dilemma

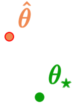
High-level idea.



- **Exploit.** Predict rewards
- **Explore.** Quantify uncertainty

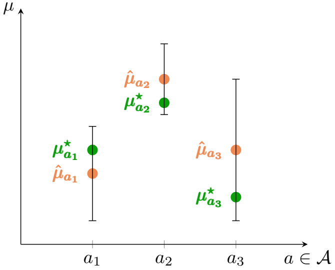
Leveraging the structure.

$\mu^\star_a = \mu(a^\top \theta_\star)$



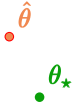
The exploration-exploitation dilemma

High-level idea.



Leveraging the structure.

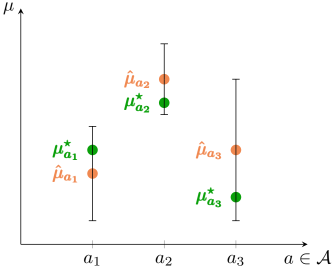
$$\mu^\star_a = \mu(a^\top \theta_\star)$$



- **Exploit.** Predict rewards $\longrightarrow$ $\hat{\mu}_a = \mu(a^\top \hat{\theta})$.
- **Explore.** Quantify uncertainty

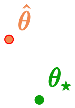
The exploration-exploitation dilemma

High-level idea.



Leveraging the structure.

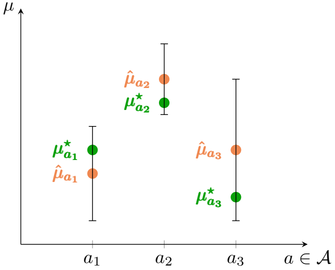
$$\mu^\star_a = \mu(a^\top \theta_\star)$$



- **Exploit.** Predict rewards $\longrightarrow$ $\hat{\mu}_a = \mu(a^\top \hat{\theta})$.
- **Explore.** Quantify uncertainty $\longrightarrow$ quantify deviation between $\hat{\theta}$ and $\theta_\star$.

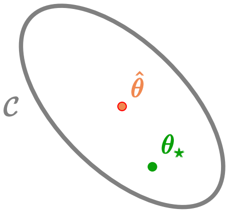
The exploration-exploitation dilemma

High-level idea.



Leveraging the structure.

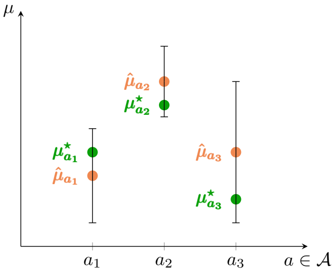
$\mu^\star_a = \mu(a^\top \theta_\star)$



- **Exploit.** Predict rewards $\longrightarrow$ $\hat{\mu}_a = \mu(a^\top \hat{\theta})$.
- **Explore.** Quantify uncertainty $\longrightarrow$ quantify deviation between $\hat{\theta}$ and $\theta_\star$.

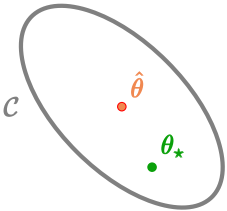
The exploration-exploitation dilemma

High-level idea.



Leveraging the structure.

$\mu^\star_a = \mu(a^\top \theta_\star)$



- **Exploit.** Predict rewards $\longrightarrow$ $\hat{\mu}_a = \mu(a^\top \hat{\theta})$.
- **Explore.** Quantify uncertainty $\longrightarrow$ quantify deviation between $\hat{\theta}$ and $\theta_\star$.

Challenge: design *sharp* and *tractable* confidence set $\mathcal{C}$

[Filippi et al. NIPS2010]

- Based on the maximum-likelihood estimator and linear design matrix $\boldsymbol{V}_t$:

$$\begin{cases} \hat{\theta}_t = \sum_{s=1}^{t} \ell_s(\theta) \quad \text{where } \ell_s(\theta) \leftarrow \text{cross-entropy on } (a_s, r_s) \ , \\ \boldsymbol{V}_t = \sum_{s=1}^{t} a_s a_s^\top \ . \end{cases}$$

[Filippi et al. NIPS2010]

- Based on the maximum-likelihood estimator and linear design matrix $\boldsymbol{V}_t$:

$$\begin{cases} \hat{\theta}_t = \sum_{s=1}^{t} \ell_s(\theta) \quad \text{where } \ell_s(\theta) \leftarrow \text{cross-entropy on } (a_s, r_s)\ , \\ \boldsymbol{V}_t = \sum_{s=1}^{t} a_s {a_s}^\top\ . \end{cases}$$

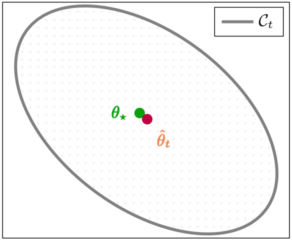
$$\mathcal{C}_t = \left\{ \theta \in \mathbb{R}^d,\ \|\theta - \hat{\theta}_t\|_{\boldsymbol{V}_t} \leq \boldsymbol{\kappa} \right\}$$

[Filippi et al. NIPS2010]

- Based on the maximum-likelihood estimator and linear design matrix $\boldsymbol{V}_t$:

$$\begin{cases} \hat{\theta}_t = \sum_{s=1}^{t} \ell_s(\theta) \quad \text{where } \ell_s(\theta) \leftarrow \text{cross-entropy on } (a_s, r_s)\ , \\ \boldsymbol{V}_t = \sum_{s=1}^{t} a_s {a_s}^\top\ . \end{cases}$$

$$\mathcal{C}_t = \left\{ \theta \in \mathbb{R}^d,\ \|\theta - \hat{\theta}_t\|_{\boldsymbol{V}_t} \leq \boldsymbol{\kappa} \right\}$$

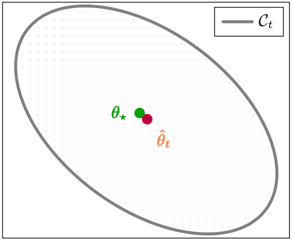


[Filippi et al. NIPS2010]

- Based on the maximum-likelihood estimator and linear design matrix $\boldsymbol{V}_t$:

$$\begin{cases} \hat{\theta}_t = \sum_{s=1}^{t} \ell_s(\theta) \quad \text{where } \ell_s(\theta) \leftarrow \text{cross-entropy on } (a_s, r_s)\ , \\ \boldsymbol{V}_t = \sum_{s=1}^{t} a_s a_s^\top\ . \end{cases}$$

$$\mathcal{C}_t = \left\{ \theta \in \mathbb{R}^d,\ \|\theta - \hat{\theta}_t\|_{\boldsymbol{V}_t} \le \boldsymbol{\kappa} \right\}$$



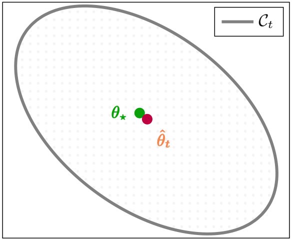
We play $a \approx 10^4$ times.

[Filippi et al. NIPS2010]

- Based on the maximum-likelihood estimator and linear design matrix $\boldsymbol{V}_t$:

$$\begin{cases} \hat{\theta}_t = \sum_{s=1}^{t} \ell_s(\theta) \quad \text{where } \ell_s(\theta) \leftarrow \text{cross-entropy on } (a_s, r_s)\,, \\ \boldsymbol{V}_t = \sum_{s=1}^{t} a_s a_s^\top\,. \end{cases}$$

$$\mathcal{C}_t = \left\{\theta \in \mathbb{R}^d,\ \|\theta - \hat{\theta}_t\|_{\boldsymbol{V}_t} \leq \boldsymbol{\kappa}\right\}$$



We play $a \approx 10^4$ times.

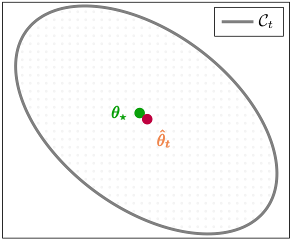
$$\begin{aligned} \hat{\mu}_a - \mu_a^\star &\approx \boldsymbol{\kappa}\|a\|_{\boldsymbol{V}_t^{-1}} \\ &\approx \frac{\boldsymbol{\kappa}}{\sqrt{\text{number of times a was played}}} \\ &\approx 10^2/\sqrt{10^4} \\ &\approx 1 \qquad \leftarrow \text{trivial bound!} \end{aligned}$$

[Filippi et al. NIPS2010]

- Based on the maximum-likelihood estimator and linear design matrix $\boldsymbol{V}_t$:

$$\begin{cases} \hat{\theta}_t = \sum_{s=1}^{t} \ell_s(\theta) \quad \text{where } \ell_s(\theta) \leftarrow \text{cross-entropy on } (a_s, r_s) \,, \\ \boldsymbol{V}_t = \sum_{s=1}^{t} a_s a_s^\top \,. \end{cases}$$

$$\mathcal{C}_t = \left\{ \theta \in \mathbb{R}^d,\ \|\theta - \hat{\theta}_t\|_{\boldsymbol{V}_t} \leq \boldsymbol{\kappa} \right\}$$



We play $a \approx 10^4$ times.

$$\begin{aligned} \hat{\mu}_a - \mu_a^\star &\approx \boldsymbol{\kappa} \|a\|_{\boldsymbol{V}_t^{-1}} \\ &\approx \frac{\boldsymbol{\kappa}}{\sqrt{\text{number of times a was played}}} \\ &\approx 10^2 / \sqrt{10^4} \\ &\approx 1 \qquad \leftarrow \text{trivial bound!} \end{aligned}$$

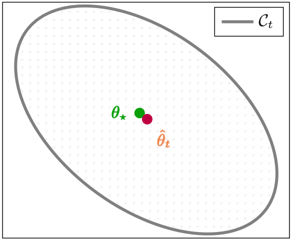
✘ Aggressive exploration

[Filippi et al. NIPS2010]

- Based on the maximum-likelihood estimator and linear design matrix $\boldsymbol{V}_t$:

$$\begin{cases} \hat{\theta}_t = \sum_{s=1}^{t} \ell_s(\theta) \quad \text{where } \ell_s(\theta) \leftarrow \text{cross-entropy on } (a_s, r_s)\ , \\ \boldsymbol{V}_t = \sum_{s=1}^{t} a_s a_s^\top\ . \end{cases}$$

$$\mathcal{C}_t = \left\{ \theta \in \mathbb{R}^d,\ \|\theta - \hat{\theta}_t\|_{\boldsymbol{V}_t} \leq \boldsymbol{\kappa} \right\}$$



We play $a \approx 10^4$ times.

$$\begin{aligned} \hat{\mu}_a - \mu_a^\star &\approx \boldsymbol{\kappa} \|a\|_{\boldsymbol{V}_t^{-1}} \\ &\approx \frac{\boldsymbol{\kappa}}{\sqrt{\text{number of times a was played}}} \\ &\approx 10^2/\sqrt{10^4} \\ &\approx 1 \qquad \leftarrow \text{trivial bound!} \end{aligned}$$

✘ Aggressive exploration

✘ Computationally Expensive

[Faury et al. ICML20, Abeille et al. AISTATS21]

- Same estimator $\hat{\theta}_t$ but new metric of deviation:

$$\boldsymbol{H}_t = \sum_{s=1}^{t} \dot{\mu}({a_s}^\top \hat{\theta}_t) a_s {a_s}^\top \ .$$

[Faury et al. ICML20, Abeille et al. AISTATS21]

- Same estimator $\hat{\theta}_t$ but new metric of deviation:

$$\boldsymbol{H}_t = \sum_{s=1}^{t} \dot{\mu}(a_s{}^\top \hat{\theta}_t) a_s a_s{}^\top \; .$$

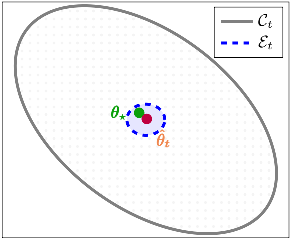
$$\mathcal{E}_t = \left\{ \theta \in \mathbb{R}^d, \|\theta - \hat{\theta}_t\|_{\boldsymbol{H}_t} \leq 1 \right\}$$

[Faury et al. ICML20, Abeille et al. AISTATS21]

- Same estimator $\hat{\theta}_t$ but new metric of deviation:

$$\boldsymbol{H}_t = \sum_{s=1}^{t} \dot{\mu}(a_s{}^\top \hat{\theta}_t) a_s a_s{}^\top \ .$$

$$\mathcal{E}_t = \left\{ \theta \in \mathbb{R}^d, \|\theta - \hat{\theta}_t\|_{\boldsymbol{H}_t} \leq 1 \right\}$$

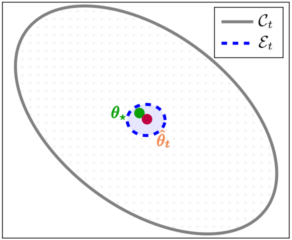


[Faury et al. ICML20, Abeille et al. AISTATS21]

- Same estimator $\hat{\theta}_t$ but new metric of deviation:

$$\boldsymbol{H}_t = \sum_{s=1}^{t} \dot{\mu}({a_s}^\top \hat{\theta}_t) a_s {a_s}^\top \ .$$

$$\mathcal{E}_t = \left\{\theta \in \mathbb{R}^d, \|\theta - \hat{\theta}_t\|_{\boldsymbol{H}_t} \leq 1\right\}$$



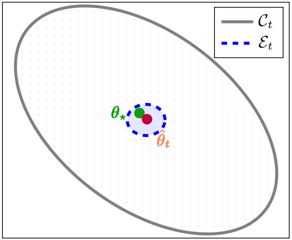
New concentration tools.

[Faury et al. ICML20, Abeille et al. AISTATS21]

- Same estimator $\hat{\theta}_t$ but new metric of deviation:

$$\boldsymbol{H}_t = \sum_{s=1}^{t} \dot{\mu}({a_s}^\top \hat{\theta}_t) a_s {a_s}^\top \ .$$

$$\mathcal{E}_t = \left\{ \theta \in \mathbb{R}^d, \|\theta - \hat{\theta}_t\|_{\boldsymbol{H}_t} \leq 1 \right\}$$



New concentration tools.

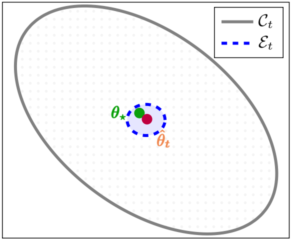
We play $a \approx 10^4$ times.

[Faury et al. ICML20, Abeille et al. AISTATS21]

- Same estimator $\hat{\theta}_t$ but new metric of deviation:

$$\boldsymbol{H}_t = \sum_{s=1}^{t} \dot{\mu}(a_s{}^\top \hat{\theta}_t) a_s a_s{}^\top \ .$$

$$\mathcal{E}_t = \left\{ \theta \in \mathbb{R}^d, \|\theta - \hat{\theta}_t\|_{\boldsymbol{H}_t} \leq 1 \right\}$$



New concentration tools.

We play $a \approx 10^4$ times.

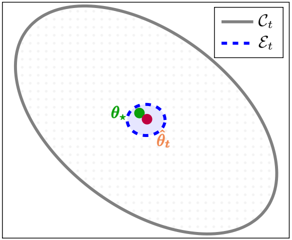
$$\begin{aligned}\hat{\mu}_a - \mu_a^\star &\approx \dot{\mu}(a^\top \hat{\theta}_t) \|a\|_{\boldsymbol{H}_t^{-1}} \\ &\approx \frac{1}{\sqrt{\text{number of times } a \text{ was played}}} \\ &\approx 10^{-2}\end{aligned}$$

[Faury et al. ICML20, Abeille et al. AISTATS21]

- Same estimator $\hat{\theta}_t$ but new metric of deviation:

$$\boldsymbol{H}_t = \sum_{s=1}^{t} \dot{\mu}(a_s{}^\top \hat{\theta}_t) a_s a_s{}^\top \ .$$

$$\mathcal{E}_t = \left\{ \theta \in \mathbb{R}^d, \|\theta - \hat{\theta}_t\|_{\boldsymbol{H}_t} \leq 1 \right\}$$



New concentration tools.

We play $a \approx 10^4$ times.

$$\hat{\mu}_a - \mu_a^\star \approx \dot{\mu}(a^\top \hat{\theta}_t) \|a\|_{\boldsymbol{H}_t^{-1}}$$
$$\approx \frac{1}{\sqrt{\text{number of times a was played}}}$$
$$\approx 10^{-2}$$

✔ Efficient exploration

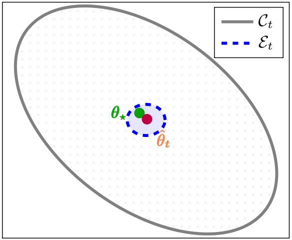
✔ Optimal (Cramer-Rao)

[Faury et al. ICML20, Abeille et al. AISTATS21]

- Same estimator $\hat{\theta}_t$ but new metric of deviation:

$$\boldsymbol{H}_t = \sum_{s=1}^{t} \dot{\mu}(a_s{}^\top \hat{\theta}_t) a_s a_s{}^\top \, .$$

$$\mathcal{E}_t = \left\{\theta \in \mathbb{R}^d, \|\theta - \hat{\theta}_t\|_{\boldsymbol{H}_t} \leq 1\right\}$$



New concentration tools.

We play $a \approx 10^4$ times.

$$\begin{aligned} \hat{\mu}_a - \mu_a^\star &\approx \dot{\mu}(a^\top \hat{\theta}_t) \|a\|_{\boldsymbol{H}_t^{-1}} \\ &\approx \frac{1}{\sqrt{\text{number of times } a \text{ was played}}} \\ &\approx 10^{-2} \end{aligned}$$

✔ Efficient exploration

✔ Optimal (Cramer-Rao)

✘ Computationally expensive!

Fast and Optimal Approach (submitted)

- Main bottleneck: computation of $\hat{\theta}_t$ and $\boldsymbol{H}_t$.
  - $\Omega(t)$ operations at each round!
  - In practice, **very** slow.

Fast and Optimal Approach (submitted)

- Main bottleneck: computation of $\hat{\theta}_t$ and $\boldsymbol{H}_t$.
  - $\Omega(t)$ operations at each round!
  - In practice, **very** slow.

- Efficient alternative through recursive-least-squares-like operations:

ECOLog procedure: $$\begin{cases} \theta_t = \operatorname{argmin}_\theta \left\|\theta - \theta_{t-1}\right\|^2_{\boldsymbol{W}_{t-1}} + \ell_t(\theta) \,, \\ \boldsymbol{W}_t = \boldsymbol{W}_{t-1} + \dot{\mu}({a_t}^\top \theta_t) a_t {a_t}^\top \,. \end{cases}$$

Fast and Optimal Approach (submitted)

- Main bottleneck: computation of $\hat{\theta}_t$ and $\boldsymbol{H}_t$.
  - $\Omega(t)$ operations at each round!
  - In practice, **very** slow.

- Efficient alternative through recursive-least-squares-like operations:

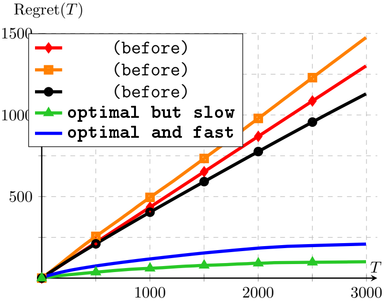
$$\textbf{ECOLog procedure: } \begin{cases} \theta_t = \operatorname{argmin}_\theta \left\|\theta - \theta_{t-1}\right\|^2_{\boldsymbol{W}_{t-1}} + \ell_t(\theta) , \\ \boldsymbol{W}_t = \boldsymbol{W}_{t-1} + \dot{\mu}({a_t}^\top \theta_t) a_t {a_t}^\top . \end{cases}$$

- Best of both world:

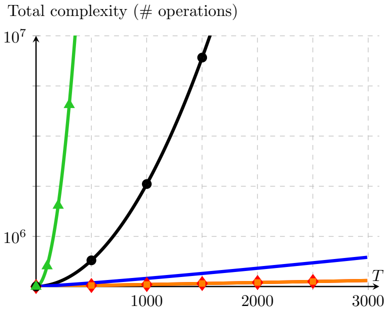
✔ Online computations: $\widetilde{\mathcal{O}}(1)$ operations!

✔ Statistical tightness: $\mathcal{E}'_t = \left\{\theta,\ \left\|\theta - \theta_t\right\|_{\boldsymbol{W}_t} \leq 1\right\} \approx \mathcal{E}_t$.

Can we see some curves?

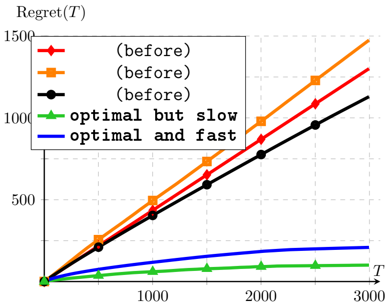


Regret: the smaller the better.

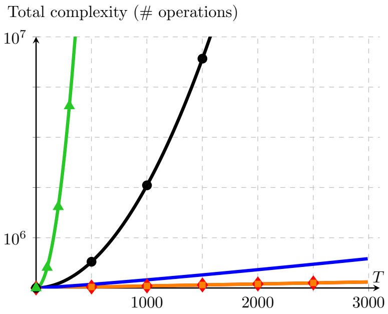


Complexity: the smaller the better

Can we see some curves?



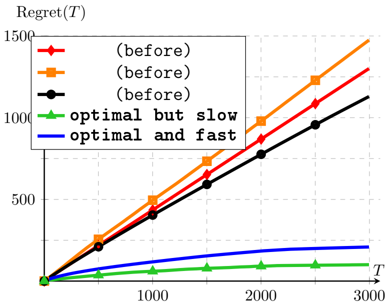
Regret: the smaller the better.



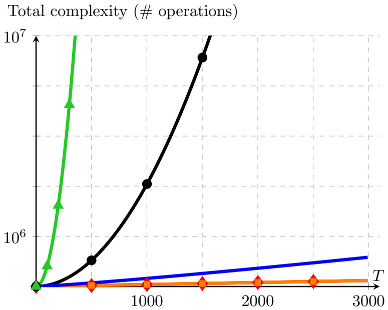
Complexity: the smaller the better

- **Blue curve**: Best of both world behavior.

Can we see some curves?



Regret: the smaller the better.



Complexity: the smaller the better

- **Blue curve**: Best of both world behavior.
- In short: **mature** for deployment in real-life situations.

What's in it for you?

- **Learning.** Principled and efficient **estimation** procedure:
  - Same convergence guarantee as MLE.
  - Fully online.
  - Compatible with non-stationary environments.

What's in it for you?

- **Learning.** Principled and efficient **estimation** procedure:
  - Same convergence guarantee as MLE.
  - Fully online.
  - Compatible with non-stationary environments.

- **Planning.** Principled **exploration**:
  - Readily usable confidence sets / predictions errors.
  - Long-term optimal without burning cash.
  - Compatible with randomized exploration.

What's in it for you?

- **Learning.** Principled and efficient **estimation** procedure:
  - Same convergence guarantee as MLE.
  - Fully online.
  - Compatible with non-stationary environments.

- **Planning.** Principled **exploration**:
  - Readily usable confidence sets / predictions errors.
  - Long-term optimal without burning cash.
  - Compatible with randomized exploration.

Thank you! Questions?