Variance-Sensitive Confidence Regions for Parametric Bandits

Louis Faury

TélécomParis and Criteo

supervised by **Olivier Fercoq** and co-supervised by **Marc Abeille**.

Paris, October 11, 2021.

Outline

- **Motivation.** Study **non-linearity** in Generalized Linear Bandits (GLBs).

Outline

- **Motivation.** Study **non-linearity** in Generalized Linear Bandits (GLBs).
  - Sequential decision making problems

Outline

- **Motivation.** Study **non-linearity** in Generalized Linear Bandits (GLBs).
  - Sequential decision making problems
  - Isolate the interaction non-linearity $\leftrightarrow$ exploration/exploitation trade-off.

Outline

- **Motivation.** Study **non-linearity** in Generalized Linear Bandits (GLBs).
  - Sequential decision making problems
  - Isolate the interaction non-linearity $\leftrightarrow$ exploration/exploitation trade-off.

- **Previous studies.** Non-linearity is harmful.

Outline

- **Motivation.** Study **non-linearity** in Generalized Linear Bandits (GLBs).
  - Sequential decision making problems
  - Isolate the interaction non-linearity $\leftrightarrow$ exploration/exploitation trade-off.

- **Previous studies.** Non-linearity is harmful.
  - The more non-linear, the **harder** the problem.

Outline

- **Motivation.** Study **non-linearity** in Generalized Linear Bandits (GLBs).
  - Sequential decision making problems
  - Isolate the interaction non-linearity $\leftrightarrow$ exploration/exploitation trade-off.

- **Previous studies.** Non-linearity is harmful.
  - The more non-linear, the **harder** the problem.
  - Status-quo since $\approx$ 10 years.

Outline

- **Motivation.** Study **non-linearity** in Generalized Linear Bandits (GLBs).
  - Sequential decision making problems
  - Isolate the interaction non-linearity $\leftrightarrow$ exploration/exploitation trade-off.

- **Previous studies.** Non-linearity is harmful.
  - The more non-linear, the **harder** the problem.
  - Status-quo since ≈ 10 years.

- **Today.** A different story:

Outline

- **Motivation.** Study **non-linearity** in Generalized Linear Bandits (GLBs).
  - Sequential decision making problems
  - Isolate the interaction non-linearity $\leftrightarrow$ exploration/exploitation trade-off.
- **Previous studies.** Non-linearity is harmful.
  - The more non-linear, the **harder** the problem.
  - Status-quo since $\approx$ 10 years.
- **Today.** A different story:
  - Improved algorithms and refined analysis.

Outline

- **Motivation.** Study **non-linearity** in Generalized Linear Bandits (GLBs).
  - Sequential decision making problems
  - Isolate the interaction non-linearity $\leftrightarrow$ exploration/exploitation trade-off.

- **Previous studies.** Non-linearity is harmful.
  - The more non-linear, the **harder** the problem.
  - Status-quo since $\approx$ 10 years.

- **Today.** A different story:
  - Improved algorithms and refined analysis.
  - Effects of non-linearity are **short-term** (and not always detrimental).

Outline

- **Motivation.** Study **non-linearity** in Generalized Linear Bandits (GLBs).
  - Sequential decision making problems
  - Isolate the interaction non-linearity $\leftrightarrow$ exploration/exploitation trade-off.

- **Previous studies.** Non-linearity is harmful.
  - The more non-linear, the **harder** the problem.
  - Status-quo since $\approx$ 10 years.

- **Today.** A different story:
  - Improved algorithms and refined analysis.
  - Effects of non-linearity are **short-term** (and not always detrimental).
  - Even better! in some cases:

Outline

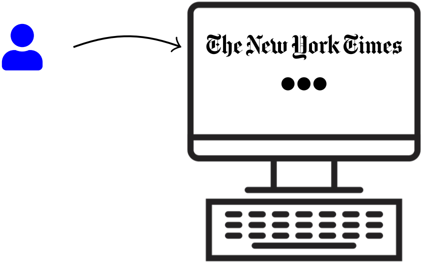
- **Motivation.** Study **non-linearity** in Generalized Linear Bandits (GLBs).
  - Sequential decision making problems
  - Isolate the interaction non-linearity $\leftrightarrow$ exploration/exploitation trade-off.
- **Previous studies.** Non-linearity is harmful.
  - The more non-linear, the **harder** the problem.
  - Status-quo since $\approx$ 10 years.
- **Today.** A different story:
  - Improved algorithms and refined analysis.
  - Effects of non-linearity are **short-term** (and not always detrimental).
  - Even better! in some cases:

> The more non-linear, the **easier** the problem.

Motivation and Setting

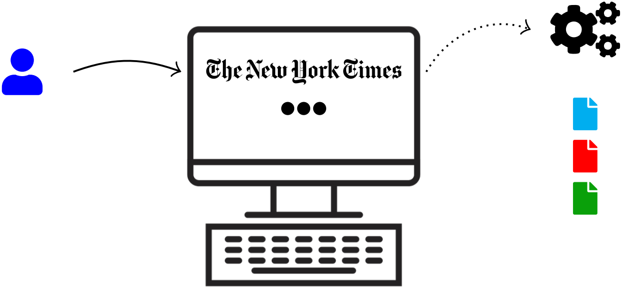
A motivating example (1/3)

- **News recommandation**. Among others (*e.g* clinical trials, ..).



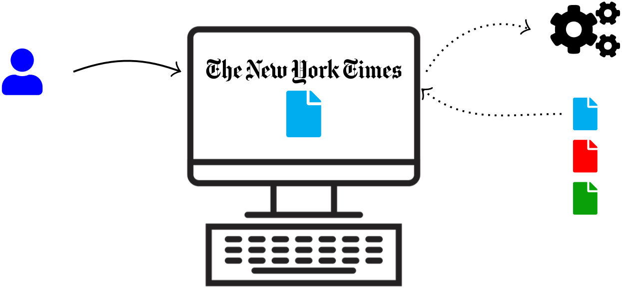
A motivating example (1/3)

- **News recommandation**. Among others (*e.g* clinical trials, ..).



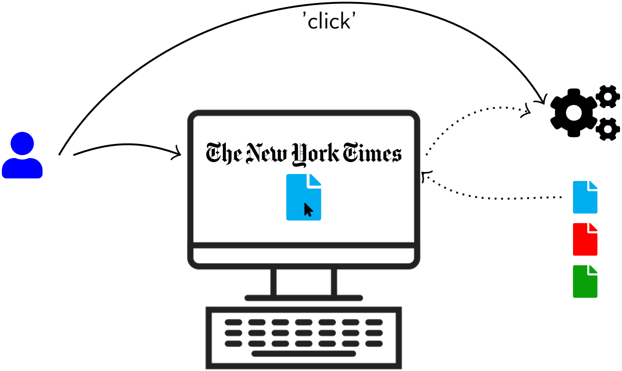
A motivating example (1/3)

- **News recommandation**. Among others (*e.g* clinical trials, ..).



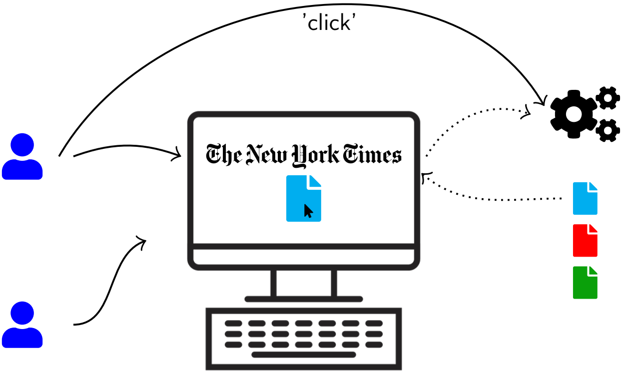
A motivating example (1/3)

- **News recommandation**. Among others (*e.g* clinical trials, ..).



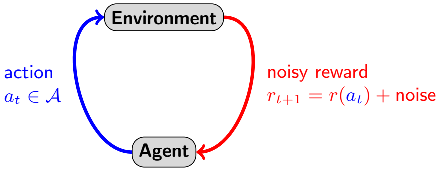
A motivating example (1/3)

- **News recommandation**. Among others (*e.g* clinical trials, ..).



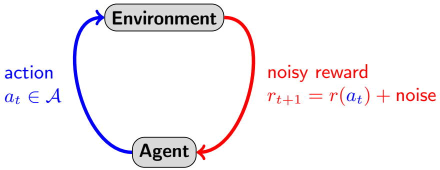
A motivating example (2/3)

- **Formalization**: stochastic bandit framework.



A motivating example (2/3)

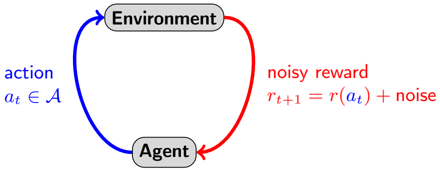
- **Formalization**: stochastic bandit framework.



- **Goal**: minimize $\text{Regret}(T) = T \max_a r(a) - \sum_{t=1}^{T} r(a_t)$.

A motivating example (2/3)

- **Formalization**: stochastic bandit framework.

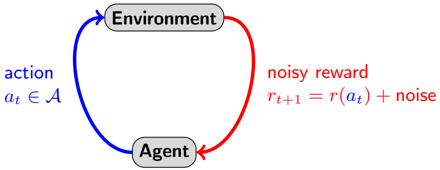


- **Goal**: minimize $\mathrm{Regret}(T) = T \max_a r(a) - \sum_{t=1}^{T} r(a_t)$.

- **Challenge**: observe (noisy) reward only for the action we play.

A motivating example (2/3)

- **Formalization**: stochastic bandit framework.



- **Goal**: minimize $\text{Regret}(T) = T \max_a r(a) - \sum_{t=1}^{T} r(a_t)$.

- **Challenge**: observe (noisy) reward only for the action we play.
  - exploration/exploitation dilemma.

A motivating example (3/3)

Modelling the reward function.

A motivating example (3/3)

Modelling the reward function.

- **Challenges.**

(1) Large action space $\mathcal{A}$ but interrelated payoffs.

A motivating example (3/3)

Modelling the reward function.

- **Challenges.**

(1) Large action space $\mathcal{A}$ but interrelated payoffs.

(2) Learn from continuous / binary / ordinal / categorical feedback.

A motivating example (3/3)

Modelling the reward function.

- **Challenges.**

(1) Large action space $\mathcal{A}$ but interrelated payoffs.

(2) Learn from continuous / binary / ordinal / categorical feedback.

(3) Theoretical guarantees: $\text{Regret}(T) = o(T)$.

A motivating example (3/3)

Modelling the reward function.

- **Challenges.**

(1) Large action space $\mathcal{A}$ but interrelated payoffs.
(2) Learn from continuous / binary / ordinal / categorical feedback.
(3) Theoretical guarantees: $\text{Regret}(T) = o(T)$.

- Solution for (1): establish **structure** through parametric model.

A motivating example (3/3)

Modelling the reward function.

- **Challenges.**

(1) Large action space $\mathcal{A}$ but interrelated payoffs.
(2) Learn from continuous / binary / ordinal / categorical feedback.
(3) Theoretical guarantees: Regret$(T) = o(T)$.

- Solution for (1): establish **structure** through parametric model.
  - Embed $\mathcal{A}$ in $\mathbb{R}^d$ with $d \ll |\mathcal{A}|$ (feature map)

A motivating example (3/3)

Modelling the reward function.

- **Challenges.**

(1) Large action space $\mathcal{A}$ but interrelated payoffs.
(2) Learn from continuous / binary / ordinal / categorical feedback.
(3) Theoretical guarantees: Regret$(T) = o(T)$.

- Solution for (1): establish **structure** through parametric model.
  - Embed $\mathcal{A}$ in $\mathbb{R}^d$ with $d \ll |\mathcal{A}|$ (feature map)
  - Reward function belongs to known parametric family:

$$\mathbb{E}\left[r_{t+1} | a_t\right] = f_{\theta_\star}(a_t) \quad \text{where } f_{\theta_\star} \in \left\{ f_\theta : \mathbb{R}^d \mapsto \mathbb{R},\, \theta \in \Theta \right\} ,$$

where $\theta_\star$ is shared but **unknown**.

Generalized Linear Bandits *[Filippi et al. 2010]*

- **Expected reward** follows a Generalized Linear model:

$$\boxed{\mathbb{E}\left[r_{t+1} | a_t\right] = \mu({a_t}^\top \theta_\star)}$$

where μ is continuously differentiable, strictly increasing.

Generalized Linear Bandits *[Filippi et al. 2010]*

- **Expected reward** follows a Generalized Linear model:

$$\boxed{\mathbb{E}\left[r_{t+1}|a_t\right] = \mu({a_t}^\top\theta_\star)}$$

where μ is continuously differentiable, strictly increasing.

- **Reward distribution.** Exponential family with underlying **linear** structure:

$$d\mathbb{P}(r|a) \propto \exp(r\,a^\top\theta_\star - b(a^\top\theta_\star))d\nu(r)$$

covers Gaussian, Bernoulli, Poisson, .. distributions.

Generalized Linear Bandits *[Filippi et al. 2010]*

- **Expected reward** follows a Generalized Linear model:

$$\boxed{\mathbb{E}\left[r_{t+1}|a_t\right] = \mu({a_t}^\top\theta_\star)}$$

where μ is continuously differentiable, strictly increasing.

- **Reward distribution.** Exponential family with underlying **linear** structure:

$$d\mathbb{P}(r|a) \propto \exp(r\, a^\top\theta_\star - b(a^\top\theta_\star))d\nu(r)$$

covers Gaussian, Bernoulli, Poisson, .. distributions. $\rightarrow$ challenge (2).

Generalized Linear Bandits *[Filippi et al. 2010]*

- **Expected reward** follows a Generalized Linear model:

$$\boxed{\mathbb{E}\left[r_{t+1}|a_t\right] = \mu({a_t}^\top \theta_\star)}$$

where μ is continuously differentiable, strictly increasing.

- **Reward distribution.** Exponential family with underlying **linear** structure:

$$d\mathbb{P}(r|a) \propto \exp(r\, a^\top \theta_\star - b(a^\top \theta_\star))d\nu(r)$$

covers Gaussian, Bernoulli, Poisson, .. distributions. $\rightarrow$ challenge (2).

- **Learnability.** Maximum-likelihood principle

$$\hat{\theta}_t := \operatorname{argmin}_\theta \sum_{s=1}^{t-1} -\log d\mathbb{P}(r_{t+1}|a_t)/d\nu(r) + \lambda\|\theta\|^2/2 \ ,$$

Generalized Linear Bandits *[Filippi et al. 2010]*

- **Expected reward** follows a Generalized Linear model:

$$\boxed{\mathbb{E}\left[r_{t+1}|a_t\right] = \mu({a_t}^\top \theta_\star)}$$

where μ is continuously differentiable, strictly increasing.

- **Reward distribution.** Exponential family with underlying **linear** structure:

$$d\mathbb{P}(r|a) \propto \exp(r\, a^\top \theta_\star - b(a^\top \theta_\star))d\nu(r)$$

covers Gaussian, Bernoulli, Poisson, .. distributions. $\rightarrow$ challenge (2).

- **Learnability.** Maximum-likelihood principle $\rightarrow$ challenge (3)

$$\hat{\theta}_t := \mathrm{argmin}_\theta \sum_{s=1}^{t-1} -\log d\mathbb{P}(r_{t+1}|a_t)/d\nu(r) + \lambda\|\theta\|^2/2 \; ,$$

An illustration: the Logistic Bandit

- **Logistic Bandit.** Structured **binary** feedback:

$$r_{t+1} \sim \text{Bernoulli}(\mu({a_t}^\top \theta_\star))$$

where $\mu(z) = (1 + \exp(-z))^{-1}$ is the logistic function.

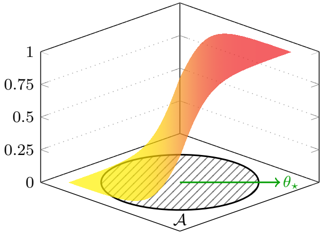
An illustration: the Logistic Bandit

- **Logistic Bandit.** Structured **binary** feedback:

$$r_{t+1} \sim \text{Bernoulli}(\mu(a_t{}^\top \theta_\star))$$

where $\mu(z) = (1 + \exp(-z))^{-1}$ is the logistic function.

- Two-dimensional illustration:



$$\mathbb{E}\left[r_{t+1} \middle| a_t\right] = (1 + \exp(-a_t{}^\top \theta_\star))^{-1}$$

Generalized Linear Bandits: beyond linearity

- **Linear Bandit** (LB). Special case with $\mu = \mathrm{Id}$:

$$\boxed{\mathbb{E}[r_{t+1} | a_t] = a_t^\top \theta_\star \ .}$$

Generalized Linear Bandits: beyond linearity

- **Linear Bandit** (LB). Special case with $\mu = \text{Id}$:

$$\boxed{\mathbb{E}[r_{t+1} | a_t] = {a_t}^\top \theta_\star \ .}$$

  - Well-understood: *[Auer. 2002, Dani et al. 2008, Abbasi-Yadkori et al. 2011, ..]*.
  - Minimax-optimal and efficient algorithms:

$$\text{Regret}(T) = \widetilde{\mathcal{O}}(d\sqrt{T}) \ .$$

Generalized Linear Bandits: beyond linearity

- **Linear Bandit** (LB). Special case with $\mu = \text{Id}$:

$$\boxed{\mathbb{E}[r_{t+1} | a_t] = {a_t}^\top \theta_\star \ .}$$

  - Well-understood: *[Auer. 2002, Dani et al. 2008, Abbasi-Yadkori et al. 2011, ..]*.
  - Minimax-optimal and efficient algorithms:

$$\text{Regret}(T) = \widetilde{\mathcal{O}}(d\sqrt{T}) \ .$$

- **GLBs**. Beyond linearity;

$$\boxed{\mathbb{E}[r_{t+1} | a_t] = \boldsymbol{\mu}({a_t}^\top \theta_\star) \ .}$$

Generalized Linear Bandits: beyond linearity

- **Linear Bandit** (LB). Special case with $\mu = \text{Id}$:

$$\boxed{\mathbb{E}[r_{t+1}|a_t] = a_t^\top \theta_\star \ .}$$

  - Well-understood: *[Auer. 2002, Dani et al. 2008, Abbasi-Yadkori et al. 2011, ..]*.
  - Minimax-optimal and efficient algorithms:

$$\text{Regret}(T) = \widetilde{\mathcal{O}}(d\sqrt{T}) \ .$$

- **GLBs**. Beyond linearity;

$$\boxed{\mathbb{E}[r_{t+1}|a_t] = \boldsymbol{\mu}(a_t^\top \theta_\star) \ .}$$

  - minimalistic **non-linear** extension of LB.

Generalized Linear Bandits: beyond linearity

- **Linear Bandit** (LB). Special case with $\mu = \mathrm{Id}$:

$$\boxed{\mathbb{E}[r_{t+1}|a_t] = {a_t}^\top \theta_\star \ .}$$

  - Well-understood: *[Auer. 2002, Dani et al. 2008, Abbasi-Yadkori et al. 2011, ..]*.
  - Minimax-optimal and efficient algorithms:

$$\text{Regret}(T) = \widetilde{\mathcal{O}}(d\sqrt{T}) \ .$$

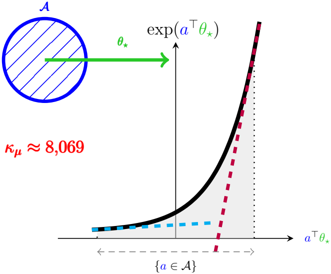
- **GLBs**. Beyond linearity;

$$\boxed{\mathbb{E}[r_{t+1}|a_t] = \boldsymbol{\mu}({a_t}^\top \theta_\star) \ .}$$

  - minimalistic **non-linear** extension of LB.
  - first step towards richer reward signal.

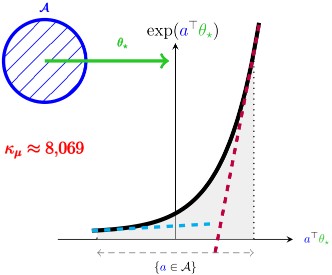
GLBs: quantifying non-linearity

- **Level of non-linearity** = conditioning of the reward signal = $\boldsymbol{\kappa}_\mu(\theta_\star, \mathcal{A})$.



GLBs: quantifying non-linearity

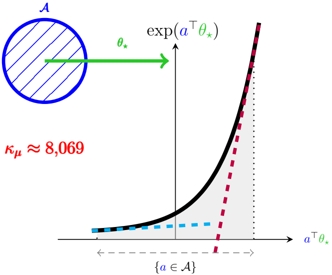
- **Level of non-linearity** = conditioning of the reward signal = $\boldsymbol{\kappa}_\mu(\theta_\star, \mathcal{A})$.



$$\boldsymbol{\kappa}_\mu(\theta_\star, \mathcal{A}) := \frac{\max_{a\in\mathcal{A}} \dot{\mu}(a^\top\theta_\star)}{\min_{a\in\mathcal{A}} \dot{\mu}(a^\top\theta_\star)} =: \frac{\mathbf{L}_\mu}{\boldsymbol{\ell}_\mu}$$

GLBs: quantifying non-linearity

- **Level of non-linearity** = conditioning of the reward signal = $\boldsymbol{\kappa}_\mu(\theta_\star, \mathcal{A})$.

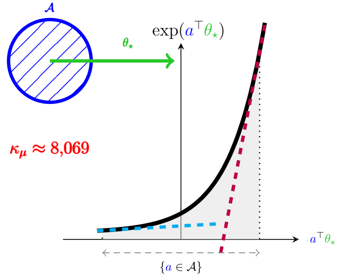


$$\boldsymbol{\kappa}_\mu(\theta_\star, \mathcal{A}) := \frac{\max_{a\in\mathcal{A}} \dot{\mu}(a^\top\theta_\star)}{\min_{a\in\mathcal{A}} \dot{\mu}(a^\top\theta_\star)} =: \frac{\mathbf{L}_\mu}{\boldsymbol{\ell}_\mu}$$

- the more non-linear the reward signal, the larger $\boldsymbol{\kappa}_\mu$.

GLBs: quantifying non-linearity

- **Level of non-linearity** = conditioning of the reward signal = $\boldsymbol{\kappa}_\mu(\theta_\star, \mathcal{A})$.

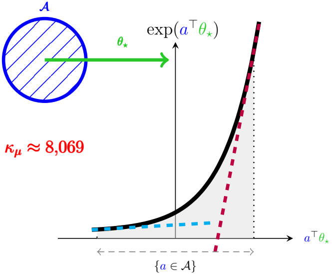


$$\boxed{\boldsymbol{\kappa}_\mu(\theta_\star, \mathcal{A}) := \frac{\max_{a\in\mathcal{A}} \dot{\mu}(a^\top\theta_\star)}{\min_{a\in\mathcal{A}} \dot{\mu}(a^\top\theta_\star)} =: \frac{\mathbf{L}_\mu}{\boldsymbol{\ell}_\mu}}$$

- the more non-linear the reward signal, the larger $\boldsymbol{\kappa}_\mu$.
- "distance" from the linear model ($\boldsymbol{\kappa}_\mu = 1$ for Linear Bandit).

GLBs: quantifying non-linearity

- **Level of non-linearity** = conditioning of the reward signal = $\boldsymbol{\kappa}_\mu(\theta_\star, \mathcal{A})$.



$$\boldsymbol{\kappa}_\mu(\theta_\star, \mathcal{A}) := \frac{\max_{a\in\mathcal{A}} \dot{\mu}(a^\top \theta_\star)}{\min_{a\in\mathcal{A}} \dot{\mu}(a^\top \theta_\star)} =: \frac{\mathbf{L}_\mu}{\boldsymbol{\ell}_\mu}$$

- the more non-linear the reward signal, the larger $\boldsymbol{\kappa}_\mu$.
- "distance" from the linear model ($\boldsymbol{\kappa}_\mu = 1$ for Linear Bandit).
- numerically very large ($\boldsymbol{\kappa}_\mu \propto \exp(\|\theta_\star\|) \approx 10^3!$)

Previous work, limitations, contributions.

GLBs: previous work

- First studied in the seminal work of *[Filippi et al. 2010]*.
  - many extensions: *[Li et al. 2017, Jun et al. 2017, Kveton et al. 2019, ..]*

GLBs: previous work

- First studied in the seminal work of *[Filippi et al. 2010]*.
  - many extensions: *[Li et al. 2017, Jun et al. 2017, Kveton et al. 2019, ..]*

- **Regret upper-bound.** With high probability:

$$\text{Regret}(T) = \widetilde{\mathcal{O}}(d\sqrt{T}) \; .$$

GLBs: previous work

- First studied in the seminal work of *[Filippi et al. 2010]*.
  - many extensions: *[Li et al. 2017, Jun et al. 2017, Kveton et al. 2019, ..]*

- **Regret upper-bound.** With high probability:

$$\text{Regret}(T) = \widetilde{\mathcal{O}}(d\sqrt{T}) \ .$$

✅ Extend LB tools to generic GLBs.

GLBs: previous work

- First studied in the seminal work of *[Filippi et al. 2010]*.
  - many extensions: *[Li et al. 2017, Jun et al. 2017, Kveton et al. 2019, ..]*

- **Regret upper-bound.** With high probability:

$$\text{Regret}(T) = \widetilde{\mathcal{O}}(\boldsymbol{\kappa}_\mu d\sqrt{T}) \ .$$

✅ Extend LB tools to generic GLBs.

GLBs: previous work

- First studied in the seminal work of *[Filippi et al. 2010]*.
  - many extensions: *[Li et al. 2017, Jun et al. 2017, Kveton et al. 2019, ..]*

- **Regret upper-bound.** With high probability:

$$\text{Regret}(T) = \widetilde{\mathcal{O}}(\boldsymbol{\kappa}_\mu d\sqrt{T}) \ .$$

✅ Extend LB tools to generic GLBs.

❌ Large regret upper bound.

GLBs: previous work

- First studied in the seminal work of *[Filippi et al. 2010]*.
  - many extensions: *[Li et al. 2017, Jun et al. 2017, Kveton et al. 2019, ..]*

- **Regret upper-bound.** With high probability:

$$\text{Regret}(T) = \widetilde{\mathcal{O}}(\boldsymbol{\kappa}_\mu d\sqrt{T}) \ .$$

  - ✅ Extend LB tools to generic GLBs.
  - ❌ Large regret upper bound.
  - ❌ Over-exploratory algorithms, poor empirical performance.

GLBs: previous work

- First studied in the seminal work of *[Filippi et al. 2010]*.
  - many extensions: *[Li et al. 2017, Jun et al. 2017, Kveton et al. 2019, ..]*

- **Regret upper-bound.** With high probability:

$$\text{Regret}(T) = \widetilde{\mathcal{O}}(\boldsymbol{\kappa}_\mu d\sqrt{T}) \; .$$

  - ✔ Extend LB tools to generic GLBs.
  - ✖ Large regret upper bound.
  - ✖ Over-exploratory algorithms, poor empirical performance.

- **Learning-theoretic**: non-linearity is **detrimental**!

GLBs: previous work

- First studied in the seminal work of *[Filippi et al. 2010]*.
  - many extensions: *[Li et al. 2017, Jun et al. 2017, Kveton et al. 2019, ..]*

- **Regret upper-bound.** With high probability:

$$\text{Regret}(T) = \widetilde{\mathcal{O}}(\boldsymbol{\kappa}_\mu d\sqrt{T})\ .$$

  - ✅ Extend LB tools to generic GLBs.
  - ❌ Large regret upper bound.
  - ❌ Over-exploratory algorithms, poor empirical performance.

- **Learning-theoretic**: non-linearity is **detrimental**!
  - the more non-linear the problem, the worse the performance.

GLBs: previous approach (1/2)

- **Algorithmic design.** Two main ingredients; at each round t:

GLBs: previous approach (1/2)

- **Algorithmic design.** Two main ingredients; at each round t:

  1. Confidence set $\mathcal{E}_t(\delta)$ for $\theta_\star$;

$$\mathbb{P}\Big(\forall t \geq 1,\, \theta_\star \in \mathcal{E}_t(\delta)\Big) \geq 1 - \delta\ .$$

GLBs: previous approach (1/2)

- **Algorithmic design.** Two main ingredients; at each round t:
  1. Confidence set $\mathcal{E}_t(\delta)$ for $\theta_\star$;

$$\mathbb{P}\Big(\forall t \geq 1,\, \theta_\star \in \mathcal{E}_t(\delta)\Big) \geq 1 - \delta\ .$$

  2. Optimism in face of uncertainty:

$$\text{play } a_t = \text{argmax}_{a\in\mathcal{A}} \max_{\theta\in\mathcal{E}_t(\delta)} \mu(a^\top\theta)\ .$$

GLBs: previous approach (1/2)

- **Algorithmic design.** Two main ingredients; at each round t:
  1. Confidence set $\mathcal{E}_t(\delta)$ for $\theta_\star$;
  $$\mathbb{P}\Big(\forall t \geq 1,\ \theta_\star \in \mathcal{E}_t(\delta)\Big) \geq 1 - \delta\ .$$
  2. Optimism in face of uncertainty:
  $$\text{play } a_t = \text{argmax}_{a\in\mathcal{A}} \max_{\theta\in\mathcal{E}_t(\delta)} \mu(a^\top\theta)\ .$$

- **Confidence set** from previous works; with $\boldsymbol{V}_t = \sum_{s=1}^{t-1} a_s a_s^\top + \lambda \boldsymbol{I}_d$:

$$\boxed{\mathcal{E}_t(\delta) = \left\{\theta,\ \left\|\theta - \hat{\theta}_t\right\|_{\boldsymbol{V}_t} \leq \sqrt{d\log(t/\delta)}/\boldsymbol{\ell}_\mu\right\}}$$

GLBs: previous approach (1/2)

- **Algorithmic design.** Two main ingredients; at each round t:
  1. Confidence set $\mathcal{E}_t(\delta)$ for $\theta_\star$;
  $$\mathbb{P}\Big(\forall t \geq 1,\ \theta_\star \in \mathcal{E}_t(\delta)\Big) \geq 1 - \delta\ .$$
  2. Optimism in face of uncertainty:
  $$\text{play } a_t = \mathrm{argmax}_{a \in \mathcal{A}} \max_{\theta \in \mathcal{E}_t(\delta)} \mu(a^\top \theta)\ .$$

- **Confidence set** from previous works; with $\boldsymbol{V}_t = \sum_{s=1}^{t-1} a_s a_s^\top + \lambda \boldsymbol{I}_d$:

$$\boxed{\mathcal{E}_t(\delta) = \left\{\theta,\ \left\|\theta - \hat{\theta}_t\right\|_{\boldsymbol{V}_t} \leq \sqrt{d \log(t/\delta)}/\boldsymbol{\ell}_\mu\right\}}$$

  - Radius $\propto \boldsymbol{\ell}_\mu^{-1} \Rightarrow$ large confidence region $\Rightarrow$ aggressive exploration.

GLBs: previous approach (1/2)

- **Algorithmic design.** Two main ingredients; at each round t:
  1. Confidence set $\mathcal{E}_t(\delta)$ for $\theta_\star$;
  $$\mathbb{P}\Big(\forall t \geq 1,\ \theta_\star \in \mathcal{E}_t(\delta)\Big) \geq 1 - \delta\ .$$
  2. Optimism in face of uncertainty:
  $$\text{play } a_t = \text{argmax}_{a\in\mathcal{A}} \max_{\theta\in\mathcal{E}_t(\delta)} \mu(a^\top\theta)\ .$$

- **Confidence set** from previous works; with $\boldsymbol{V}_t = \sum_{s=1}^{t-1} a_s a_s^\top + \lambda \boldsymbol{I}_d$:

$$\boxed{\mathcal{E}_t(\delta) = \left\{\theta,\ \left\|\theta - \hat{\theta}_t\right\|_{\boldsymbol{V}_t} \leq \sqrt{d\log(t/\delta)/\boldsymbol{\ell}_\mu}\right\}}$$

  - Radius $\propto \boldsymbol{\ell}_\mu{}^{-1}$ $\Rightarrow$ large confidence region $\Rightarrow$ aggressive exploration.
  - Learn as slow as in the flattest region, in every direction.

GLBs: previous approach (2/2)

- **Analysis.** Upper linear bound:

$$\text{Regret}(T) = \sum_{t=1}^{T} \mu(a_\star^\top \theta_\star) - \mu(a_t^\top \theta_\star)$$

GLBs: previous approach (2/2)

- **Analysis.** Upper linear bound:

$$\mathrm{Regret}(T) \leq \sum_{t=1}^{T} \mu(a_t^\top \theta_t) - \mu(a_t^\top \theta_\star) \qquad\qquad (\theta_t \text{ opt. param})$$

GLBs: previous approach (2/2)

- **Analysis.** Upper linear bound:

$$
\begin{aligned}
\text{Regret}(T) &\leq \sum_{t=1}^{T} \mu(a_t^\top \theta_t) - \mu(a_t^\top \theta_\star) && (\theta_t \text{ opt. param}) \\
&\leq \mathbf{L}_\mu \sum_{t=1}^{T} a_t^\top (\theta_t - \theta_\star)
\end{aligned}
$$

GLBs: previous approach (2/2)

- **Analysis.** Upper linear bound:

$$
\begin{aligned}
\text{Regret}(T) &\leq \sum_{t=1}^{T} \mu(a_t^\top \theta_t) - \mu(a_t^\top \theta_\star) && (\theta_t \text{ opt. param}) \\
&\leq \mathbf{L}_\mu \sum_{t=1}^{T} a_t^\top (\theta_t - \theta_\star) \\
&\leq \mathbf{L}_\mu \sum_{t=1}^{T} \|a_t\|_{\boldsymbol{V}_t^{-1}} \|\theta_t - \theta_\star\|_{\boldsymbol{V}_t} && (\text{C.S ineq.})
\end{aligned}
$$

GLBs: previous approach (2/2)

- **Analysis.** Upper linear bound:

$$
\begin{aligned}
\mathrm{Regret}(T) &\leq \sum_{t=1}^{T} \mu(a_t^\top \theta_t) - \mu(a_t^\top \theta_\star) && (\theta_t \text{ opt. param}) \\
&\leq \mathbf{L}_\mu \sum_{t=1}^{T} a_t^\top (\theta_t - \theta_\star) \\
&\leq \mathbf{L}_\mu \sum_{t=1}^{T} \|a_t\|_{\boldsymbol{V}_t^{-1}} \|\theta_t - \theta_\star\|_{\boldsymbol{V}_t} && \text{(C.S ineq.)} \\
&\leq \mathbf{L}_\mu / \boldsymbol{\ell}_\mu \sqrt{d \log(T/\delta)} \sum_{t=1}^{T} \|a_t\|_{\boldsymbol{V}_t^{-1}}
\end{aligned}
$$

GLBs: previous approach (2/2)

- **Analysis.** Upper linear bound:

$$\begin{aligned}
\text{Regret}(T) &\leq \sum_{t=1}^{T} \mu(a_t^\top \theta_t) - \mu(a_t^\top \theta_\star) && (\theta_t \text{ opt. param}) \\
&\leq \mathbf{L}_\mu \sum_{t=1}^{T} a_t^\top(\theta_t - \theta_\star) \\
&\leq \mathbf{L}_\mu \sum_{t=1}^{T} \|a_t\|_{\boldsymbol{V}_t^{-1}} \|\theta_t - \theta_\star\|_{\boldsymbol{V}_t} && (\text{C.S ineq.}) \\
&\leq \mathbf{L}_\mu / \boldsymbol{\ell}_\mu \sqrt{d \log(T/\delta)} \sum_{t=1}^{T} \|a_t\|_{\boldsymbol{V}_t^{-1}} \\
&= \boldsymbol{\kappa}_\mu d\sqrt{T} \log(T/\delta)
\end{aligned}$$

GLBs: previous approach (2/2)

- **Analysis.** Upper linear bound:

$$
\begin{aligned}
\text{Regret}(T) &\le \sum_{t=1}^{T} \mu(a_t^\top \theta_t) - \mu(a_t^\top \theta_\star) && (\theta_t \text{ opt. param}) \\
&\le \mathbf{L}_\mu \sum_{t=1}^{T} a_t^\top (\theta_t - \theta_\star) \\
&\le \mathbf{L}_\mu \sum_{t=1}^{T} \|a_t\|_{\boldsymbol{V}_t^{-1}} \|\theta_t - \theta_\star\|_{\boldsymbol{V}_t} && (\text{C.S ineq.}) \\
&\le \mathbf{L}_\mu / \boldsymbol{\ell}_\mu \sqrt{d \log(T/\delta)} \sum_{t=1}^{T} \|a_t\|_{\boldsymbol{V}_t^{-1}} \\
&= \boldsymbol{\kappa}_\mu d \sqrt{T} \log(T/\delta)
\end{aligned}
$$

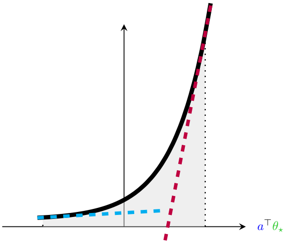
- We pay errors as in the sharpest linear case.

GLBs: previous approach (2/2)

- **Analysis.** Upper linear bound:

$$\begin{aligned}
\text{Regret}(T) &\leq \sum_{t=1}^{T} \mu(a_t^\top \theta_t) - \mu(a_t^\top \theta_\star) && (\theta_t \text{ opt. param}) \\
&\leq \mathbf{L}_\mu \sum_{t=1}^{T} a_t^\top (\theta_t - \theta_\star) \\
&\leq \mathbf{L}_\mu \sum_{t=1}^{T} \|a_t\|_{\boldsymbol{V}_t^{-1}} \|\theta_t - \theta_\star\|_{\boldsymbol{V}_t} && \text{(C.S ineq.)} \\
&\leq \mathbf{L}_\mu / \boldsymbol{\ell}_\mu \sqrt{d \log(T/\delta)} \sum_{t=1}^{T} \|a_t\|_{\boldsymbol{V}_t^{-1}} \\
&= \boldsymbol{\kappa}_\mu d \sqrt{T} \log(T/\delta)
\end{aligned}$$

- We pay errors as in the sharpest linear case.
- Worst-case errors / worst-case learning.



GLBs: our approach

Local treatment of non-linearity for **improved** regret bounds.

GLBs: our approach

Local treatment of non-linearity for **improved** regret bounds.

- **New confidence set**.
  - sensitive to effective reward sensitivity ($\neq$ worst-case)
  - provably tighter.

GLBs: our approach

Local treatment of non-linearity for **improved** regret bounds.

- **New confidence set**.
  - sensitive to effective reward sensitivity ($\neq$ worst-case)
  - provably tighter.
- **Locality-sensitive analysis** under generalized self-concordance *[Bach. 2010]*:

$$|\ddot{\mu}| \leq c\dot{\mu}$$

  - allows exact Taylor control with local quantities.

GLBs: our approach

Local treatment of non-linearity for **improved** regret bounds.

- **New confidence set**.
  - sensitive to effective reward sensitivity ($\neq$ worst-case)
  - provably tighter.

- **Locality-sensitive analysis** under generalized self-concordance *[Bach. 2010]*:

$$|\ddot{\mu}| \leq c\dot{\mu}$$

  - allows exact Taylor control with local quantities.
  - not restrictive: Logistic and Poisson Bandits (c=1).

Variance-Sensitive Confidence Sets for GLBs

Improved confidence set: asymptotic intuition

- **Objective.** Dependence to effective reward sensitivity:
  - measured through the **variance** of the reward signal:

$$\mathbb{V}\mathrm{ar}(r_{t+1}|a_t) = \dot{\mu}({a_t}^\top \theta_\star) \; .$$

Improved confidence set: asymptotic intuition

- **Objective.** Dependence to effective reward sensitivity:
  - measured through the **variance** of the reward signal:

$$\mathbb{V}\text{ar}(r_{t+1}|a_t) = \dot{\mu}({a_t}^\top \theta_\star) \ .$$

  - Variance-sensitive concentration tools.

Improved confidence set: asymptotic intuition

- **Objective.** Dependence to effective reward sensitivity:
  - measured through the **variance** of the reward signal:

$$\mathbb{V}\text{ar}(r_{t+1}|a_t) = \dot{\mu}({a_t}^\top \theta_\star) \; .$$

  - Variance-sensitive concentration tools.

- **Asymptotic intuition.** Let $\boldsymbol{H}_t(\theta) = \sum_{s=1}^{t} \dot{\mu}(a_s^\top \theta) a_s a_s^\top$.

Improved confidence set: asymptotic intuition

- **Objective.** Dependence to effective reward sensitivity:
  - measured through the **variance** of the reward signal:

$$\mathbb{V}\mathrm{ar}(r_{t+1}|a_t) = \dot{\mu}({a_t}^\top \theta_\star) \; .$$

  - Variance-sensitive concentration tools.

- **Asymptotic intuition.** Let $\boldsymbol{H}_t(\theta) = \sum_{s=1}^{t} \dot{\mu}(a_s^\top \theta) a_s a_s^\top$.

$$\lim_{t\to\infty} \mathbb{P}\left(\|\hat{\theta}_t - \theta_\star\|^2_{\boldsymbol{H}_t(\theta_\star)} \leq d \log(1/\delta) \right) \geq 1 - \delta \; .$$

  under **random** design.

Improved confidence set: asymptotic intuition

- **Objective.** Dependence to effective reward sensitivity:
  - measured through the **variance** of the reward signal:

$$\mathbb{V}\mathrm{ar}(r_{t+1}|a_t) = \dot{\mu}(a_t{}^\top \theta_\star) \ .$$

  - Variance-sensitive concentration tools.

- **Asymptotic intuition.** Let $\boldsymbol{H}_t(\theta) = \sum_{s=1}^{t} \dot{\mu}(a_s^\top \theta) a_s a_s^\top$.

$$\lim_{t\to\infty} \mathbb{P}\left(\|\hat{\theta}_t - \theta_\star\|^2_{\boldsymbol{H}_t(\theta_\star)} \leq d \log(1/\delta) \right) \geq 1 - \delta \ .$$

  under **random** design.

- **Challenge.** Generalization for:
  - finite-time (non-asymptotic).
  - **adaptive** design ($\{a_1, \ldots, a_s\}_s$ are not independent).

A novel concentration inequality (1/2)

Theorem (F., Abeille, Calauzènes and Fercoq, 2020.)

Let $\{\mathcal{F}_t\}_{t\in\mathbb{N}}$ be a filtration and:

- $\{a_t\}_{t\in\mathbb{N}}$ a $\mathcal{F}_t$-measurable stochastic process.

A novel concentration inequality (1/2)

Theorem (F., Abeille, Calauzènes and Fercoq, 2020.)

Let $\{\mathcal{F}_t\}_{t\in\mathbb{N}}$ be a filtration and:

- $\{a_t\}_{t\in\mathbb{N}}$ a $\mathcal{F}_t$-measurable stochastic process.
- $\{\eta_{t+1}\}_{t\in\mathbb{N}}$ a $\mathcal{F}_{t+1}$-measurable martingale difference sequence s.t:
    - $|\eta_{t+1}| \leq \sigma$ almost surely and ${\upsilon_t}^2 := \mathbb{V}\text{ar}(\eta_{t+1}|\mathcal{F}_t)$.

A novel concentration inequality (1/2)

Theorem (F., Abeille, Calauzènes and Fercoq, 2020.)

Let $\{\mathcal{F}_t\}_{t\in\mathbb{N}}$ be a filtration and:

- $\{a_t\}_{t\in\mathbb{N}}$ a $\mathcal{F}_t$-measurable stochastic process.
- $\{\eta_{t+1}\}_{t\in\mathbb{N}}$ a $\mathcal{F}_{t+1}$-measurable martingale difference sequence s.t:
  - $|\eta_{t+1}| \leq \sigma$ almost surely and ${\upsilon_t}^2 := \mathbb{V}\mathrm{ar}(\eta_{t+1}|\mathcal{F}_t)$.
- Let $\lambda > 0$ and define for $t \in \mathbb{N}$:

$$S_{t+1} := \sum_{s=1}^{t} \eta_{s+1} a_s \text{ and } \boldsymbol{H}_t := \sum_{s=1}^{t} {\upsilon_s}^2 a_s a_s^\top + \lambda \boldsymbol{I}_d$$

A novel concentration inequality (1/2)

Theorem (F., Abeille, Calauzènes and Fercoq, 2020.)

For $\delta \in (0, 1]$ *the event:*

$$\forall t \geq 1, \ \|S_{t+1}\|_{\boldsymbol{H}_t^{-1}} \leq \frac{\sqrt{\lambda}}{2\sigma} + \frac{2\sigma}{\sqrt{\lambda}} d \log\left(\frac{4(1+\sigma^2 t/(d\lambda))}{\delta}\right) ,$$

holds with probability at least $1 - \delta$.

Let $\{\mathcal{F}_t\}_{t\in\mathbb{N}}$ be a filtration and:

- $\{a_t\}_{t\in\mathbb{N}}$ a $\mathcal{F}_t$-measurable stochastic process.
- $\{\eta_{t+1}\}_{t\in\mathbb{N}}$ a $\mathcal{F}_{t+1}$-measurable martingale difference sequence s.t:
  - $|\eta_{t+1}| \leq \sigma$ almost surely and ${\upsilon_t}^2 := \mathbb{V}\text{ar}(\eta_{t+1}|\mathcal{F}_t)$.
- Let $\lambda > 0$ and define for $t \in \mathbb{N}$:

$$S_{t+1} := \sum_{s=1}^{t} \eta_{s+1} a_s \text{ and } \boldsymbol{H}_t := \sum_{s=1}^{t} {\upsilon_s}^2 a_s a_s^\top + \lambda \boldsymbol{I}_d$$

A novel concentration inequality (1/2)

Theorem (F., Abeille, Calauzènes and Fercoq, 2020.)

For $\delta \in (0,1]$ the event:

$$\forall t \geq 1, \; \|S_{t+1}\|_{\boldsymbol{H}_t^{-1}} \leq \mathcal{O}\left(\sqrt{d \log(t/\delta)}\right),$$

holds with probability at least $1-\delta$ (setting $\lambda_t = d\log(t/\delta)$).

Let $\{\mathcal{F}_t\}_{t\in\mathbb{N}}$ be a filtration and:

- $\{a_t\}_{t\in\mathbb{N}}$ a $\mathcal{F}_t$-measurable stochastic process.
- $\{\eta_{t+1}\}_{t\in\mathbb{N}}$ a $\mathcal{F}_{t+1}$-measurable martingale difference sequence s.t:
  - $|\eta_{t+1}| \leq \sigma$ almost surely and ${\upsilon_t}^2 := \mathbb{V}\text{ar}(\eta_{t+1}|\mathcal{F}_t)$.
- Let $\lambda > 0$ and define for $t \in \mathbb{N}$:

$$S_{t+1} := \sum_{s=1}^{t} \eta_{s+1} a_s \text{ and } \boldsymbol{H}_t := \sum_{s=1}^{t} {\upsilon_s}^2 a_s a_s^\top + \lambda \boldsymbol{I}_d$$

A novel concentration inequality (2/2)

Theorem (F., Abeille, Calauzènes and Fercoq, 2020.)

For $\delta \in (0,1]$ the event:

$$\forall t \geq 1,\ \|S_{t+1}\|_{\boldsymbol{H}_t^{-1}} \leq \mathcal{O}\left(\sqrt{d\log(t/\delta)}\right),$$

holds with probability at least $1-\delta$ (setting $\lambda_t = d\log(t/\delta)$).

- **Sketch of proof.**

A novel concentration inequality (2/2)

Theorem (F., Abeille, Calauzènes and Fercoq, 2020.)

For $\delta \in (0,1]$ *the event:*

$$\forall t \geq 1, \; \|S_{t+1}\|_{\boldsymbol{H}_t^{-1}} \leq \mathcal{O}\left(\sqrt{d\log(t/\delta)}\right),$$

holds with probability at least $1-\delta$ *(setting* $\lambda_t = d\log(t/\delta)$*).*

- **Sketch of proof.**
  - Pseudo-maximization (methods of mixture) *[de la Peña. 2007]*.

A novel concentration inequality (2/2)

Theorem (F., Abeille, Calauzènes and Fercoq, 2020.)

For $\delta \in (0, 1]$ the event:

$$\forall t \geq 1,\ \|S_{t+1}\|_{\boldsymbol{H}_t^{-1}} \leq \mathcal{O}\left(\sqrt{d \log(t/\delta)}\right) ,$$

holds with probability at least $1 - \delta$ (setting $\lambda_t = d\log(t/\delta)$).

- **Sketch of proof.**
  - Pseudo-maximization (methods of mixture) *[de la Peña. 2007]*.
  - Similar to *[Abbasi-Yadkori et al. 2011]*, different base super-martingale:

$$M_t(\xi) = \xi^\top S_{t+1} - \|\xi\|^2_{\boldsymbol{H}_t} \quad \text{for } \|\xi\| \leq 1 \, .$$

  - Bernstein vs. Hoeffding conditions.

Application to GLBs (1/2)

- Using optimality of $\hat{\theta}_t$ for the regularized log-loss $\mathcal{L}_t^{\lambda_t}$:

$$\forall t \geq 1, \quad \left\| \theta_\star - \hat{\theta}_t \right\|_{\boldsymbol{H}_t(\theta_\star)} \leq \left\| \sum_{s=1}^{t-1} \eta_{s+1} a_s \right\|_{\boldsymbol{H}_t^{-1}(\theta_\star)}$$

where:

$$\eta_{s+1} = r_{s+1} - \mu(a_s^\top \theta_\star) \quad \text{and} \quad \boldsymbol{H}_t(\theta) = \sum_{s=1}^{t} \dot{\mu}(a_s^\top \theta) a_s a_s^\top + \lambda_t \boldsymbol{I}_d$$

Application to GLBs (1/2)

- Using optimality of $\hat{\theta}_t$ for the regularized log-loss $\mathcal{L}_t^{\lambda_t}$:

$$\forall t \geq 1, \quad \left\| \theta_\star - \hat{\theta}_t \right\|_{\boldsymbol{H}_t(\theta_\star)} \leq \left\| \sum_{s=1}^{t-1} \eta_{s+1} a_s \right\|_{\boldsymbol{H}_t^{-1}(\theta_\star)} = \mathcal{O}\left(\sqrt{d \log(t/\delta)} \right)$$

where:

$$\eta_{s+1} = r_{s+1} - \mu(a_s^\top \theta_\star) \quad \text{and} \quad \boldsymbol{H}_t(\theta) = \sum_{s=1}^{t} \dot{\mu}(a_s^\top \theta) a_s a_s^\top + \lambda_t \boldsymbol{I}_d$$

Application to GLBs (1/2)

- Using optimality of $\hat{\theta}_t$ for the regularized log-loss $\mathcal{L}_t^{\lambda_t}$:

$$\forall t \geq 1, \quad \left\| \theta_\star - \hat{\theta}_t \right\|_{\boldsymbol{H}_t(\theta_\star)} \leq \left\| \sum_{s=1}^{t-1} \eta_{s+1} a_s \right\|_{\boldsymbol{H}_t^{-1}(\theta_\star)} = \mathcal{O}\left(\sqrt{d \log(t/\delta)}\right)$$

where:

$$\eta_{s+1} = r_{s+1} - \mu(a_s^\top \theta_\star) \quad \text{and} \quad \boldsymbol{H}_t(\theta) = \sum_{s=1}^{t} \dot{\mu}(a_s^\top \theta) a_s a_s^\top + \lambda_t \boldsymbol{I}_d$$

- New confidence set:

Proposition (F., Abeille, Calauzènes and Fercoq, 2020.)

For $\delta \in (0,1]$ *let:*

$$\mathcal{C}_t(\delta) := \left\{ \left\| \theta - \hat{\theta}_t \right\|_{\boldsymbol{H}_t(\theta)} \leq \mathcal{O}\left(\sqrt{d \log(t/\delta)}\right) \right\} .$$

Then $\mathbb{P}\left(\forall t \geq 1,\, \theta_\star \in \mathcal{C}_t(\delta)\right) \geq 1 - \delta.$

Application to GLBs (2/2)

$$\mathcal{C}_t(\delta) = \left\{ \left\|\theta - \hat{\theta}_t\right\|_{\boldsymbol{H}_t(\theta)} \leq \mathcal{O}\left(\sqrt{d\log(t/\delta)}\right)\right\} , \qquad \text{(ours)}$$

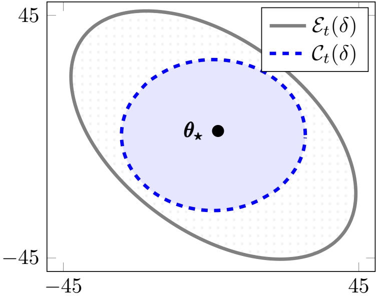
$$\mathcal{E}_t(\delta) = \left\{ \left\|\theta - \hat{\theta}_t\right\|_{\boldsymbol{V}_t} \leq \mathcal{O}\left(\sqrt{d\log(t/\delta)}/\boldsymbol{\ell}_\mu\right)\right\} . \qquad \text{[Filippi et al.]}$$

Application to GLBs (2/2)

$$\mathcal{C}_t(\delta) = \left\{ \left\| \theta - \hat{\theta}_t \right\|_{\boldsymbol{H}_t(\theta)} \leq \mathcal{O}\left(\sqrt{d \log(t/\delta)}\right) \right\}, \qquad \text{(ours)}$$

$$\mathcal{E}_t(\delta) = \left\{ \left\| \theta - \hat{\theta}_t \right\|_{\boldsymbol{V}_t} \leq \mathcal{O}\left(\sqrt{d \log(t/\delta)}/\boldsymbol{\ell}_\mu\right) \right\}. \qquad \textit{[Filippi et al.]}$$

- Illustration for Logistic Bandit:



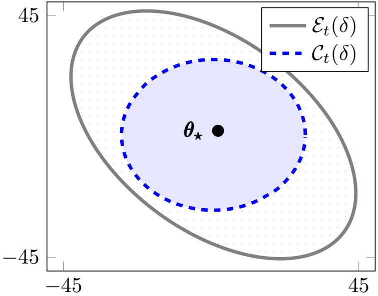
$\|\theta_\star\| = 2$ and $1/\ell_\mu \approx 10$

Application to GLBs (2/2)

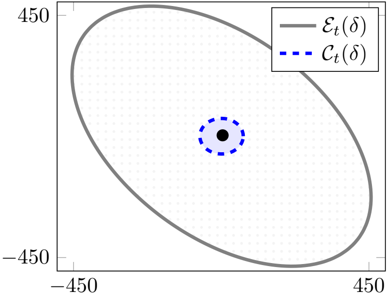
$$\mathcal{C}_t(\delta) = \left\{ \left\| \theta - \hat{\theta}_t \right\|_{\boldsymbol{H}_t(\theta)} \leq \mathcal{O}\left(\sqrt{d\log(t/\delta)}\right) \right\}, \qquad \text{(ours)}$$

$$\mathcal{E}_t(\delta) = \left\{ \left\| \theta - \hat{\theta}_t \right\|_{\boldsymbol{V}_t} \leq \mathcal{O}\left(\sqrt{d\log(t/\delta)}/\boldsymbol{\ell}_\mu\right) \right\}. \qquad \textit{[Filippi et al.]}$$

- Illustration for Logistic Bandit:



$\|\theta_\star\| = 2$ and $1/\ell_\mu \approx 10$

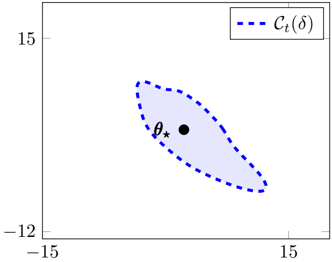


$\|\theta_\star\| = 5$ and $1/\ell_\mu \approx 150$

Extension: convex relaxation

$$\mathcal{C}_t(\delta) = \left\{ \left\| \theta - \hat{\theta}_t \right\|_{\boldsymbol{H}_t(\theta)} \leq \mathcal{O}\left(\sqrt{d \log(t/\delta)}. \right) \right\}$$

- Non-convex, burdensome to manipulate.

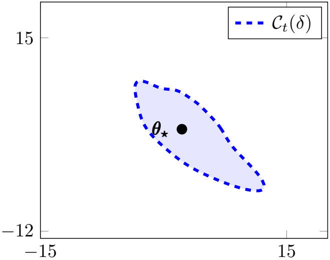


Extension: convex relaxation

$$\mathcal{C}_t(\delta) = \left\{ \left\| \theta - \hat{\theta}_t \right\|_{\boldsymbol{H}_t(\theta)} \leq \mathcal{O}\left(\sqrt{d \log(t/\delta).} \right) \right\}$$

- Non-convex, burdensome to manipulate.
- **Convex** relaxation based on log-loss $\mathcal{L}_t$:

$$\mathcal{C}_t^c(\delta) = \left\{ \mathcal{L}_t(\theta) - \mathcal{L}_t(\hat{\theta}_t) \leq \mathcal{O}\left(d \log(t/\delta) \right) \right\} .$$

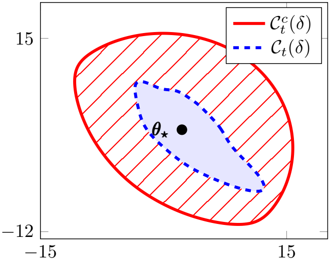


Extension: convex relaxation

$$\mathcal{C}_t(\delta) = \left\{ \left\| \theta - \hat{\theta}_t \right\|_{\boldsymbol{H}_t(\theta)} \leq \mathcal{O}\left(\sqrt{d \log(t/\delta)}. \right) \right\}$$

- Non-convex, burdensome to manipulate.
- **Convex** relaxation based on log-loss $\mathcal{L}_t$:

$$\mathcal{C}_t^c(\delta) = \left\{ \mathcal{L}_t(\theta) - \mathcal{L}_t(\hat{\theta}_t) \leq \mathcal{O}\left(d \log(t/\delta) \right) \right\} .$$

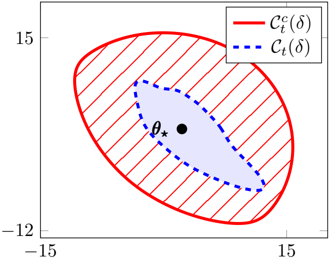


Extension: convex relaxation

$$\mathcal{C}_t(\delta) = \left\{ \left\| \theta - \hat{\theta}_t \right\|_{\boldsymbol{H}_t(\theta)} \leq \mathcal{O}\left(\sqrt{d \log(t/\delta)}.\right) \right\}$$

- Non-convex, burdensome to manipulate.
- **Convex** relaxation based on log-loss $\mathcal{L}_t$:

$$\mathcal{C}_t^c(\delta) = \left\{ \mathcal{L}_t(\theta) - \mathcal{L}_t(\hat{\theta}_t) \leq \mathcal{O}\left(d \log(t/\delta)\right) \right\} .$$



Proposition (Abeille, F. and Calauzènes, 2021)

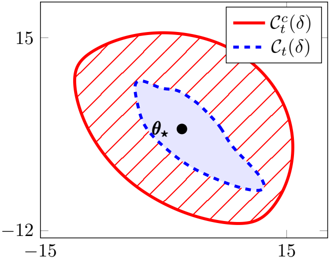
The following holds:

Extension: convex relaxation

$$\mathcal{C}_t(\delta) = \left\{ \left\| \theta - \hat{\theta}_t \right\|_{\boldsymbol{H}_t(\theta)} \leq \mathcal{O}\left(\sqrt{d \log(t/\delta)}.\right) \right\}$$

- Non-convex, burdensome to manipulate.
- **Convex** relaxation based on log-loss $\mathcal{L}_t$:

$$\mathcal{C}_t^c(\delta) = \left\{ \mathcal{L}_t(\theta) - \mathcal{L}_t(\hat{\theta}_t) \leq \mathcal{O}\left(d \log(t/\delta)\right) \right\} .$$



Proposition (Abeille, F. and Calauzènes, 2021)

The following holds:

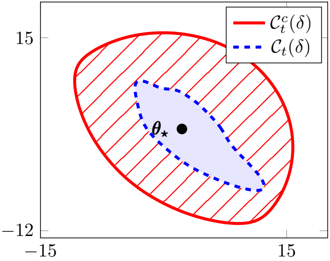
- *For all* $t \geq 1$, $\mathcal{C}_t(\delta) \subseteq \mathcal{C}_t^c(\delta)$ *i.e* $\mathcal{C}_t^c(\delta)$ *is a confidence set for* $\theta_\star$.

Extension: convex relaxation

$$\mathcal{C}_t(\delta) = \left\{ \left\| \theta - \hat{\theta}_t \right\|_{\boldsymbol{H}_t(\theta)} \leq \mathcal{O}\left(\sqrt{d \log(t/\delta)}. \right) \right\}$$

- Non-convex, burdensome to manipulate.
- **Convex** relaxation based on log-loss $\mathcal{L}_t$:

$$\mathcal{C}_t^c(\delta) = \left\{ \mathcal{L}_t(\theta) - \mathcal{L}_t(\hat{\theta}_t) \leq \mathcal{O}\left(d \log(t/\delta)\right) \right\} .$$



Proposition (Abeille, F. and Calauzènes, 2021)

The following holds:

- *For all $t \geq 1$, $\mathcal{C}_t(\delta) \subseteq \mathcal{C}_t^c(\delta)$ i.e $\mathcal{C}_t^c(\delta)$ is a confidence set for $\theta_\star$.*
- *With proba. at least $1 - \delta$:*

$$\forall \theta \in \mathcal{C}_t^c(\delta),\, \forall t \geq 1,\ \|\theta - \theta_\star\|_{\boldsymbol{H}_t(\theta_\star)} \leq \mathcal{O}\left(\sqrt{d \log(t/\delta)} \right)$$

Algorithm and regret bounds

OFU-GLB

- **Algorithm.** New ingredients, same recipe.

$$\text{play } a_t = \mathrm{argmax}_{a \in \mathcal{A}} \max_{\theta \in \boldsymbol{\mathcal{C}_t(\delta)}} a^\top \theta \ .$$

OFU-GLB

- **Algorithm.** New ingredients, same recipe.

$$\text{play } a_t = \text{argmax}_{a \in \mathcal{A}} \max_{\theta \in \mathcal{C}_t(\delta)} a^\top \theta \; .$$

- Pseudo-code.

Algorithm OFU-GLB

---

input: Arm set $\mathcal{A}$, regularizations $\{\lambda_t\}_t$, failure level δ, norm upper-bound S.
  Set $\boldsymbol{H}_1 \leftarrow \lambda_1 \boldsymbol{I}_d$, $\hat{\theta}_1 \leftarrow 0_d$.
  for $t \in [1, T]$ **do**
    Solve $a_t \in \text{argmax}_{\mathcal{A}} \max_{\theta \in \mathcal{C}_t(\delta)} a^\top \theta$. ▷ *planning*
    Play the arm a_t and observe reward r_{t+1}.
    Update the estimator $\hat{\theta}_{t+1}$ and the confidence interval $\mathcal{C}_t(\delta)$. ▷ *learning*
  end for

OFU-GLB: analysis

- **Sketch of proof.**

$$\text{Regret}(T) = \sum_{t=1}^{T} \mu(a_\star^\top \theta_\star) - \mu(a_t^\top \theta_\star)$$

OFU-GLB: analysis

- **Sketch of proof.**

$$\text{Regret}(T) \leq \sum_{t=1}^{T} \mu(a_t^\top \theta_t) - \mu(a_t^\top \theta_\star) \qquad (\theta_t \text{ optim.})$$

OFU-GLB: analysis

- **Sketch of proof.**

$$\begin{aligned}
\text{Regret}(T) &\leq \sum_{t=1}^{T} \mu(a_t^\top \theta_t) - \mu(a_t^\top \theta_\star) && (\theta_t \text{ optim.}) \\
&\leq \sum_{t=1}^{T} \dot{\mu}(a_t^\top \theta_\star) a_t^\top (\theta_t - \theta_\star) + \mathbf{L}_{\boldsymbol{\mu}} \sum_{t=1}^{T} (a_t^\top (\theta_t - \theta_\star))^2 && (\text{Taylor})
\end{aligned}$$

OFU-GLB: analysis

- **Sketch of proof.**

$$\text{Regret}(T) \leq \sum_{t=1}^{T} \mu(a_t^\top \theta_t) - \mu(a_t^\top \theta_\star) \qquad (\theta_t \text{ optim.})$$

$$\leq \underbrace{\sum_{t=1}^{T} \dot{\mu}(a_t^\top \theta_\star) a_t^\top (\theta_t - \theta_\star)}_{①} + \mathbf{L}_{\boldsymbol{\mu}} \underbrace{\sum_{t=1}^{T} (a_t^\top (\theta_t - \theta_\star))^2}_{②} \qquad (\text{Taylor})$$

OFU-GLB: analysis

- **Sketch of proof.**

$$\begin{aligned}\text{Regret}(T) &\leq \sum_{t=1}^{T} \mu(a_t^\top \theta_t) - \mu(a_t^\top \theta_\star) && (\theta_t \text{ optim.}) \\ &\leq \underbrace{\sum_{t=1}^{T} \dot{\mu}(a_t^\top \theta_\star) a_t^\top (\theta_t - \theta_\star)}_{①} + \boldsymbol{\kappa}_\mu d^2 \log(T/\delta)^2\end{aligned}$$

OFU-GLB: analysis

- **Sketch of proof.**

$$\begin{aligned}\text{Regret}(T) &\leq \sum_{t=1}^{T} \mu(a_t^\top \theta_t) - \mu(a_t^\top \theta_\star) && (\theta_t \text{ optim.})\\ &\leq \underbrace{\sum_{t=1}^{T} \dot{\mu}(a_t^\top \theta_\star) a_t^\top (\theta_t - \theta_\star)}_{①} + \boldsymbol{\kappa}_\mu d^2 \log(T/\delta)^2\end{aligned}$$

- Bounding ①:

OFU-GLB: analysis

- **Sketch of proof.**

$$\text{Regret}(T) \leq \sum_{t=1}^{T} \mu(a_t^\top \theta_t) - \mu(a_t^\top \theta_\star) \qquad (\theta_t \text{ optim.})$$

$$\leq \underbrace{\sum_{t=1}^{T} \dot{\mu}(a_t^\top \theta_\star) a_t^\top (\theta_t - \theta_\star)}_{①} + \boldsymbol{\kappa}_\mu d^2 \log(T/\delta)^2$$

- Bounding ①:

$$① \leq d \log(T/\delta) \sqrt{\sum_{t=1}^{T} \dot{\mu}(a_t^\top \theta_\star)} \qquad \text{(Ell. Pot.)}$$

OFU-GLB: analysis

- **Sketch of proof.**

$$\text{Regret}(T) \leq \sum_{t=1}^{T} \mu(a_t^\top \theta_t) - \mu(a_t^\top \theta_\star) \qquad (\theta_t \text{ optim.})$$

$$\leq \underbrace{\sum_{t=1}^{T} \dot{\mu}(a_t^\top \theta_\star) a_t^\top (\theta_t - \theta_\star)}_{①} + \boldsymbol{\kappa}_\mu d^2 \log(T/\delta)^2$$

- Bounding ①:

$$① \leq d \log(T/\delta) \sqrt{\sum_{t=1}^{T} \dot{\mu}(a_t^\top \theta_\star)} \qquad (\text{Ell. Pot.})$$

$$\leq d \log(T/\delta) \sqrt{T \dot{\mu}({a_\star}^\top \theta_\star) + \text{Regret}(T)} \qquad (\text{s.c})$$

where $a_\star = \text{argmax}_{\mathcal{A}}\, a^\top \theta_\star$ (best action in hindsight)

OFU-GLB: analysis

- **Sketch of proof.**

$$
\begin{aligned}
\text{Regret}(T) &\le \sum_{t=1}^{T} \mu(a_t^\top \theta_t) - \mu(a_t^\top \theta_\star) && (\theta_t \text{ optim.})\\
&\le \underbrace{\sum_{t=1}^{T} \dot{\mu}(a_t^\top \theta_\star) a_t^\top (\theta_t - \theta_\star)}_{①} + \boldsymbol{\kappa}_\mu d^2 \log(T/\delta)^2 \\
&\le d \log(T/\delta) \sqrt{T \dot{\mu}(a_\star{}^\top \theta_\star) + \text{Regret}(T)} + \boldsymbol{\kappa}_\mu d^2 \log(T/\delta)^2
\end{aligned}
$$

- Bounding ①:

$$
\begin{aligned}
① &\le d \log(T/\delta) \sqrt{\sum_{t=1}^{T} \dot{\mu}(a_t^\top \theta_\star)} && (\text{Ell. Pot.})\\
&\le d \log(T/\delta) \sqrt{T \dot{\mu}(a_\star{}^\top \theta_\star) + \text{Regret}(T)} && (\text{s.c})
\end{aligned}
$$

where $a_\star = \text{argmax}_{\mathcal{A}}\, a^\top \theta_\star$ (best action in hindsight)

OFU-GLB: new regret upper-bound

Theorem (Extends [F., Abeille, Calauzènes, Fercoq (2020))

For all self-concordant GLBs, OFU-GLB satisfies:

$$\text{Regret}(T) = \widetilde{\mathcal{O}}\left(d\sqrt{\dot{\mu}({a_\star}^\top \theta_\star)T} + \boldsymbol{\kappa}_\mu d^2\right) \ ,$$

with probability at least $1-\delta$.

OFU-GLB: new regret upper-bound

Theorem (Extends [F., Abeille, Calauzènes, Fercoq (2020))

For all self-concordant GLBs, OFU-GLB satisfies:

$$\text{Regret}(T) = \widetilde{\mathcal{O}}\left(d\sqrt{\dot{\mu}({a_\star}^\top\theta_\star)T} + \boldsymbol{\kappa}_\mu d^2\right) ,$$

with probability at least $1-\delta$.

- **Non-linearity** deferred to **second-order term**.

$$\text{for } T \gg {\boldsymbol{\kappa}_\mu}^2, \quad \text{Regret}(T) = \widetilde{\mathcal{O}}\left(d\sqrt{\dot{\mu}({a_\star}^\top\theta_\star)T}\right) .$$

OFU-GLB: new regret upper-bound

Theorem (Extends [F., Abeille, Calauzènes, Fercoq (2020))

For all self-concordant GLBs, OFU-GLB satisfies:

$$\text{Regret}(T) = \widetilde{\mathcal{O}}\left(d\sqrt{\dot{\mu}({a_\star}^\top\theta_\star)T} + \boldsymbol{\kappa}_\mu d^2\right) \,,$$

with probability at least $1-\delta$*.*

- **Non-linearity** deferred to **second-order term**.

$$\text{for } T \gg {\boldsymbol{\kappa}_\mu}^2, \quad \text{Regret}(T) = \widetilde{\mathcal{O}}\left(d\sqrt{\dot{\mu}({a_\star}^\top\theta_\star)T}\right) \,.$$

- **Exponential improvement** over previous work: *e.g* Logistic Bandit:

$$\begin{aligned} \text{(before)} \quad & \text{Regret}(T) \lessapprox \boldsymbol{\kappa}_\mu d\sqrt{T} \,, \\ \text{(now)} \quad & \text{Regret}(T) \lessapprox \exp(-\|\theta_\star\|/2)d\sqrt{T} \,. \end{aligned}$$

Transitory and permanent regimes

- Making sense of the regret bound:

$$\text{Regret}(T) = \widetilde{\mathcal{O}}\left(d\sqrt{\dot{\mu}({a_\star}^\top \theta_\star)T} + \boldsymbol{\kappa}_\mu d^2\right) ,$$

Transitory and permanent regimes

- Making sense of the regret bound:

$$\mathrm{Regret}(T) = \widetilde{\mathcal{O}}\left(d\sqrt{\dot{\mu}({a_\star}^\top\theta_\star)T} + \boldsymbol{\kappa}_\mu d^2\right) \;,$$

$$= R_T^{\mathsf{perm}} + R_T^{\mathsf{trans}} \;.$$

  - each term associated to a different **regime** of algorithm behavior.

Transitory and permanent regimes

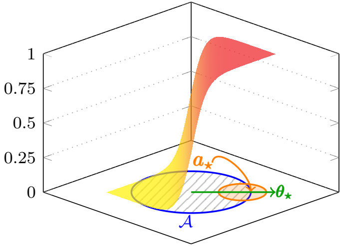
- Making sense of the regret bound:

$$\text{Regret}(T) = \widetilde{\mathcal{O}}\left(d\sqrt{\dot\mu({a_\star}^\top\theta_\star)T} + \boldsymbol{\kappa}_\mu d^2\right) ,$$

$$= R_T^{\textsf{perm}} + R_T^{\textsf{trans}} .$$

  - each term associated to a different **regime** of algorithm behavior.

- Permanent regret $\longleftrightarrow$ $a_\star$ $\approx$ located.



Transitory and permanent regimes

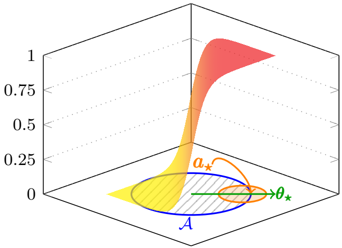
- Making sense of the regret bound:

$$\mathrm{Regret}(T) = \widetilde{\mathcal{O}}\left(d\sqrt{\dot{\mu}({a_\star}^\top\theta_\star)T} + \boldsymbol{\kappa}_\mu d^2\right) ,$$

$$= R_T^{\mathsf{perm}} + R_T^{\mathsf{trans}} .$$

  - each term associated to a different **regime** of algorithm behavior.

- Permanent regret $\longleftrightarrow$ $a_\star \approx$ located.
- locally *linear* with slope $\dot{\mu}({a_\star}^\top\theta_\star)$.



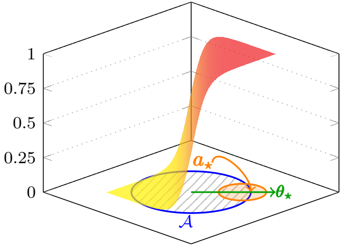
Transitory and permanent regimes

- Making sense of the regret bound:

$$\text{Regret}(T) = \widetilde{\mathcal{O}}\left(d\sqrt{\dot{\mu}({a_\star}^\top\theta_\star)T} + \boldsymbol{\kappa}_\mu d^2\right) ,$$

$$= R_T^{\textsf{perm}} + R_T^{\textsf{trans}} .$$

  - each term associated to a different **regime** of algorithm behavior.



- Permanent regret $\longleftrightarrow$ $a_\star$ $\approx$ located.
  - locally *linear* with slope $\dot{\mu}({a_\star}^\top\theta_\star)$.
  - *e.g* **flat** $\Rightarrow$ small regret.

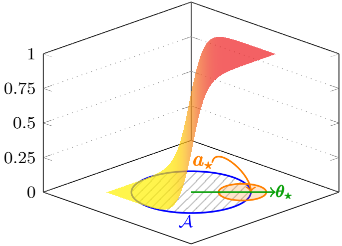
Transitory and permanent regimes

- Making sense of the regret bound:

$$\text{Regret}(T) = \widetilde{\mathcal{O}}\left(d\sqrt{\dot{\mu}({a_\star}^\top\theta_\star)T} + \boldsymbol{\kappa}_\mu d^2\right) ,$$
$$= R_T^{\text{perm}} + R_T^{\text{trans}} .$$

  - each term associated to a different **regime** of algorithm behavior.

- Permanent regret $\longleftrightarrow$ $a_\star \approx$ located.
  - locally *linear* with slope $\dot{\mu}({a_\star}^\top\theta_\star)$.
  - *e.g* **flat** $\Rightarrow$ small regret.



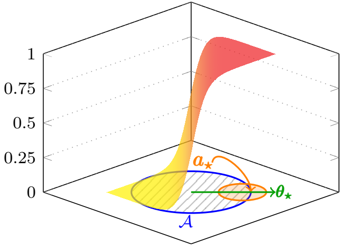
- Transitory regret: how long to find "good" regions of $\mathcal{A}$.
  - can be hard because of non-linearity since $R_T^{\text{trans}} \propto \boldsymbol{\kappa}_\mu$.

Transitory and permanent regimes

- Making sense of the regret bound:

$$\text{Regret}(T) = \widetilde{\mathcal{O}}\left(d\sqrt{\dot\mu({a_\star}^\top\theta_\star)T} + \boldsymbol{\kappa}_\mu d^2\right) \,,$$
$$= R_T^{\textsf{perm}} + R_T^{\textsf{trans}} \,.$$

  - each term associated to a different **regime** of algorithm behavior.

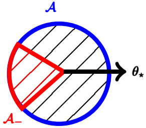


- Permanent regret $\longleftrightarrow$ $a_\star$ $\approx$ located.
  - locally *linear* with slope $\dot\mu({a_\star}^\top\theta_\star)$.
  - *e.g* **flat** $\Rightarrow$ small regret.

- Transitory regret: how long to find "good" regions of $\mathcal{A}$.
  - can be hard because of non-linearity since $R_T^{\textsf{trans}} \propto \boldsymbol{\kappa}_\mu$.
  - coherent with the Bayesian lower-bound of *[Dong et al. 2019]*.

Transitory regret on Logistic Bandit

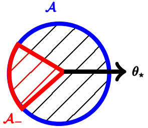
- Detrimental arms: large sub-optimality gap and small information.



Transitory regret on Logistic Bandit

- Detrimental arms: large sub-optimality gap and small information.
- Transitory regret = how many times $\mathcal{A}_-$ is played:

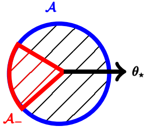
$$R_T^{\text{trans}} \leq \sum_{t=1}^{T} \mathbb{1}\left\{a_t \in \mathcal{A}_-\right\} \ .$$



Transitory regret on Logistic Bandit

- Detrimental arms: large sub-optimality gap and small information.
- Transitory regret = how many times $\mathcal{A}_-$ is played:

$$R_T^{\text{trans}} \leq \sum_{t=1}^{T} \mathbb{1}\left\{a_t \in \mathcal{A}_-\right\} \ .$$



Proposition (Abeille, F. and Calauzènes (2021))

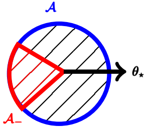
For Logistic Bandit if $\mathcal{A} = \mathcal{B}_d$ the transitory regret satisfies:

$$R_T^{\textit{trans}} = \widetilde{\mathcal{O}}\left(d^3\right)$$

Transitory regret on Logistic Bandit

- Detrimental arms: large sub-optimality gap and small information.
- Transitory regret = how many times $\mathcal{A}_-$ is played:

$$R_T^{\textsf{trans}} \leq \sum_{t=1}^{T} \mathbb{1}\left\{a_t \in \mathcal{A}_-\right\} .$$



Proposition (Abeille, F. and Calauzènes (2021))

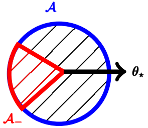
For Logistic Bandit if $\mathcal{A} = \mathcal{B}_d$ the transitory regret satisfies:

$$R_T^{\textit{trans}} = \widetilde{\mathcal{O}}\left(d^3\right) \leftarrow \textbf{\textit{independent of}}\ \boldsymbol{\kappa}_\mu!$$

Transitory regret on Logistic Bandit

- Detrimental arms: large sub-optimality gap and small information.
- Transitory regret = how many times $\mathcal{A}_-$ is played:

$$R_T^{\text{trans}} \leq \sum_{t=1}^{T} \mathbb{1}\left\{a_t \in \mathcal{A}_-\right\} .$$



Proposition (Abeille, F. and Calauzènes (2021))

For Logistic Bandit if $\mathcal{A} = \mathcal{B}_d$ the transitory regret satisfies:

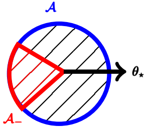
$$R_T^{\textit{trans}} = \widetilde{\mathcal{O}}\left(d^3\right) \leftarrow \textbf{\textit{independent of }} \boldsymbol{\kappa}_\mu \textit{!}$$

✅ "Good" case where non-linearity has **no effect** on the regret bound.

Transitory regret on Logistic Bandit

- Detrimental arms: large sub-optimality gap and small information.
- Transitory regret = how many times $\mathcal{A}_-$ is played:

$$R_T^{\text{trans}} \leq \sum_{t=1}^{T} \mathbb{1}\left\{a_t \in \mathcal{A}_-\right\} \ .$$



Proposition (Abeille, F. and Calauzènes (2021))

For Logistic Bandit if $\mathcal{A} = \mathcal{B}_d$ the transitory regret satisfies:

$$R_T^{\textit{trans}} = \widetilde{\mathcal{O}}\left(d^3\right) \leftarrow \textbf{\textit{independent of }} \boldsymbol{\kappa}_\mu!$$

- "Good" case where non-linearity has **no effect** on the regret bound.
- For any T, we have Regret$(T) \lessapprox d\sqrt{\dot{\mu}(a_\star{}^\top\theta_\star)T} + d^3$.

Non-linearity: a blessing?

- The Logistic Bandit case on $\mathcal{A} = \mathcal{B}_d \implies$ no more $\boldsymbol{\kappa}_\mu$.

$$\text{Regret}(T) \lessapprox d\sqrt{\dot{\mu}({a_\star}^\top \theta_\star)T}$$

Non-linearity: a blessing?

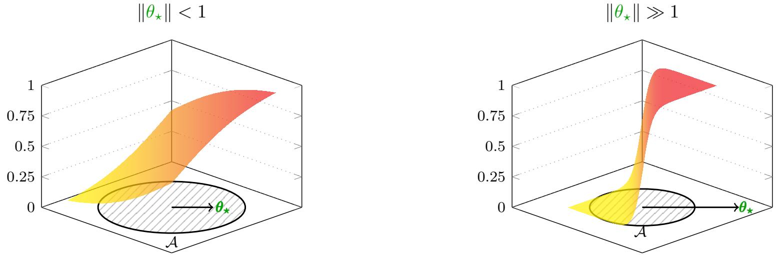
- The Logistic Bandit case on $\mathcal{A} = \mathcal{B}_d \implies$ no more $\boldsymbol{\kappa}_\mu$.

$$\mathsf{Regret}(T) \lessapprox d \exp(-\|\theta_\star\|/2)\sqrt{T}$$

Non-linearity: a blessing?

- The Logistic Bandit case on $\mathcal{A} = \mathcal{B}_d \implies$ no more $\boldsymbol{\kappa}_\mu$.

$$\text{Regret}(T) \lessapprox d\exp(-\|\theta_\star\|/2)\sqrt{T}$$

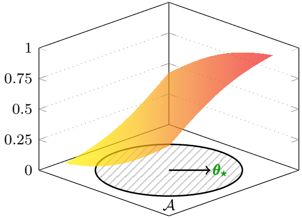


Non-linearity: a blessing?

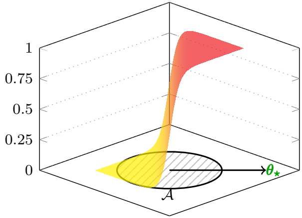
- The Logistic Bandit case on $\mathcal{A} = \mathcal{B}_d \implies$ no more $\boldsymbol{\kappa}_\mu$.

$$\text{Regret}(T) \lessapprox d \exp(-\|\theta_\star\|/2)\sqrt{T}$$

$\|\theta_\star\| < 1$



$\|\theta_\star\| \gg 1$



The more non-linear the problem, the smaller the regret.

Numerical simulations

A tractable algorithm

- The planning step of OFU-GLB is intractable:

$$\text{play } a_t = \text{argmax}_{a\in\mathcal{A}} \max_{\theta\in\mathcal{C}_t(\delta)} a^\top\theta \; .$$

A tractable algorithm

- The planning step of OFU-GLB is intractable:

$$\text{play } a_t = \mathrm{argmax}_{a\in\mathcal{A}} \max_{\theta\in\mathcal{C}_t(\delta)} a^\top\theta \,.$$

  - the constraint $\theta \in \mathcal{C}_t(\delta)$ is **non-convex**.
  - no principled way to solve (even approximately).

A tractable algorithm

- The planning step of OFU-GLB is intractable:

$$\text{play } a_t = \operatorname{argmax}_{a \in \mathcal{A}} \max_{\theta \in \mathcal{C}_t(\delta)} a^\top \theta \,.$$

  - the constraint $\theta \in \mathcal{C}_t(\delta)$ is **non-convex**.
  - no principled way to solve (even approximately).

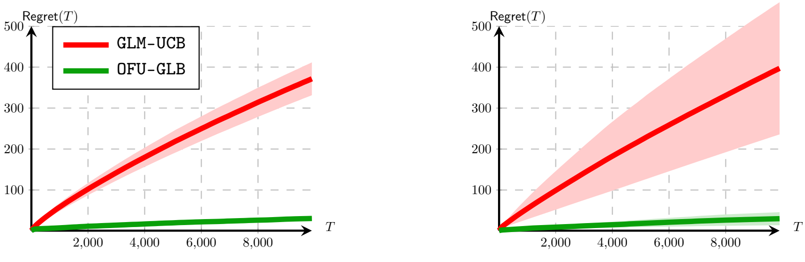
- Use the convex relaxation $\mathcal{C}_t^c(\delta)$:

$$\text{play } a_t = \operatorname{argmax}_{a \in \mathcal{A}} \max_{\theta \in \boldsymbol{\mathcal{C}_t^c(\delta)}} a^\top \theta \,.$$

  - tractable when $|\mathcal{A}| < \infty$ (solve $|\mathcal{A}|$ convex programs).
  - same theoretical guarantees.

Empirical performances

- Improved performances compared to GLM-UCB *[Filippi et al. 2010]*

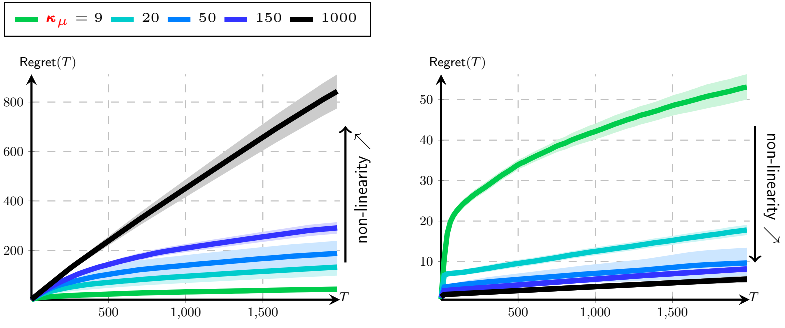


(a) $\boldsymbol{\kappa}_\mu = 50$ (b) $\boldsymbol{\kappa}_\mu = 400$

Comparing `GLM-UCB` and `OFU-GLB` on toy Logistic Bandit experiments.

Empirical performances (ctn'd)

- Check the impact of non-linearity:



(a) GLM-UCB (b) OFU-GLB

Figure: Comparing the effect of non-linearity on GLM-UCB and OFU-GLB by varying the level of non-linearity in a Logistic Bandit setting.

Logistic Bandit
Regret Lower-Bound

Optimality (1/3)

- Are these new regret upper-bounds **optimal**?
  - can we show that for any algorithms, there exist situations where:

$$\text{Regret}(T) \geq d\sqrt{\dot{\mu}(a_\star^\top \theta_\star)T} + \boldsymbol{\kappa}_\mu d^2 \ .$$

Optimality (1/3)

- Are these new regret upper-bounds **optimal**?
  - can we show that for any algorithms, there exist situations where:

$$\text{Regret}(T) \geq d\sqrt{\dot{\mu}(a_\star^\top \theta_\star)T} + \boldsymbol{\kappa}_\mu d^2 \ .$$

- Why is it challenging?
  - involves **problem-dependent** constants.
  - describe a **continuum** of hard situations.
  - existing approaches from LB; typically use $\|\theta_\star\| \propto 1/T$

Optimality (1/3)

- Are these new regret upper-bounds **optimal**?
  - can we show that for any algorithms, there exist situations where:

$$\text{Regret}(T) \geq d\sqrt{\dot{\mu}(a_\star^\top \theta_\star)T} + \boldsymbol{\kappa}_\mu d^2 \ .$$

- Why is it challenging?
  - involves **problem-dependent** constants.
  - describe a **continuum** of hard situations.
  - existing approaches from LB; typically use $\|\theta_\star\| \propto 1/T$

- Notion of **local** minimax-regret *[Simchowitz and Foster. 2021]*:

$$\text{MinimaxRegret}_{\theta_\star}(T, \varepsilon) := \min_\pi \max_{\|\theta - \theta_\star\| \leq \varepsilon} \text{Regret}_\theta^\pi(T)$$

for a given (arbitrary) reference $\theta_\star$.

Optimality (2/3)

Theorem (Abeille, F., Calauzènes (2021))

For the Logistic Bandit with $\mathcal{A} = \mathcal{S}_d$*:* $\forall \theta_\star$, $\exists \varepsilon$ *small s.t:*

$$\textit{MinimaxRegret}_{\theta_\star}(T, \varepsilon) \geq d\sqrt{\dot{\mu}(a_\star^\top \theta_\star) T}\ .$$

whenever $T \geq d^2 \kappa(\theta_\star)$.

Optimality (2/3)

Theorem (Abeille, F., Calauzènes (2021))

For the Logistic Bandit with $\mathcal{A} = \mathcal{S}_d$: $\forall \theta_\star$, $\exists \varepsilon$ small s.t:

$$\textit{MinimaxRegret}_{\theta_\star}(T, \varepsilon) \geq d\sqrt{\dot{\mu}(a_\star^\top \theta_\star) T}\ .$$

whenever $T \geq d^2 \kappa(\theta_\star)$.

- **Discussion.** $\theta_\star$ arbitrary reference point, π a given algorithm

Optimality (2/3)

Theorem (Abeille, F., Calauzènes (2021))

For the Logistic Bandit with $\mathcal{A} = \mathcal{S}_d$: $\forall \theta_\star$, $\exists \varepsilon$ small s.t:

$$\textit{MinimaxRegret}_{\theta_\star}(T, \varepsilon) \geq d\sqrt{\dot{\mu}(a_\star^\top \theta_\star)T}\ .$$

whenever $T \geq d^2 \kappa(\theta_\star)$.

- **Discussion.** $\theta_\star$ arbitrary reference point, π a given algorithm
  - θ the "hardest" nearby instance in $\{\|\theta' - \theta_\star\| \leq \varepsilon\}$.

Optimality (2/3)

Theorem (Abeille, F., Calauzènes (2021))

For the Logistic Bandit with $\mathcal{A} = \mathcal{S}_d$: $\forall \theta_\star$, $\exists \varepsilon$ small s.t:

$$\textit{MinimaxRegret}_{\theta_\star}(T, \varepsilon) \geq d\sqrt{\dot{\mu}(a_\star^\top \theta_\star) T}\ .$$

whenever $T \geq d^2 \kappa(\theta_\star)$.

- **Discussion.** $\theta_\star$ arbitrary reference point, π a given algorithm
  - θ the "hardest" nearby instance in $\{\|\theta' - \theta_\star\| \leq \varepsilon\}$.
  - the regret of π against θ is:

$$\text{Regret}_\theta^\pi(T) \geq d\sqrt{\dot{\mu}(a_\star^\top \theta_\star) T}$$

Optimality (2/3)

Theorem (Abeille, F., Calauzènes (2021))

For the Logistic Bandit with $\mathcal{A} = \mathcal{S}_d$: $\forall \theta_\star$, $\exists \varepsilon$ small s.t:

$$\textit{MinimaxRegret}_{\theta_\star}(T, \varepsilon) \geq d\sqrt{\dot{\mu}(a_\star^\top \theta_\star) T} \, .$$

whenever $T \geq d^2 \kappa(\theta_\star)$.

- **Discussion.** $\theta_\star$ arbitrary reference point, π a given algorithm
  - θ the "hardest" nearby instance in $\{\|\theta' - \theta_\star\| \leq \varepsilon\}$.
  - the regret of π against θ is:

$$\text{Regret}_\theta^\pi(T) \geq d\sqrt{\dot{\mu}(a_\star^\top \theta_\star) T}$$

  - small $\varepsilon \implies \theta$ and $\theta_\star$ share same problem-dependent constants:

$$\dot{\mu}(a_\star^\top \theta_\star) \approx \dot{\mu}(a_\star(\theta)^\top \theta) \, .$$

Optimality (2/3)

Theorem (Abeille, F., Calauzènes (2021))

For the Logistic Bandit with $\mathcal{A} = \mathcal{S}_d$: $\forall \theta_\star$, $\exists \varepsilon$ small s.t:

$$\mathit{MinimaxRegret}_{\theta_\star}(T, \varepsilon) \geq d\sqrt{\dot{\mu}(a_\star^\top \theta_\star) T} \;.$$

whenever $T \geq d^2 \kappa(\theta_\star)$.

- **Discussion.** $\theta_\star$ arbitrary reference point, π a given algorithm
  - θ the "hardest" nearby instance in $\{\|\theta' - \theta_\star\| \leq \varepsilon\}$.
  - the regret of π against θ is:

$$\text{Regret}_\theta^\pi(T) \geq d\sqrt{\dot{\mu}(a_\star^\top \theta_\star) T} \approx d\sqrt{\dot{\mu}(a_\star(\theta)^\top \theta) T}$$

  - small $\varepsilon \implies \theta$ and $\theta_\star$ share same problem-dependent constants:

$$\dot{\mu}(a_\star^\top \theta_\star) \approx \dot{\mu}(a_\star(\theta)^\top \theta) \;.$$

Optimality (2/3)

Theorem (Abeille, F., Calauzènes (2021))

For the Logistic Bandit with $\mathcal{A} = \mathcal{S}_d$: $\forall \theta_\star$, $\exists \varepsilon$ small s.t:

$$\mathit{MinimaxRegret}_{\theta_\star}(T, \varepsilon) \geq d\sqrt{\dot{\mu}(a_\star^\top \theta_\star) T} \; .$$

whenever $T \geq d^2 \kappa(\theta_\star)$. The ***permanent*** *regret is* ***minimax-optimal***

- **Discussion.** $\theta_\star$ arbitrary reference point, π a given algorithm
  - θ the "hardest" nearby instance in $\{\|\theta' - \theta_\star\| \leq \varepsilon\}$.
  - the regret of π against θ is:

$$\text{Regret}_\theta^\pi(T) \geq d\sqrt{\dot{\mu}(a_\star^\top \theta_\star) T} \approx d\sqrt{\dot{\mu}(a_\star(\theta)^\top \theta) T}$$

  - small $\varepsilon \implies \theta$ and $\theta_\star$ share same problem-dependent constants:

$$\dot{\mu}(a_\star^\top \theta_\star) \approx \dot{\mu}(a_\star(\theta)^\top \theta) \; .$$

Optimality (3/3)

- **Proof sketch.** To find a hard nearby instance θ for $\theta_\star$:

Optimality (3/3)

- **Proof sketch.** To find a hard nearby instance θ for $\theta_\star$:

(1) π must behave similarly on $\theta_\star$ and θ.

Optimality (3/3)

- **Proof sketch.** To find a hard nearby instance θ for $\theta_\star$:

(1) π must behave similarly on $\theta_\star$ and θ.

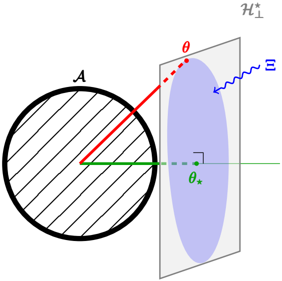
(2) the best arm differs: $a_\star(\theta) \neq a_\star(\theta_\star)$.

Optimality (3/3)

- **Proof sketch.** To find a hard nearby instance θ for $\theta_\star$:

(1) π must behave similarly on $\theta_\star$ and θ.

(2) the best arm differs: $a_\star(\theta) \neq a_\star(\theta_\star)$.



- discrepancy measure: for $\theta' \in \mathcal{H}_\perp^\star$:

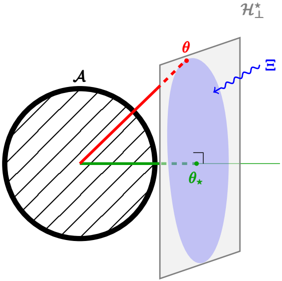
$$\boxed{d(\theta', \theta_\star) := \sqrt{T\dot{\mu}(a_\star^\top \theta_\star)}\|\theta' - \theta_\star\|^2}$$

Optimality (3/3)

- **Proof sketch.** To find a hard nearby instance θ for $\theta_\star$:

(1) π must behave similarly on $\theta_\star$ and θ.

(2) the best arm differs: $a_\star(\theta) \neq a_\star(\theta_\star)$.



- discrepancy measure: for $\theta' \in \mathcal{H}_\perp^\star$:

$$\boxed{d(\theta', \theta_\star) := \sqrt{T\dot{\mu}(a_\star^\top \theta_\star)} \|\theta' - \theta_\star\|^2}$$

- compromise between (1) and (2):

$$\theta \in \Xi := \{\theta' \in \mathcal{H}_\perp^\star,\ d(\theta', \theta_\star) = 1\} .$$

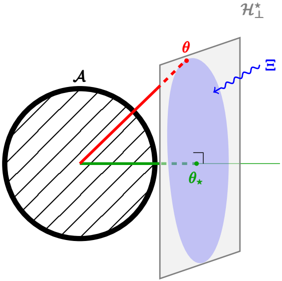
$$\implies \|\theta - \theta_\star\|^2 = (T\dot{\mu}(a_\star^\top \theta_\star))^{-1/2}$$

Optimality (3/3)

- **Proof sketch.** To find a hard nearby instance θ for $\theta_\star$:

(1) π must behave similarly on $\theta_\star$ and θ.

(2) the best arm differs: $a_\star(\theta) \neq a_\star(\theta_\star)$.



- discrepancy measure: for $\theta' \in \mathcal{H}_\perp^\star$:

$$\boxed{d(\theta', \theta_\star) := \sqrt{T\dot\mu(a_\star^\top \theta_\star)}\|\theta' - \theta_\star\|^2}$$

- compromise between (1) and (2):

$$\theta \in \boldsymbol{\Xi} := \{\theta' \in \mathcal{H}_\perp^\star,\ d(\theta', \theta_\star) = 1\} .$$

$$\implies \|\theta - \theta_\star\|^2 = (T\dot\mu(a_\star^\top \theta_\star))^{-1/2}$$

- what about regret?

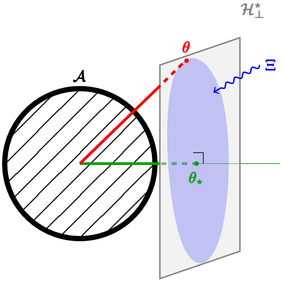
$$\text{Regret}_\theta{}^\pi(T) \approx \dot\mu(a_\star(\theta)^\top \theta) \sum_{t=1}^{T} \|a_t - a_\star(\theta)\|^2$$

Optimality (3/3)

- **Proof sketch.** To find a hard nearby instance θ for $\theta_\star$:

(1) π must behave similarly on $\theta_\star$ and θ.

(2) the best arm differs: $a_\star(\theta) \neq a_\star(\theta_\star)$.



- discrepancy measure: for $\theta' \in \mathcal{H}_\perp^\star$:

$$\boxed{d(\theta', \theta_\star) := \sqrt{T\dot{\mu}(a_\star^\top \theta_\star)}\|\theta' - \theta_\star\|^2}$$

- compromise between (1) and (2):

$$\theta \in \Xi := \{\theta' \in \mathcal{H}_\perp^\star,\ d(\theta', \theta_\star) = 1\}\ .$$

$$\implies \|\theta - \theta_\star\|^2 = (T\dot{\mu}(a_\star^\top \theta_\star))^{-1/2}$$

- what about regret?

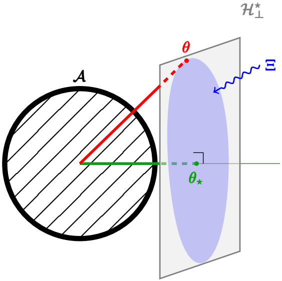
$$\text{Regret}_\theta{}^\pi(T) \approx \dot{\mu}(a_\star(\theta_\star)^\top \theta_\star) \sum_{t=1}^{T} \|a_\star(\theta_\star) - a_\star(\theta)\|^2$$

Optimality (3/3)

- **Proof sketch.** To find a hard nearby instance θ for $\theta_\star$:

(1) π must behave similarly on $\theta_\star$ and θ.

(2) the best arm differs: $a_\star(\theta) \neq a_\star(\theta_\star)$.



- discrepancy measure: for $\theta' \in \mathcal{H}_\perp^\star$:

$$\boxed{d(\theta', \theta_\star) := \sqrt{T\dot{\mu}(a_\star^\top \theta_\star)}\|\theta' - \theta_\star\|^2}$$

- compromise between (1) and (2):

$$\theta \in \Xi := \{\theta' \in \mathcal{H}_\perp^\star,\ d(\theta', \theta_\star) = 1\}\ .$$

$$\implies \|\theta - \theta_\star\|^2 = (T\dot{\mu}(a_\star^\top \theta_\star))^{-1/2}$$

- what about regret?

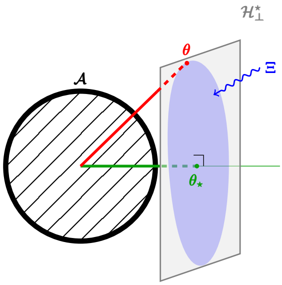
$$\text{Regret}_\theta{}^\pi(T) \approx \dot{\mu}(a_\star(\theta_\star)^\top \theta_\star) T \|\theta - \theta_\star\|^2$$

Optimality (3/3)

- **Proof sketch.** To find a hard nearby instance θ for $\theta_\star$:

(1) π must behave similarly on $\theta_\star$ and θ.

(2) the best arm differs: $a_\star(\theta) \neq a_\star(\theta_\star)$.



- discrepancy measure: for $\theta' \in \mathcal{H}_\perp^\star$:

$$\boxed{d(\theta', \theta_\star) := \sqrt{T\dot{\mu}(a_\star^\top \theta_\star)}\|\theta' - \theta_\star\|^2}$$

- compromise between (1) and (2):

$$\theta \in \Xi := \{\theta' \in \mathcal{H}_\perp^\star,\ d(\theta', \theta_\star) = 1\} .$$

$$\implies \|\theta - \theta_\star\|^2 = (T\dot{\mu}(a_\star^\top \theta_\star))^{-1/2}$$

- what about regret?

$$\text{Regret}_\theta{}^\pi(T) \approx \sqrt{\dot{\mu}(a_\star(\theta_\star)^\top \theta_\star)T}$$

Extensions

Contextual bandits

- Reward is also a function of exogenous context $x_t \in \mathcal{X}$:

$$\mathbb{E}\left[r_{t+1} \,|a_t\right] = \mu(\phi(a_t, x_t)^\top \theta_\star) \ .$$

for some $\phi : \mathcal{A} \times \mathcal{X} \mapsto \mathbb{R}^d$.

Contextual bandits

- Reward is also a function of exogenous context $x_t \in \mathcal{X}$:

$$\mathbb{E}\left[r_{t+1} \,|a_t\right] = \mu(\phi(a_t, x_t)^\top \theta_\star) \; .$$

  for some $\phi : \mathcal{A} \times \mathcal{X} \mapsto \mathbb{R}^d$.

- Similar regret upper-bounds:

$$\text{Regret}(T) = \widetilde{\mathcal{O}}\left(d\sqrt{T}\sqrt{\frac{1}{T}\sum_{t=1}^{T}\dot{\mu}(\phi(a_{\star,t}, x_t)^\top \theta_\star)}\right)$$

  where $a_{\star,t} = \operatorname{argmax}_{a\in\mathcal{A}} \phi(a, x_t)$ best arm at round t.

Contextual bandits

- Reward is also a function of exogenous context $x_t \in \mathcal{X}$:

$$\mathbb{E}\left[r_{t+1} \,|a_t\right] = \mu(\phi(a_t, x_t)^\top \theta_\star) \ .$$

for some $\phi : \mathcal{A} \times \mathcal{X} \mapsto \mathbb{R}^d$.

- Similar regret upper-bounds:

$$\text{Regret}(T) = \widetilde{\mathcal{O}}\left(d\sqrt{T}\sqrt{\frac{1}{T}\sum_{t=1}^{T}\dot{\mu}(\phi(a_{\star,t}, x_t)^\top\theta_\star)}\right)$$

where $a_{\star,t} = \operatorname{argmax}_{a\in\mathcal{A}} \phi(a, x_t)$ best arm at round t.

- Same goes for time-varying arm-sets.

Non-stationary bandits

- Piece-wise stationary environment:

$$\mathbb{E}\left[r_{t+1} \left| a_t\right.\right] = \mu({a_t}^\top \theta_\star^t) \qquad \text{where } \sum_{t=2}^{T} \mathbb{1}\left(\theta_\star^t \neq \theta_\star^{t-1}\right) = \Gamma_T$$

Non-stationary bandits

- Piece-wise stationary environment:

$$\mathbb{E}\left[r_{t+1}\,|a_t\right] = \mu({a_t}^\top \theta_\star^t) \qquad \text{where } \sum_{t=2}^{T} \mathbb{1}\left(\theta_\star^t \neq \theta_\star^{t-1}\right) = \Gamma_T$$

- Change the estimation process to forget the past:

$$\hat{\theta}_t = \mathrm{argmin}_\theta - \sum_{s=1}^{t} \gamma^{t-s} \log d\mathbb{P}(r_{t+1}|a_t)/d\nu(r) + \lambda\|\theta\|^2/2\,.$$

Non-stationary bandits

- Piece-wise stationary environment:

$$\mathbb{E}\left[r_{t+1}\,|a_t\right] = \mu({a_t}^\top \theta_\star^t) \qquad \text{where} \sum_{t=2}^{T} \mathbb{1}\left(\theta_\star^t \neq \theta_\star^{t-1}\right) = \Gamma_T$$

- Change the estimation process to forget the past:

$$\hat{\theta}_t = \operatorname{argmin}_\theta - \sum_{s=1}^{t} \gamma^{t-s} \log d\mathbb{P}(r_{t+1}|a_t)/d\nu(r) + \lambda\|\theta\|^2/2\ .$$

- Similar conclusion:

Theorem (improves (Russac, F., Cappé and Garivier , 2021))

There exists an algorithm on the piece-wise stationary GLB problem s.t:

$$\mathit{DynamicRegret}(T) = \widetilde{\mathcal{O}}\left(T^{2/3}\Gamma_T^{1/3}\sqrt{\ell_\mu^\star} + \boldsymbol{\kappa}_\mu T^{1/3}\Gamma_T^{2/3}\right)\ .$$

where $\ell_\mu^\star := \frac{1}{T}\sum_{t=1}^{T} \dot{\mu}(a_{\star,t}^\top \theta_\star^t)$ *is averaged sensitivity at best-arm.*

Towards computationally efficient algorithms

- **Computationally** hungry algorithms;

$$\text{total computational cost } = \widetilde{\mathcal{O}}\left(|\mathcal{A}|T^2\right)$$

Towards computationally efficient algorithms

- **Computationally** hungry algorithms;

$$\text{total computational cost } = \widetilde{\mathcal{O}}\left(|\mathcal{A}|T^2\right)$$

- Two computational bottlenecks; at each round:
  - *(learning)* compute maximum likelihood (up to precision $\varepsilon = 1/T$).
  - *(planning)* solve $|\mathcal{A}|$ likelihood-based convex programs.

Towards computationally efficient algorithms

- **Computationally** hungry algorithms;

$$\text{total computational cost } = \widetilde{\mathcal{O}}\left(|\mathcal{A}|T^2\right)$$

- Two computational bottlenecks; at each round:
  - *(learning)* compute maximum likelihood (up to precision $\varepsilon = 1/T$).
  - *(planning)* solve $|\mathcal{A}|$ likelihood-based convex programs.

- Simultaneously computationally and statistically efficient GLB algorithms?
  - *(learning)* confidence sets with $\widetilde{\mathcal{O}}(1)$ sufficient statistic
  - *(planning)* Thompson Sampling alternative.

Towards computationally efficient algorithms

- **Computationally** hungry algorithms;

$$\text{total computational cost } = \widetilde{\mathcal{O}}\left(|\mathcal{A}|T^2\right)$$

- Two computational bottlenecks; at each round:
  - *(learning)* compute maximum likelihood (up to precision $\varepsilon = 1/T$).
  - *(planning)* solve $|\mathcal{A}|$ likelihood-based convex programs.

- Simultaneously computationally and statistically efficient GLB algorithms?
  - *(learning)* confidence sets with $\widetilde{\mathcal{O}}(1)$ sufficient statistic
  - *(planning)* Thompson Sampling alternative.

✔ tools from online convex optimization literature (*[Jézéquel et al. 2020]*).

Towards computationally efficient algorithms

- **Computationally** hungry algorithms;

$$\text{total computational cost } = \widetilde{\mathcal{O}}\left(|\mathcal{A}|T^2\right)$$

- Two computational bottlenecks; at each round:
  - *(learning)* compute maximum likelihood (up to precision $\varepsilon = 1/T$).
  - *(planning)* solve $|\mathcal{A}|$ likelihood-based convex programs.

- Simultaneously computationally and statistically efficient GLB algorithms?
  - *(learning)* confidence sets with $\widetilde{\mathcal{O}}(1)$ sufficient statistic
  - *(planning)* Thompson Sampling alternative.

✔ tools from online convex optimization literature (*[Jézéquel et al. 2020]*).

$\Longrightarrow$ same regret guarantees and computationally efficient algorithm.

Conclusion

Key Take-Aways.

- **Generalized Linear Bandits**:
  - Flexible yet simple model for many real-word situations.

Conclusion

Key Take-Aways.

- **Generalized Linear Bandits**:
  - Flexible yet simple model for many real-word situations.
  - Neat study of non-linearity in parametric bandits.

Conclusion

Key Take-Aways.

- **Generalized Linear Bandits**:
  - Flexible yet simple model for many real-word situations.
  - Neat study of non-linearity in parametric bandits.

- **Contributions**:
  - Improved algorithms (much smaller regret).

Conclusion

Key Take-Aways.

- **Generalized Linear Bandits**:
  - Flexible yet simple model for many real-word situations.
  - Neat study of non-linearity in parametric bandits.

- **Contributions**:
  - Improved algorithms (much smaller regret).
  - Refined analysis tool for **local** treatment.

Conclusion

Key Take-Aways.

- **Generalized Linear Bandits**:
  - Flexible yet simple model for many real-word situations.
  - Neat study of non-linearity in parametric bandits.

- **Contributions**:
  - Improved algorithms (much smaller regret).
  - Refined analysis tool for **local** treatment.
  - Not harder to solve than Linear Bandit!

Conclusion

Key Take-Aways.

- **Generalized Linear Bandits**:
  - Flexible yet simple model for many real-word situations.
  - Neat study of non-linearity in parametric bandits.
- **Contributions**:
  - Improved algorithms (much smaller regret).
  - Refined analysis tool for **local** treatment.
  - Not harder to solve than Linear Bandit!
- **Limitations and Perspectives.**
  - Towards richer reward models?
  - Adversarial bandits.

Thank you!

Bibliography

- Sarah Filippi, Olivier Cappé, Aurélien Garivier, Csaba Szepesvári. *Parametric Bandits: The Generalized Linear Case*, NeurIPS, 2010.

- Francis Bach. *Self-Concordant Analysis for Logistic Regression*, EJS, 2010.

- Yasin Abbasi-Yadkori, Csaba Szepesvári, David Pal. Improved Algorithms for Linear Stochastic Bandits, NeurIPS, 2011.

- Lihong Li, Yu Lu, Dengyong Zhou. Provably Optimal Algorithms for Generalized Linear Contextual Bandits, ICML, 2017.

- Kwang-Sung Jun, Aniruddha Bhargava, Robert Nowak, Rebecca Willett. Scalable Generalized Linear Bandits: Online Computation and Hashing, NeurIPS, 2017.

- Shi Dong, Tengyu Ma, Benjamin Van Roy. *On the Performance of Thompson Sampling on Logistic Bandits, COLT, 2019*.

Relevant Publications

- L.F, Marc Abeille, Clément Calauzènes, Olivier Fercoq. *Improved Optimistic Algorithms for Logistic Bandits*, ICML, 2020.
- Marc Abeille, L.F, Clément Calauzènes. *Instance-Wise Minimax-Optimal Algorithms for Logistic Bandits*, AISTATS, 2021.
- Yoan Russac, L.F, Olivier Cappé, Aurélien Garivier. *Self-Concordant Analysis of Generalized Linear Bandits with Forgetting*, AISTATS, 2021.
- L.F, Yoan Russac, Marc Abeille, Clément Calauzènes. *A Technical Note on Non-Stationary Parametric Bandits: Existing Mistakes and Preliminary Solutions*, ALT, 2021.
- L.F, Yoan Russac, Marc Abeille, Clément Calauzènes. *Regret Bounds for Generalized Linear Bandits under Parameter Drift*, preprint, 2021.

What about self-concordance?

- Mostly used for the learning process. Actually, concentration is given by:

$$\left\| \sum_{s=1}^{t-1} \left[\mu(a_s^\top \hat{\theta}_t) - \mu(a_s^\top \theta_\star) \right] a_s \right\|_{\boldsymbol{H}_t^{-1}(\theta_\star)} = \left\| \sum_{s=1}^{t-1} \eta_{s+1} a_s \right\|_{\boldsymbol{H}_t^{-1}(\theta_\star)} = \mathcal{O}\sqrt{d \log(t/\delta)}\,.$$

What about self-concordance?

- Mostly used for the learning process. Actually, concentration is given by:

$$\left\| \sum_{s=1}^{t-1} \left[\mu(a_s^\top \hat{\theta}_t) - \mu(a_s^\top \theta_\star) \right] a_s \right\|_{\boldsymbol{H}_t^{-1}(\theta_\star)} = \left\| \sum_{s=1}^{t-1} \eta_{s+1} a_s \right\|_{\boldsymbol{H}_t^{-1}(\theta_\star)} = \mathcal{O}\sqrt{d \log(t/\delta)} \,.$$

  - To be used for bounding regret, need to be tied to $\left\| \hat{\theta}_t - \theta_\star \right\|_{\boldsymbol{H}_t^{-1}(\theta_\star)}$

What about self-concordance?

- Mostly used for the learning process. Actually, concentration is given by:

$$\left\| \sum_{s=1}^{t-1} \left[\mu(a_s^\top \hat{\theta}_t) - \mu(a_s^\top \theta_\star) \right] a_s \right\|_{\boldsymbol{H}_t^{-1}(\theta_\star)} = \left\| \sum_{s=1}^{t-1} \eta_{s+1} a_s \right\|_{\boldsymbol{H}_t^{-1}(\theta_\star)} = \mathcal{O}\sqrt{d \log(t/\delta)} \,.$$

  - To be used for bounding regret, need to be tied to $\left\| \hat{\theta}_t - \theta_\star \right\|_{\boldsymbol{H}_t^{-1}(\theta_\star)}$

- By the mean-value theorem:

$$\sum_{s=1}^{t-1} \left[\mu(a_s^\top \hat{\theta}_t) - \mu(a_s^\top \theta_\star) \right] a_s = \boldsymbol{G}_t(\hat{\theta}_t, \theta_\star)(\hat{\theta}_t - \theta_\star)$$

where $\boldsymbol{G}_t(\hat{\theta}_t, \theta_\star) = \sum_{s=1}^{t} \left[\int_{v=0}^{1} \dot{\mu}(a_s^\top \theta_\star + v a_s^\top (\hat{\theta}_t - \theta_\star) dv \right] a_s a_s^\top + \lambda \boldsymbol{I}_d$.

What about self-concordance?

- Mostly used for the learning process. Actually, concentration is given by:

$$\left\| \sum_{s=1}^{t-1} \left[\mu(a_s^\top \hat{\theta}_t) - \mu(a_s^\top \theta_\star) \right] a_s \right\|_{\boldsymbol{H}_t^{-1}(\theta_\star)} = \left\| \sum_{s=1}^{t-1} \eta_{s+1} a_s \right\|_{\boldsymbol{H}_t^{-1}(\theta_\star)} = \mathcal{O}\sqrt{d \log(t/\delta)} \,.$$

  - To be used for bounding regret, need to be tied to $\left\| \hat{\theta}_t - \theta_\star \right\|_{\boldsymbol{H}_t^{-1}(\theta_\star)}$

- By the mean-value theorem:

$$\sum_{s=1}^{t-1} \left[\mu(a_s^\top \hat{\theta}_t) - \mu(a_s^\top \theta_\star) \right] a_s = \boldsymbol{G}_t(\hat{\theta}_t, \theta_\star)(\hat{\theta}_t - \theta_\star)$$

where $\boldsymbol{G}_t(\hat{\theta}_t, \theta_\star) = \sum_{s=1}^{t} \left[\int_{v=0}^{1} \dot{\mu}(a_s^\top \theta_\star + v a_s^\top (\hat{\theta}_t - \theta_\star) dv \right] a_s a_s^\top + \lambda \boldsymbol{I}_d$.

- Self-concordance to the rescue:

$$\int_{v=0}^{1} \dot{\mu}(a_s^\top \theta_\star + v a_s^\top (\hat{\theta}_t - \theta_\star) dv \geq (1 + 2S)^{-1} \dot{\mu}(a_s^\top \theta_\star)$$

so $\boldsymbol{G}_t(\hat{\theta}_t, \theta_\star) \geq (1 + 2S)^{-1} \boldsymbol{H}_t(\theta_\star)$.

Bonus vs parameter based optimism

- For the Linear Bandit:
  - Bonus: play $a_t = \operatorname{argmax}_{a \in \mathcal{A}} a^\top \hat{\theta}_t + \sqrt{d \log(t)} \|a\|_{\boldsymbol{V}_t^{-1}}$.
  - Parameter: play $a_t = \operatorname{argmax}_{a \in \mathcal{A}} \max_{\theta \in \mathcal{C}_t(\delta)} a^\top \theta$

  are exactly **equivalent**.

Bonus vs parameter based optimism

- For the Linear Bandit:
  - Bonus: play $a_t = \mathrm{argmax}_{a \in \mathcal{A}} a^\top \hat{\theta}_t + \sqrt{d \log(t)} \|a\|_{\boldsymbol{V}_t^{-1}}$.
  - Parameter: play $a_t = \mathrm{argmax}_{a \in \mathcal{A}} \max_{\theta \in \mathcal{C}_t(\delta)} a^\top \theta$

  are exactly **equivalent**.

- No longer true with non-ellipsoidal confidence sets.

Bonus vs parameter based optimism

- For the Linear Bandit:
  - Bonus: play $a_t = \text{argmax}_{a\in\mathcal{A}}\, a^\top \hat{\theta}_t + \sqrt{d\log(t)}\|a\|_{\boldsymbol{V}_t^{-1}}$.
  - Parameter: play $a_t = \text{argmax}_{a\in\mathcal{A}}\, \max_{\theta\in\mathcal{C}_t(\delta)}\, a^\top\theta$

  are exactly **equivalent**.

- No longer true with non-ellipsoidal confidence sets.

- Bonus-based exploration:
  - much more complicated bonus function.
  - requires additional projection.
  - non-tight analysis $\Rightarrow$ non-tight design.
  - typically much less performant for GLBs.

Bonus vs parameter based optimism

- For the Linear Bandit:
  - Bonus: play $a_t = \mathrm{argmax}_{a\in\mathcal{A}}\, a^\top \hat{\theta}_t + \sqrt{d\log(t)}\|a\|_{\boldsymbol{V}_t^{-1}}$.
  - Parameter: play $a_t = \mathrm{argmax}_{a\in\mathcal{A}} \max_{\theta\in\mathcal{C}_t(\delta)} a^\top\theta$

  are exactly **equivalent**.

- No longer true with non-ellipsoidal confidence sets.

- Bonus-based exploration:
  - much more complicated bonus function.
  - requires additional projection.
  - non-tight analysis $\Rightarrow$ non-tight design.
  - typically much less performant for GLBs.

- Parameter-based:
  - non-tight analysis remains at analysis time
  - more adaptive algorithms (*e.g* second-order term)

Fast yet optimal algorithms (1/2)

- Replace $\hat{\theta}_t$ by:

$$\theta_t = \mathrm{argmin}_\Theta \, \|\theta - \theta_{t-1}\|^2_{\boldsymbol{W}_{t-1}} + \ell(a_t^\top \theta, r_{t+1}) \ .$$

  where:
  - Θ is a "small" convex set around $\theta_\star$ (forced-exploration)
  - and $\boldsymbol{W}_t = \sum_{s=1}^{t-1} \dot{\mu}(a_s^\top \theta_{s+1}) a_s a_s$.

Fast yet optimal algorithms (1/2)

- Replace $\hat{\theta}_t$ by:

$$\theta_t = \mathrm{argmin}_{\Theta} \|\theta - \theta_{t-1}\|^2_{\boldsymbol{W}_{t-1}} + \ell(a_t^\top \theta, r_{t+1}) \ .$$

  where:
  - Θ is a "small" convex set around $\theta_\star$ (forced-exploration)
  - and $\boldsymbol{W}_t = \sum_{s=1}^{t-1} \dot{\mu}(a_s^\top \theta_{s+1}) a_s a_s$.

- Yields the confidence set:

$$\left\{\theta, \ \|\theta - \theta_t\|_{\boldsymbol{W}_t} \leq \sqrt{d \log(t/\delta)}\right\} \ .$$

  - $\boldsymbol{W}_t$ mimics $\boldsymbol{H}_t(\theta) \implies$ reward sensitivity.
  - Sufficient statistics can be maintained at $\widetilde{\mathcal{O}}(1)$ cost.

Fast yet optimal algorithms (1/2)

- Replace $\hat{\theta}_t$ by:

$$\theta_t = \operatorname{argmin}_\Theta \|\theta - \theta_{t-1}\|^2_{\boldsymbol{W}_{t-1}} + \ell(a_t^\top \theta, r_{t+1}) \ .$$

  where:
  - Θ is a "small" convex set around $\theta_\star$ (forced-exploration)
  - and $\boldsymbol{W}_t = \sum_{s=1}^{t-1} \dot{\mu}(a_s^\top \theta_{s+1}) a_s a_s$.

- Yields the confidence set:

$$\left\{\theta,\ \|\theta - \theta_t\|_{\boldsymbol{W}_t} \leq \sqrt{d \log(t/\delta)}\right\} \ .$$

  - $\boldsymbol{W}_t$ mimics $\boldsymbol{H}_t(\theta) \implies$ reward sensitivity.
  - Sufficient statistics can be maintained at $\widetilde{\mathcal{O}}(1)$ cost.

- Same regret bounds!

Fast yet optimal algorithms (1/2)

- Replace $\hat{\theta}_t$ by:

$$\theta_t = \operatorname{argmin}_{\Theta} \|\theta - \theta_{t-1}\|^2_{\boldsymbol{W}_{t-1}} + \ell(a_t^\top \theta, r_{t+1}) \ .$$

  where:
  - Θ is a "small" convex set around $\theta_\star$ (forced-exploration)
  - and $\boldsymbol{W}_t = \sum_{s=1}^{t-1} \dot{\mu}(a_s^\top \theta_{s+1}) a_s a_s$.

- Yields the confidence set:

$$\left\{\theta,\ \|\theta - \theta_t\|_{\boldsymbol{W}_t} \leq \sqrt{d \log(t/\delta)}\right\} \ .$$

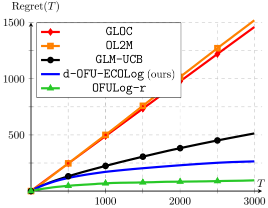
  - $\boldsymbol{W}_t$ mimics $\boldsymbol{H}_t(\theta) \implies$ reward sensitivity.
  - Sufficient statistics can be maintained at $\widetilde{\mathcal{O}}(1)$ cost.

- Same regret bounds!

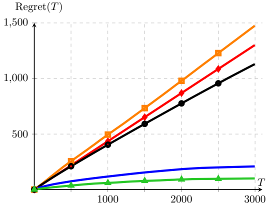
- Forced exploration can be dropped through at data-dependent approach.

Fast yet optimal algorithms (2/2)

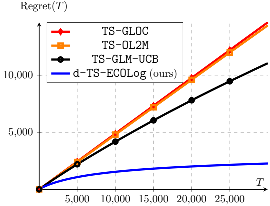
- Some experimental results:



$d = 2$, $|\mathcal{A}| = 20$, $\kappa = 150$



$d = 2$, $|\mathcal{A}| = 20$, $\kappa = 400$



$d = 5$, $\mathcal{A} = \mathcal{B}_5$, $\kappa = 400$.

Application to GLBs

- Using optimality of $\hat{\theta}_t$ for the regularized log-loss $\mathcal{L}_t^{\lambda_t}$:

$$\forall t \geq 1, \quad \left\| \theta_\star - \hat{\theta}_t \right\|_{\boldsymbol{H}_t(\theta_\star)} \leq \left\| \sum_{s=1}^{t-1} \eta_{s+1} a_s \right\|_{\boldsymbol{H}_t^{-1}(\theta_\star)}$$

where:

$$\eta_{s+1} = r_{s+1} - \mu(a_s^\top \theta_\star) \quad \text{and} \quad \boldsymbol{H}_t(\theta) = \sum_{s=1}^{t} \dot{\mu}(a_s^\top \theta) a_s a_s^\top + \lambda_t \boldsymbol{I}_d$$

Application to GLBs

- Using optimality of $\hat{\theta}_t$ for the regularized log-loss $\mathcal{L}_t^{\lambda_t}$:

$$\forall t \geq 1, \quad \left\| \theta_\star - \hat{\theta}_t \right\|_{\boldsymbol{H}_t(\theta_\star)} \leq \left\| \sum_{s=1}^{t-1} \eta_{s+1} a_s \right\|_{\boldsymbol{H}_t^{-1}(\theta_\star)}$$

where:

$$\eta_{s+1} = r_{s+1} - \mu(a_s^\top \theta_\star) \quad \text{and} \quad \boldsymbol{H}_t(\theta) = \sum_{s=1}^{t} \dot{\mu}(a_s^\top \theta) a_s a_s^\top + \lambda_t \boldsymbol{I}_d$$

- Define $\mathcal{F}_s = \sigma\left(a_1, r_2, \ldots, r_s, a_s\right)$; Exponential family distribution:

Application to GLBs

- Using optimality of $\hat{\theta}_t$ for the regularized log-loss $\mathcal{L}_t^{\lambda_t}$:

$$\forall t \geq 1, \quad \left\|\theta_\star - \hat{\theta}_t\right\|_{\boldsymbol{H}_t(\theta_\star)} \leq \left\|\sum_{s=1}^{t-1} \eta_{s+1} a_s\right\|_{\boldsymbol{H}_t^{-1}(\theta_\star)}$$

where:

$$\eta_{s+1} = r_{s+1} - \mu(a_s^\top \theta_\star) \quad \text{and} \quad \boldsymbol{H}_t(\theta) = \sum_{s=1}^{t} \dot{\mu}(a_s^\top \theta) a_s a_s^\top + \lambda_t \boldsymbol{I}_d$$

- Define $\mathcal{F}_s = \sigma\left(a_1, r_2, \ldots, r_s, a_s\right)$; Exponential family distribution:
  - $\mathbb{E}\left[\eta_{s+1} | \mathcal{F}_s\right] = \mathbb{E}\left[r_{s+1} - \mu(a_s^\top \theta_\star)\right]$

Application to GLBs

- Using optimality of $\hat{\theta}_t$ for the regularized log-loss $\mathcal{L}_t^{\lambda_t}$:

$$\forall t \geq 1, \quad \left\| \theta_\star - \hat{\theta}_t \right\|_{\boldsymbol{H}_t(\theta_\star)} \leq \left\| \sum_{s=1}^{t-1} \eta_{s+1} a_s \right\|_{\boldsymbol{H}_t^{-1}(\theta_\star)}$$

where:

$$\eta_{s+1} = r_{s+1} - \mu(a_s^\top \theta_\star) \quad \text{and} \quad \boldsymbol{H}_t(\theta) = \sum_{s=1}^{t} \dot{\mu}(a_s^\top \theta) a_s a_s^\top + \lambda_t \boldsymbol{I}_d$$

- Define $\mathcal{F}_s = \sigma\left(a_1, r_2, \ldots, r_s, a_s\right)$; Exponential family distribution:
    - $\mathbb{E}\left[\eta_{s+1} | \mathcal{F}_s\right] = 0$.

Application to GLBs

- Using optimality of $\hat{\theta}_t$ for the regularized log-loss $\mathcal{L}_t^{\lambda_t}$:

$$\forall t \geq 1, \quad \left\| \theta_\star - \hat{\theta}_t \right\|_{\boldsymbol{H}_t(\theta_\star)} \leq \left\| \sum_{s=1}^{t-1} \eta_{s+1} a_s \right\|_{\boldsymbol{H}_t^{-1}(\theta_\star)}$$

where:

$$\eta_{s+1} = r_{s+1} - \mu(a_s^\top \theta_\star) \quad \text{and} \quad \boldsymbol{H}_t(\theta) = \sum_{s=1}^{t} \dot{\mu}(a_s^\top \theta) a_s a_s^\top + \lambda_t \boldsymbol{I}_d$$

- Define $\mathcal{F}_s = \sigma\left(a_1, r_2, \ldots, r_s, a_s\right)$; Exponential family distribution:
  - $\mathbb{E}\left[\eta_{s+1} | \mathcal{F}_s\right] = 0.$
  - $\mathbb{V}\text{ar}(\eta_{s+1} | \mathcal{F}_s) = \mathbb{V}\text{ar}(r_{s+1} | \mathcal{F}_s)$

Application to GLBs

- Using optimality of $\hat{\theta}_t$ for the regularized log-loss $\mathcal{L}_t^{\lambda_t}$:

$$\forall t \geq 1, \quad \left\| \theta_\star - \hat{\theta}_t \right\|_{\boldsymbol{H}_t(\theta_\star)} \leq \left\| \sum_{s=1}^{t-1} \eta_{s+1} a_s \right\|_{\boldsymbol{H}_t^{-1}(\theta_\star)}$$

where:

$$\eta_{s+1} = r_{s+1} - \mu(a_s^\top \theta_\star) \quad \text{and} \quad \boldsymbol{H}_t(\theta) = \sum_{s=1}^{t} \dot{\mu}(a_s^\top \theta) a_s a_s^\top + \lambda_t \boldsymbol{I}_d$$

- Define $\mathcal{F}_s = \sigma\left(a_1, r_2, \ldots, r_s, a_s\right)$; Exponential family distribution:
  - $\mathbb{E}\left[\eta_{s+1} | \mathcal{F}_s\right] = 0.$
  - $\mathbb{V}\text{ar}(\eta_{s+1} | \mathcal{F}_s) = \dot{\mu}(a_s^\top \theta_\star)$

Application to GLBs

- Using optimality of $\hat{\theta}_t$ for the regularized log-loss $\mathcal{L}_t^{\lambda_t}$:

$$\forall t \geq 1, \quad \left\|\theta_\star - \hat{\theta}_t\right\|_{\boldsymbol{H}_t(\theta_\star)} \leq \left\|\sum_{s=1}^{t-1} \eta_{s+1} a_s\right\|_{\boldsymbol{H}_t^{-1}(\theta_\star)}$$

where:

$$\eta_{s+1} = r_{s+1} - \mu(a_s^\top \theta_\star) \quad \text{and} \quad \boldsymbol{H}_t(\theta) = \sum_{s=1}^{t} \dot{\mu}(a_s^\top \theta) a_s a_s^\top + \lambda_t \boldsymbol{I}_d$$

- Define $\mathcal{F}_s = \sigma\left(a_1, r_2, \ldots, r_s, a_s\right)$; Exponential family distribution:
  - $\mathbb{E}\left[\eta_{s+1} | \mathcal{F}_s\right] = 0$.
  - $\mathbb{V}\text{ar}(\eta_{s+1} | \mathcal{F}_s) = \dot{\mu}(a_s^\top \theta_\star)$ so $\boldsymbol{H}_t(\theta_\star) = \sum_{s=1}^t \mathbb{V}\text{ar}(\eta_{s+1} | \mathcal{F}_s) a_s a_s^\top + \lambda \boldsymbol{I}_d$.

Application to GLBs

- Using optimality of $\hat{\theta}_t$ for the regularized log-loss $\mathcal{L}_t^{\lambda_t}$:

$$\forall t \geq 1, \quad \left\| \theta_\star - \hat{\theta}_t \right\|_{\boldsymbol{H}_t(\theta_\star)} \leq \left\| \sum_{s=1}^{t-1} \eta_{s+1} a_s \right\|_{\boldsymbol{H}_t^{-1}(\theta_\star)}$$

where:

$$\eta_{s+1} = r_{s+1} - \mu(a_s^\top \theta_\star) \quad \text{and} \quad \boldsymbol{H}_t(\theta) = \sum_{s=1}^{t} \dot{\mu}(a_s^\top \theta) a_s a_s^\top + \lambda_t \boldsymbol{I}_d$$

- Define $\mathcal{F}_s = \sigma(a_1, r_2, \ldots, r_s, a_s)$; Exponential family distribution:
  - $\mathbb{E}\left[\eta_{s+1} | \mathcal{F}_s\right] = 0$.
  - $\mathbb{V}\text{ar}(\eta_{s+1}|\mathcal{F}_s) = \dot{\mu}(a_s^\top \theta_\star)$ so $\boldsymbol{H}_t(\theta_\star) = \sum_{s=1}^t \mathbb{V}\text{ar}(\eta_{s+1}|\mathcal{F}_s) a_s a_s^\top + \lambda \boldsymbol{I}_d$.

Application to GLBs

- Using optimality of $\hat{\theta}_t$ for the regularized log-loss $\mathcal{L}_t^{\lambda_t}$:

$$\forall t \geq 1, \quad \left\| \theta_\star - \hat{\theta}_t \right\|_{\boldsymbol{H}_t(\theta_\star)} \leq \left\| \sum_{s=1}^{t-1} \eta_{s+1} a_s \right\|_{\boldsymbol{H}_t^{-1}(\theta_\star)} = \mathcal{O}\left(\sqrt{d \log(t/\delta)} \right)$$

where:

$$\eta_{s+1} = r_{s+1} - \mu(a_s^\top \theta_\star) \quad \text{and} \quad \boldsymbol{H}_t(\theta) = \sum_{s=1}^{t} \dot{\mu}(a_s^\top \theta) a_s a_s^\top + \lambda_t \boldsymbol{I}_d$$

- Define $\mathcal{F}_s = \sigma\left(a_1, r_2, \ldots, r_s, a_s\right)$; Exponential family distribution:
  - $\mathbb{E}\left[\eta_{s+1} | \mathcal{F}_s\right] = 0$.
  - $\mathbb{V}\text{ar}(\eta_{s+1} | \mathcal{F}_s) = \dot{\mu}(a_s^\top \theta_\star)$ so $\boldsymbol{H}_t(\theta_\star) = \sum_{s=1}^{t} \mathbb{V}\text{ar}(\eta_{s+1} | \mathcal{F}_s) a_s a_s^\top + \lambda \boldsymbol{I}_d$.

Application to GLBs

- Using optimality of $\hat{\theta}_t$ for the regularized log-loss $\mathcal{L}_t^{\lambda_t}$:

$$\forall t \geq 1, \quad \left\| \theta_\star - \hat{\theta}_t \right\|_{\boldsymbol{H}_t(\theta_\star)} \leq \left\| \sum_{s=1}^{t-1} \eta_{s+1} a_s \right\|_{\boldsymbol{H}_t^{-1}(\theta_\star)} = \mathcal{O}\left(\sqrt{d \log(t/\delta)}\right)$$

where:

$$\eta_{s+1} = r_{s+1} - \mu(a_s^\top \theta_\star) \quad \text{and} \quad \boldsymbol{H}_t(\theta) = \sum_{s=1}^{t} \dot{\mu}(a_s^\top \theta) a_s a_s^\top + \lambda_t \boldsymbol{I}_d$$

Application to GLBs

- Using optimality of $\hat{\theta}_t$ for the regularized log-loss $\mathcal{L}_t^{\lambda_t}$:

$$\forall t \geq 1, \quad \left\| \theta_\star - \hat{\theta}_t \right\|_{\boldsymbol{H}_t(\theta_\star)} \leq \left\| \sum_{s=1}^{t-1} \eta_{s+1} a_s \right\|_{\boldsymbol{H}_t^{-1}(\theta_\star)} = \mathcal{O}\left(\sqrt{d \log(t/\delta)}\right)$$

where:

$$\eta_{s+1} = r_{s+1} - \mu(a_s^\top \theta_\star) \quad \text{and} \quad \boldsymbol{H}_t(\theta) = \sum_{s=1}^{t} \dot{\mu}(a_s^\top \theta) a_s a_s^\top + \lambda_t \boldsymbol{I}_d$$

- New confidence set:

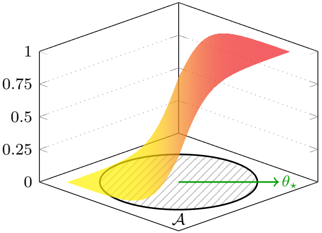
Proposition (F., Abeille, Calauzènes and Fercoq, 2020.)

For $\delta \in (0, 1]$ *let:*

$$\mathcal{C}_t(\delta) := \left\{ \|\theta\| \leq S, \left\| \theta - \hat{\theta}_t \right\|_{\boldsymbol{H}_t(\theta)} \leq \mathcal{O}\left(\sqrt{d \log(t/\delta)}\right) \right\} .$$

Then $\mathbb{P}\left(\forall t \geq 1, \, \theta_\star \in \mathcal{C}_t(\delta)\right) \geq 1 - \delta$.

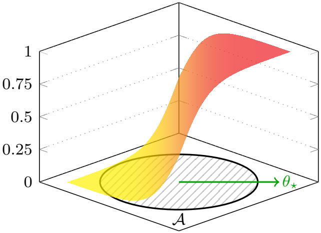
GLBs: information vs. regret



$$\mathbb{E}[r_{t+1} | a_t] = (1 + \exp(-a_t{}^\mathsf{T}\theta_\star))^{-1}$$

- Varying reward sensitivity:

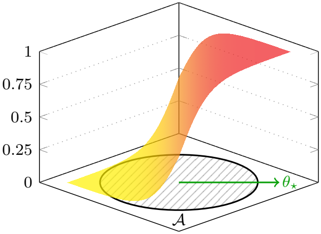
GLBs: information vs. regret



$$\mathbb{E}\left[r_{t+1} \middle| a_t\right] = (1 + \exp(-a_t^\mathsf{T} \theta_\star))^{-1}$$

- Varying reward sensitivity:
  - low-sensitivity:
    - information is hard to get
    - small regret

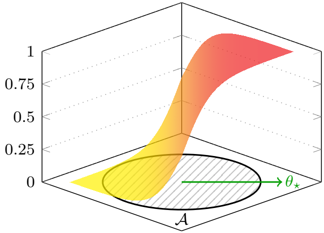
GLBs: information vs. regret



$$\mathbb{E}\left[r_{t+1} \middle| a_t\right] = (1 + \exp(-a_t{}^\mathsf{T}\theta_\star))^{-1}$$

- Varying reward sensitivity:
  - low-sensitivity:
    - information is hard to get
    - small regret
  - high-sensitivity:
    - information is easy to get
    - large regret

GLBs: information vs. regret

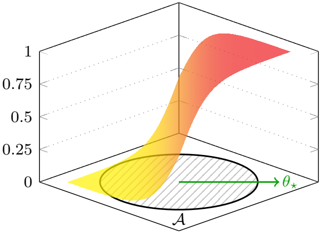


$$\mathbb{E}\left[r_{t+1} \middle| a_t\right] = (1 + \exp(-a_t^\mathsf{T}\theta_\star))^{-1}$$

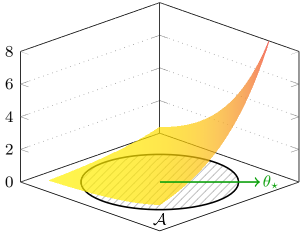
- Varying reward sensitivity:
  - low-sensitivity:
    - information is hard to get
    - small regret
  - high-sensitivity:
    - information is easy to get
    - large regret

$\Rightarrow$ linearization: worst of both world.

GLBs: information vs. regret



$$\mathbb{E}[r_{t+1} | a_t] = (1 + \exp(-a_t^\mathsf{T} \theta_\star))^{-1}$$



$$\mathbb{E}[r_{t+1} | a_t] = \exp(a_t^\mathsf{T} \theta_\star)$$

- Varying reward sensitivity:
  - low-sensitivity:
    - information is hard to get
    - small regret
  - high-sensitivity:
    - information is easy to get
    - large regret

$\Rightarrow$ linearization: worst of both world.

Minimax rates in general non-stationary settings

- Beyond piece-wise stationarity thanks to *variation-budget*:

$$B_T := \sum_{t=2}^{T} \|\theta_\star^t - \theta_\star^{t-1}\| \ .$$

  - describe broader non-stationary environments.

Minimax rates in general non-stationary settings

- Beyond piece-wise stationarity thanks to *variation-budget*:

$$B_T := \sum_{t=2}^{T} \|\theta_\star^t - \theta_\star^{t-1}\| \ .$$

  - describe broader non-stationary environments.

- Forgetting mechanisms to the rescue?
  - minimax-optimal for the MAB setting

Minimax rates in general non-stationary settings

- Beyond piece-wise stationarity thanks to *variation-budget*:

$$B_T := \sum_{t=2}^{T} \|\theta_\star^t - \theta_\star^{t-1}\| \, .$$

  - describe broader non-stationary environments.

- Forgetting mechanisms to the rescue?
  - minimax-optimal for the MAB setting
  - not so well understood in LB! *[F. et al, 2021a]*

Minimax rates in general non-stationary settings

- Beyond piece-wise stationarity thanks to *variation-budget*:

$$B_T := \sum_{t=2}^{T} \|\theta_\star^t - \theta_\star^{t-1}\| \; .$$

  - describe broader non-stationary environments.

- Forgetting mechanisms to the rescue?
  - minimax-optimal for the MAB setting
  - not so well understood in LB! *[F. et al, 2021a]*
  - best know regret bound for GLBs *[F. et al. 2021b]*:

$$\text{DynamicRegret}(T) = \widetilde{\mathcal{O}}\left(\boldsymbol{\kappa}_\mu B_T^{1/5} T^{4/5}\right) \; .$$

Minimax rates in general non-stationary settings

- Beyond piece-wise stationarity thanks to *variation-budget*:

$$B_T := \sum_{t=2}^{T} \|\theta_\star^t - \theta_\star^{t-1}\| \, .$$

  - describe broader non-stationary environments.

- Forgetting mechanisms to the rescue?
  - minimax-optimal for the MAB setting
  - not so well understood in LB! *[F. et al, 2021a]*
  - best know regret bound for GLBs *[F. et al. 2021b]*:

$$\text{DynamicRegret}(T) = \widetilde{\mathcal{O}}\left(\boldsymbol{\kappa}_\mu B_T^{1/5} T^{4/5}\right) \, .$$

- Room for improvement! $\rightarrow$ *[Wei and Luo, 2021]*