Instance-Wise Minimax-Optimal Algorithms for Logistic Bandits

Marc Abeille[1], **Louis Faury**[1,2], **Clément Calauzènes**[1]

[1] Criteo AI Lab [2] LTCI Telecom Paris

Presentation Outline

- **Goal**.
  - Study non-linearity in sequential decision making.
  - A simple problem: the Logistic Bandit.
    - $\rightsquigarrow$ Compact non-linear extension to the Linear Bandit.
    - $\rightsquigarrow$ Very relevant in practical problems with **binary** feedback.

Presentation Outline

- **Goal**.
  - Study non-linearity in sequential decision making.
  - A simple problem: the Logistic Bandit.
    - $\rightsquigarrow$ Compact non-linear extension to the Linear Bandit.
    - $\rightsquigarrow$ Very relevant in practical problems with **binary** feedback.

- **Logistic Bandit: high-level contributions.**
  - [Filippi et al. 2010, Faury et al. 2020]: non-linearity is harmful. Actually:

Non-linearity can make the problem **easier**.

Presentation Outline

- **Goal**.
  - Study non-linearity in sequential decision making.
  - A simple problem: the Logistic Bandit.
    - $\rightsquigarrow$ Compact non-linear extension to the Linear Bandit.
    - $\rightsquigarrow$ Very relevant in practical problems with **binary** feedback.

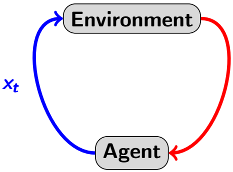
- **Logistic Bandit: high-level contributions.**
  - [Filippi et al. 2010, Faury et al. 2020]: non-linearity is harmful. Actually:

Non-linearity can make the problem **easier**.

  - Identify two distinct regimes:
    - $\rightsquigarrow$ Short-term $\leftrightarrow$ early exploration phase: neutral (most often).
    - $\rightsquigarrow$ Long-term $\leftrightarrow$ exploration-exploitation phase: beneficial.

The Learning Problem

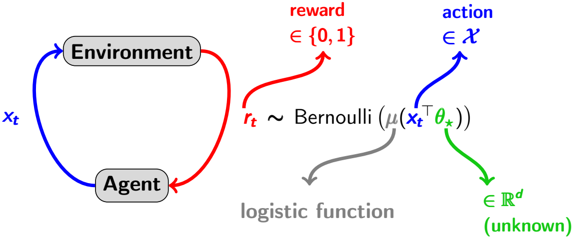
- Repeated game with **structured binary** feedback.



$$\overbrace{r_t}^{\text{reward } \in \{0,1\}} \sim \text{Bernoulli}\left(\underbrace{\mu}_{\text{logistic function}}\big(\overbrace{x_t}^{\text{action } \in \mathcal{X}}{}^\top \underbrace{\theta_\star}_{\in \mathbb{R}^d \text{ (unknown)}}\big)\right)$$

The Learning Problem

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

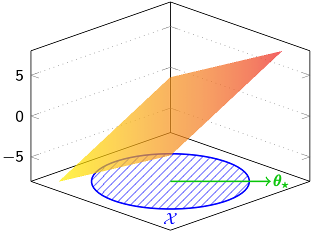


- **Regret.** The agent tries to minimize its cumulative pseudo-regret:

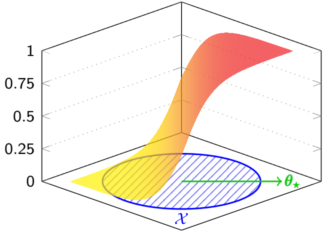
$$\text{Regret}_{\theta_\star}(T) := T \max_{x \in \mathcal{X}} \mu(x^\top \theta_\star) - \sum_{t=1}^{T} \mu(x_t{}^\top \theta_\star) \ .$$

The Learning Problem (ctn'd)

- **Reward model.** Minimalist non-linear extension from the linear bandit.



$$\mathbb{E}[r_t | x_t] = x_t^\mathsf{T}\theta_\star$$

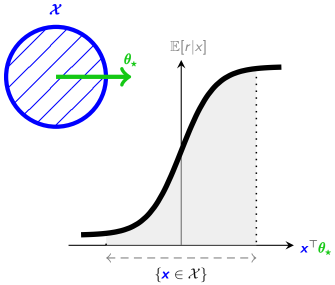


$$\mathbb{E}[r_t | x_t] = (1 + \exp(-x_t^\mathsf{T}\theta_\star))^{-1}$$

- **Exploration-exploitation.** Same recipe:
  - Learning: maximum likelihood.
  - Planning: Optimism through confidence sets.

- **Additional challenge.** Non-linearity: information vs. regret.

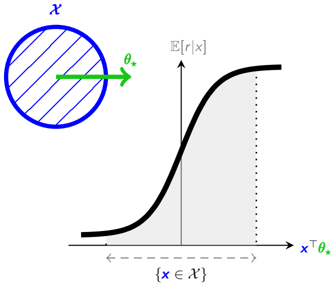
Quantifying non-linearity



- **Level of non-linearity = conditioning.**
  - How **flat** are the tails.

- **Important quantities.** The level of non-linearity is problem-dependent.

Quantifying non-linearity

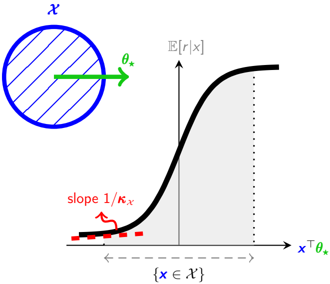


- **Level of non-linearity = conditioning.**
  - How **flat** are the tails.

- **Important quantities.** The level of non-linearity is problem-dependent.
  - Historically characterized by a constant $\kappa_{\mathcal{X}}$:

$$\kappa_{\mathcal{X}} := \frac{1}{\min_{x \in \mathcal{X}} \dot{\mu}(x^\top \theta_\star)} .$$

Quantifying non-linearity



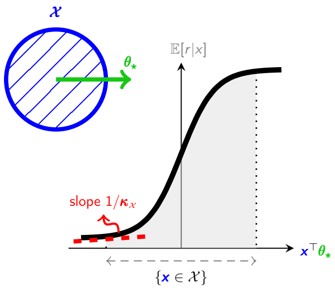
- **Level of non-linearity = conditioning.**
  - How **flat** are the tails.

- **Important quantities.** The level of non-linearity is problem-dependent.
  - Historically characterized by a constant $\kappa_{\mathcal{X}}$:

$$\kappa_{\mathcal{X}} := \frac{1}{\min_{x\in\mathcal{X}} \dot{\mu}(x^\top\theta_\star)} .$$

Quantifying non-linearity

- **Level of non-linearity = conditioning.**
    - How **flat** are the tails.



- **Important quantities.** The level of non-linearity is problem-dependent.
    - Historically characterized by a constant $\kappa_{\mathcal{X}}$:

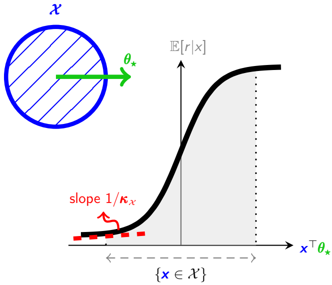
$$\kappa_{\mathcal{X}} := \frac{1}{\min_{x\in\mathcal{X}} \dot{\mu}(x^\top\theta_\star)}.$$

the more non-linear the bigger

$\propto \exp(\|\theta_\star\|)$

Quantifying non-linearity

- **Level of non-linearity = conditioning.**
  - How **flat** are the tails.



- **Important quantities.** The level of non-linearity is problem-dependent.
  - Historically characterized by a constant $\kappa_\mathcal{X}$:

$$\kappa_\mathcal{X} := \frac{1}{\min_{x\in\mathcal{X}} \dot{\mu}(x^\top\theta_\star)} .$$

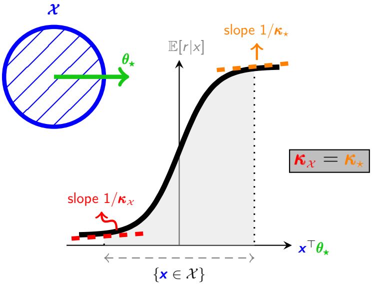
the more non-linear the bigger

$\propto \exp(\|\theta_\star\|)$

  - Inverse slope at the optimum; letting $x_\star = \operatorname{argmax}_{x\in\mathcal{X}} x^\top\theta_\star$:

$$\kappa_\star := \frac{1}{\dot{\mu}(x_\star^\top\theta_\star)} .$$

Quantifying non-linearity



- **Level of non-linearity = conditioning.**
  - How **flat** are the tails.

- **Important quantities.** The level of non-linearity is problem-dependent.
  - Historically characterized by a constant $\kappa_{\mathcal{X}}$:

$$\kappa_{\mathcal{X}} := \frac{1}{\min_{x\in\mathcal{X}} \dot{\mu}(x^\top\theta_\star)} \,.$$

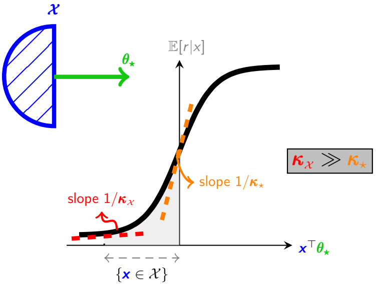
the more non-linear the bigger

$\propto \exp(\|\theta_\star\|)$

  - Inverse slope at the optimum; letting $x_\star = \operatorname{argmax}_{x\in\mathcal{X}} x^\top\theta_\star$:

$$\kappa_\star := \frac{1}{\dot{\mu}(x_\star^\top\theta_\star)} \,.$$

Quantifying non-linearity



- **Level of non-linearity = conditioning.**
  - How **flat** are the tails.

- **Important quantities.** The level of non-linearity is problem-dependent.
  - Historically characterized by a constant $\kappa_\mathcal{X}$:

$$\kappa_\mathcal{X} := \frac{1}{\min_{x \in \mathcal{X}} \dot{\mu}(x^\top \theta_\star)} \, .$$

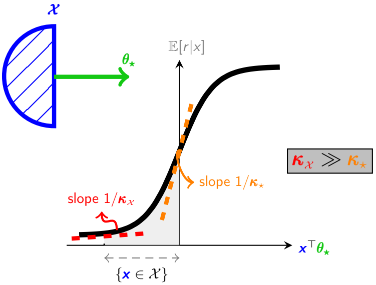
the more non-linear the bigger

$\propto \exp(\|\theta_\star\|)$

  - Inverse slope at the optimum; letting $x_\star = \operatorname{argmax}_{x \in \mathcal{X}} x^\top \theta_\star$:

$$\kappa_\star := \frac{1}{\dot{\mu}(x_\star^\top \theta_\star)} \, .$$

Quantifying non-linearity



- **Level of non-linearity = conditioning.**
  - How **flat** are the tails.

- **Important quantities.** The level of non-linearity is problem-dependent.
  - Historically characterized by a constant $\kappa_{\mathcal{X}}$:

$$\kappa_{\mathcal{X}} := \frac{1}{\min_{x\in\mathcal{X}} \dot{\mu}(x^\top\theta_\star)} .$$

the more non-linear the bigger

$\propto \exp(\|\theta_\star\|)$

  - Inverse slope at the optimum; letting $x_\star = \text{argmax}_{x\in\mathcal{X}}\, x^\top\theta_\star$:

$$\kappa_\star := \frac{1}{\dot{\mu}(x_\star^\top\theta_\star)} .$$

$\in [4, \kappa_{\mathcal{X}}]$

Non-linearity vs. regret: previous work

Approach	Regret
[Filippi et al. 2010] Linearization (global)	$\tilde{\mathcal{O}}\left(\boldsymbol{\kappa}_{\mathcal{X}} d\sqrt{T}\right)$
[Faury et al. 2020] Self-concordance (local)	$\tilde{\mathcal{O}}\left(d\sqrt{T} + \boldsymbol{\kappa}_{\mathcal{X}}\right)$
This work Refined local approach	$\tilde{\mathcal{O}}\left(d\sqrt{T/\boldsymbol{\kappa}_{\star}}\,(+\boldsymbol{\kappa}_{\mathcal{X}})\right)$

Non-linearity vs. regret: previous work

Approach	Regret
[Filippi et al. 2010] Linearization (global)	$\tilde{\mathcal{O}}\left(\boldsymbol{\kappa}_{\mathcal{X}} d\sqrt{T}\right)$
[Faury et al. 2020] Self-concordance (local)	$\tilde{\mathcal{O}}\left(d\sqrt{T} + \boldsymbol{\kappa}_{\mathcal{X}}\right)$
This work Refined local approach	$\tilde{\mathcal{O}}\left(d\sqrt{T/\boldsymbol{\kappa}_\star}\ (+\boldsymbol{\kappa}_{\mathcal{X}})\right)$

- **Exponential improvement**. If $\boldsymbol{\mathcal{X}} = \{\|x\| \leq 1\}$ then $\boldsymbol{\kappa}_{\mathcal{X}} = \boldsymbol{\kappa}_\star \geq e^{\|\boldsymbol{\theta}_\star\|}$ then regret:

$$\boxed{\tilde{\mathcal{O}}(e^{\|\boldsymbol{\theta}_\star\|} d\sqrt{T}) \longrightarrow \tilde{\mathcal{O}}(d\sqrt{T} + e^{\|\boldsymbol{\theta}_\star\|}) \longrightarrow \tilde{\mathcal{O}}(e^{-\|\boldsymbol{\theta}_\star\|/2} d\sqrt{T})}$$

Regret Upper-Bound

- **Effects of non-linearity**: transitory and permanent regime.

$$\text{Regret}_{\theta_\star}(T) = \underbrace{R^{\text{perm}}(T)}_{\tilde{\mathcal{O}}(\sqrt{T})} + \underbrace{R^{\text{trans}}(T)}_{\tilde{\mathcal{O}}(1)}$$

Regret Upper-Bound

- **Effects of non-linearity**: transitory and permanent regime.

$$\text{Regret}_{\boldsymbol{\theta}_\star}(T) = \underbrace{R^{\text{perm}}(T)}_{\tilde{\mathcal{O}}(\sqrt{T})} + \underbrace{R^{\text{trans}}(T)}_{\tilde{\mathcal{O}}(1)}$$

- **Permanent regime.** For $t \gg 1$, only the local slope around $x_\star$ matters.
  - Conceptually:
    - Sub-linear regret $\rightsquigarrow$ play mostly $x_t \approx x_\star$ for large t.
    - Linear bandit with slope $\dot{\mu}(x_\star^\top \boldsymbol{\theta}_\star) = \frac{1}{\kappa_\star}$ (potentially $\ll 1$).

Regret Upper-Bound

- **Effects of non-linearity**: transitory and permanent regime.

$$\text{Regret}_{\theta_\star}(T) = \underbrace{R^{\text{perm}}(T)}_{\tilde{\mathcal{O}}(\sqrt{T})} + \underbrace{R^{\text{trans}}(T)}_{\tilde{\mathcal{O}}(1)}$$

- **Permanent regime.** For $t \gg 1$, only the local slope around $x_\star$ matters.
  - Conceptually:
    - Sub-linear regret $\rightsquigarrow$ play mostly $x_t \approx x_\star$ for large t.
    - Linear bandit with slope $\dot{\mu}(x_\star^\top \theta_\star) = \frac{1}{\kappa_\star}$ (potentially $\ll 1$).
  - The smaller this local slope, the easier the problem:

$$R^{\text{perm}}(T) = \tilde{\mathcal{O}}\left(d\sqrt{T/\kappa_\star}\right) .$$

    - Formal proof: self-concordance.

Regret Upper-Bound

- **Effects of non-linearity**: transitory and permanent regime.

$$\text{Regret}_{\boldsymbol{\theta}_\star}(T) = \underbrace{R^{\text{perm}}(T)}_{\tilde{\mathcal{O}}(\sqrt{T})} + \underbrace{R^{\text{trans}}(T)}_{\tilde{\mathcal{O}}(1)}$$

- **Permanent regime.** For $t \gg 1$, only the local slope around $x_\star$ matters.
  - Conceptually:
    - Sub-linear regret $\rightsquigarrow$ play mostly $x_t \approx x_\star$ for large t.
    - Linear bandit with slope $\dot{\mu}(x_\star^\top \boldsymbol{\theta}_\star) = \frac{1}{\boldsymbol{\kappa}_\star}$ (potentially $\ll 1$).
  - The smaller this local slope, the easier the problem:

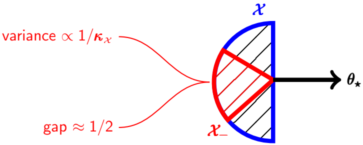
$$R^{\text{perm}}(T) = \tilde{\mathcal{O}}\left(d\sqrt{T/\boldsymbol{\kappa}_\star}\right) .$$

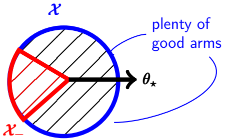
$\approx \exp(\|\boldsymbol{\theta}_\star\|)$!

    - Formal proof: self-concordance.
  - Question: how long to reach it?

Regret Upper Bounds (ctn'd)

- **Transitory Regret.** Also linked to the problem's geometry..
  - Proportion of **detrimental** arms: little information and large sub-optimality.



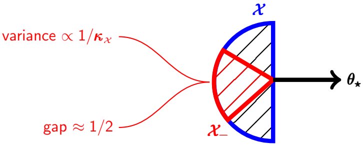


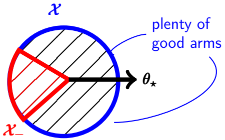
  - Transitory regret = how long are we stuck playing **detrimental** arms?

$$R^{\text{trans}}(T) \propto \sum_{t=1}^{T} \mathbb{1}(x_t \in \mathcal{X}_-)$$

Regret Upper Bounds (ctn'd)

- **Transitory Regret.** Also linked to the problem's geometry..
  - Proportion of **detrimental** arms: little information and large sub-optimality.





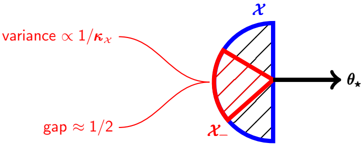
$$\boxed{R^{\text{trans}}(T) = \tilde{\mathcal{O}}(\kappa_{\mathcal{X}})}$$

  - Transitory regret = how long are we stuck playing **detrimental** arms?

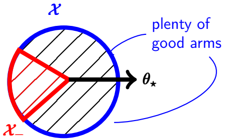
$$R^{\text{trans}}(T) \propto \sum_{t=1}^{T} \mathbb{1}(x_t \in \mathcal{X}_-)$$

Regret Upper Bounds (ctn'd)

- **Transitory Regret.** Also linked to the problem's geometry..
  - Proportion of **detrimental** arms: little information and large sub-optimality.



$$\boxed{R^{\text{trans}}(T) = \tilde{\mathcal{O}}(\boldsymbol{\kappa}_{\mathcal{X}})}$$



$$\boxed{R^{\text{trans}}(T) = \tilde{\mathcal{O}}(1)}$$

- Transitory regret = how long are we stuck playing **detrimental** arms?

$$R^{\text{trans}}(T) \propto \sum_{t=1}^{T} \mathbb{1}(x_t \in \mathcal{X}_-)$$

Regret Upper Bounds (ctn'd)

- **Wrapping up.**

Theorem (Regret upper-bound)

With high probability:

$$\text{Regret}_{\boldsymbol{\theta}_\star}(T) = \tilde{\mathcal{O}}\left(d\sqrt{T/\boldsymbol{\kappa}_\star} + (\boldsymbol{\kappa}_{\mathcal{X}})\right)$$

- Refined problem-dependent bounds:

Regret Upper Bounds (ctn'd)

- **Wrapping up.**

Theorem (Regret upper-bound)

With high probability:

$$\text{Regret}_{\boldsymbol{\theta}_\star}(T) = \tilde{\mathcal{O}}\left(d\sqrt{T/\boldsymbol{\kappa}_\star} + (\boldsymbol{\kappa}_{\mathcal{X}})\right)$$

- Refined problem-dependent bounds:
  - **Worst configuration.**

$$\text{Regret}_{\boldsymbol{\theta}_\star}(T) = \tilde{\mathcal{O}}(d\sqrt{T} + \boldsymbol{\kappa}_{\mathcal{X}})$$

Regret Upper Bounds (ctn'd)

- **Wrapping up.**

Theorem (Regret upper-bound)

With high probability:

$$\text{Regret}_{\theta_\star}(T) = \tilde{\mathcal{O}}\left(d\sqrt{T/\kappa_\star} + (\kappa_{\mathcal{X}})\right)$$

- Refined problem-dependent bounds:
  - **Worst configuration.**

$$\text{Regret}_{\theta_\star}(T) = \tilde{\mathcal{O}}(d\sqrt{T} + \kappa_{\mathcal{X}})$$

  - **Best configuration.**

$$\text{Regret}_{\theta_\star}(T) = \tilde{\mathcal{O}}(d\sqrt{T/\kappa_{\mathcal{X}}})$$

$\rightsquigarrow$ **Is this optimal?**

Problem-dependent lower-bound

- **Challenge.** Study optimality w.r.t problem-dependent constants $\boldsymbol{\kappa}_{\mathcal{X}}$.
  - Lower-bound for a *continuum* of problems, each with different $\boldsymbol{\kappa}_{\mathcal{X}}$.
  - Traditional lower-bound technique fails.

Problem-dependent lower-bound

- **Challenge.** Study optimality w.r.t problem-dependent constants $\boldsymbol{\kappa}_{\mathcal{X}}$.
  - Lower-bound for a *continuum* of problems, each with different $\boldsymbol{\kappa}_{\mathcal{X}}$.
  - Traditional lower-bound technique fails.

Theorem (A local lower-bound)

Let $\mathcal{X} = \{\|x\| = 1\}$, fix $\theta_\star \in \mathbb{R}^d$ and denote $\kappa = \kappa_\star(\theta_\star)$. For any policy

$$\boxed{\max_{\|\theta' - \theta_\star\| \leq \varepsilon} \text{Regret}_{\theta'}(T) = \Omega\left(d\sqrt{T/\kappa}\right)} \quad \text{if } T \geq \kappa$$

where ε is such that $\forall \theta' \in \{\|\theta' - \theta\| \leq \epsilon$ we have $\kappa_\star(\theta') = \Theta(\kappa)$.

Problem-dependent lower-bound

- **Challenge.** Study optimality w.r.t problem-dependent constants $\boldsymbol{\kappa}_{\mathcal{X}}$.
  - Lower-bound for a *continuum* of problems, each with different $\boldsymbol{\kappa}_{\mathcal{X}}$.
  - Traditional lower-bound technique fails.

Theorem (A local lower-bound)

Let $\mathcal{X} = \{\|x\| = 1\}$, fix $\theta_\star \in \mathbb{R}^d$ and denote $\kappa = \kappa_\star(\theta_\star)$. For any policy

$$\boxed{\max_{\|\theta' - \theta_\star\| \leq \varepsilon} \text{Regret}_{\theta'}(T) = \Omega\left(d\sqrt{T/\kappa}\right)} \quad \text{if } T \geq \kappa$$

where ε is such that $\forall \theta' \in \{\|\theta' - \theta\| \leq \epsilon$ we have $\kappa_\star(\theta') = \Theta(\kappa)$.

- **Interpretation.** For *any* problem:
  - Consider the hardest alternative in nearby instances.
  - That share the same problem-dependent constant $\boldsymbol{\kappa}_\star$.

- **Conclusion.** The long-term regret is **tight**.

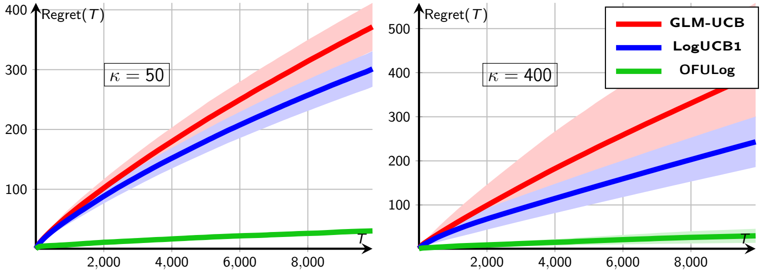
Algorithm

- **Algorithm**. `OFULog`:
  - Relies on the confidence set $\mathcal{C}_t(\delta)$ of [Faury et al. 2020].[1]
  - Parameter-based optimism (vs. bonus-based)

$$\boxed{\mathbf{x_t} = \max_{x \in \mathcal{X}} \max_{\theta \in \mathcal{C}_t(\delta)} x^\top \theta} \qquad \left(\max_{x \in \mathcal{X}} \mu(x^\top \hat{\theta}_t) + \varepsilon_t(x)\right)$$

  - More adaptive to the problem effective's hardness.
  - Tractable algorithm (no non-convex optimization routines).

- **In practice.** Large improvement on the regret.



[1] We also introduce a convex relaxation which leads to a fully tractable algorithm

Bibliography

- Sarah Filippi, Olivier Cappé, Aurélien Garivier, Csaba Szepesvári. *Parametric Bandits: The Generalized Linear Case*, 2010.
- Francis Bach. *Self-Concordant Analysis for Logistic Regression*, 2010.
- Shi Dong, Tengyu Ma, Benjamin Van Roy. *On the Performance of Thompson Sampling on Logistic Bandits, 2019*.
- Louis Faury, Marc Abeille, Clément Calauzènes, Olivier Fercoq. *Improved Optimistic Algorithms for Logistic Bandits*, 2020.

See you at the Q&A!