Self-Concordant Perturbations for Linear Bandits

Lucas Lévy[1,2,†] Jean-Lou Valeau[1,3] Arya Akhavan[1]
Patrick Rebeschini[1]

[1]Department of Statistics, University of Oxford
[2]École Polytechnique, IP Paris
[3]ENSAE Paris, IP Paris
[†]Currently at Inria, Ecole Normale Supérieure, Paris

Conference on Learning Theory, July 1st 2026

DEPARTMENT OF
STATISTICS

Linear Bandits

Given an horizon n, an action set $K \subset \mathbb{R}^d$, and an unknown sequence of loss vectors $(y_t)_{t=1}^n$, sequentially choose an action $a_t \in K$ then observe punctual loss $\langle y_t, a_t \rangle$.

We define the *regret*:

$$R_n := \mathbb{E}\left[\sum_{t=1}^{n} \langle y_t, a_t \rangle\right] - \inf_{u \in K} \sum_{t=1}^{n} \langle y_t, u \rangle$$

FTRL & FTPL

FTRL: $a_t = \arg\min_{a \in K} \{ \sum_{s<t} \langle a, y_s \rangle + \psi(a) \}$

FTPL: $a_t = \arg\min_{a \in K} \{ \sum_{s<t} \langle a, y_s \rangle + \langle a, \xi_t \rangle \}$ with $\xi_t \sim D$

FTRL & FTPL

FTRL: $a_t = \arg\min_{a \in K} \{ \sum_{s<t} \langle a, y_s \rangle + \psi(a) \}$

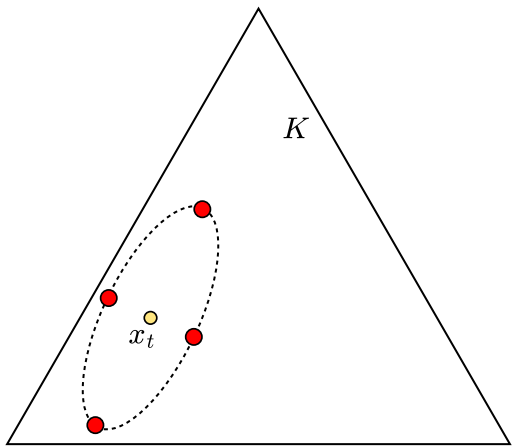
FTPL: $a_t = \arg\min_{a \in K} \{ \sum_{s<t} \langle a, y_s \rangle + \langle a, \xi_t \rangle \}$ with $\xi_t \sim D$

- Duality when $(y_s)_{s<t}$ are known: both part of the GBPA framework (Abernethy et al., '14)
- Issue: we do not know the y_s. Instead, use estimates $\hat{y}_s$ of y_s using only $\langle y_s, a_s \rangle$. Need *randomization* when taking our action.

Exploration

- For FTRL-based alg., need for *additional exploration mechanism*

SCRiBLe: x_t given by FTRL step, then
sample $a_t \sim \mathcal{U}(\text{poles of } W(x_t))$

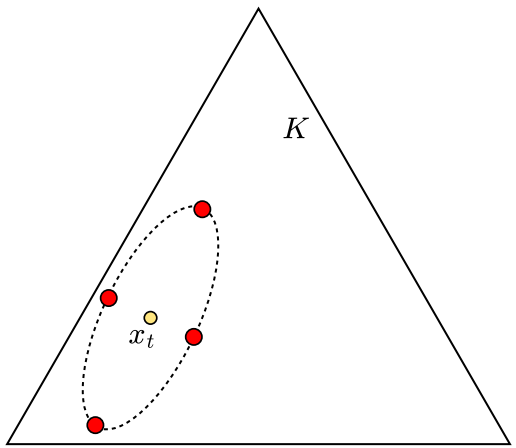


- Uses self-concordant barriers
- Worst-case regret in $O(d^{3/2}\sqrt{n \ln n})$

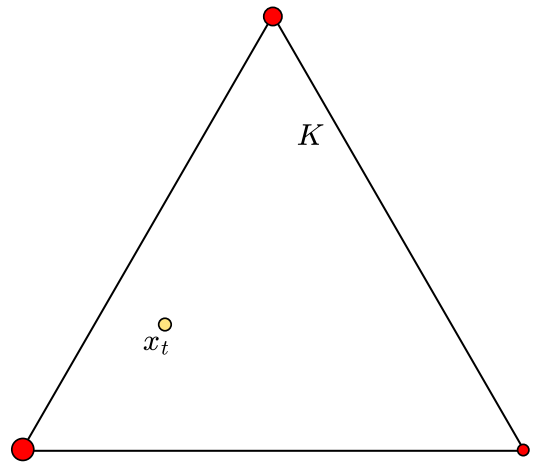
Exploration

- For FTRL-based alg., need for *additional exploration mechanism*

SCRiBLe: x_t given by FTRL step, then sample $a_t \sim \mathcal{U}(\text{poles of } W(x_t))$



FTPL: a_t solves a stochastically perturbed LP

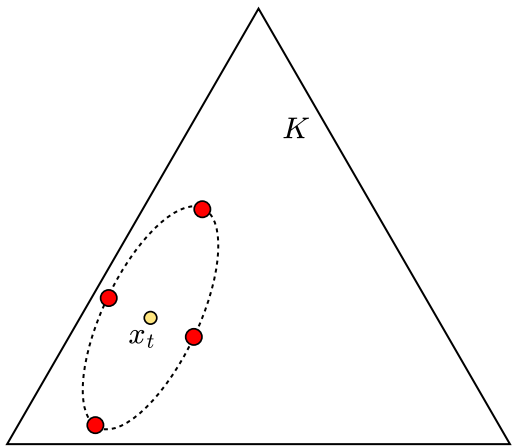


- Uses self-concordant barriers
- Worst-case regret in $O(d^{3/2}\sqrt{n \ln n})$

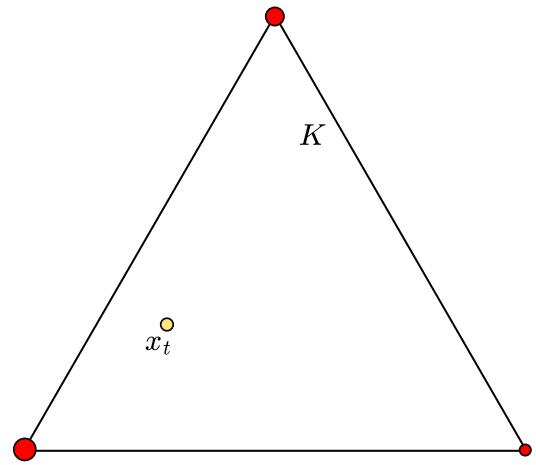
Exploration

- For FTRL-based alg., need for *additional exploration mechanism*

SCRiBLe: x_t given by FTRL step, then sample $a_t \sim \mathcal{U}(\text{poles of } W(x_t))$



FTPL: a_t solves a stochastically perturbed LP



- Uses self-concordant barriers
- Worst-case regret in $O(d^{3/2}\sqrt{n \ln n})$

Question: Can we get optimal regret using self-concordance and no additional exploration mechanism using FTPL ?

SC-FTPL

A distribution $\mathcal{D}$ is a ϑ-**self-concordant perturbation** for K if

$$\nabla \mathcal{R}^*(\theta) = \mathbb{E}_{\xi \sim \mathcal{D}}[\nabla \phi_K(\theta + \xi)], \quad \forall \theta \in \mathbb{R}^d$$

with $\mathcal{R}$ a ϑ-self-conc. barrier and ϕ_K the support function of K.

SC-FTPL: Given a ϑ-self-concordant perturbation $\mathcal{D}$, for all t,

- Sample $\xi_t \sim \mathcal{D}$
- Play $a_t = \arg\min_{a \in K} \langle a, \hat{Y}_{t-1} - \frac{1}{\eta}\xi_t \rangle$
- Estimate $\hat{y}_t = \langle y_t, a_t \rangle Q_t^{-1} a_t$, where $Q_t = \mathbb{E}_{t-1}[a_t a_t^\top]$

SC-FTPL

A distribution $\mathcal{D}$ is a ϑ-**self-concordant perturbation** for K if

$$\nabla \mathcal{R}^*(\theta) = \mathbb{E}_{\xi \sim \mathcal{D}}[\nabla \phi_K(\theta + \xi)], \quad \forall \theta \in \mathbb{R}^d$$

with $\mathcal{R}$ a ϑ-self-conc. barrier and ϕ_K the support function of K.

SC-FTPL: Given a ϑ-self-concordant perturbation $\mathcal{D}$, for all t,

- Sample $\xi_t \sim \mathcal{D}$
- Play $a_t = \arg\min_{a \in K} \langle a, \hat{Y}_{t-1} - \frac{1}{\eta}\xi_t \rangle$
- Estimate $\hat{y}_t = \langle y_t, a_t \rangle Q_t^{-1} a_t$, where $Q_t = \mathbb{E}_{t-1}[a_t a_t^\top]$

Lemma (Th. 5): If $2\eta \|\hat{y}_t\|_t \leq 1$ a.s, then

$$R_n \leq \frac{\vartheta \ln n}{\eta} + \eta \sum_{t=1}^{n} \mathbb{E}\|\hat{y}_t\|_t^2 + 2$$

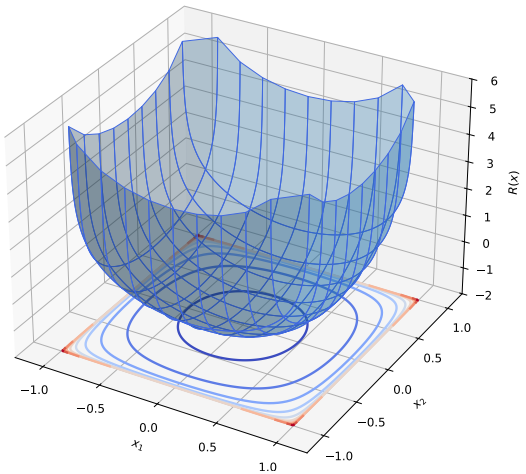
- For SCRiBLe $\mathbb{E}\|\hat{y}_t\|_t^2 \leq d^2$, yields $R_n = O(d\sqrt{\vartheta\, n \ln n})$

Hypercube

Entropic barrier

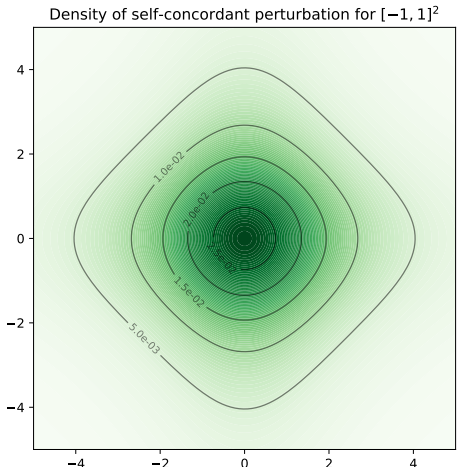
$\mathcal{R}^* : \theta \mapsto \ln \int_K e^{\langle \theta, x \rangle} dx$

d-self-concordant on $[-1, 1]^d$



Replicating d-self-concordant perturbation with density

$f(x) = \prod_{i=1}^{d} \left(\frac{1}{2x_i^2} - \frac{1}{2\sinh(x_i)^2} \right)$

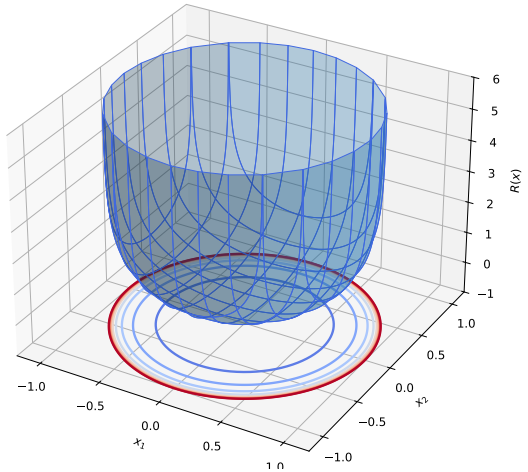


- Bounds on the estimator $\|\hat{y}_t\|_t^2 \leq 3d$ and $\mathbb{E}\|\hat{y}_t\|_t^2 \leq \frac{1}{3}d$
- Improvement over SCRiBLe by a $3d$ factor
- **Optimal regret rate**

$$R_n \leq \frac{2}{\sqrt{3}} d \sqrt{n \ln n} + 2$$

Euclidean Ball

Log-barrier $\mathcal{R} : x \mapsto -\ln(1 - \|x\|^2)$
1-self-concordant on $\mathbb{B}^d$



Replicating 1-self-concordant perturbation given by

$$\xi = \mathbf{N}/ZU$$

with $\mathbf{N} \sim \mathcal{N}(0, I_d)$, $Z \sim \chi_{d+1}$, and $U \sim \mathcal{U}([0,1])$

- Bounds $\|\hat{y}_t\|_t^2 \leq d^2\eta\|\hat{Y}_{t-1}\| + 4d^2$ and $\mathbb{E}\|\hat{y}_t\|_t^2 \leq d^2$
- No improvement over SCRiBLe
- **Suboptimal** **regret rate**

$$R_n \leq 2d\sqrt{n \ln n} + 2 + O\left(\frac{\ln^3 n}{d}\right)$$

(Bubeck et al., '12) achieved $O(\sqrt{dn \ln n})$ regret on the ℓ_2 ball

Open Questions

- Can a FTPL scheme yield $O(\sqrt{dn})$ regret on the ℓ_2 ball ?
- Do self-concordant perturbations exist for every convex body $K \subset \mathbb{R}^d$?

  In particular, can we replicate the entropic barrier (Bubeck and Eldan, '15): $\mathcal{R}^* : \theta \mapsto \ln \int_K e^{\langle \theta, x \rangle} dx$, which is a universal d-self-conc. barrier

Open Questions

- Can a FTPL scheme yield $O(\sqrt{dn})$ regret on the ℓ_2 ball ?
- Do self-concordant perturbations exist for every convex body $K \subset \mathbb{R}^d$?
  In particular, can we replicate the entropic barrier (Bubeck and Eldan, '15): $\mathcal{R}^* : \theta \mapsto \ln \int_K e^{\langle \theta, x \rangle} dx$, which is a universal d-self-conc. barrier

Thank you for your attention

paper & slides