Sampling-Based Regularisation for Adversarial Linear Bandits

Lucas Lévy

in place of Arya Akhavan

Inria Paris, Ecole Normale Supérieure
research done while at the University of Oxford

IMS Annual Meeting, July 6th 2026

Self-Concordant Perturbations for Linear Bandits

Lucas Lévy, Jean-Lou Valeau, Arya Akhavan, Patrick Rebeschini

Proceedings of 39th Conference on Learning Theory (COLT), PMLR 336:4635-4673, 2026.




Outline

Linear Bandits

Instance defined by an horizon n, an action set $K \subset \mathbb{R}^d$ (convex body), and an unknown sequence of loss vectors $(y_t)_{t=1}^n$.

At each time $t \in [n]$, the learner

- Picks an action $a_t \in K$,
- Observe its punctual loss $\langle y_t, a_t \rangle$

Linear Bandits

Instance defined by an horizon n, an action set $K \subset \mathbb{R}^d$ (convex body), and an unknown sequence of loss vectors $(y_t)_{t=1}^n$.

At each time $t \in [n]$, the learner

- Picks an action $a_t \in K$,
- Observe its punctual loss $\langle y_t, a_t \rangle$

We define the *regret*:

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

Linear Bandits

Instance defined by an horizon n, an action set $K \subset \mathbb{R}^d$ (convex body), and an unknown sequence of loss vectors $(y_t)_{t=1}^n$.

At each time $t \in [n]$, the learner

- Picks an action $a_t \in K$,
- Observe its punctual loss $\langle y_t, a_t \rangle$

We define the *regret*:

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

Problem has links with sequential decision-making, reinforcement learning, stochastic optimization (SGD), ...

Follow-the-Leader

Let's first suppose $(y_s)_{s<t}$ are known at time t.

Naive algorithm (FTL): At each time t, choose

$$a_t \in \arg\min_{a\in K} \Big\{ \sum_{s<t} \langle a, y_s \rangle \Big\} = \partial \phi_K(-Y_{t-1})$$

where $\phi_K : \theta \mapsto \max_{x\in K} \langle x, \theta \rangle$ is the support function of K

Follow-the-Leader

Let's first suppose $(y_s)_{s<t}$ are known at time t.

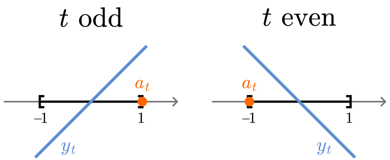
Naive algorithm (FTL): At each time t, choose

$$a_t \in \arg\min_{a \in K} \Big\{ \sum_{s<t} \langle a, y_s \rangle \Big\} = \partial \phi_K(-Y_{t-1})$$

Failure of FTL:

$$K = [-1, 1], \quad y_t = \begin{cases} 1/2 & \text{if } t = 1 \\ -1 & \text{if } t \text{ even} \\ 1 & \text{otherwise} \end{cases}$$

$\sum_{t=1}^{n} \langle y_t, a_t \rangle - \sum_{t=1}^{n} \langle y_t, 0 \rangle \geq n - 1$



FTRL & FTPL

FTRL: Given a strictly convex regularizer $\boxed{\psi} : K \to \mathbb{R}$,

$$a_t = \arg\min_{a \in K} \left\{ \sum_{s<t} \langle a, y_s \rangle + \boxed{\psi(a)} \right\} = \boxed{\nabla \psi^*}(-Y_{t-1})$$

FTPL: Given a probability distribution $\boxed{D}$ on $\mathbb{R}^d$, at each time t, sample $\boxed{\xi_t \sim D}$ and choose

$$a_t \in \arg\min_{a \in K} \left\{ \sum_{s<t} \langle a, y_s \rangle - \langle a, \boxed{\xi_t} \rangle \right\} = \partial \phi_K(\boxed{\xi_t} - Y_{t-1})$$

- For a wide variety of regularizers ψ or perturbations D, it has been shown that regret scales in $\sqrt{n}$

GBPA

FTRL: $a_t = \nabla\psi^*(-Y_{t-1})$

FTPL: $a_t = \nabla\phi_K(\xi_t - Y_{t-1})$ with $\xi_t \sim D$

- Both part of the GBPA framework [Abe+14] where

$$a_t = \nabla\Phi(-Y_{t-1})$$

for some potential function $\Phi : \mathbb{R}^d \to \mathbb{R}$ satisfying $\mathrm{Im}(\nabla\Phi) \subseteq K$

GBPA

FTRL: $a_t = \nabla\psi^*(-Y_{t-1})$

FTPL: $a_t = \nabla\phi_K(\xi_t - Y_{t-1})$ with $\xi_t \sim D$

- Both part of the GBPA framework [Abe+14] where

$$a_t = \nabla\Phi(-Y_{t-1})$$

  for some potential function $\Phi : \mathbb{R}^d \to \mathbb{R}$ satisfying $\mathrm{Im}(\nabla\Phi) \subseteq \mathrm{K}$

- For FTPL,

$$\Phi : \theta \mapsto \mathbb{E}_{\xi\sim D}[\phi_K(\theta + \xi)]$$

  Φ can be seen as a smoothing of ϕ_K

- Some instances of FTRL and FTPL are equivalent, e.g
  $\psi : x \mapsto \sum_i x_i \ln x_i$ and $D = \mathrm{Gumbel}(0,1)^{\otimes d}$ on the simplex

Exploration

- We do not know the past loss vectors $(y_s)_{s<t}$, only $\langle y_s, a_s \rangle$
- Need *exploration (randomization)* to estimate them
- General recipe: FTRL with estimated loss vectors + additional exploration mechanism

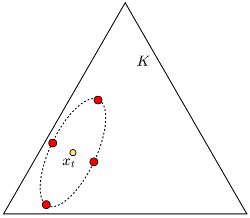
Exploration

- We do not know the past loss vectors $(y_s)_{s<t}$, only $\langle y_s, a_s \rangle$
- Need *exploration (randomization)* to estimate them
- General recipe: FTRL with estimated loss vectors + additional exploration mechanism

SCRiBLe [Abe+08]:

Let $\mathcal{R}$ be a self-concordant barrier on K

- Let $x_t = \nabla\mathcal{R}^*(-\sum_{s<t} \hat{y}_s)$
- Sample $a_t \sim \mathcal{U}\{\text{poles of } W\}$ where W is the ellipsoid defined by $H = \nabla^2\mathcal{R}(x_t)$
- Estimate $\hat{y}_t = d\langle y_t, a_t\rangle H(a_t - x_t)$



Self-concordant barriers

Let $\mathcal{R} : \operatorname{int} K \to \mathbb{R}$ be a ϑ-self-concordant barrier. It satisfies

- **Dikin ellipsoid**: For $x \in \operatorname{int} K$,

$$W(x) = \{y : \|x - y\|_{\nabla^2 \mathcal{R}(x)} < 1\} \subset K$$

- **Local strong convexity**:

$$\|\theta - \theta'\|_{\nabla^2 \mathcal{R}^*(\theta)} \leq \frac{1}{2} \implies B_{\mathcal{R}^*}(\theta', \theta) \leq \|\theta - \theta'\|^2_{\nabla^2 \mathcal{R}^*(\theta)}$$

Note: A function f is μ-strongly convex w.r.t $\|\cdot\|$ iff $B_{f^*}(x, y) \leq \frac{1}{2\mu}\|x - y\|_*^2$

Self-concordant barriers

Let $\mathcal{R} : \operatorname{int} K \to \mathbb{R}$ be a ϑ-self-concordant barrier. It satisfies

- **Dikin ellipsoid**: For $x \in \operatorname{int} K$,

$$W(x) = \{y : \|x - y\|_{\nabla^2 \mathcal{R}(x)} < 1\} \subset K$$

- **Local strong convexity**:

$$\|\theta - \theta'\|_{\nabla^2 \mathcal{R}^*(\theta)} \leq \frac{1}{2} \implies B_{\mathcal{R}^*}(\theta', \theta) \leq \|\theta - \theta'\|^2_{\nabla^2 \mathcal{R}^*(\theta)}$$

- **Controlled divergence** to $+\infty$ on the boundary: for $x, y \in \operatorname{int} K$

$$\mathcal{R}(y) \leq \mathcal{R}(x) + \vartheta \ln\left(\frac{1}{1 - \pi_x(y)}\right)$$

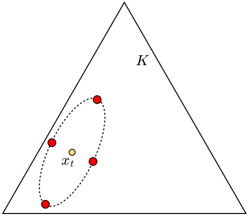
where $\pi_x(y) = \inf\{t > 0 : x + \frac{1}{t}(y - x) \in K\}$. Informally, $\pi_x(y) \to 1$ iff $y \to \partial K$

Existence [LY21]: For any convex body $K \subset \mathbb{R}^d$, there exists a d-self-concordant barrier.

FTPL exploration scheme

SCRiBLe: additional exploration

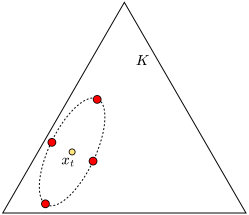
$a_t \sim \mathcal{U}(\text{poles of } W(x_t))$



- Suboptimal exploration mechanism
- Regret for arbitrary convex body in $O(d^{3/2}\sqrt{n \ln n})$

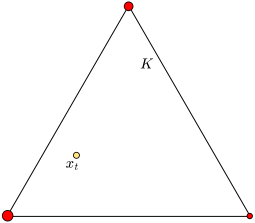
FTPL exploration scheme

SCRiBLe: additional exploration
$a_t \sim \mathcal{U}(\text{poles of } W(x_t))$



- Suboptimal exploration mechanism
- Regret for arbitrary convex body in $O(d^{3/2}\sqrt{n \ln n})$

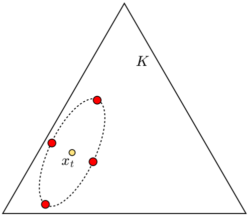
FTPL: a_t solves a stochastically perturbed LP



- action already chosen at random
- $a_t \in \text{extr}\, K$, could lead to better exploration

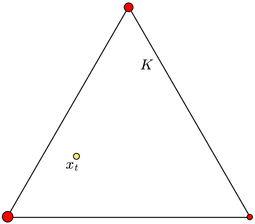
FTPL exploration scheme

SCRiBLe: additional exploration
$a_t \sim \mathcal{U}(\text{poles of } W(x_t))$



- Suboptimal exploration mechanism
- Regret for arbitrary convex body in $O(d^{3/2}\sqrt{n \ln n})$

FTPL: a_t solves a stochastically perturbed LP



- action already chosen at random
- $a_t \in \operatorname{extr} K$, could lead to better exploration

Question: Can we get optimal regret using self-concordance and FTPL natural exploration scheme ?

Self-concordant perturbation

A distribution $\mathcal{D}$ is a ϑ-**self-concordant perturbation** for K if there exists $\mathcal{R}$ a ϑ-self-concordant barrier s.t

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

Self-concordant perturbation

A distribution $\mathcal{D}$ is a ϑ-**self-concordant perturbation** for K if there exists $\mathcal{R}$ a ϑ-self-concordant barrier s.t

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

SC-FTPL: Given a ϑ-self-concordant perturbation $\mathcal{D}$ and a learning rate $\eta > 0$, 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 $Q_t = \mathbb{E}_{t-1}[a_t a_t^\top]$

Regret decomposition

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

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

where $\|\cdot\|_t$ is the Mahalanobis norm w.r.t $\nabla^2 \mathcal{R}(x_t)^{-1}$.

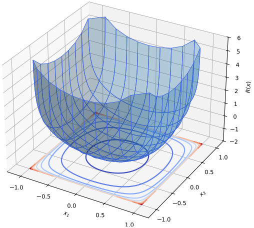
- For SCRiBLe, $\mathbb{E}\|\hat{y}_t\|_t^2 \leq d^2$, yields $R_n = O(d\sqrt{\vartheta\, n \ln n})$
- For arbitrary convex bodies, self-concordant barriers with $\vartheta = d$
- What variance guarantees for SC-FTPL ?

Hypercube

Entropic barrier

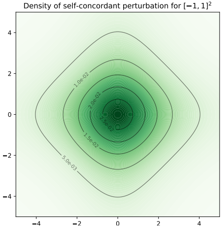
$\mathcal{R}^* : \theta \mapsto \ln \int_K e^{\langle \theta, x \rangle} dx$

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



Replicating d-self-concordant perturbation with density

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

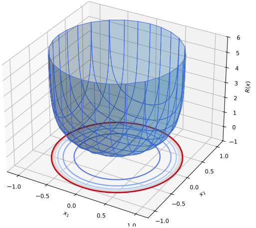


- Bounds on the estimator $\|\hat{y}_t\|_t^2 \leq 3d$ and $\mathbb{E}\|\hat{y}_t\|_t^2 \leq \frac{1}{3}d$
- Improvement over SCRiBLe by a $3d$ factor
- **Optimal regret rate**

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

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**

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

[Bub+12] achieved $O(\sqrt{dn \ln n})$ regret on the ℓ_2 ball

Open Questions

- Do self-concordant perturbations exist for every convex body $K \subset \mathbb{R}^d$? What regret rate would we obtain ?
- In particular, can we replicate the entropic barrier [BE15]: $\mathcal{R}^* : \theta \mapsto \ln \int_K e^{\langle \theta, x \rangle} dx$, which is a universal d-self-conc. barrier

Open Questions

- Do self-concordant perturbations exist for every convex body $K \subset \mathbb{R}^d$? What regret rate would we obtain ?
- In particular, can we replicate the entropic barrier [BE15]: $\mathcal{R}^* : \theta \mapsto \ln \int_K e^{\langle \theta, x \rangle} dx$, which is a universal d-self-conc. barrier

paper & slides

Thank you for your attention

References

[Abe+08] Abernethy, Hazan, and Rakhlin. Competing in the dark: An efficient algorithm for bandit linear optimization. *COLT 2008*

[Abe+14] Abernethy, Lee, Sinha, and Tewari. Online linear optimization via smoothing. *COLT 2014*

[Bub+12] Bubeck, Cesa-Bianchi, and Kakade. Towards minimax policies for online linear optimization with bandit feedback. *COLT 2012*

[BE15] Bubeck and Eldan. The entropic barrier: a simple and optimal universal self-concordant barrier. *COLT 2015*

[LY21] Lee and Yue. Universal barrier is n-self-concordant. *Math. Oper. Res.* 46.3. 2021