Gamification of Pure Exploration for Linear Bandits

ICML 2020


Rémy Degenne


Pierre Ménard


Xuedong Shang


Michal Valko

Inria

DeepMind

Linear Bandits

Finite set $\mathcal{A}$ (size A) of vectors in $\mathbb{R}^d$.

At time t: choose $a_t \in \mathcal{A}$, observe

$$Y_t = \langle \theta, a_t \rangle + \eta_t \quad \text{where } \eta_t \sim \mathcal{N}(0, 1) \; .$$

$\theta \in \mathbb{R}^d$ **unknown parameter.**

Goal: sample arms, then answer a query about θ.

Pure exploration for linear bandits

Question

- Which arm $a \in \mathcal{A}$ has highest mean $\langle \theta, a \rangle$? $\rightarrow$ answer set $\mathcal{A}$
- Is there $a \in \mathcal{A}$ with mean < 0 ? $\rightarrow$ answer set $\{$yes, no$\}$
- In general, finite answer set $\mathcal{I}$

$$\begin{aligned} i^\star : \mathbb{R}^d &\rightarrow \mathcal{I} \\ \theta &\mapsto i^\star(\theta) \end{aligned}$$

Pure exploration

- *sampling rule* $(a_t)_{t \geq 1}$
- *stopping rule* τ_δ, a stopping time for the filtration $(\mathcal{F}_t)_{t \geq 1}$
- *decision rule* $\widehat{i} \in \mathcal{I}$ which is $\mathcal{F}_{\tau_\delta}$-measurable.

Objective **Minimize** $\mathbb{E}_\theta[\tau_\delta]$ under the constraint $\mathbb{P}_\theta\left(\widehat{i} \neq i^\star(\theta)\right) \leq \delta$

Contributions

Insight on complexities used in linear bandits

Saddle-point approach with a convexified point of view for simpler proofs

Two algorithms with

- **asymptotically optimal sample complexity** (as $\delta \to 0$)
- **competitive empirical performance**
- **small computational cost**

Lower Bound

Alternative $\neg i := \{\theta \in \mathbb{R}^d : \; i \neq i^\star(\theta)\}$

Design matrix $V_w := \sum_{a \in \mathcal{A}} w^a a a^\intercal \quad (\|x\|_V := \sqrt{x^\intercal V x})$

Lower Bound

Alternative $\neg i := \{\theta \in \mathbb{R}^d : \; i \neq i^\star(\theta)\}$
Design matrix $V_w := \sum_{a \in \mathcal{A}} w^a a a^\intercal \quad (\|x\|_V := \sqrt{x^\intercal V x})$

Asymptotic lower bound:

$$\liminf_{\delta \to 0} \frac{\mathbb{E}_\theta[\tau_\delta]}{\log(1/\delta)} \geq T^\star(\theta)$$

where the *characteristic time* $T^\star(\theta)$ is defined by

$$T^\star(\theta)^{-1} := \max_{w \in \Sigma_A} \inf_{\lambda \in \neg i^\star(\theta)} \frac{1}{2} \|\theta - \lambda\|_{V_w}^2$$

Lower Bound

Alternative $\neg i := \{\theta \in \mathbb{R}^d : \; i \neq i^\star(\theta)\}$
Design matrix $V_w := \sum_{a \in \mathcal{A}} w^a a a^\mathsf{T} \quad (\|x\|_V := \sqrt{x^\mathsf{T} V x})$

Asymptotic lower bound:

$$\liminf_{\delta \to 0} \frac{\mathbb{E}_\theta[\tau_\delta]}{\log(1/\delta)} \geq T^\star(\theta)$$

where the *characteristic time* $T^\star(\theta)$ is defined by

$$T^\star(\theta)^{-1} := \max_{w \in \Sigma_A} \inf_{\lambda \in \neg i^\star(\theta)} \frac{1}{2} \|\theta - \lambda\|_{V_w}^2$$

Asymptotically optimal algorithm if

$$\limsup_{\delta \to 0} \frac{\mathbb{E}_\theta[\tau_\delta]}{\log(1/\delta)} \leq T^\star(\theta)$$

Pure Exploration as a Game

$T^\star(\theta)^{-1}$ **value of a zero-sum game** between the **agent** playing action $a \sim w$ and the **nature** playing alternative λ

$$T^\star(\theta)^{-1} = \max_{w \in \Sigma_A} \inf_{\lambda \in \neg i^\star(\theta)} \frac{1}{2} \sum_{a \in \mathcal{A}} w^a \|\theta - \lambda\|^2_{aa^\mathsf{T}}$$

Pure Exploration as a Game

$T^\star(\theta)^{-1}$ **value of a zero-sum game** between the **agent** playing action $a \sim w$ and the **nature** playing alternative λ

$$T^\star(\theta)^{-1} = \max_{w \in \Sigma_A} \inf_{\lambda \in \neg i^\star(\theta)} \frac{1}{2} \sum_{a \in \mathcal{A}} w^a \|\theta - \lambda\|^2_{aa^\mathsf{T}}$$

Nature plays in $\neg i^\star(\theta)$ unkown!

Pure Exploration as a Game

$T^\star(\theta)^{-1}$ **value of a zero-sum game** between the **agent** playing action $a \sim w$ and the **nature** playing alternative λ

$$T^\star(\theta)^{-1} = \max_{w \in \Sigma_A} \inf_{\lambda \in \neg i^\star(\theta)} \frac{1}{2} \sum_{a \in \mathcal{A}} w^a \|\theta - \lambda\|^2_{aa^\intercal}$$

Nature plays in $\neg i^\star(\theta)$ unkown!

Convexified Game: the **agent** plays action and answer $(a, i) \sim \widetilde{w}$ and the **nature** plays vector of alternatives $\widetilde{\lambda}$

$$T^\star(\theta)^{-1} = \max_{\widetilde{w} \in \Sigma_{AI}} \inf_{\widetilde{\lambda} \in \prod_i (\neg i)} \frac{1}{2} \sum_{(a,i) \in \mathcal{A} \times \mathcal{I}} \widetilde{w}^{a,i} \|\theta - \widetilde{\lambda}^i\|^2_{aa^\intercal}$$

Example: Best Arm Identification

Question: Which arm $a \in \mathcal{A}$ has highest mean $\langle \theta, a \rangle$?

Example: Best Arm Identification

Question: Which arm $a \in \mathcal{A}$ has highest mean $\langle \theta, a \rangle$?

Estimate uniformly the means of the arms: Optimal Design

$$\mathcal{A}\mathcal{A} = \min_{w \in \Sigma_A} \max_{a \in \mathcal{A}} \|a\|^2_{V_w^{-1}}$$

Example: Best Arm Identification

Question: Which arm $a \in \mathcal{A}$ has highest mean $\langle \theta, a \rangle$?

Estimate uniformly the means of the arms: Optimal Design

$$\mathcal{A}\mathcal{A} = \min_{w \in \Sigma_A} \max_{a \in \mathcal{A}} \|a\|^2_{V_w^{-1}}$$

Estimate uniformly the mean of the directions: Transductive design

$$\mathcal{A}\mathcal{B}_{\texttt{dir}} = \min_{w \in \Sigma_A} \max_{b \in \mathcal{B}_{\texttt{dir}}} \|b\|^2_{V_w^{-1}} \qquad \mathcal{B}_{\texttt{dir}} := \{a - a' : (a, a') \in \mathcal{A} \times \mathcal{A}\}$$

Example: Best Arm Identification

Question: Which arm $a \in \mathcal{A}$ has highest mean $\langle \theta, a \rangle$?

Estimate uniformly the means of the arms: Optimal Design

$$\mathcal{A}\mathcal{A} = \min_{w \in \Sigma_A} \max_{a \in \mathcal{A}} \|a\|^2_{V_w^{-1}}$$

Estimate uniformly the mean of the directions: Transductive design

$$\mathcal{A}\mathcal{B}_{\texttt{dir}} = \min_{w \in \Sigma_A} \max_{b \in \mathcal{B}_{\texttt{dir}}} \|b\|^2_{V_w^{-1}} \qquad \mathcal{B}_{\texttt{dir}} := \{a - a' :\ (a, a') \in \mathcal{A} \times \mathcal{A}\}$$

Estimate the mean of the gap-weighted directions: Best Arm Identification

$$\mathcal{A}\mathcal{B}^{\star}(\theta) := \min_{w \in \Sigma_A} \max_{b \in \mathcal{B}^{\star}(\theta)} \|b\|^2_{V_w^{-1}} \qquad \mathcal{B}^{\star} := \left\{(a^{\star}(\theta) - a)/|\langle \theta, a^{\star}(\theta) - a \rangle| : a \in \mathcal{A}/\{a^{\star}(\theta)\}\right\}$$

Example: Best Arm Identification

Question: Which arm $a \in \mathcal{A}$ has highest mean $\langle \theta, a \rangle$?

Estimate uniformly the means of the arms: Optimal Design

$$\mathcal{AA} = \min_{w \in \Sigma_A} \max_{a \in \mathcal{A}} \|a\|^2_{V_w^{-1}}$$

Estimate uniformly the mean of the directions: Transductive design

$$\mathcal{AB}_{\mathtt{dir}} = \min_{w \in \Sigma_A} \max_{b \in \mathcal{B}_{\mathtt{dir}}} \|b\|^2_{V_w^{-1}} \qquad \mathcal{B}_{\mathtt{dir}} := \{a - a' : (a, a') \in \mathcal{A} \times \mathcal{A}\}$$

Estimate the mean of the gap-weighted directions: Best Arm Identification

$$\mathcal{AB}^\star(\theta) := \min_{w \in \Sigma_A} \max_{b \in \mathcal{B}^\star(\theta)} \|b\|^2_{V_w^{-1}} \qquad \mathcal{B}^\star := \left\{(a^\star(\theta) - a)/|\langle \theta, a^\star(\theta) - a \rangle| : a \in \mathcal{A}/\{a^\star(\theta)\}\right\}$$

Ordering

$$T^\star(\theta) = 2\mathcal{AB}^\star(\theta) \le 2\frac{\mathcal{AB}_{\mathtt{dir}}}{\Delta_{\min}(\theta)^2} \le 8\frac{\mathcal{AA}}{\Delta_{\min}(\theta)^2}$$

Designing algorithms with a game

When do we stop?

At t, each arm a is played N_t^a times. $\hat{\theta}_t$ is our estimate for θ.

Concentration result: with probability $1-\delta$,

$$\log \frac{1}{\delta} > \frac{1}{2} \|\hat{\theta}_t - \theta\|^2_{V_{N_t}}$$

Conclusion: if we have

$$\log \frac{1}{\delta} \leq \inf_{\lambda \in \neg i^*(\hat{\theta}_t)} \frac{1}{2} \|\hat{\theta}_t - \lambda\|^2_{V_{N_t}}$$

then w.p. $1-\delta$, $\theta \notin \neg i^*(\hat{\theta}_t)$, which means $i^*(\theta) = i^*(\hat{\theta}_t)$.

Designing algorithms with a game

What should we pull?

When not stopped:

$$\log\frac{1}{\delta} > \inf_{\lambda\in\neg i^\star(\hat\theta_t)} \frac{1}{2}\|\hat\theta_t - \lambda\|^2_{V_{N_t}}$$
$$= \inf_{\lambda\in\neg i^\star(\hat\theta_t)} \sum_{s=1}^{t} \frac{1}{2}\|\hat\theta_t - \lambda\|^2_{V_{w_s}}$$

Goal:

$$\inf_{\lambda\in\neg i^\star(\hat\theta_t)} \sum_{s=1}^{t} \frac{1}{2}\|\hat\theta_t - \lambda\|^2_{V_{w_s}} \geq t \max_{w\in\Sigma_A} \inf_{\lambda\in\neg i^\star(\theta)} \frac{1}{2}\|\theta - \lambda\|^2_{V_w} - o(t)$$
$$= t T^\star(\theta)^{-1} - o(t)$$

$\rightarrow$ **asymptotic optimality.**

Designing algorithms with a game

Ingredients

- Algorithm (1) playing **arm proportions** w_t.
- Algorithm (2) playing **alternatives** $\lambda_t \in \neg i^*(\hat{\theta}_t)$.
- **Optimism** for added exploration.
- Optional: (1) plays over both w_t and answer $i_t \to$ simpler proof.

(1) and (2) ensure saddle point property:

$$\inf_{\lambda \in \neg i^\star(\hat{\theta}_t)} \sum_{s=1}^{t} \frac{1}{2} \|\hat{\theta}_t - \lambda\|_{V_{w_s}}^2 \approx \sum_{s=1}^{t} \frac{1}{2} \|\hat{\theta}_t - \lambda_s\|_{V_{w_s}}^2 \approx \max_{w \in \Sigma_K} \sum_{s=1}^{t} \frac{1}{2} \|\theta - \lambda_s\|_{V_w}^2$$

$$\geq t \max_{w \in \Sigma_A} \inf_{\lambda \in \neg i^\star(\theta)} \frac{1}{2} \|\theta - \lambda\|_{V_w}^2$$

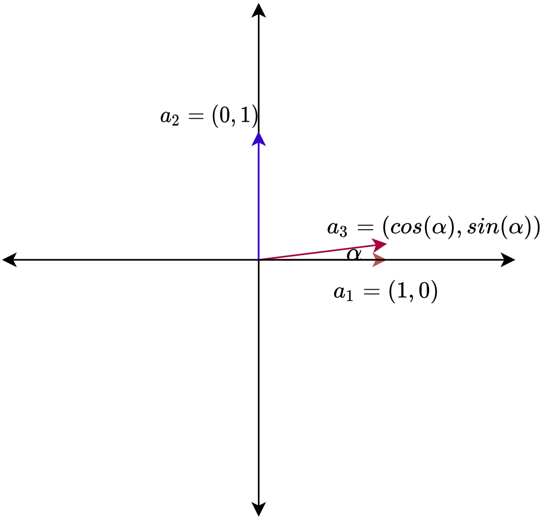
$$= t T^\star(\theta)^{-1}$$

Convexified version: LinGame-C

Algorithm 1 LinGame-C

Input: Agent learner $\mathcal{L}_{\widetilde{w}}$, threshold $\beta(\cdot, \delta)$
for t = 1 ... **do**
  if $\max_{i \in \mathcal{I}} \inf_{\lambda \in \neg i} \frac{1}{2} \|\widehat{\theta}_{t-1} - \lambda\|^2_{V_{N_{t-1}}} \geq \beta(t-1, \delta)$ **then**
    stop and **return** $\widehat{i} = i^\star(\hat{\theta}_{t-1})$
  end if
  Get $\widetilde{w}_t$ **from** $\mathcal{L}_{\widetilde{w}}$ and update $\widetilde{W}_t = \widetilde{W}_{t-1} + \widetilde{w}_t$
  For all $i \in \mathcal{I}$, $\widetilde{\lambda}^i_t \in \arg\min_{\lambda \in \neg i} \|\widehat{\theta}_{t-1} - \lambda\|^2_{V_{\widetilde{w}^i_t}}$
  Feed learner $\mathcal{L}_{\widetilde{w}}$ with $g_t(\widetilde{w}) = \sum_{(a,i) \in \mathcal{A} \times \mathcal{I}} \widetilde{w}^{a,i} U^{a,i}_t / 2$
  Pull a_t such that $(a_t, i_t) \in \arg\min_{(a,i) \in \mathcal{A} \times \mathcal{I}} N^{a,i}_{t-1} - \widetilde{W}^{a,i}_t$
end for

The usual hard instance



The usual hard instance

	LinGame	LinGame-C	DKM
a_1	1912	1959	1943
a_2	5119	4818	4987
a_3	104	77	1775
Total	7135	**6854**	8705

Table: Average number of pulls of each arm.

Comparison with other algorithms: The usual hard instance ($\delta = 0.1, 0.01, 0.0001$)

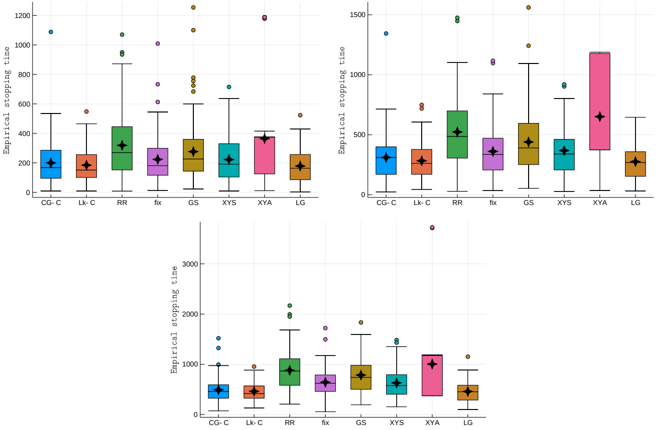
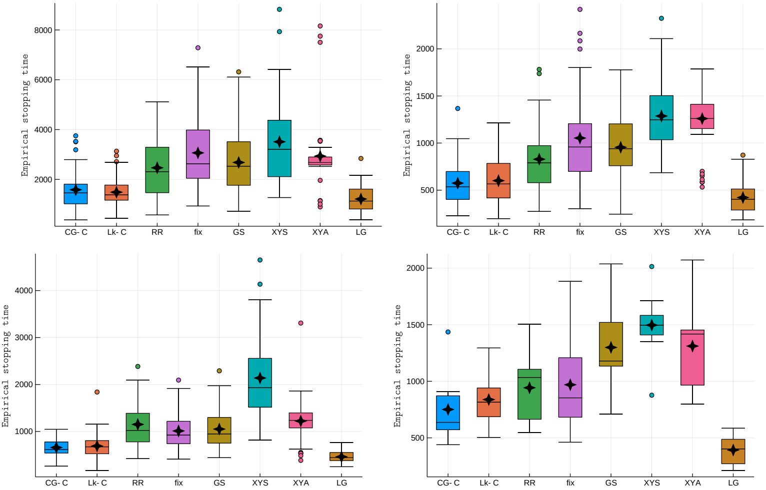
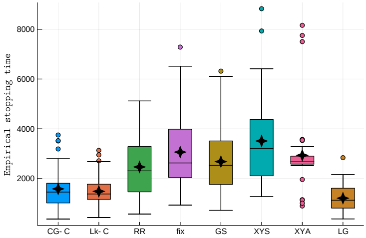


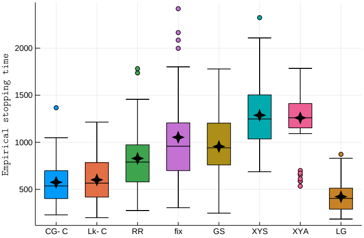
Figure: CG = `LinGame-C`, Lk = `LinGame`, RR = uniform sampling, fix = tracking the fixed weights, GS = $\mathcal{XY}$-`Static` with $\mathcal{AA}$-allocation, XYS = $\mathcal{XY}$-`Static` with $\mathcal{AB}_{\texttt{dir}}$-allocation, LG = `LinGapE`.

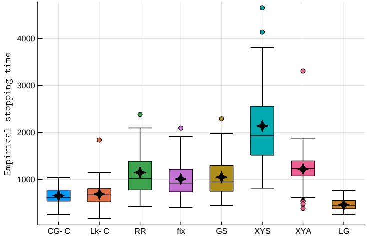
Comparison with other algorithms: Random unit sphere vectors ($d = 6, 8, 10, 12$)



Comparison with other algorithms: Random unit sphere vectors ($d = 6, 8, 10, 12$)







Thank you!