# Parameter-Free Dynamic Regret for Unconstrained Linear Bandits

**Alberto Rumi**[⋆] Andrew Jacobsen[⋆]
Nicolò Cesa-Bianchi Fabio Vitale

AISTATS 2026

Intesa Sanpaolo AI Research · Università degli Studi di Milano

# Unconstrained linear bandits, dynamic comparator

**The game.** For $t = 1, \ldots, T$:

- Learner plays $\boldsymbol{w}_t \in \mathbb{R}^d$ (unconstrained)
- Adversary picks loss vector $\boldsymbol{\ell}_t \in \mathbb{R}^d$
- Learner observes only $\langle \boldsymbol{\ell}_t, \boldsymbol{w}_t \rangle$ (bandit feedback)

**Dynamic regret** against an arbitrary sequence $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_T$:

$$R_T(\boldsymbol{u}_1, \ldots, \boldsymbol{u}_T) \;=\; \sum_{t=1}^{T} \langle \boldsymbol{\ell}_t,\, \boldsymbol{w}_t - \boldsymbol{u}_t \rangle.$$

**Complexity of the comparator** — number of switches:

$$S_T \;=\; \sum_{t=2}^{T} \mathbb{I}\{\boldsymbol{u}_t \neq \boldsymbol{u}_{t-1}\}.$$

# The open problem

Lower bound: $\Omega\left(\sqrt{d(1+S_T)T}\right)$.

Existing bandit algorithms achieve $\mathcal{O}\left(\sqrt{d(1+S_T)T}\right)$ only when $S_T$ is given as input.

**Objective: Parameter-free:** hit the optimal rate without prior knowledge of $S_T$.

- Full-information setting: solved. (Cutkosky '19, Zhang et al. '18, ...)
- Bandit setting: open since Auer et al. 2002.

# Our result

**Theorem (informal)**
For *any* oblivious comparator sequence $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_T$,

$$\mathbb{E}[R_T(\boldsymbol{u}_1, \ldots, \boldsymbol{u}_T)] = \widetilde{\mathcal{O}}\left(M\sqrt{d\,(1+S_T)\,T}\right),$$

where $M = \max_t \|\boldsymbol{u}_t\|$, with no prior knowledge of $S_T$ or $M$.

- First optimal parameter-free dynamic regret for adversarial linear bandits.
- Resolves the long-standing open problem for unconstrained linear bandits.

# Starting point: Cutkosky's iterate-adding trick

**Full-info OCO (Cutkosky, 2019).** Given $N$ comparator-adaptive algos $\mathcal{A}_1, \ldots, \mathcal{A}_N$:

$$\text{play } \boldsymbol{w}_t = \sum_{i=1}^{N} \boldsymbol{w}_t^{(i)} \quad \Longrightarrow \quad R_T(\boldsymbol{u}) \le \min_i R_T^{\mathcal{A}_i}(\boldsymbol{u}) + \mathcal{O}(N).$$

*Iterate-adding* — simple, competitive with the best base algorithm.

**Breaks under bandit feedback.** We only observe the scalar

$$\left\langle \boldsymbol{\ell}_t, \textstyle\sum_i \boldsymbol{w}_t^{(i)} \right\rangle.$$

We don't observe each learner's contribution $\Rightarrow$ no proper feedback to individual $\mathcal{A}_i$'s.

# The fix: uniform sampling

Instead of *summing* iterates, sample one:

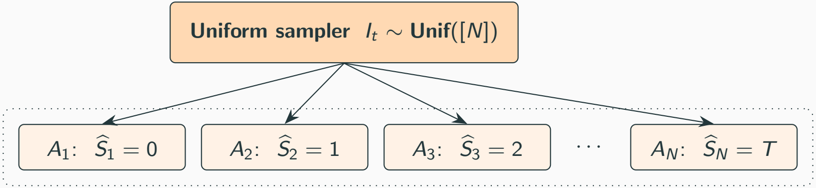
$$I_t \sim \mathrm{Unif}([N]), \qquad \boldsymbol{w}_t \;=\; \boldsymbol{w}_t^{(I_t)}.$$

In expectation this plays the *average*:

$$\mathbb{E}[\boldsymbol{w}_t] \;=\; \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{w}_t^{(i)}.$$

**Cost.** Comparator *rescaled* by $N$.

# Uniform Sampling in practice




$N = \mathcal{O}(\log T)$ *tuned instances*

**Base algo** $\Rightarrow \mathcal{A}_i$ that guarantees $\widetilde{\mathcal{O}}\left(\sqrt{d(1+S_T)T}\right)$ *when given* $S_T$.
(Luo et al. '22 + van der Hoeven et al. '20)

Guess on a geometric grid

$$\widehat{S} \in \{0, 1, 2, 4, 8, \ldots, T\}, \qquad N = \mathcal{O}(\log T) \text{ guesses}$$

Some $\widehat{S}_n$ always satisfies $\widehat{S}_n \le S_T \le 2\widehat{S}_n$.

**Uniform sampling** $\Rightarrow$ inherit the best instance's guarantee, up to log factors.

# Takeaways

- **First** $\widetilde{\mathcal{O}}(\sqrt{d(1+S_T)T})$ **parameter-free dynamic regret** for unconstrained linear bandits.
- **Tool:** uniform sampling — a bandit-friendly replacement for iterate-adding.
- **Next Questions?**
  - Fully adaptive adversary (over comparators too)?
  - Path-length $P_T$ instead of switches $S_T$?
  - Constrained domains?

**Come at poster #174!**