High-Dimensional Sparse Linear Bandits

Botao Hao, Tor Lattimore, Mengdi Wang

Contribution

- Provide a fresh view of sparse linear bandits from a **high-dimensional regime** perspective.
- Derive the first $\Theta(n^{2/3})$ minimax optimal regret bound.
- Provide an example where carefully balancing the trade-off between **information** and **regret** is necessary, in terms of **minimax regret**.

Problem Setting

- **Model:**
  At each round t, the agent chooses an action $A_t \in \mathcal{A} \subseteq \mathbb{R}^d$ (finite, fixed action set) and receives a reward:
  $$Y_t = \langle A_t, \theta^* \rangle + \eta_t,\ t \in [n],$$
  where $\|\theta^*\|_0 = s \ll d$.
  Interested in **high-dimensional regime:** $d > n$.
- **Hardness result:**
  Unfortunately, there exists a $\Omega(\sqrt{dsn})$ **minimax lower bound** in general.

However, minimax bounds Do Not tell the whole story!

- **Why?**
  A crude maximisation over **all environments** hides much of the rich structure of linear bandits with sparsity.
- **What we want to tell:**
  Derive a **sharp** $\Omega(\text{poly}(s)n^{2/3})$ lower bound in high-dimensional regime under the condition that

  "the feature vectors admit a well-conditioned exploration distribution".

A Novel Minimax Lower Bound

Definition. Let $\mathcal{P}(\mathcal{A})$ be the space of probability measures over $\mathcal{A}$. Then we define
$$C_{\min}(\mathcal{A}) = \sup_{\mu \in \mathcal{P}(\mathcal{A})} \sigma_{\min}\Big(\mathbb{E}_{A\sim\mu}[AA^\top]\Big).$$

Remark 0.1. • When $C_{\min}(\mathcal{A})$ is independent of d, n, we say *"feature vectors admit a well-conditioned exploration distribution"*.

- $C_{\min}(\mathcal{A}) > 0$ if and only if $\mathcal{A}$ spans $\mathbb{R}^d$. Two illustrative examples are the hypercube and probability simplex. Sampling uniformly from the corners of each set shows that $C_{\min}(\mathcal{A}) \geq 1$ for the former and $C_{\min}(\mathcal{A}) \geq 1/d$ for the latter.

Theorem (Minimax Lower Bound). For any policy π, there exists s-sparse parameter $\theta \in \mathbb{R}^d$ and an action set $\mathcal{A}$ where $C_{\min}(\mathcal{A})$ is independent of d, n such that
$$R_\theta(n) \gtrsim \min\Big(C_{\min}^{-\frac{1}{3}}(\mathcal{A})s^{\frac{1}{3}}n^{\frac{2}{3}}, \sqrt{dn}\Big),$$
where $\gtrsim$ just hides universal constants.

Remark 0.2. • When $d > n^{1/3}s^{2/3}$ the bound is $\Omega(n^{2/3})$, which is **independent of the dimension**.

- When $d \leq n^{1/3}s^{2/3}$, we recover the standard $\Omega(\sqrt{sdn})$ dimension-dependent lower bound up to a $\sqrt{s}$-factor, even though feature vectors admit a well-conditioned exploration distribution.

Hard Problem Instance

- Construct a **low regret** action set $\mathcal{S}$ (**sparse**) and an **informative** action set $\mathcal{H}$ (**half of the hypercube**) as follows:
  $$\mathcal{S} = \Big\{x \in \mathbb{R}^d \Big| x_j \in \{-1, 0, 1\} \text{ for } j \in [d-1], \|x\|_1 = s-1, x_d = 0\Big\},$$
  $$\mathcal{H} = \Big\{x \in \mathbb{R}^d \Big| x_j \in \{-1, 1\} \text{ for } j \in [d-1], x_d = 1\Big\}.$$
- True parameter θ:
  $$\theta = \big(\underbrace{\varepsilon, \ldots, \varepsilon}_{s-1}, 0, \ldots, 0, -1\big),$$
  for some small $\varepsilon > 0$.
- Pull $\mathcal{H}$: provide more information to infer θ but suffer high regret due to the last coordinate -1.

Matching Upper Bound

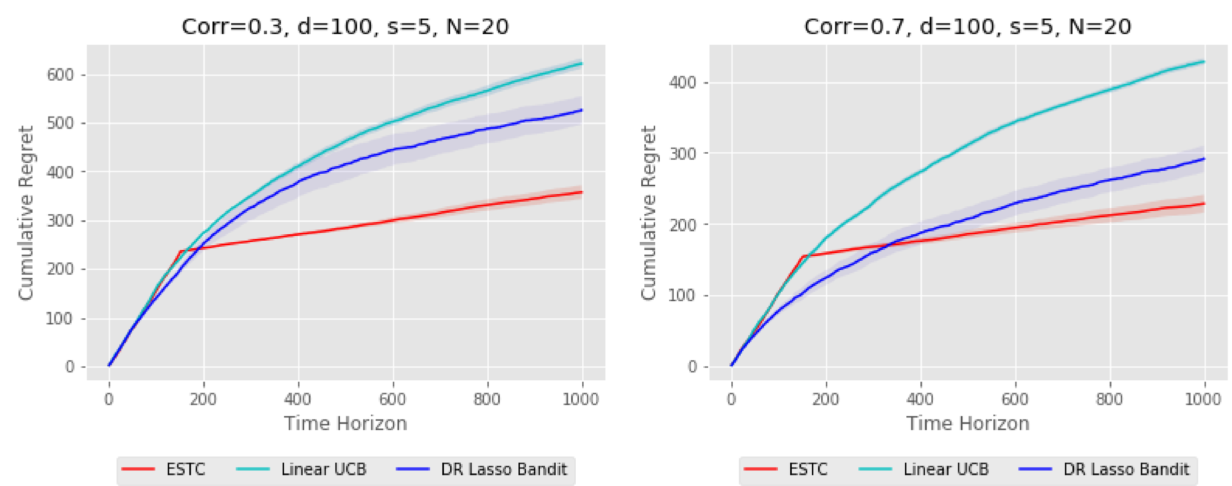
Theorem. Assume the action set $\mathcal{A}$ spans $\mathbb{R}^d$. The regret upper bound of explore-the-sparsity-then-commit (ESTC) algorithm satisfies
$$R_\theta(n) \lesssim C_{\min}^{-\frac{2}{3}}(\mathcal{A})s^{\frac{2}{3}}n^{\frac{2}{3}}.$$

Algorithm.

1. Given an action set $\mathcal{A}$, we first solve the following optimization problem to find the most informative design:
$$\hat{\mu} = \max_{\mu \in \mathcal{P}(\mathcal{A})} \sigma_{\min}\Big(\int_{x\in\mathcal{A}} xx^\top d\mu(x)\Big).$$
2. Independently pulls arms following $\hat{\mu}$ by n_1 rounds and denote the collected samples as $\{(A_1, y_1), \ldots, (A_{n_1}, y_{n_1})\}$. Then we calculate the lasso estimator $\hat{\theta}_{n_1}$.
3. Executes the greedy action $A_t = \text{argmax}_{x\in\mathcal{A}}\langle x, \hat{\theta}_{n_1}\rangle$ for the rest $n - n_1$ rounds.

Experiments

We compare ESTC (our algorithm) with LinUCB [1] and doubly-robust (DR) lasso bandits [2] on a linear contextual bandits: action set from $N(0_N, V)$, where N is the number of arms, $V_{ii} = 1$ and $V_{ik} = \rho^2$ for every $i \neq k$. Larger ρ favorable to DR-lasso.



[1]. Improved algorithms for linear stochastic bandits. [2]. Doubly-robust lasso bandit.