High-Dimensional Sparse Linear Bandits

Botao Hao, Tor Lattimore, Mengdi Wang
Deepmind and Princeton University

Neural Information Processing Systems (2020)

Problem Setting

- 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 $\|\theta^*\|_0 = s \ll d$, η_t is 1-sub-Gaussian noise and $|\mathcal{A}| = K$.

- Interested in **high-dimensional regime:** $d > n$.

Problem Setting

- 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 $\|\theta^*\|_0 = s \ll d$ and η_t is 1-sub-Gaussian noise.

- Unfortunately, there exists a $\Omega(\sqrt{dsn})$[1] **minimax lower bound** in general (no additional assumption on $\mathcal{A}$ and θ^*).
- **High-dimensional regime** ($d > n$) leads to **linear regret!**

[1] Section 24.3 of Bandit Algorithms

Problem Setting

- 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 $\|\theta^*\|_0 = s \ll d$ and η_t is 1-sub-Gaussian noise.

- Unfortunately, there exists a $\Omega(\sqrt{dsn})$[2] **minimax lower bound** in general (no additional assumption on $\mathcal{A}$ and θ^*).
- **High-dimensional regime** ($d > n$) leads to **linear regret!**

But, minimax bounds Do Not tell the whole story!

[2] Section 24.3 of Bandit Algorithms

Why 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.
- **Contribution:** derive a **sharp** $\Omega(\text{poly}(s)n^{2/3})$ lower bound in high-dimensional regime *where the feature vectors admit a well-conditioned exploration distribution*.
- **Implication:** provide an example where carefully balancing the trade-off between **information** and **regret** is necessary, in terms of **worse-case regret**.

A Novel Minimax Lower Bound

Definition. Let $\mathcal{P}(\mathcal{A})$ be the space of probability measures over $\mathcal{A}$. Then we define

$$C_{\min}(\mathcal{A}) = \sup_{\mu \in \mathcal{P}(\mathcal{A})} \sigma_{\min}\Big(\mathbb{E}_{A \sim \mu}\left[AA^\top\right]\Big).$$

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

$$R_\theta(n) \gtrsim \min\left(C_{\min}^{-\frac{1}{3}}(\mathcal{A}) s^{\frac{1}{3}} n^{\frac{2}{3}}, \sqrt{dn}\right),$$

where $\gtrsim$ just hides universal constants.


thank you!