Information Directed Sampling for Sparse Linear Bandits

Botao Hao
Deepmind

Joint works with Tor Lattimore (Deepmind), Wei Deng (Purdue University)

Stochastic Sparse Linear Bandits

- At each round $t \in [n]$, the agent chooses an action $A_t \in \mathcal{A} \subseteq \mathbb{R}^d$ and receives a reward:

$$Y_t = \langle A_t, \theta^* \rangle + \eta_t.$$

where η_t is 1-sub-Gaussian noise. Assume for any $a \in \mathcal{A}$, $\|a\|_\infty \leq 1$ and $|\mathcal{A}| = K$. The notion of sparsity can be defined through the parameter space Θ:

$$\Theta = \left\{ \theta \in \mathbb{R}^d \middle| \sum_{j=1}^{d} \mathbb{1}\{\theta_j \neq 0\} \leq s, \|\theta\|_2 \leq 1 \right\}.$$

Stochastic Sparse Linear Bandits

- At each round $t \in [n]$, the agent chooses an action $A_t \in \mathcal{A} \subseteq \mathbb{R}^d$ and receives a reward:

$$Y_t = \langle A_t, \theta^* \rangle + \eta_t.$$

  where η_t is 1-sub-Gaussian noise. Assume for any $a \in \mathcal{A}$, $\|a\|_\infty \leq 1$ and $|\mathcal{A}| = K$. The notion of sparsity can be defined through the parameter space Θ:

$$\Theta = \left\{ \theta \in \mathbb{R}^d \middle| \sum_{j=1}^{d} \mathbb{1}\{\theta_j \neq 0\} \leq s, \|\theta\|_2 \leq 1 \right\}.$$

- **Data-poor regime:** $d \gtrsim n$; **data-rich regime:** $d \lesssim n$.
- Cumulative regret for bandit θ^*:

$$\mathfrak{R}_{\theta^*}(n; \pi) = \mathbb{E}\left[\sum_{t=1}^{n} \langle x^*, \theta^* \rangle - \sum_{t=1}^{n} Y_t \right],$$

  where x^* is the optimal action.
- Worse-case regret: $\sup_{\theta^*} \mathfrak{R}_{\theta^*}(n; \pi)$; Bayesian regret: $\mathbb{E}_{\theta^*}[\mathfrak{R}_{\theta^*}(n; \pi)]$.

Does Sparsity Help?

- If the action set is **arbitrary**, there exists a $\Omega(\sqrt{dsn})$ minimax lower bound.
- If the action set is **exploratory**, there exists a $\Omega(\min(s^{1/3}n^{2/3}, \sqrt{dn}))$ minimax lower bound[1].
- Carefully balancing the trade-off between **information and regret** is necessary in sparse linear bandits.

[1] High-Dimensional Sparse Linear Bandits. NeurIPS 2020.

Does Sparsity Help?

- Those lower bounds are (nearly) sharp:
  - $\widetilde{O}(s^{2/3}n^{2/3})$ achieved by explore-then-commit[2].
    Optimal in **data-poor** regime but sub-optimal in **data-rich** regime.
  - $\widetilde{O}(\sqrt{dsn})$ achieved by optimism-based algorithm[3].
    Optimal in **data-rich** regime but sub-optimal in **data-poor** regime.

Q: Can we have an algorithm that is optimal in both regimes?

---

[2] High-Dimensional Sparse Linear Bandits. NeurIPS 2020.

[3] Online-to-Confidence-Set Conversions and Application to Sparse Stochastic Bandits. AISTATS 2012.

Our Contribution

- We prove that optimism-based algorithms fail to optimally address the information-regret trade-off in sparse linear bandits, which results in a sub-optimal regret bound.
- We provide the first analysis using information theory for sparse linear bandits and derive a class of nearly optimal Bayesian regret bounds for IDS that can adapt to information-regret structures.
- To approximate the information ratio, we develop an empirical Bayesian approach for sparse posterior sampling using spike-and-slab Gaussian-Laplace prior.

Optimism-Based Algorithms

Q: Does the optimism optimally balance information and regret?

In general, optimism-based algorithms π^{opt} choose

$$A_t = \operatorname*{argmax}_{a \in \mathcal{A}} \max_{\widetilde{\theta} \in \mathcal{C}_t} \langle a, \widetilde{\theta} \rangle,$$

where $\mathcal{C}_t$ is some sparsity-aware confidence set that can be constructed through online-to-confidence-set conversions.

Claim. Let π^{opt} be such an optimism-based algorithm. There exists a sparse linear bandit instance characterized by θ such that for the **data-poor** regime, we have

$$\mathfrak{R}_\theta(n; \pi^{\mathsf{opt}}) \gtrsim n / (\log(n) s \log(ed/s)) \,.$$

Exploratory Policy

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

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

Remarks.

- When $C_{\min}(\mathcal{A})$ is a constant, we say

  "action set $\mathcal{A}$ admits a well-conditioned exploratory policy".

- What is **information**? Pulling arms according to this exploratory policy, we collect information (well-conditioned data).

Information Directed Sampling

IDS (Russo and Van Roy (2014)) balances the information gain about the optimal action and single-round regret:

- Assume θ^* is from some sparse prior distribution.
- $\mathbb{P}_t(\cdot) = \mathbb{P}(\cdot|\mathcal{F}_t)$ as the posterior measure.
- Information gain $I_t(x^*; Y_{t,a})$: the mutual information between the optimal action and the reward the agent receives for taking action a.
- Expected single-round regret $\Delta_t(a) := \mathbb{E}_t[\langle x^*, \theta^* \rangle - \langle a, \theta^* \rangle]$.
- IDS takes the action according to

$$\pi_t = \operatorname*{argmin}_{\pi} \Psi_t(\pi) = \frac{(\Delta_t^\top \pi)^2}{I_t^\top \pi}.$$

Bayesian Regret Bound

Theorem. For an arbitrary action set, the following regret bound holds

$$\mathfrak{BR}(n;\pi^{\text{IDS}}) \lesssim \sqrt{nds}\,.$$

When $\mathcal{A}$ is exploratory and has sparse optimal actions, the following regret bound holds

$$\mathfrak{BR}(n;\pi^{\text{IDS}}) \lesssim \min\left\{\sqrt{nds}, \frac{sn^{2/3}}{(2C_{\min}(\mathcal{A}))^{1/3}}\right\}.$$

Great adaptivity of IDS for sparse linear bandits in the sense that a single policy adapts to different information-regret structures.

Table 1: Regret bounds of IDS for different regimes.

	Arbitrary action set	Exploratory (data-rich)	Exploratory (data-poor)
Large K	$O(\sqrt{nds})$	$O(\sqrt{nds})$	$O(sn^{2/3})$
Small K	$O(\sqrt{nd\log(K)})$	$O(\sqrt{nd\log(K)})$	$O(s^{2/3}n^{2/3}\log^{1/3}(K))$

Bonus: efficient implementation is available through an empirical Bayesian approach for sparse posterior sampling.

Bayesian Regret Bound for Sparse TS

Corollary. For an arbitrary action set, the following regret bound holds for some absolute constant $C > 0$

$$\mathfrak{BR}(n;\pi^{\mathsf{TS}}) \le \sqrt{\frac{1}{2} nd \min(\log(K), 2s\log(Cdn^{1/2}/s))}\,.$$


thank you!