High-Dimensional Sparse Linear Bandits

Lizhang Miao

"Hao B, Lattimore T, Wang M. High-Dimensional Sparse Linear Bandits."

March 5, 2021

Problem Setting

- Compact action set $\mathcal{A} \in \mathbb{R}^d$
- Choose an action $A_t \in \mathcal{A}$, recieve a reward $Y_t = \langle A_t, \theta \rangle + \eta_t$. For all $x \in \mathcal{A}$, $\|x\|_\infty \leq 1$
- The parameter θ is assumed to be sparse, $\|\theta\|_0 \leq s$
- Regret $R_\theta(n) = \mathbb{E}\left[\sum_{t=1}^{n} \langle x^*, \theta \rangle - \sum_{t=1}^{n} Y_t\right]$
- Let $\mathcal{P}(\mathcal{A})$ be the space of probability measures over $\mathcal{A}$ with the Borel σ-algebra and define

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

Results

Upper Bound	Regret	Assumptions	Regime
Abbasi-Yadkori et al. [2012]	$\mathcal{O}(\sqrt{sdn})$	none	rich
Sivakumar et al. [2020]	$\mathcal{O}(\sqrt{sdn})$	adver. + Gaussian noise	rich
Bastani and Bayati [2020]	$\mathcal{O}(\tau K s^2 (\log(n))^2)$	compatibility condition	rich
Wang et al. [2018]	$\mathcal{O}(\tau K s^3 \log(n))$	compatibility condition	rich
Kim and Paik [2019]	$\mathcal{O}(\tau s \sqrt{n})$	compatibility condition	rich
Lattimore et al. [2015]	$\mathcal{O}(s \sqrt{n})$	action set is hypercube	rich
This paper (Thm. 4.2)	$\mathcal{O}(C_{\min}^{-2/3} s^{2/3} n^{2/3})$	action set spans $\mathbb{R}^d$	poor
This paper (Thm. 5.2)	$\mathcal{O}(C_{\min}^{-1/2} \sqrt{sn})$	action set spans $\mathbb{R}^d$ + mini. signal	rich
Lower Bound			
Multi-task bandits[1]	$\Omega(\sqrt{sdn})$	N.A.	rich
This paper (Thm. 3.3)	$\Omega(C_{\min}^{-1/3} s^{1/3} n^{2/3})$	N.A.	poor

Lower Bound

Theorem 3.3. Consider the sparse linear bandits described in Eq. (2.1). Then for any policy π there exists an action set $\mathcal{A}$ with $C_{\min}(\mathcal{A}) > 0$ and s-sparse parameter $\theta \in \mathbb{R}^d$ such that

$$R_\theta(n) \geq \frac{\exp(-4)}{4} \min\left(C_{\min}^{-\frac{1}{3}}(\mathcal{A}) s^{\frac{1}{3}} n^{\frac{2}{3}}, \sqrt{dn}\right). \tag{3.1}$$

Theorem 3.3 holds for any data regime and suggests an intriguing transition between $n^{2/3}$ and $n^{1/2}$ regret, depending on the relation between the horizon and the dimension. When $d > n^{1/3}s^{2/3}$ the bound is $\Omega(n^{2/3})$, which is independent of the dimension. On the other hand, 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. In Section 4, we prove that the $\Omega(n^{2/3})$ minimax lower bound is tight by presenting a nearly matching upper bound in the data-poor regime.

Lower Bound

Step 1: construct a hard instance. We first construct a low regret action set $\mathcal{S}$ and an informative action set $\mathcal{H}$ as follows:

$$\begin{aligned} \mathcal{S} &= \left\{ x \in \mathbb{R}^d \middle| x_j \in \{-1, 0, 1\} \text{ for } j \in [d-1], \|x\|_1 = s-1, x_d = 0 \right\}, \\ \mathcal{H} &= \left\{ x \in \mathbb{R}^d \middle| x_j \in \{-\kappa, \kappa\} \text{ for } j \in [d-1], x_d = 1 \right\}, \end{aligned} \tag{3.2}$$

where $0 < \kappa \leq 1$ is a constant. The action set is the union $\mathcal{A} = \mathcal{S} \cup \mathcal{H}$ and let

$$\theta = \big(\underbrace{\varepsilon, \ldots, \varepsilon}_{s-1}, 0, \ldots, 0, -1\big),$$

where $\varepsilon > 0$ is a small constant to be tuned later. Because $\theta_d = -1$, actions in $\mathcal{H}$ are associated with a large regret. On the other hand, actions in $\mathcal{H}$ are also highly informative, which hints towards an interesting tradeoff between regret and information. Note that $\mathcal{H}$ is nearly the whole binary hypercube, while actions in $\mathcal{S}$ are $(s-1)$-sparse. The optimal action is in the action set $\mathcal{A}$:

$$x^* = \operatorname*{argmax}_{x \in \mathcal{A}} \langle x, \theta \rangle = \big(\underbrace{1, \cdots, 1}_{s-1}, 0, \ldots, 0\big) \in \mathcal{A}. \tag{3.3}$$

Lower Bound

Step 2: construct an alternative bandit

$$\mathcal{S}' = \Big\{x \in \mathbb{R}^d \Big| x_j \in \{-1, 0, 1\} \text{ for } j \in \{s, s+1, \dots, d-1\}, \\ x_j = 0 \text{ for } j = \{1, \dots, s-1, d\}, \|x\|_1 = s-1\Big\}.$$

Then, we denote $\tilde{x} = \arg\min_{x \in \mathcal{S}'} \mathbb{E}_\theta[\sum_{t=1}^n \langle A_t, x \rangle]$. Construct the alternative bandit as

$$\tilde{\theta} = \theta + 2\epsilon\tilde{x}.$$

Define an event

$$\mathcal{D} = \left\{ \sum_{t=1}^n \mathbb{1}(A_t \in \mathcal{S}) \sum_{j=1}^{s-1} A_{tj} \leq \frac{n(s-1)}{2} \right\}.$$

Lower Bound

Claim 3.5. Regret lower bounds with respect to event $\mathcal{D}$:

$$R_\theta(n) \geq \frac{n(s-1)\varepsilon}{2}\mathbb{P}_\theta(\mathcal{D}) \qquad \text{and} \qquad R_{\widetilde{\theta}}(n) \geq \frac{n(s-1)\varepsilon}{2}\mathbb{P}_{\widetilde{\theta}}(\mathcal{D}^c)\,.$$

By the Bretagnolle–Huber inequality (Lemma C.1 in the appendix),

$$R_\theta(n) + R_{\widetilde{\theta}}(n) \geq \frac{n(s-1)\varepsilon}{2}\Big(\mathbb{P}_\theta(\mathcal{D}) + \mathbb{P}_{\widetilde{\theta}}(\mathcal{D}^c)\Big) \geq \frac{n(s-1)\varepsilon}{4}\exp\Big(-\mathrm{KL}\big(\mathbb{P}_\theta, \mathbb{P}_{\widetilde{\theta}}\big)\Big)\,,$$

where $\mathrm{KL}(\mathbb{P}_\theta, \mathbb{P}_{\widetilde{\theta}})$ is the KL divergence between probability measures $\mathbb{P}_\theta$ and $\mathbb{P}_{\widetilde{\theta}}$.

Lower Bound

$$R_\theta(n) = \mathbb{E}_\theta\left[\sum_{t=1}^{n}\langle x^*, \theta\rangle\right] - \mathbb{E}_\theta\left[\sum_{t=1}^{n}\langle A_t, \theta\rangle\right]$$
$$= \mathbb{E}_\theta\left[n(s-1)\varepsilon - \sum_{t=1}^{n}\mathbb{1}(A_t \in \mathcal{H})\langle A_t, \theta\rangle - \sum_{t=1}^{n}\mathbb{1}(A_t \in \mathcal{S})\langle A_t, \theta\rangle\right],$$

$$\sum_{t=1}^{n}\mathbb{1}(A_t \in \mathcal{H})\langle A_t, \theta\rangle \leq T_n(\mathcal{H})(\kappa(s-1)\varepsilon - 1) \leq 0, \tag{B.1}$$

where $T_n(\mathcal{H}) = \sum_{t=1}^{n}\mathbb{1}(A_t \in \mathcal{H})$. Since $\langle A_t, \theta\rangle = \sum_{j=1}^{s} A_{tj}\varepsilon$ for $A_t \in \mathcal{S}$, then it holds that

$$R_\theta(n) \geq \mathbb{E}_\theta\left[n(s-1)\varepsilon - \sum_{t=1}^{n}\mathbb{1}(A_t \in \mathcal{S})\sum_{j=1}^{s-1}A_{tj}\varepsilon\right]$$
$$\geq \mathbb{E}_\theta\left[\left(n(s-1)\varepsilon - \sum_{t=1}^{n}\mathbb{1}(A_t \in \mathcal{S})\sum_{j=1}^{s-1}A_{tj}\varepsilon\right)\mathbb{1}(\mathcal{D})\right] \tag{B.2}$$
$$\geq \left(n(s-1)\varepsilon - \frac{n(s-1)\varepsilon}{2}\right)\mathbb{P}_\theta(\mathcal{D})$$
$$= \frac{n(s-1)\varepsilon}{2}\mathbb{P}_\theta(\mathcal{D}).$$

Lower Bound

Claim 3.6. Define $T_n(\mathcal{H}) = \sum_{t=1}^{n} \mathbb{1}(A_t \in \mathcal{H})$. The KL divergence between $\mathbb{P}_\theta$ and $\mathbb{P}_{\tilde{\theta}}$ is upper bounded by

$$\mathrm{KL}\left(\mathbb{P}_\theta, \mathbb{P}_{\tilde{\theta}}\right) \leq 2\varepsilon^2 \left(\frac{n(s-1)^2}{d} + \kappa^2 (s-1)\mathbb{E}_\theta[T_n(\mathcal{H})] \right) . \tag{3.7}$$

Then

$T_n(\mathcal{H}) < 1/(\kappa^2(s-1)\varepsilon^2)$, it is easy to see

$$R_\theta(n) + R_{\tilde{\theta}}(n) \geq \frac{n(s-1)\varepsilon}{4} \exp\left(-\frac{2n\varepsilon^2(s-1)^2}{d} \right) \exp(-2) . \tag{3.8}$$

On the other hand, when $T_n(\mathcal{H}) > 1/(\kappa^2\varepsilon^2(s-1))$, we have

$$R_\theta(n) \geq \mathbb{E}_\theta[T_n(\mathcal{H})] \min_{x \in \mathcal{H}} \Delta_x \geq \frac{1}{\kappa^2\varepsilon^2(s-1)} + \frac{1-\kappa}{\kappa^2\varepsilon} , \tag{3.9}$$

Lower Bound

Step 4: conclusion. Combining the above two cases together, we have

$$R_\theta(n) + R_{\widetilde{\theta}}(n) \geq \min\left(\left(\frac{ns\varepsilon}{4}\right)\exp\left(-\frac{2\varepsilon^2 s^2 n}{d}\right)\exp(-2), \frac{1}{\kappa^2\varepsilon^2 s} + \frac{1-\kappa}{\kappa^2\varepsilon}\right), \tag{3.10}$$

where we replaced $s-1$ by s in the final result for notational simplicity. Consider a sampling distribution μ that uniformly samples actions from $\mathcal{H}$. A simple calculation shows that $C_{\min}(\mathcal{A}) \geq C_{\min}(\mathcal{H}) \geq \kappa^2 > 0$. This is due to

$$\sigma_{\min}\left(\sum_{x\in\mathcal{H}} \mu(x)xx^\top\right) = \sigma_{\min}\left(\mathbb{E}_{X\sim\mu}[XX^\top]\right) = \kappa^2,$$

where each coordinate of the random vector $X \in \mathbb{R}^d$ is sampled independently uniformly from $\{-1, 1\}$. In the data poor regime when $d \geq n^{1/3}s^{2/3}$, we choose $\varepsilon = \kappa^{-2/3}s^{-2/3}n^{-1/3}$ such that

$$\begin{aligned}\max(R_\theta(n), R_{\widetilde{\theta}}(n)) &\geq R_\theta(n) + R_{\widetilde{\theta}}(n) \\ &\geq \frac{\exp(-4)}{4}\kappa^{-\frac{2}{3}}s^{\frac{1}{3}}n^{\frac{2}{3}} \geq \frac{\exp(-4)}{4}C_{\min}^{-\frac{1}{3}}(\mathcal{A})s^{\frac{1}{3}}n^{\frac{2}{3}}.\end{aligned}$$

Finally, in the data rich regime when $d < n^{1/3}s^{2/3}$ we choose $\varepsilon = \sqrt{d/(ns^2)}$ such that the exponential term is a constant, and then

$$\max(R_\theta(n), R_{\widetilde{\theta}}(n)) \geq R_\theta(n) + R_{\widetilde{\theta}}(n) \geq \frac{\exp(-4)}{4}\sqrt{dn}.$$

□

Upper Bound

Algorithm 1 Explore the sparsity then commit (ESTC)

Input: time horizon n, action set $\mathcal{A}$, exploration length n_1, regularization parameter λ_1;
Solve the optimization problem in Eq. (4.1) and denote the solution as $\widehat{\mu}$.
for $t = 1, \cdots, n_1$ **do**
    Independently pull arm A_t according to $\widehat{\mu}$ and receive a reward: $Y_t = \langle A_t, \theta \rangle + \eta_t$.
end for
Calculate the Lasso estimator [Tibshirani, 1996]:

$$\widehat{\theta}_{n_1} = \operatorname*{argmin}_{\theta \in \mathbb{R}^d} \Big(\frac{1}{n_1} \sum_{t=1}^{n_1} \big(Y_t - \langle A_t, \theta \rangle \big)^2 + \lambda_1 \|\theta\|_1 \Big). \tag{4.2}$$

for $t = n_1 + 1$ to n **do**
    Take greedy actions $A_t = \operatorname{argmin}_{x \in \mathcal{A}} \langle \widehat{\theta}_{n_1}, x \rangle$.
end for

Upper Bound

Theorem 4.2. Consider the sparse linear bandits described in Eq. (2.1) and assume the action set $\mathcal{A}$ spans $\mathbb{R}^d$. Suppose $R_{\max}$ is an upper bound of maximum expected reward such that $\max_{x\in\mathcal{A}}\langle x, \theta\rangle \leq R_{\max}$. In Algorithm 1, we choose

$$n_1 = n^{2/3}(s^2\log(2d))^{1/3}R_{\max}^{-2/3}(2/C_{\min}^2(\mathcal{A}))^{1/3}, \tag{4.3}$$

and $\lambda_1 = 4\sqrt{\log(d)/n_1}$. Then the following regret upper bound holds,

$$R_\theta(n) \leq (2\log(2d)R_{\max})^{\frac{1}{3}}C_{\min}^{-\frac{2}{3}}(\mathcal{A})s^{\frac{2}{3}}n^{\frac{2}{3}} + 3nR_{\max}\exp(-c_1n_1). \tag{4.4}$$

Upper Bound

Step 1: regret decomposition. Suppose $R_{\max}$ is an upper bound of maximum expected reward such that $\max_{x \in \mathcal{A}} \langle x, \theta \rangle \le R_{\max}$. We decompose the regret of ESTC as follows:

$$\begin{aligned} R_\theta(n) &= \mathbb{E}_\theta\Big[\sum_{t=1}^{n} \langle \theta, x^* - A_t \rangle\Big] \\ &= \mathbb{E}_\theta\Big[\sum_{t=1}^{n_1} \langle \theta, x^* - A_t \rangle + \sum_{t=n_1+1}^{n} \langle \theta, x^* - A_t \rangle\Big] \\ &\le \mathbb{E}_\theta\Big[2n_1 R_{\max} + \sum_{t=n_1+1}^{n} \langle \theta - \widehat{\theta}_{n_1}, x^* - A_t \rangle + \sum_{t=n_1+1}^{n} \langle \widehat{\theta}_{n_1}, x^* - A_t \rangle\Big]. \end{aligned}$$

$$\begin{aligned} R_\theta(n) &\le \mathbb{E}_\theta\Big[2n_1 R_{\max} + \sum_{t=n_1+1}^{n} \langle \theta - \widehat{\theta}_{n_1}, x^* - A_t \rangle\Big] \\ &\le \mathbb{E}_\theta\Big[2n_1 R_{\max} + \sum_{t=n_1+1}^{n} \big\|\theta - \widehat{\theta}_{n_1}\big\|_1 \big\|x^* - A_t\big\|_\infty\Big]. \end{aligned}$$

Upper Bound

Step 2: sparse learning

$$\left\|\widehat{\theta}_{n_1} - \theta^*\right\|_1 \leq \frac{2}{C_{\min}} \sqrt{\frac{2s^2(\log(2d) + \log(n_1))}{n_1}}.$$

with probability at least $1 - \exp(-n_1)$.

Upper Bound

Step 3: optimize the length of exploration. Define an event $\mathcal{E}$ as follows:

$$\mathcal{E} = \left\{\phi(\widehat{\Sigma}, s, 3) \geq \frac{C_{\min}^{1/2}}{2}, \left\|\widehat{\theta}_{n_1} - \theta^*\right\|_1 \leq \frac{2}{C_{\min}}\sqrt{\frac{2s^2(\log(2d) + \log(n_1))}{n_1}}\right\}.$$

We know that $\mathbb{P}(\mathcal{E}) \geq 1 - 3\exp(-c_1 n_1)$. Note that $\|x^* - A_t\|_\infty \leq 2$. According to Eq. (B.15), we have

$$R_\theta(n) \leq \mathbb{E}_\theta\Big[\Big(2n_1 R_{\max} + \sum_{t=n_1+1}^{n} \left\|\theta - \widehat{\theta}_{n_1}\right\|_1 \left\|x^* - A_t\right\|_\infty\Big) \mathbb{1}(\mathcal{E})\Big] + nR_{\max}\mathbb{P}(\mathcal{E}^c)$$

$$\leq n_1 R_{\max} + (n - n_1)\frac{4}{C_{\min}}\sqrt{\frac{2s^2(\log(2d) + \log(n_1))}{n_1}}2 + 3nR_{\max}\exp(-c_1 n_1)$$

with probability at least $1 - \delta$. By choosing $n_1 = n^{2/3}(s^2\log(2d))^{1/3}R_{\max}^{-2/3}(2/C_{\min}^2)^{1/3}$, we have

$$R_n \leq (sn)^{2/3}(\log(2d))^{1/3}R_{\max}^{1/3}(\frac{2}{C_{\min}^2})^{1/3} + 3nR_{\max}\exp(-c_1 n_1).$$

We end the proof.

Improved Algorithm

Algorithm 2 Restricted phase elimination

Input: time horizon n, action set $\mathcal{A}$, exploration length n_2, regularization parameter λ_2;
Solve the optimization problem Eq. (4.1) and denote the solution as $\widehat{\mu}$.
for $t = 1, \cdots, n_2$ **do**
    Independently pull arm A_t according to $\widehat{\mu}$ and receive a reward: $Y_t = \langle A_t, \theta \rangle + \eta_t$.
end for
Calculate the Lasso estimator $\widehat{\theta}_{n_2}$ as in Eq. (4.2) with λ_2.
Identify the support: $\widehat{S} = \mathrm{supp}(\widehat{\theta}_{n_2})$.
for $t = n_2 + 1$ to n **do**
    Invoke phased elimination algorithm for linear bandits on $\widehat{S}$.
end for

experiments