We thank the reviewers for their valuable comments! We have added comprehensive empirical studies and hope you are satisfied with our point-by-point responses and increase your scores!

(Experiments): We agree that adding experiments is a good idea and have completed an extensive empirical evaluation. Given the space limitation in the response, only a subset is included below. All the experiments will be added in the revised version. We compare ESTC (our algorithm) with LinUCB [2] and doubly-robust (DR) lasso bandits [1]. For ESTC, we use the theoretically suggested length of exploration stage. For LinUCB, we use the theoretically suggested confidence interval. For DR-lasso, we use the code made available by the authors on-line.

Case 1: linear contextual bandits. We use the setting in Section 5 of [1] with $N = 20$ arms, dimension $d = 100$, sparsity $s = 5$. At round t, we generate the action set from $N(0_N, V)$, where $V_{ii} = 1$ and $V_{ik} = \rho^2$ for every $i \neq k$. Larger ρ corresponds to high correlation setting that is more favorable to DR-lasso. The noise is from $N(0, 1)$ and $\|\theta\|_0 = s$. **Case 2: hard problem instance.** Consider the hard problem instance in the proof of minimax lower bound (Thm 3.3), including an informative action set and an uninformative action set.

Conclusion: The experiments confirm our theoretical findings. Although our theory focuses on the fixed action set setting, ESTC works well in the contextual setting. DR-lasso bandits heavily rely on context distribution assumption and almost fail for the hard instance. LinUCB suffers in the data-poor regime since it ignores the sparsity information. We do not evaluate [3] since it is not a polynomial-time algorithm.

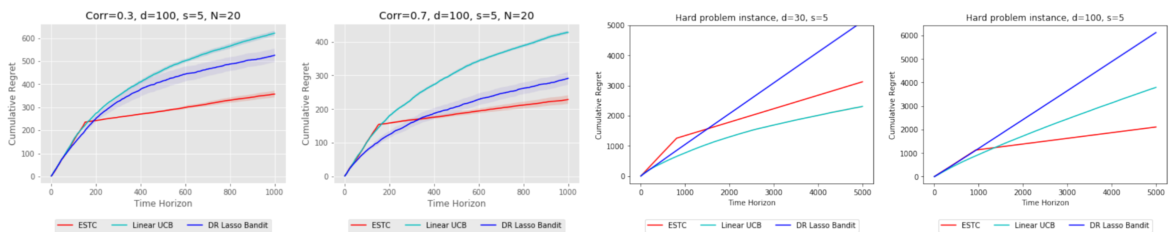

Figure 1: The left two figures are for Case 1 and the right two figures are for Case 2.

Reviewer #1. (Compare with [4]): Thanks the reference, which will be included in a revised version. The algorithm in [4] and ESTC share the explore-then-commit template but both the exploration and exploitation stages are very different. [4] considers simple regret minimization while we focus on cumulative regret minimization. **(Dependence on $C_{\min}$):** Surprisingly, even in the classical statistical settings there are still gaps between upper and lower bounds. We speculate that the upper bound may be improvable, though at present we do not know how to do it. A discussion will be included in the revised version.

Reviewer #2. (Interpreting of claims): We agree with this comment and will make this clear up-front. **(Relation between eigenvalue and sparsity):** The optimization problem in Eq. (4.1) only depends on the action set and not the sparsity.

(Weeding out actions): This is an interesting question. As the second part of your question hints, things are already delicate when the actions do not span $\mathbb{R}^d$. One cannot expect the Lasso estimate to be close to the true parameter because there is no information in some directions. We are not aware of direction-dependent confidence bounds for Lasso that are suitable in this case. A standard idea in non-sparse settings is to change the coordinates to the low-dimensional subspace, but rotations do not preserve sparsity, so this does not work here. For the first part of your question. A small value of $C_{\min}$ sometimes happens when most actions pointing in some direction are quite short. You might hope to learn that these actions cannot be optimal and then work in the low dimensional subspace, so solving one problem may help with the other.

Reviewer #4. (Transition between $n^{2/3}$ and $n^{1/2}$): Our bounds are non-asymptotic and our intention is not to treat any quantities as constants. The bounds show that there is a rich information-tradeoff in sparse linear bandits that appears in the high-dimensional regime. In particular, for certain action sets algorithms can enjoy nearly dimension-free regret by exploring carefully while algorithms based on optimism may be very suboptimal.

(Tightness of lower bound in the data-rich regime): We do not claim our lower bound is tight in the data-rich regime ($d < n$) where a lower bound of $\Omega(\sqrt{dsn})$ is already known to be optimal. **(Solving the optimization problem):** When the number of arms is finite it can be solved using standard convex solvers since the minimum eigenvalue is a concave function. If the number of arms is infinite, things will likely be delicate in general. Hints may be found in the literature on optimal design where computational questions remain open.

[1]. Doubly-robust lasso bandit. NeurIPS 2019. [2]. Improved algorithms for linear stochastic bandits. NIPS 2011. [3]. Online-to-confidence-set conversions and application to sparse stochastic bandits. AISTATS 2012. [4]. Simple regret minimization for contextual bandits. ArXiv 2018.