We thank all reviewers for their valuable comments, and will improve our writing (like correcting grammar error, reorganizing main results, explaining relations with previous work, introducing various definitions more clear) in further revision. Below we respond to several raised issues. Here, the reference citation number is the same as main submission.

Reviewer 1:

— "report the best existing results for regular DP analogous to Table 1 (in the paper)":

Please see right table for the comparison between our results in LDP setting and the best existing results in DP setting. As one can see, in smooth BCO setting (either convex or strongly convex), we achieve improved regret bounds compared with DP counterparts (note in context-free setting, LDP bandit learning is strictly harder than DP bandit learning). For multi-point BCO, as far as we know, there is no prior study in this setting. In contextual bandit setting, LDP and DP are not comparable, and [34] proved a lower bound of linear regret for DP contextual linear bandit, which is a special case of generalized linear bandits. Given this lower bound, our results show there is a fundamental difference between LDP and DP contextual bandits learning.

Type	Problem		LDP (Ours)	DP (Existing)
Context Free	BCO	Convex	$\tilde{\mathcal{O}}\left(T^{3/4}/\varepsilon\right)$	$\tilde{\mathcal{O}}\left(T^{3/4}/\varepsilon\right)$ [36]
		Convex & Smooth	$\tilde{\mathcal{O}}\left(T^{2/3}/\varepsilon\right)$	$\tilde{\mathcal{O}}\left(T^{3/4}/\varepsilon\right)$ [36]
		S.C	$\tilde{\mathcal{O}}\left(T^{2/3}/\varepsilon\right)$	$\tilde{\mathcal{O}}\left(T^{2/3}/\varepsilon\right)$ [36]
		S.C & Smooth	$\tilde{\mathcal{O}}\left(T^{1/2}/\varepsilon\right)$	$\tilde{\mathcal{O}}\left(T^{2/3}/\varepsilon\right)$ [36]
	MP-BCO	Convex	$\tilde{\mathcal{O}}\left(T^{1/2}/\varepsilon^2\right)$	None
		S.C	$\tilde{\mathcal{O}}\left(\log T/\varepsilon^2\right)$	None
Context Based	Contextual Linear Bandits		$\tilde{\mathcal{O}}(T^{3/4}/\varepsilon)$	$\Omega(T)$ [34]
	Generalize Linear Bandits		$\tilde{\mathcal{O}}(T^{3/4}/\varepsilon)$	$\Omega(T)$ [34]

Reviewer 2:

— "For contextual bandit, this work shows a sub-linear regret for LDP, but there exists a linear regret for DP..."

As explained in lines 53-61 in our submission ("collected information" there means the quantity based on use's private data, like line 9 in Algorithm 3 without adding noise), our result doesn't contradict with the lower bound proved in DP setting [34]. In more detail, on one hand, post-processing property holds only for the output of a DP algorithm which doesn't use private data any more. However, in our algorithms for LDP contextual bandits, though we can use post-processing property to prove estimation sequence $\{\tilde{\theta}_t\}$ satisfies DP, it doesn't imply the output action sequence $\{x_t\}$ satisfies DP, as these actions are made in the local side which use private local data. On the other hand, to show the difference more intuitively, assume the true parameter θ^* is known in advance. For two users with completely different features, optimal actions for them should be different. In DP setting, it requires output actions to be close to each other, which then will cause some regret inevitably. While in LDP setting, since we know θ^* and decisions are made locally based on local features, we can choose optimal actions for each user locally without causing any regret. Hope these two explanations help you understand our results and the difference with previous DP setting better.

— "the regrets for MP are not consistent, one is over epsilon squared, the other is over epsilon. ":

Thanks for pointing out this issue! It is a typo, and the regret for MP-BCO should be over ϵ^2, since the regret in non-private setting depends on G^2 (G is the Lipschitz constant) which leads to ϵ^2 here.

— "in the proof of Theorem 6, Eq(27) ... is not trivial":

Eq(27) holds after using the non-private guarantee $\mathrm{Reg}(\mathcal{A}, \cdot)$ with respect to functions $\{\tilde{f}_t(x)\}$ defined in line 534. Since the Lipschitz constant of $\{\tilde{f}_t(x)\}$ is upper bounded by $G + \sigma\sqrt{d}$, we obtain Eq(27) accordingly.

Reviewer 3:

— "Is the dependence on eps in the reduction optimal?":

We tend to believe our dependence on ϵ in the reduction is optimal in most cases, and there are some implications. First, [Agarwal Singh 2017] achieved the form (non-private regret + $1/\epsilon$) under DP guarantee only for online linear optimization rather than bandit setting, and [Tossou and Dimitrakakis 2015] achieved similar form under a variant of DP focusing on a single output instead of regular DP focusing on sequential output. Besides, for MAB, [7] proved lower bounds for several different versions of DP (including DP and LDP), which nearly match our upper bounds.

— "For the contextual setting, it is not accurate to say ... make the comparisons clear":

We agree with your comment! Actually in our submission, we do mention LDP and DP are not comparable in contextual bandits (in lines 53-54) and only claim LDP is stronger than DP in context-free setting. Besides, LDP does require some computational resource at users which may be unavoidable if we want to protect LDP and make accurate recommendations simultaneously. We will explain all of these more clear in future version.

Reviewer 4:

— "The paper has no experiments. It is more interesting to see the trade-off ...":

Thanks for your suggestion! We have conducted some experiment for private MAB, and it indeed shows the trade-off between privacy and utility. We will add experimental part in future version.