Dear Reviewer #2:

> prior work combining cooperative and delay model together, while this paper only considers them separately.

Our work combines cooperative and delay models together as well. In the cooperative setting defined in Section 3.2., the message is sent at the end of each round, which implies that the feedback information observed by an agent is transmitted to another distant agent with a delay depending on their distance, as mentioned in Lines 245–248. Therefore, the cooperative setting in our paper includes delayed-feedback problems. This structure of delay and cooperation is the same as the prior work in COLT 2016 paper by Cesa-Bianchi et al.

> 2. it still requires more computation time than other discrete linear bandit algorithms, not to mention ..

For some discrete linear bandit problems, our algorithm is more efficient than existing algorithms. This is because existing algorithms (such as in [Cesa-Bianchi and Lugosi (2012)]) rely on sampling over a combinatorial action set, which can be computationally hard depending on the action set. For example, when the action set is the set of all maximum matchings of a given non-bipartite graph, there is no known polynomial-time algorithm for sampling over this discrete action set. In such a case, our approach, continuous relaxation and truncation, is computationally better. As the reviewer mentions, however, improving practical computational cost is important future work. We shall mention the computational weakness that the reviewer pointed out in the revised version.

> 1. It is not clear to $\hat{\ell}$ is unbiased. The point is that $S(p'_t)^{-1}$ may not exist. I think a fix is that we have ..

We can see that $S(p'_t)$ is invertible from the assumption that $\mathcal{A}$ is not contained in any proper linear subspace, which is stated at Line 140. Under this assumption, indeed, $\mathcal{B} = \mathrm{Conv}(\mathcal{A})$ is a full-dimensional convex set with a positive Lebesgue measure. Combining this and Lemma 1, we can see that the domain of p'_t is full-dimensional as well. Therefore, the distribution p'_t has a density function taking positive values over a full-dimensional convex set, which implies that $S(p'_t)$ is invertible. A similar argument can be found, e.g., in p.8 of [Ito et al., oracle-efficient algorithms for online linear optimization with bandit feedback, NeurIPS2019] (between Eq. (4) and (5)), and is implicitly used in [Bubeck, Eldan, Lee, STOC2017] as well. In the revised manuscript, we add a more clarified proof for this fact. We also would like to note that the assumption at Line 140 does not affect the generality of the problem. Indeed, if $\mathcal{A}$ is contained in a proper linear subspace, we can find such a subspace using the linear optimization oracle for $\mathcal{A}$ (e.g., from Corollary 14.1 of [Schrijver (1998)]). Hence, by reducing the entire vector space into this linear subspace, we can transform the problem so that the assumption holds.

> 2. Moreover, the computation of $S(p'_t)^{-1}$ can also be problematic. .. It is not even a log-concave distribution.

We can see that p'_t is a log-concave distribution as its domain $\{x \in \mathbb{R}^m \mid \|x\|^2_{S(p)^{-1}} \leq m\gamma^2\} \cap \mathcal{B}$ is a convex region and the density function p_t defined by (9) is log-concave. Since $\tilde{p}_t$ is log-concave, for any $\epsilon > 0$, we can get an ϵ-approximation of $S(\tilde{p}_t)$ w.h.p. by generating $(d/\epsilon)^{O(1)}$ samples from $\tilde{p}_t$, from Corollary 2.7 of [Lovaŝz and Vempara (2007)]. Samples from $\tilde{p}_t$ can be generated with their polynomial-time sampling algorithm as mentioned around Line 188. A similar discussion can be found in Lemma 5.17 of [Bubeck, Lee, Eldan (2017)]. This fact is used in [Hazan and Karnin (2016)] as well. We shall clarify this in the revised manuscript.

> 3. In Lemma 4 and equation (27), it is not trivial the inequality can be applied.

As the reviewer pointed out, in (27), we need to confirm that $y > -1$ holds for applying the inequality $\log(1+y) \leq y$. This condition $y > -1$ indeed holds since y can be expressed as $y = \mathbf{E}[-\eta\hat{\ell}_t^\top x + g(-\eta\hat{\ell}_t^\top x)] = \mathbf{E}[\exp(-\eta\hat{\ell}_t^\top x)] - 1 > -1$. We add a more clarified explanation in the revised manuscript.

Dear Reviewer #3:

> It would be nice to discuss whether it relates to the focus region of [Bubeck, Eldan, Lee, STOC 2017].

Thanks for providing an important reference. Their technique is similar to ours in that they truncate the domain using the covariance matrix, though much difference can be found as well. For example, in contrast to our truncation technique, their focus region is updated so that the new one is included in the prior one. This property seems essential for stabilizing their kernel-based estimators, but makes the algorithm and the analysis much complicated. In the revised paper, we cite this reference and add a discussion on the relation to this.

Dear Reviewer #4:

> The paper only considers the “easy” setting of fixed and known delay d for all time steps.

As the reviewer pointed out, it is a significant future work to extend the model to deal with the unknown and round-dependent delay. We shall mention this in the revised manuscript. We believe that adjusting parameters adaptively should work well for this general setting. However, we have not yet found such a sophisticated way of parameter setting.