We thank all reviewers for their comments. Minor comments will be addressed in the final version.

Reviewer 1

Comparison with related work Thanks for the references to work of Ross & Bagnell, Saha et al., and Arora et al. All of these papers investigate the relationship between regret and stability of an online learning algorithm and a comparison between different stability conditions certainly makes sense. However, note that these papers do *not* investigate any of the following (that we do in our paper): connections with differential privacy, first order regret bounds, and partial information settings (Arora et al. do consider partial information). We will add a comparison in the final version along with a pointer to different types of stability conditions existing in the literature.

Reviewer 2

Comment 1. Your questioning of the dimension dependence in Theorem 3.2 and Corollary 3.3 is valid. Indeed OGD/FTRL algorithms in these settings will not incur the dimension dependence. However, note that it is not at all clear whether this dimension dependence is due to the use of privacy tools and techniques. Indeed, even the best known *zero-order* bound (ref. [1] in our paper) for OLO in the ℓ_2/ℓ_2 setting via FTPL has a $d^{1/4}$ dependence on dimension d. To the best of our knowledge, there is no existing analysis, whether privacy/stability based or otherwise, of *any* FTPL algorithm which does not incur at least this much dependence. It is unknown whether this is an intrinsic limitation of currently available analysis tools or of FTPL methods themselves. In fact, it is somewhat surprising that using the DP based analysis we get first-order regret bounds with a dimension dependence that is the best possible given currently available techniques. Further, this dimension dependence only arises in Theorem 3.2 and Corollary 3.3. Other results in the paper, e.g. experts setting result (Theorem 3.6) and multi-armed bandit result (Theorem 4.2) do not incur any avoidable polynomial dimension dependence.

Comment 2. You're right, Algorithm 1 has to solve a convex optimization problem at each step. But it shares this property with FTRL algorithms that have to do the same. As you mentioned, deriving a perturbed OGD type algorithm with a more efficient update will be an interesting topic for future work.

Comment 3. The analysis of geometric resampling that takes error due to finite number of samples is already given by Neu & Bartok (ref. [26] in our paper). They showed that error due to drawing M samples contributes an extra KT/M term in the regret where K is the number of arms in a bandit problem. For zero-order $O(\sqrt{T})$ regret bounds, one can therefore choose M of the order of $\sqrt{T}$ (however the expected number of samples needed can be shown to be *constant* per time step, see their Theorem 2). For first-order bounds, one would have to choose a larger M of the order of T which increases the computation. We will add more discussion about this in the final version.

Reviewer 4

adding DP definition to the paper Yes, we can do that.

rank one Hessian of loss functions Yes, we are quite positive that the restriction can be removed (personal communication with one of the top experts in DP). In particular, the corresponding DP result should hold for higher ranks at the cost of some degradation in the privacy parameter ϵ. Our online learning result will immediately inherit such a future improvement when it occurs in the DP literature.

first-order bounds for non-convex problems This is an intriguing suggestion! The regret bound of the fictitious algorithm $\mathcal{A}^+$ does not use convexity! The only reason convexity is required in Theorem 3.2 is because the DP result of Kifer et al. required convexity. So if one had a DP guarantee for a perturbed ERM algorithm even with non-convex losses, everything would work. Of course, computing the perturbed ERM would involve non-convex optimization so efficient computation may not be possible.

GPBA for OCO The best in hindsight action in OLO is $\mathrm{argmin}_{x\in\mathcal{X}} \sum_t \ell_t^\top x = \mathrm{argmin}_{x\in\mathcal{X}} L_T^\top x = \mathrm{argmax}_{x\in\mathcal{X}} (-L_T)^\top x$. This is simply the gradient of the support function of the set $\mathcal{X}$, $y \mapsto \max_{x\in\mathcal{X}} y^\top x$, evaluated at $y = -L_T$. The gradient mapping does not arise in this natural way when the best in hindsight action $\mathrm{argmin}_{x\in\mathcal{X}} \sum_t f_t(x)$ is considered in the case of convex functions f_t.

DC assumption seems more applicable to the bandit setting A clarification is needed here: DC assumption was indeed introduced by Abernethy et al. (ref. [2] in our paper) in the bandit setting but note that it is actually an assumption on the *potential function*. If you look at the (short!) proof of Theorem 3.6 in Appendix B.3, you will see that DC property of a set of potentials also leads to regret bounds in the full-information experts setting. To summarize, the DC assumption is an assumption on potential functions. Certain potential functions (e.g., those that satisfy DC with a larger exponent γ) are better suited for bandit settings but the assumption itself is not at all tied to the bandit setting.