We would like to thank the reviewers for their careful revision, positive comments and constructive criticism. We address below some of the major criticism raised by reviewers. Due to space limitations, not all suggestions can be included in the conference file, but we will make sure to include them in the full version of the paper.

- *Broader Impact:* 2 reviewers commented that the paper is short on the Broader Impact section. In the updated conference version we will add a more thorough discussion on private data analysis and its societal impact.
- *Strong convexity:* Unfortunately, our analysis does not benefit from strong convexity. However, using UAS of ERM for strongly convex losses + convergence rates for strongly convex SGD leads to UAS for SGD.
- *Motivating examples:* due to space considerations, we cannot include in the NeurIPS file detailed motivating examples of nonsmooth losses arising in ML. However, we will include those discussions in the full version. We also want to point out the reviewer is right in that when proximal mappings are efficiently computable, one should prefer using them: in particular, stability bounds for proximal methods are similar to those of smooth SGD, which is known since Hardt et al.'16. We do not include this discussion since it is a well known technique, and we address general settings when this operator may not be efficiently computable.
- *Gradients vs. subgradients:* We acknowledge the ambiguity in denoting subgradients by $\nabla f(x)$. However, since $\partial f(x)$ is a set (the subdifferential), not a vector, we prefer not to use ∂. We will add a comment clarifying that in the paper we denote by $\nabla f(x)$ any subgradient of f at x.
- *About other works on stability:* We thank the reviewer for pointing out the "Fine-grained analysis" paper, which is a concurrent work to ours and appeared at ICML and on the arxiv both several weeks after the Neurips deadline. This work only addresses a weaker on-average stability notion, which in particular is not applicable to our privacy results. We will discuss this comparison in more detail in the revision. We also remark that approximate contraction may not hold for SGD in our setting. Indeed our lower bounds show that SGD trajectories can deviate $\Omega(\eta_t)$ at every step, which is the worst possible. Hence, our upper bounds are unimprovable, and they circumvent this requirement of approximate contractivity altogether.

  We were aware of "Private Stochastic Convex Optimization" paper, which also does not address uniform stability. More crucially, this paper has serious errors that impact all their main claims. We have contacted the authors of this paper, who have admitted these errors, and they will soon retract it from arXiv.
- *Privacy Analysis:* The privacy analysis follows from the prior work in almost a straightforward way. The technique in [1] applies to our case equally well. The only place where sampling is invoked in the analysis of [1] is in Lemma 3. In the proof of that lemma, it is easy to see that the only relevant condition that involves sampling is satisfied in our case. Note that in our algorithm we sample one point in each iteration, hence the distributions induced over a pair of neighboring datasets satisfy the same condition in the proof of Lemma 3 in [1] (where q in that lemma is $1/n$). As for the asymptotic moment bound in [1], we can derive explicit bounds on the constants in [1] in our case when n is sufficiently large. We will clarify all these details in the full version.
- *Optimality of SGD:* Unfortunately, we do not have space in the main file to add derivations regarding the optimal rates for multipass SGD. We will however explicitly mention this result, referring to the full version.
- *DP-SCO based on permutation SGD?* For fixed permutation SGD, we don't have privacy amplification, which we rely on in the construction based on sampling with replacement. For a random permutation (per epoch), we have some form of privacy amplification by sampling *without* replacement, but it is weaker than what we have in the sampling with replacement SGD. However, this approach could be interesting with some form of mini-batching (where batches are sampled without replacement).