We thank all the reviewers for their thoughtful reviews and for pointing out some typos in the manuscript.

Responses to Reviewer-1's comments:

"... the restriction on delta seems crude. I cannot see where these restrictions come from.... "

First, we note that our technical results remain the same without this restriction on δ. However, this restriction is made to comply with a standard requirement on δ in the literature of differential privacy to ensure the robustness of the definition. In particular, it is standard to require δ to be much smaller than $1/n$ to rule out trivial mechanisms, e.g., the one that selects one individual uniformly at random and publishes her record in the clear (note that such mechanism satisfies $(0, 1/n)$-differential privacy but is blatantly non-private). This requirement on δ is discussed in several early references on differential privacy including the textbook by Dwork & Roth (towards the end of Sec. 2.3.3) and the survey by Salil Vadhan, and several others. In fact, in some of these references δ is even assumed to be a negligible function of n (i.e., smaller than the inverse of any polynomial in n).

"There seems to be an assumption $\sigma \geq 1$ in Lemma 3 of [ACG+16], however I do not see this appearing in the article."

First, we note that the lower bound on σ in [ACG+16, Lemma 3] is because the bound on the norm of f in that lemma (which represents the Lipschitz constant) is assumed to be 1. So, when the Lipschitz constant is L, then by simple re-normalization it is easy to see that the analogous lower bound on σ is L. Second, as pointed out in the supplementary document, due to different normalization of the noise in the gradient update step (noise in our case is normalized by the batch size m), our setting of σ is smaller than that of [ACG+16] by a factor of m. Taking these different normalizations into account, we note that the analogous lower bound on σ is L/m, and it is indeed satisfied in our case. To see this, note that given the setting of m in Step 2 of Algorithm 1 and the fact that $\epsilon \leq 1$, the setting of σ in Step 1 implies that $\sigma \geq 2L/m$.

"... is it not entirely correct to say (as in Definition 1): 'for any pair of datasets S and S' differ in exactly one data point....' Then the results of [ACG+16] would not hold. Perhaps you could comment on the neighbouring relation of S and S'."

The definition of differential privacy with respect to the adjacency notion involving addition/removal of one element in the dataset is equivalent (up to a factor of 2 in the privacy parameters) to the definition w.r.t. the adjacency notion involving replacement of one element. This follows from a simple triangle-inequality style argument since replacing one element z by another element z' can be carried out via a removal step (of z) followed by an addition step (of z'). Hence, the same techniques and results of [ACG+16] still apply in our case (after renormalizing the privacy parameters by a factor of 2).

Response to Reviewer-2's comments:

"I still think this work lacks the experimental part. As I know, most of the recent work on the central (ϵ, δ) DP-ERM has experimental study such as [1-6]. ... should provide some experimental study in order to say the improvement."

Our work is the first one to focus on differentially private stochastic convex optimization (as opposed to previous works on private empirical risk minimization). Our primary contribution is theoretical and we believe that the fundamental nature of the question we resolve makes the work interesting for the community. At the same time, our analysis concerns standard algorithms such as mini-batch DP SGD and objective perturbation for which one can easily find experimental results in the literature. The only new algorithmic aspect is the use of prox step (to get Moreau-Yosida smoothing). This step is necessary in the worst case but will not make much difference for simple linear models used in most experiments. Thus we do not think that experiments are likely to give additional insights into the question we investigate.