We thank the reviewers for their valuable feedback. We first address a general comment.

1. Strict complementarity assumption: The strict complementarity assumption is used in many papers [Forsgren et al., SIAM Rev.], [Carbonetto et al.,NIPS'08], [Liang et al.,NIPS'14], [Namkoong et al., NIPS'16], [Lu et al, arxiv:1907.04450]. It is generically true (i.e. holds with probability 1) if there is a linear term in the objective function and the data is from a continuous distribution (similar to [Lu et al., arxiv:1907.04450], [29]). The weaker assumption mentioned in the appendix generically holds for robust regression problems with square loss.

Response to reviewer 1:

1. *Arguments of multi-loop methods seem quite weak?* Indeed the statement for multi-loop methods is not a rigorous claim. It is an observation that the existing algorithms all require $O(\epsilon^{-2})$ outer iterations. We will replace this statement by "the existing works on multi-loop algorithms require $\mathcal{O}(\epsilon^{-2})$ outer iterations".

2. *Adaptive version of this algorithm is more applicable in practice.* Thank you for your suggestion. We agree and will address this issue in the future work.

3. *a) Move claims of complementarity to main text. b) Motivation for the boundedness assumption is much needed.* We will move the claims as you suggest. As mentioned in Line 186, the bounded-level-set assumption is to ensure the iterates stay bounded. The bounded level set assumption, a.k.a. coerciveness assumption, is widely used in many papers [Cannelli1 et al., arXiv:160704818], [Hong et al., ICML'17], [Cannelli1 et al., arXiv:160704818]. Bounded-iterates-assumption itself is common in optimization too (see [Xu et al., arxiv:1408.2597], [Défossez et al., arxiv:2003.02395], and [Carbonetto et al., NIPS'08]). In practice, people usually add a regularizer to the objective function to make the level set and the iterates bounded (see, e.g., [Liang et al., arxiv:1912.13472] for a neural network example). We will include these motivations.

4. *Appendix is hard to read.* We included a proof sketch in Appendix B; will move to main text and add more intuition.

5. *More reference on traditional smoothing literature.* We will add references on Moreau smoothing (e.g. Section 3.1 in the book on Proximal Algorithms by Parikh and Boyd]) and Nesterov smoothing [Nesterov, Math. Program., 2005].

6. *claim of contribution on removing compact X should be removed or put in full context of what will be later assumed.* Thanks for the comment. We will add a sentence in the introduction near the claim: note that we do not need any boundness assumption for general non-convex-concave problem, but we still have an assumption of bounded iterates (which is weaker than compact X) for the point-maximum problems.

Response to reviewer 2:

1. *Comparison to recent work [A].* Thank you for your comment. We were not aware of it. [A] proposes a single-loop algorithm for min-max problems by performing GDA to a regularized version of the original min-max problem. Using primal-dual analysis, they also prove that their algorithm attains an ϵ-solution with $\mathcal{O}(\epsilon^{-4})$ iterations for nonconvex-concave problems as in our first result. Their paper differs from ours in two aspects: i) The algorithms are different: we add a proximal term centered at the auxiliary variable z, while [A] adds a regularization term $\alpha_t(\|x\|^2 - \|y\|^2)$ that is diminishing ii) They do not prove $\mathcal{O}(\epsilon^{-2})$ complexity for the pointwise maximum case. We will add a discussion.

2. *Practicality of strict complementarity assumptions.* Please refer to the discussion of the assumptions at the beginning of this response. Also, we will further study whether we can remove or relax these assumptions in the future.

3. *Additional experiments?* We have run more experiments on CIFAR10. Our algorithm achieve similar or better accuracy than [20] on the robust training tasks under different perturbation levels. As for training speed, we takes only 9 epochs to reach loss value 0.01, while the other algorithm in [20] takes at least 30 epochs to reach the same value.

Response to reviewer 3:

1. *Title is misleading.* Thank you for your comment. We will change the title to "A single-loop smoothed gradient descent-ascent algorithm for nonconvex-concave min-max problems". Further comment is welcome.

2. *Strict complementarity assumption holds in practice? How to check?* Please refer to the discussion of the assumptions at the beginning of this response and Appendix E. It is a common assumption and can not be directly checked.

3-(a) *How Assumption 3.5 impacts the claim of contribution of removing compact X?* Assumption 3.5 is indepencent of that claim. The claim "current algorithm does not require compact X" is for "general nonconvex-concave problems", as stated in the beginning of that paragraph; but Assumption 3.5 is for pointwise maximum problem class. We will clarify this; see Response 6 to Reviewer 1.

3-(b) *Does assumption 3.5 imply compact X?* There are examples for which X is unbounded while the level set is bounded. For example, consider $\min_x \max_{i\in[m]} f_i(x)$ where $f_i(x) = \|A_i x - b_i\|^2 + \sum_{j=1}^n x_j^2/(1+x_j^2)$ with $[A_1, A_2, ..., A_m]$ full column rank. Here the domain $X = \mathbb{R}^n$ is not compact but $\max_i f_i(x)$ has bounded level sets.

4. *Does relaxing the compactness of X extend the applicability of the algorithm?* i) There exist applications; for the example in 3-(b): our result still applies, but [24] does not directly apply. ii) Anyhow, we change it to "However, the current algorithm does not require the compactness of the domain X" since we did not intend to list many specific applications for this point.

5. *Proof of Lemma B.8 is missing.* Thanks for pointing it out. We will add it in the revised version.

6. *Typos.* We really appreciate your careful reading and pointing out these typos. We will address them in revised version.