We thank the reviewers for their constructive comments. However, we are afraid that some reviewers have underrated our technical contributions and the importance of the work.

Technical Contributions: While some of the ideas introduced in the paper (generalization of two-sided PL condition, algorithmic extension with variance reduction) seem intuitive, the theoretical analysis in terms of alternating GDA is nowhere near as straightforward. We emphasize several facts, which are most likely overlooked by some reviewers:

- Even in the strongly-convex-strongly-concave setting, no convergence analysis exists for alternating GDA in the literature. We provide *the first convergence result for alternating GDA* for more general problems.
- For another special case of our setting (convex-strongly-concave with bilinear term), the best known result [1] gives the $\mathcal{O}(poly(\kappa)\log(1/\epsilon))$ complexity of the simultaneous GDA, which is much worse than our result.
- Unlike simultaneous GDA, our alternating GDA uses *different learning rates* for the primal and dual variables. The ratio between these learning rates plays a critical role in establishing the convergence.
- Our analysis hinges upon novel construction of a potential function inherent from geometric property of the two-sided PL function and introducing a balancing parameter to establish the contraction. When our analysis is extended to nonconvex-PL minimax (presented in the appendix), it allows *larger stepsizes and is much neater* compared to some recent work [2], which analyzed alternating GDA in nonconvex-strongly-concave setting.

Motivation and Importance:

- **Why two-sided PL condition?** There are tons of applications whose minimization objectives satisfy PL condition, such as LQR (Fazel et al, 2018), computing matrix squareroot (Jain et al, 2017), phase retrieval (Zhou et al, 2016), etc. Significant advance has been recognized by utilizing the PL conditions in these field. Their robust counterparts would naturally lead to two-sided PL conditions. Hence, understanding this subclass of problems is of great importance.
- **Why alternating GDA?** Alternating GDA is widely used for training GANs and other minimax problems in practice; see e.g., [3; 4]. Yet, its convergence rate has rarely been analyzed in theory. *The main goal of this paper to is provide the theoretical analysis for alternating GDA, rather than designing an optimal algorithm for minimax optimization* (as some reviewers may expect).

Additional Response:

To Reviewer 1: (1) See response above. (2) It may not be obvious, but the objective in Figure 1 is nonconvex-nonconcave. We will remove these figures to avoid confusion. (3) To the best our knowledge, very little is known about *algorithm-specific lower bound* for minimax optimization. Even the well-known complexity of $\mathcal{O}(\kappa^2\log(1/\epsilon)$ of simultaneous GDA on *SC-SC setting* has never been proven to be tight. (4) LQR example is non-convex in terms of the primal variable, which is a well-known challenge in the control field, albeit the small dimension.

To Reviewer 2: Note that the analysis of minimization with PL condition (Karimi et al., 2016) is exactly the same as the classical analysis for strongly-convex objectives. However, there is no previous results on AGDA for SC-SC setting, and the analysis of SGDA for SC-SC setting (Facchinei and Pang, 2007) can be, by no means, extended to the two-side PL setting. Our analysis hinges upon novel construction of a potential function and leads to a number of interesting results that improve over existing work. See detailed discussion above.

To Reviewer 3: Thanks for the acknowledgement of our contribution! We will dedicate a section to discuss potential accelerations of the alternating GDA algorithm and open questions related to the algorithmic-specific lower bounds.

To Reviewer 4: (1) It should be noted that even in many special settings, e.g., strongly-convex-strongly-concave, convex-concave with two-sided PL, and one-sided PL+strongly-concave, our convergence analysis either improves over existing results or closes the gap in the literature. Hence, our contribution is not just limited to nonconvex-nonconcave problems with two-sided PL conditions.

(2) The example provided in Figure 1 is indeed nonconvex-nonconcave; the LQR example in Figure 4 is nonconvex-strongly-concave. Although LQR and robust least square examples include some convexity/concavity, there was no theoretical analysis before showing that AGDA achieves linear convergence for such cases.

[1] Du, S. and Hu, W. *Linear convergence of the primal-dual gradient method for convex-concave saddle point problems without strong convexity.* AISTATS, 2019.

[2] Xu, Z., Zhang, H., Xu, Y., and Lan, G. *A unified single-loop alternating gradient projection algorithm for nonconvex-concave and convex-nonconcave minimax problems.* arXiv: 2006.02032, 2020.

[3] Liu, M. and Tuzel, O. *Coupled generative adversarial networks.* NeurIPS, 2016.

[4] Metz, L., Poole, B., Pfau, D., and Sohl-Dickstein, J. *Unrolled generative adversarial networks.* ICLR, 2017.