Rebuttals for "General Proximal Incremental Aggregated Gradient Algorithms: Better and Novel Results under General Scheme"

We thank the three reviewers for their constructive feedback.

Reviewer 1.

Q1:*... a lot of references on KL property, e.g.[1][2]. But I cannot find any of them.* **A1**: Thanks for your kind suggestions. [1][2] are good references for the KL property and should be included in the further version of our paper. The detailed definitions and references of KL and semi-algebraicity will be presented in the appendix part.

Q2: *The non-asymptotic rate achieve by assuming 1) boundedness of $\{x_t\}$; 2) semi-algebraic of objective, which is a sufficient condition for KL.* **A2**: Good suggestion. As R1 says, the non-asymptotic rate can be proved under these conditions. The proofs can be presented with existing methodology given in works like [1][2]. The future version can present the results about the rates.

Q3:*The experiment is not convincing. The authors should include more comparisons in the revision, e.g. [3].* **A3**: We will add the numerical comparisons with other incremental gradient methods including the one given in [3] for convex and nonconvex regression tasks. Thank you!

Q4:*The paper is mainly based on the assumption that $\sum_i \sigma_i^2 < \infty$. However, the authors did not provide a method to guarantee the assumption in stochastic setting.* **A4**: The assumption $\sum_i \sigma_i^2 < \infty$ is like the Lipschitz assumption of gradient. If a function is nonsmooth, the vanilla gradient descent certainly cannot work. The theory built in this paper is for the algorithms satisfying the summable assumption. But if the algorithm fails to obey, the general PIAG cannot work, either. Whether the assumption is satisfied or not depends on the algorithm itself. In lines 131-136, we mentioned these facts. This is no universal way to guarantee the assumption just like that we cannot make sure all functions are smooth. For example, for SGD, SVRG and SAGA, the assumption is broken. But for the stochastic BCD algorithm and the asynchronous SBCD, the assumption holds well. We can provide the guarantee for block coordinate descents. And we will be specific on this point in future version.

Reviewer 2.

Q1: *many definitions are lacking (e.g., the semi-algebraic property and the distance used at l. 154)* **A1**: We will give the detailed definitions of KL and semi-algebraic in the appendix. The distance is denoted by $\mathrm{dist}(\mathbf{0}, \partial F(x^k)) := \min_{v\in\partial F(x^k)} \|v\|_2$. We will be specific on the definitions and notation in revision. Thanks.

Q2: *I feel some assumptions should be discussed (e.g., l. 192, is this assumption realistic when $\sigma_k = k^{-\eta}$?)* **A2**: $\sigma_k = k^{-\eta}$ cannot obey the assumption. The assumption can promise geometric decreasing parameters, which are given in Theorem 4 and 5. We currently cannot include the assumption $\sigma_k = k^{-\eta}$ for technical reasons. How to weaken the assumption will be left to future studies.

Q3: *plots are hard to read and would benefit from additional comments.* **A3**: The figure will be enlarged and additional comments (like comparison and explanations of the performances) will be given. We will add experiments with other incremental methods as mentioned in A2 for Reviewer 1.

Reviewer 3.

Q1: *The paper contains high quality results and proofs. Although, the primary focus of the paper is on theory, the quality of numerical section and the figure can be much improved.* **A1**: Thank you very much for your positive opinions. The extra numerical experiments and additional comments will be made and plots will be enlarged. Please see A3 to R1 and A3 to R2.

Q2: *Line 132: can be explained better or add citation to the result if taken from somewhere else.* **A2**: The result can be found in [Neterov Y. 2011, Efficiency of coordinate descent methods on huge-scale optimization problems, Siam J. Opt.] and [25]. We will add the citations.

Q3: *the assumption on σ_i's can be explained better with some discussion/note.* **A3**: Thank you for your suggestion. We answered part of this in A4 for R1. And we will be more specific on this assumption.

Q4: *Lemma 2: uses a quantity 't' without defining it.* **A4**: That is a typo. Actually, here $t = 1$. We will correct it.

Q5: *(the inequality should be LHS $\leq 0.5\times\sqrt{H^2+4H} - H$* **A5**: We do not compute errors here because $\frac{\sqrt{H^2+4H}-H}{2} = \frac{(\sqrt{H^2+4H}-H)(\sqrt{H^2+4H}+H)}{2(\sqrt{H^2+4H}+H)} = \frac{4H}{2(\sqrt{H^2+4H}+H)} = \frac{2H}{(\sqrt{H^2+4H}+H)}$.

All reviewers: We will address your other comments in the final version. All of your major concerns have been addressed above. We hope you can reconsider your opinion on our paper.