

1 We thank the reviewers for their time and feedback. We address the main concerns raised by different reviewers below.

## 2 **Reviewer 1**

3 We will provide more illustrations on simple data-sets. As our first goal was to provide a tight analysis of local-sgd  
4 from a theoretical perspective, we did not highlight the experiments, some are provided in the Supplementary Material.

5 We will add an illustration of the optimal behavior described on lines 343-357, and a direct illustration of Corollary 4.

6 While not directly illustrative of our theory, experiments in related papers like [1], [2] (which focus on practical aspects)  
7 are consistent with our findings, and underline similar observations.

## 8 **Reviewer 2**

9 Thank you for your detailed feedback and precise comments!

10 We will further clarify the discussion about functional convergence and also present it earlier. We will also add the  
11 suggested references and minor changes to the final draft.

12 Thank you for the suggestion regarding Assumption A3, which is indeed a bit technical: we will simplify its description  
13 as you suggested!

## 14 **Reviewer 3**

15 We would like to point out that the reviewer's major criticism is already tackled in our paper. We offer results both with  
16 and without the quadratic assumption (Q1): while Theorem 3 uses the Q1, Proposition 1,2 and Theorem 6.2 **do not**  
17 **require** this assumption.

18 Especially, Theorem 6.2 which is one of the main contributions of the paper, does not use Q1, but only the more  
19 common smoothness assumption. The propositions leading up to this theorem (placed in the appendix due to lack of  
20 space), as well as the online setting's results (Proposition 7) also do not make this assumption.

21 We chose to first provide and describe results under Q1 in the main paper only for the sake of brevity, and because they  
22 convey a very similar message more elegantly, and help to understand our approach.

23 We hope this clarifies the reviewer's main concern and that the score will be updated to reflect this.

## 24 **References**

25 [1] T. Lin, S. U. Stich, and M. Jaggi. Don't Use Large Mini-Batches, Use Local SGD. 2018.

26 [2] J. Zhang, C. De Sa, I. Mitliagkas, and C. Ré. Parallel SGD: When does averaging help? 2016.