We thank the reviewers for their helpful comments. Please find below our responses.

**Reviewer 1:** Thank you for your helpful feedback. We will add additional discussion to the paper to give more intuition for the potential function, and we will add some experiments illustrating its behavior. (In footnote 6 we give an example showing that a natural alternative to our potential, namely consecutive "gradient" norms \( \| (F_{\hat{G}}(z(t)), F_{\hat{G}}(z(t-1))) \| \), will not suffice as a potential function.)

**Reviewer 2:** Thank you for your helpful feedback. We will update the discussion around Theorem 7 to state the complementary fact (and refer to the appropriate references) that for a fixed problem instance the last iterate has linear convergence and so eventually becomes better than the averaged iterate.

**Reviewer 3:** Thank you for your helpful feedback. We will add a reference to the CDC 2017 paper.