We sincerely thank the three reviewers for their valuable feedback. Upon acceptance, we will include in the final version (a) *improved notations* and (b) *an improved presentation of our bounds and proofs*.

Reviewer 1: We thank the reviewer for the valuable comments.
Notations: We will include the suggested notation for T. We will also define the dual norm in the Notations paragraph in the revision of our paper.

Reviewer 3: We thank the reviewer for the thorough analysis. Our remarks are listed below:
Assumption H3:

Thanks for your constructive comments. It is clear that in convex case a better prediction reduces the bound. In the non-convex case it holds as well, with some careful analysis. For **H3**, if we alternatively consider $0 < m_t^T g_t = a\|g_t\|^2$ and $\|m_t\| \leq \|g_t\|$ (i.e. m_t lies in the hemisphere with g_t as its midline), we can show that $\tilde{C}_2$ reaches minimum when $a = 1$ (i.e. $m_t = g_t$). Also, $\tilde{C}_1$ is minimized at $a = 1$ under some conditions on the parameters (β_1, β_2 etc.). **That means the bound for non-convex case is tighter when m_t predicts g_t well, similar to the convex analysis.** We will adjust our discussion and presentation in the paper to address this point.
Proof Theorem 1: As rightly mentioned by the reviewer, we use Eq (18) the inequality $\|w_t - \tilde{w}_{t+1}\|\|g_t - \tilde{m}_t\| \leq (1/2\eta)\|w_t - \tilde{w}_{t+1}\|^2 + (\eta/2)\|g_t - \tilde{m}_t\|^2$ which stems from an application of Young's inequality as explained page 18 under Eq (18). m_t should read $\tilde{m}_t$, this is a typo since we can notice in the final bound that only $\tilde{m}_t$ appears. This typo is fixed and does not change the final bound.

When $B_1 = 0$, $h_t = \tilde{m}_t$ indeed. We will include that.

Equation (20) is a one line calculation. It corresponds to a weighted sum of squares bounded by the largest term times the sum of weights. We will add an intermediate line in the revision.

Lemma 3: Note that Eq (36) uses the intermediate equality on the quantity $\overline{w}_{t+1} - \overline{w}_t$ (and not the upperbound) that is used in the proof of Lemma 3. We will clarify this point in our proof. Hence, as we are using an equality, the problem you raised is no longer one.

We thank the reviewer for the typo (bad placement of subscript) in eq (6) that will be fixed in our revision.