We sincerely thank the four reviewers for their valuable feedback. We first discuss a few common concerns shared by **reviewer 1**, **reviewer 2**, **reviewer 3** and **reviewer 4**.

●● **Non-convex bound:** empirical edge of the optimistic update but also a theoretical. 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.

Reviewer 1: We thank the reviewer for valuable comments. Our point-to-point response is as follows:
Convex regret bound:

Reviewer 2: We thank the reviewer for valuable comments. A proofreading is being done we clarify that:
Novelty of the contribution:

Reviewer 3: We thank the reviewer for the thorough analysis. Our remarks are listed below:
Gradient prediction algorithm: