Reviewer 1

Related Work Concern. Thank you for emphasizing the importance of including related work in the main paper. We fully agree that discussing connections to prior studies is a vital component of scientific work. In the later version, we will relocate the related work section to the main body, ensuring its prominence. To accommodate this, we will shift less critical content, such as supplementary experiments, to the appendix.

Minor issues. Page 1, Edge of Stability Citation We will add a citation to clarify the origin of the Edge of Stability concept.

Minor issues. Page 6, Definition 5 We appreciate your observation that constraining $\mathbf{w}$ to the span of $\{\mathbf{x}_1, \cdots, \mathbf{x}_n\}$ is unnecessary, as components outside the span do not affect the linear constraints and only increase the norm. We will refine this definition to align with the standard definition and correct our disscussions.

Minor issues. Page 15, Citation Correction We sincerely apologize for our oversight in the related work review. We regret missing the key contribution of Wu et al. (NeurIPS 2018, "How SGD Selects the Global Minima in Over-parameterized Learning") on the Edge of Stability. In the later version, we will correct the citations to properly credit this work.

Reviewer 3

Minor remarks: in the definition of ρ_t, what if $w_t = 0$? $\mathbf{w}_t$ cannot be zero in either practical scenarios or theoretical analysis because of the normalization $\mathbf{w}/\|\mathbf{w}\|$.

Minor remarks: Assumption 4, couldn't we just need the data to be linearly separable? It is possible for Assumption 4 to be replaced by linearly separable condition. The reason we use Assumption 4 is by this, we have cleaner technical details.

Minor remarks: Theorem 7: point 3, what do you mean by "the peak is at most ..." ? The peak means the spike value. "the peak is at most ..." means that spike value can be no larger than ..., that is it is upper-bounded by some values. We will clarify this in the further vision.

Minor remarks: Figure 1: typo in the caption, it should be a minus instead of plus sign We regret the typo and will correct the plus sign to a minus sign.

Minor remarks: after Lemma 12: typo, the bound should be on the gradient of R, not R itself We apologize for the error: the bound applies to the gradient of R, not R itself. This will be fixed.