We thank the reviewers for their efforts and overall positive feedback. Below we address their main concerns.

1. Bounded distributions: We note that *any* whitened distribution is 1-bounded. Hence, $O(1)$ bounded do arise in practice (as whitening is a very popular reprocessing). We will add a comment about that.
2. Tightness of the analysis: The analysis is tight in the sense that the bound in theorem 4 is optimal, up to constant. To the best we know, previews results do not imply that, despite significant efforts in recent years. Moreover, we disagree that the "convergence is directly proved based on Hoeffding's bound". This is far from being true. In order to establish our result we developed a new methodology to analyze vector random features, and used the boundedness of the distribution in a delicate way. The best evidence that the analysis is not trivial is that the result is new despite very significant research in recent years by top researchers – there were more than 20 papers devoted to memorization and NTK, and none of them derived such a convergence result. It is clear that an effort has been made to derive such a result, as the rate in which the NTK converge is central in the analysis of most of these papers.
3. "it is not well explained in what sense are the results considered using "no over-parameterization": No over parametrization means that $\tilde{O}(m)$ parameters are enough to memorize m points. This is standard terminology in related literature. We will add details about this and will make it clearer in the final version.
4. Presentation and details: We will make any effort to improve the writing and add proof details along the lines raised by the reviewers.