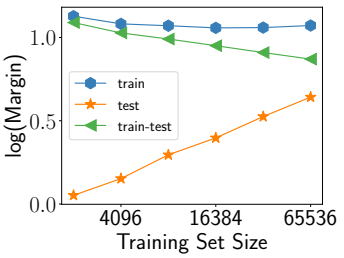
We thank the reviewers for their feedback, especially R1, R4 and R5 for appreciating the significance of our work at this juncture in deep learning theory. Below, we respond to the reviewers in order.

R1: Again, we thank the reviewer for their encouraging comments. As indicated in their suggested improvements, we will certainly add a table that lists all existing uniform convergence bounds, in future versions of the paper.

R3: Based on our understanding, R3 believes that our negative result about the overparameterized linear model in Thm 3.1 – although correct – is trivial because it is already known that "uniform convergence (u.c.) holds only when dataset size is proportionally larger than dimension". This is a strong claim which implies that u.c. is known to not hold in *any* overparameterized linear setting (hence implying Thm 3.1). **First**, there is no such strong statement in learning theory. (Nor is there a specific result like ours which shows that even the *tightest* u.c. bound, namely $\epsilon_{\text{unif-alg}}$, can fail for *some* overparameterized linear models). **Second**, **this strong claim is incorrect as it contradicts fundamental results** like margin-based u.c. bounds for SVMs (Theorem 4.4. in [18]), which are known to be meaningful even in infinite dimensions. **Third**, we must emphasize our ReLU example in Sec 3.2, which is not mentioned in the review. This example is nearly identical in terms of its "dimensionality" to common settings like MNIST (parameter count $\gg$ dataset size, input dimensionality is $\approx 10^3$, the dataset size is $\approx 50k$). If it is indeed trivial that for these dimensions & dataset size, u.c. would fail, then it follows immediately that the u.c. bounds proposed for these settings by the dozens of post-Zhang-et-al. papers, are all *obviously* pointless – which is clearly not the case. **Finally**, we'd also like to draw R3's attention to the empirical contributions in Sec. 2 that have not been mentioned in the review. These are new and constitute half the paper. Keeping these facts in mind, we politely request a fair and complete re-evaluation of the paper.

R4: We thank the reviewer for their suggestions, and also for stating their main concern precisely. R4 believes our examples are somewhat a strawman because they show failure of uniform convergence (u.c.) only in cases without explicit regularization. We must strongly emphasize that **this is not a strawman: our unregularized setting is in fact, the precise setting that has been the focus of Zhang et al., '17 and all the other dozens of follow-up work.** The key surprising phenomenon in deep learning that has gained significant theoretical interest is the fact that deep networks generalize *even when there's no explicit control*, either on the parameter count or on the norms – the lack of regularization is pivotal to the "surprise" here. To this end, post-Zhang et al., works developed u.c. bounds with the goal of explaining generalization in this *unregularized* setting – and this goal has been elusive. Our work is a warning that this particular active, ongoing pursuit may after all be a futile exercise unless we go beyond u.c. Finally, indeed, the reviewer is right in noting that u.c. may still hold in other settings (with regularization, compression, SRM etc.,). These settings, however, are somewhat orthogonal to the main generalization puzzle (and we make no claims about these settings). To conclude, we hope our response explains why our examples are certainly not a strawman in the context of current deep learning theory research, and thus, we hope this helps in re-evaluating the paper.

R5: We thank R5 for their careful reading & thorough summary. The main concern of R5 is that, while we claim deep networks "do not suffer from pseudo-overfitting" (by which we mean, the gap between the mean test and training margins of deep networks does decrease with training data size), it seems that our examples do suffer from pseudo-overfitting. Hence, R5 is wary of the relevance of our examples to deep learning. This is an interesting point, and we argue why this is actually not of concern. And we will certainly add the following discussion to the paper.



First, in our hypersphere example, as shown in the accompanied figure, the mean margins on the test data (orange line) and on training data (blue line) do converge to each other with more training data size m i.e., **the gap in the mean test and training margins (green line) does decrease with** m. Thus our setup exhibits a behavior similar to deep networks on MNIST in Figure 11 in our paper. As noted in lines 565-570, since the rate of decrease of the mean margin gap in MNIST is not as large as the decrease in test error itself, there should be "a small amount" of psuedo-overfitting in MNIST. The same holds in this setting, although, here we observe an even milder decrease, implying a larger amount of pseudo-overfitting. (Nevertheless, uniform convergence cannot capture even this decrease with m.) **Secondly**, we must note that here we train an actual ReLU-based network using vanilla SGD just like in the MNIST example. Hence, if any bound claims to "explain generalization in deep learning", it *should* explain generalization in our example – and we establish that uniform convergence bounds cannot do the trick. This makes our example relevant to deep learning regardless of pseudo-overfitting. **Third**, the reviewer is certainly right in that our linear example does suffer from pseudo-overfitting. However, we must emphasize that this psuedo-overfitting in itself does not imply Theorem 3.1's lower bound on $\epsilon_{\text{unif-alg}}$ as noted in lines 565-570 (as pseudo-overfitting implies a lower bound only on a specific class of uniform convergence bounds). We do think that pseudo-overfitting is a phenomenon worth exploring better; however, we also believe that there is a phenomenon beyond pseudo-overfitting that is at play in deep learning, which our examples elucidate. We hope these three points inform the reviewer as to why pseudo-overfitting is not of concern while drawing implications from our examples.