To Reviewer 1:

We thank the reviewer for providing constructive feedback and suggestions. Also, in the final version we will fix the typos and make the minor improvements the reviewer suggests.

On the assumptions. As you also noted, the source condition and the eigenvalue decay are fairly standard in the nonparametric regression setting. By now, the number of papers (and books!) using these two parametrizations is quite large. So, for example, the eigenvalue decay of the Gaussian Kernel is well-known (see the very recent results in Belkin, COLT'18 and references therein). In particular, as briefly explained in Section 2, the source condition and the eigenvalue decay parametrize the difficulty of the problem and the "finiteness" of the space. Hence, our rate improves the known rates for "difficult" problems with low effective dimension. However, given that our result is based on a minor modification of regularized least square, it is likely that the previous analyses were loose, not that our algorithm is intrinsically better! In this view, the specific regime in which our rates are better is not really important. On the other hand, in our opinion, closing the gap between upper and lower bounds and pointing out possible major problems in previous work through a completely novel analysis are major contributions.

On "strong" assumptions. Our final comments on the bias of the community towards "weak assumptions" was exactly to provoke a discussion in this sense and a better judgment on these issues, rather than justifying our results. So, we are happy that the reviewer engaged with us in this discussion! In this view, we abstain from judging how "strong" is the case of zero Bayes error w.r.t. the square loss: It is completely a problem-dependent judgment rather than a universal one. Instead, we just consider it an interesting setting that researchers have ignored for a long time. Moreover, we do plan to extend the results we presented to smooth classification losses, as the squared hinge loss. In that setting, the same results are expected through an Online-Newton-Step analysis. Indeed, the work in Orabona (2014) already shows an acceleration for zero Bayes error for any smooth and Lipschitz loss, (even classification ones like the smoothed hinge loss), but the acceleration appears inferior to the one we can show for the square loss. So, we believe this is an interesting area to explore.

To Reviewer 2:

We thank the reviewer for providing constructive feedback. We will improve accordingly in the final version.

To Reviewer 3:

We thank the reviewer for raising interesting questions and suggestions.

On the lower bound. We actually believe that the lower bound is known and matching our upper bound. However, we did use the wrong citation, thanks for pointing it out! In particular, the lower bound is widely discussed in Section 4.2 of Pillaud-Vivien et al. (2018), that in turn is based on the theorems in

S. Fischer and I. Steinwart. Sobolev norm learning rates for regularized least-squares algorithm. Fakultät für Mathematik und Physik, Universität Stuttgart, 2017.

We will make it clear in the final version.

On the experiments. We are *not* doing the full sample version, and we do plot the *expected* value of the risk, as written on the y axis. So, we exactly compute the expectation with respect to the randomization of the algorithm using k from 0 to $n-1$, while we estimate the test error for each k using a finite test set. We will make this clear in the final version.

On the effect of randomization. We have not performed a thorough empirical comparison with the full sample version, and we are not sure of the exact effect of randomization (besides the fact that it gives us a way to obtain a good theoretical bound). You have raised an interesting question, which would be an interesting future work. We believe this is quite nontrivial for the following reason: If the error rate of the kernel ridge regression is monotonic with the data size, then the randomization would necessarily harm the prediction error. However, in general it turns out the error rate can be non-monotonic by a recent study by Viering et al., Open Problem: Monotonicity of Learning, COLT, 2019 (see example III therein). Specifically, the error rate of ridge regression can even *increase* with the data size in some regime.