We thank the reviewers for their careful reading of the manuscript, and for their constructive comments. We feel that the reviews are largely positive. In the remainder, we want to address some of the issues raised, and we will address them in detail in the revision:

Reviewer 1: "Could you comment on the possibility to generalize these results, for example, to regularized LS, LASSO, elastic net regularization, TLS, etc.?" *Generalizing to ridge regression is possible using random matrix theory, but very nontrivial, and in fact we are currently pursuing this. In our view, this would be a different project. Generalizing to Lasso, elastic net, TLS seems like a much more difficult problem. It may be possible to handle lasso using approximate message passing (AMP), but this would be a new project.*

"In the simulation experiment the input matrix X should be studied for other distributions." *We have simulations with correlated t-distributed data in Appendix A.14*

"Also a finite-time analysis would be more convincing." *For Gaussian and iid projections, it may be possible to obtain convergence rates using known results on the convergence rate of Stieltjes transforms. However, for Hadamard transforms, the only results we are aware of are asymptotic, as they are based on free probability theory. Thus finite-sample results may be hard.*

Reviewer 2: "'we get more accurate results for the performance of sketching' – what exactly does this mean? How is accuracy measured? *It means that our results are more accurate in simulations, and "get the right constant". See Appendix 16.*

Maybe this is common in statistics but why is it reasonable to assume that p and n would grow together with the aspect ratio converging? Why do we inherently expect more parameters with more data points? *This is actually just a model for "large n, large p". We do not really think that the number of parameters is growing.*

Table 1 is very hard to read. Not sure if lines are divisions or line breaks. Many entries are missing and alignments makes it impossible to tell which values are supposed to fall in which column. Variables in leverage score line are undefined. *We have made the table easier to read: add separators, copy values in multiple columns.*

What talking about comparison to Raskutti and Mahoney in related work what sketching method does the 1 + 44p/r term apply to? Also what are the stronger assumptions? I don't see any nontrivial assumptions listed for the random projection case in the table or in Theorem 2.1. *It refers to their subsampling and subgaussian projection results. The stronger conditions refer to subsampling, when we need ortho-invariance*

In Theorem 2.3 what does it mean that X's esd converges if X is just a matrix? Doe you mean that there is a function mapping n and p to deterministic matrices and the esps of these matrices converge as n and p increase? *We mean the esd of* $X^T X$

Line 233: what does it mean for a matrix to be orthogonally invariant? *This is the same as rotationally invariant, defined in lines 239-242*

The rotationally invariant assumption seems super strong. This isn't even the case when each row is a p-dimensional Gaussian with some fixed covariance Sigma right? *Indeed, this condition is quite strong and does not hold for correlated Gaussians. However, it seems that the current proof technique (asymptotic freeness, Theorem 4.3.11 of Hiai and Petz (2006)) requires it*

Line 255: what does it mean it doesn't 'introduce enough randomness?' How is this measured? *This is an intuitive claim and we do not know how to measure it*

I don't see how the line in 280 gives leverage score sampling. Where does the eta-transform come up? Would be helpful to explain a bit more. *We will explain more, but the eta-transform is the limit of the leverage score*

Although would be a lot more convincing if more than just a single dataset were tested. *Figure 2 has two datasets*

What about at least X with Gaussian rows with a non identity covariance? *We have simulations with correlated t-distributed data in Appendix A.14, "Simulation for nonuniform data"*

Was leverage score sampling just not tested because you couldn't compute the theoretical bounds in closed form? *See Appendix 13 for simulations with leverage scores*

Under Prop A.1, what is eps_t? Under Proof of Theorem 2.1 $tr((X^T X)^{-}1)$ shouldn't have an expectation around it. X is fixed here right? eps_t *is test noise. Indeed no expectation*

"I think the biggest thing is to push forward the proof techniques and give some intuition. Too much of the paper is spent just describing the results and explaining why they are good. More should be spent actually explaining the results." *We will work on explaining the results and proof techniques, and giving intuition.*