Reviewer #1 Thanks for your comments; we'll clarify our usage of the following terms in the revised paper.

- We use **"consistency"** to mean $\mathbb{E}[L_{\mathcal{D}}(\hat{w}) - L_{\mathcal{D}}(w^*)] \to 0$. Traditionally, as in [27], this limit means $n \to \infty$ for a fixed problem, but in that setting linear models do not interpolate. Instead, for asymptotic interpolation we study a sequence of distributions changing with n, with the noise magnitude λ_n possibly increasing. In a more typical "high-dimensional" regime, p would also increase with n, e.g. $p = \gamma n$ in [13]; we instead take $p \to \infty$ for each n.
- By **"interpolation learning"** we mean achieving "good" $L_{\mathcal{D}}(w)$ while $L_S(w) = 0$ in a noisy, non-realizable setting.

Reviewer #2 Thanks for your feedback; we'll add more intuition, details, and reorganize proofs in revision.

- **Min-risk interpolator:** Thm. 4.5 decomposes as risk of one interpolator, plus gap to worst; $\hat{w}_{MR}$ minimizes risk.
- **Restricted eigenvalue:** It arises naturally from (7)'s dual; it measures how of Σ is unobserved by X, and is the generalization gap for $y = 0_n$, $B = 1$. It also relates to [3]: the "malignant" covariance I_p has $\kappa_X(I_p) \overset{a.s.}{=} 1$, while the benign covariance of Setting B has $\kappa_X(\Sigma) \approx \lambda_n / n \to 0$. We expect it might play the role of ξ_n in $(\star)$.
- **Finite degrees of freedom:** It is true that Setting B is simple in this way. Our approach also allows for analysis where d_S increases with n, though we know it must be $o(n)$ for consistency to be possible.
- **Consistency $\to$ 1-sided unif. conv.:** Take $\mathcal{S}_{n,\delta} = \{(X, y) : L_{\mathcal{D}}(A(X, y)) \leq \sigma^2 + \epsilon_{n,\delta}\}$. (We'll clarify footnote 5.)

Reviewer #3 Thanks for your writing suggestions (converting some discussion into lemmas, substantially re-focusing the abstract, and clarifying e.g. line 230), which we agree will improve the presentation.

- **Portable insights:** The main takeaway we believe to be broadly relevant is that when analyzing using "uniform convergence," especially in the context of interpolation learning, it is important to use "relative" or "optimistic" bounds which take L_S into account. Our approach of bounding the generalization gap via duality may also be widely applicable: even in complex settings without strong duality, upper bounds should still be possible from weak duality. We will emphasize these more throughout the paper.
- **Comparison to [3]/[23]:** While prior work almost fully characterizes consistency in this class of problems, it is quite different from most existing work in statistical learning theory. Our theorem 4.5 attempts to be more like popular Rademacher bounds, although to develop this connection further (and compare with existing conditions), more calculations are required in general – even if the speculative bound $(\star)$ holds. We'll increase our discussion of the relationship to the benign/weakly benign conditions, e.g. with the examples above. Our approach also explains non-minimal-norm predictors, and it may be easier to numerically check $\kappa_X(\Sigma)$ and $\|X\|$ in practice.
- **1- vs 2-sided uniform convergence:** For predictors with $L_S(w) = 0$, these modes of convergence are indeed identical. These restricted uniform bounds sidestep entirely the two-sided failure mode of Section 3.2, with high L_S but low $L_{\mathcal{D}}$. This is not the only difference between the standard and restricted settings, however: we strongly expect that norm balls do not exhibit one-sided uniform convergence either (line 125), due to cases where L_S is large but $L_{\mathcal{D}}$ is even larger. We will add more discussion of this relationship in the revision.
- **Restricted eigenvalue under interpolation:** We are not aware of any previous use of $\kappa_X(\Sigma)$ in the literature.
- **Is low norm key?** As any low-norm interpolator generalizes, we believe we've shown that the answer to this question is "yes." We agree that this belongs in the abstract and should be highlighted more in the paper body as well.
- **Restricted convergence bounds:** As we mention around line 177, bounds like (7) are very standard in realizable PAC analyses. Generally, (7) will *always* be small for consistent predictors – even if, as in Section 3, unrestricted bounds fail – because taking $B = \|\hat{w}_{MN}\|$ makes (7) just $L_{\mathcal{D}}(\hat{w}_{MN})$. The questions are whether we can usefully bound the analogue of (7), and how large B can be; we answer these questions for Setting B in Section 4.1.
- **DCT in footnote 3:** Since $L_{\mathcal{D}}$ is also an expectation, there are two interchanges of limit and expectation, and finding a dominator seems nontrivial; it seems to essentially boil down to the proof of Proposition 4.6.

Reviewer #4 Thank you for your questions and suggestions, which we will emphasize in revision.

- **Connection to deep learning:** High-dimensional linear models can serve as a simpler test bed to help develop methods useful for deep nets: we need to walk before we can run. Knowing which techniques can explain many of the surprising phenomena of deep learning, e.g. double descent, in linear models, helps us narrow down which tools to try in the harder setting. (See also the *portable insights* comment to Reviewer 3.)
- **Why uniform convergence?** (1) It is in many ways *the* standard toolkit in statistical learning theory. (2) A direct bound on $L_{\mathcal{D}}(\hat{w}_{MN})$ may not tell us *why* $\hat{w}_{MN}$ works; a uniform bound based on norm strongly indicates norm is the "reason." (3) In practice we may not find the exactly minimal-norm interpolator; uniform bounds are more "robust."
- **Restricted problem setting:** Indeed, Theorem 4.1 is limited to a very particular setting, but we use it mainly to demonstrate the success of our style of analysis. We emphasize that Theorem 4.5 holds quite generally.
- **Comparison to LASSO:** Here, we simply make the point that in a sparse setting, there exists a consistent learning rule, but *no* interpolation method – including the minimal l_1 norm interpolator – can be consistent for $p = \mathcal{O}(n)$.
- **Related statistics papers:** If you have any particular work in mind, we are eager to consider it.