**Motivation**

In high dimensions, models that interpolate noisy training data can still generalize well [1]. How come?

"Benign overfitting" explanation [2]: min-\(\ell_2\)-norm interpolation is consistent when covariates are effectively low-dimensional, i.e. \(d_{\text{eff}} = \text{tr}(\Sigma)/\|\Sigma\|_{\text{op}} < n\).

- What about effectively high-dimensional covariates \(d_{\text{eff}} = d \gg n\)?
- What about other interpolating models?

Can we consistently learn sparse ground truths with minimum-norm interpolators on high-dimensional features?

This work: YES for isotropic covariates \(x \sim \mathcal{N}(0, I_d)\), sparse ground truth \(\|w^*\|_0 \leq O(n)\), and min-\(\ell_1\)-norm interpolation.

**Previous Results**

Min-\(\ell_1\)-norm interpolation (Basis Pursuit) in our setting was known to

- achieve consistency for zero noise \(\sigma = 0\);
- have statistical rate \(\|\hat{w} - w^*\|^2 \leq O(\sigma^2)\) as \(d/n \to \infty\) [3];
- have statistical rate \(\|\hat{w} - w^*\|^2 \geq \Omega\left(\frac{\sigma^2}{\log(d/n)}\right)\) [4].

We close the gap between upper and lower bound, showing \(\|\hat{w} - w^*\|^2 \sim \frac{\sigma^2}{\log(d/n)}\). In particular, Basis Pursuit is consistent even in the presence of noise.

**Remark.** In practice, \(\ell_1\)-norm penalization (LASSO) is preferable to interpolation when noise is present.

**Main Result**

**Problem setting:**

- Data model: covariates \(x \sim \mathcal{N}(0, I_d)\), noisy observations \(y = (w^*, x) + \xi\) where \(\xi \sim \mathcal{N}(0, \sigma^2)\).
- Prediction error \(\mathbb{E}_{x,y}(\hat{w}, x) - y)^2 = \|\hat{w} - w^*\|^2 + \sigma^2\).
- We study the min-\(\ell_1\)-norm interpolator defined by
  \[
  \hat{w} = \argmin_w \|w\|_1 \text{ such that } \forall i, \langle x_i, w \rangle = y_i.
  \]

**Main result:** Non-asymptotic matching upper and lower bounds for prediction error of min-\(\ell_1\)-norm interpolator.

**Theorem.** Suppose \(\|w^*\|_0 \leq \kappa_1 \frac{n}{\log(d/n)}\) for some constant \(\kappa_1 > 0\). There exist constants \(\kappa_2, \kappa_3, \kappa_4, c_1, c_2, c_3 \geq 0\) such that, if \(n \geq \kappa_2\) and \(\kappa_3 n \log(n) \leq d \leq \exp(\kappa_4 n^{1/6})\),

\[
\frac{\|\hat{w} - w^*\|_2^2 - \sigma^2}{\log(d/n)} \leq c_1 \frac{\sigma^2}{\log(d/n)^{3/2}}
\]

with probability \(\geq 1 - c_2 \exp\left(-\frac{n}{\log(d/n)}\right) - d \exp(-c_3 n)\).

**Experimental Validation**

Dashed curve: theoretical rate
Orange squares: experimental rate for Normal-distributed features (our setting)
Conjecture: min-\(\ell_1\)-norm interpolation also has statistical rate \(\frac{\sigma^2}{\log(d/n)}\) for certain heavy-tailed feature distributions.

**Comparison to Min-\(\ell_2\) Interpolation**

Min-\(\ell_1\)-norm interpolation is sensitive to the noise level \(\sigma^2\); min-\(\ell_2\)-norm interpolation has similar (non-vanishing) prediction error across all values of \(\sigma^2\).

Trade-off between structural bias vs. sensitivity to noise:

- Min-\(\ell_1\)-norm interpolation (squares):
  - ✓ strong structural bias,
  - ✓ efficient noiseless recovery of sparse signals,
  - ✗ but poor rate in the presence of noise.
- Min-\(\ell_2\)-norm interpolation (diamonds):
  - ✗ no structural bias (except towards zero),
  - ✗ fails to recover any non-zero signal even in the absence of noise,
  - ✓ but does not suffer from overfitting of the noise.

**References**