

Reading group discussion of Wainwright's “Sharp thresholds”

Dean P Foster

February 1, 2026

1 Review of regression

Facts one should know about doing regression:

- Fact: Adding p variables to a n observation regression will generate a perfect fit.
- Heuristic: Adding one variable will improve R-squared by $1/n$ on average.
- Fact: $\Phi^{-1}(1/p) \approx \sqrt{2\log p}$
- Fact: $\Delta R^2 = t^2/n$
- Heuristic: Best feature out of p improves R^2 by $2\log(p)/n$.
- Fact: Random vectors in \mathfrak{R}^n are typically orthogonal
- Heuristic: One can “hope” for approximate orthogonality up to $p \approx 2^n$. Clearly exact orthogonality is only possible up to size n .

2 Implications

- If the model we want has dimension k which is bigger than $O(n/\log(p))$ variables, there will be a random model that fits better with fewer variables.
- This is what is called a threshold.
- Needing to search might make this worse.

3 Now on to Wainwright's paper

4 Introduction: L0 vs L1

- Goal: low dimensional recovery
- Solution: good fit with few's variables
- Problem: NP hard
- Approximation: Relax to L1
- This is equation 3 and 4

5 previous work

- equation 5: amazingly, regression heuristics help here
- If you can get close using random vectors, you can make it exact by adding in n more small vectors

6 Gaussian model

- $X_{ij} \sim N(0, 1)$
- good way of generating almost orthogonal vectors

7 Our contribution

- equation 6: crude recovery of the β 's not just the subspace.

8 Figure 1

- Note: it gets sharper and sharper on rescaled axis. So only the “mean” is right—not the “SD”.

9 Section II: Primal-Dual witness construction

Constructs properties around the optimum point. We have enough statistics to cover—that I’m going to skip the optimization.

10 III: Deterministic designs

Equation 14a, and 15:

- Recall $\hat{\beta} = (X'X)^{-1}X'Y$ for least squares regression
- equation 15 then says: Every regression equation to predict every left out variable has all its beta's less than 1.

- Totally weird as far as L2 matrix concepts go.
- Problem: what if incoherence parameter $\gamma < 0$:
 - the shadow larger than the object (Think starwars 1).
 - $X_{fake} = 2 * X_{real}$ doesn't work since now fake is same as real
 - $X_{fake} = 2 * X_{real} + noise$ – this shadow is dangerous. It moves things closer at a lower β cost

Theorem 1:

- equation 15 keeps the shadows away
- equation 16 keeps the truth from being too colinear
- equation 17 keeps the impostors out of the equation

Then statement (a) says we don't over fit.

equation (b) requires a strong signal—and it then says we don't under fit.

11 Necessary conditions

Theorem 2:

- equation 19 says we have a long shadow for a beta we care about
- equation 20 then says we will use that beta instead.

12 Skip to bottom of 2192

Typical Oracle results have a $\log(p)$ term in them.

This holds here.

13 Skip to figure 2

same shape. same ideas. Notice, steeper as sample sizes increases. So he only is capturing the cross point—not the shape of the transition.