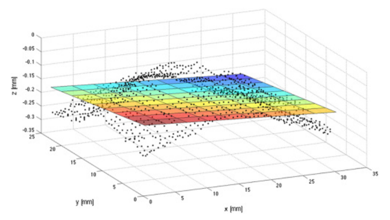
High Dimensional Linear Regression with Binary Coefficients.

Mean Squared Error and Phase Transitions

Ilias Zadik & David Gamarnik

Massachussets Insititute of Technology (MIT)

The problem of recovering the sparsity pattern of an unknown vector β^* based on noisy observations arises in a broad variety of contexts including subset selection in regression, structure estimation in graphical models and signal denoising.



Our Model

Setup: Let $\beta^* \in \{0,1\}^p$ be a **binary** k**-sparse** vector. For

- $X \in \mathbb{R}^{n \times p}$ consisting of entries i.i.d $N(0,1)$ **random variables**
- $W \in \mathbb{R}^n$ consisting of entries i.i.d. $N(0,\sigma^2)$ **random variables** with $\sigma^2 = o(k)$

we get n noisy linear samples of β^*, $Y \in \mathbb{R}^n$, given by,

$$Y := X\beta^* + W.$$

Goal: Given (Y, X), recover **w.h.p.** β^* with the minimum n possible.

Why binary?

- Discrete structure $\Rightarrow$ easier to analyze.
- Keeps the challenge of **support recovery** (highly nontrivial)
- Best known information theoretic lower bound is **much smaller** than the best known algorithmic upper bound.

Literature Review

- Best known **positive** results
  (e.g. [Donoho '06],[Wainwright '09]) If
  $$n > 2k \log p$$
  many efficient algorithms (including LASSO) recover exactly β^* w.h.p.
- Best known **negative** result
  ([Wang et al '10]) If
  $$n < n^* := \frac{2k}{\log\left(\frac{2k}{\sigma^2}+1\right)} \log p,$$
  then there is no recovery mechanism of β^* which succeeds w.h.p.

Main Question

There is a **gap** in the literature when $n^* < n < 2k \log p$.
Is there **enough information**/ **efficient algorithms** to recover β^* in this regime?

Maximum Likelihood Estimator - All or Nothing result

It has a **simple-to-state form:** the MLE $\hat{\beta}$ is the optimal solution of

$$(\Phi_2) \quad \min_{\beta \in \{0,1\}^p, \sum_{i=1}^p \beta_i = k} ||Y - X\beta||_2.$$

Definition 1 *For $\beta \in \{0,1\}^p$, k-sparse we define*

$$\text{Overlap}(\beta) := |\text{Support}(\beta^*) \cap \text{Support}(\beta)|.$$

Theorem 1 ("All or nothing") *Set $n^* := \frac{2k}{\log\left(\frac{2k}{\sigma^2}+1\right)} \log p$ and let $\epsilon > 0$ be arbitrary.*

- *If $n < (1-\epsilon) n^*$, then w.h.p. $\frac{1}{k}\text{Overlap}(\hat{\beta}) \to 0$, as $n, p, k \to +\infty$.*
- *If $n > (1+\epsilon) n^*$, then w.h.p. $\frac{1}{k}\text{Overlap}(\hat{\beta}) \to 1$, as $n, p, k \to +\infty$.*

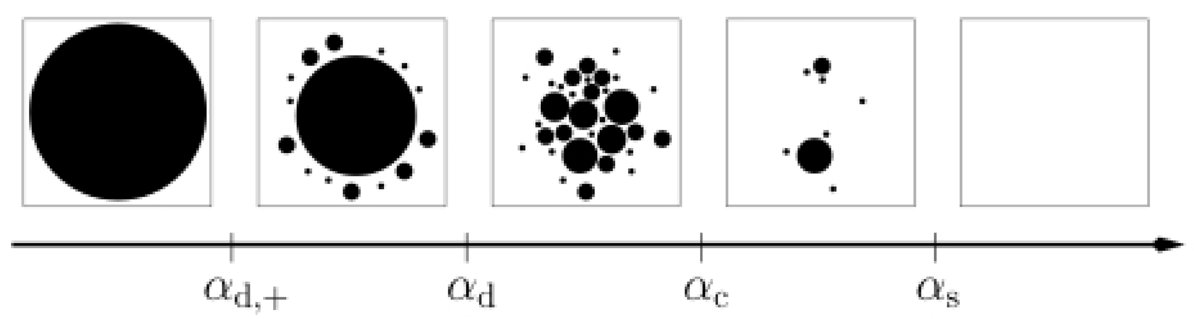
So, when $n > n^*$ **information exists** and n^* is a **sharp phase transition point**.

Algorithmic Difficulty

Why all known efficient algorithms seem to fail when $n^* < n < 2k \log p$ and work only if $n > 2k \log p$?

The picture from the analysis of randoms CSPs and spin glass theory suggests that a usual reason is an **"important change in the geometry of the space of solutions"** between the two regimes.[Achlioptas et al, 2008]

Such a geometrical property has been established for many problems such as *random k-SAT, k-coloring of a random graph, maximum independent set in a sparse random graph* and many others. (Figure below by [Krzakala et al' 07])



Overlap Gap Property in Linear Regression

We prove a geometrical property for the near-optimal feasible solutions of the problem (Φ_2). We call the property **Overlap Gap Property (OGP)** for high-dimensional linear regression. For $r > 0$, set

$$S_r := \{\beta \in \{0,1\}^p : ||\beta||_0 = k, n^{-\frac{1}{2}}||Y - X\beta||_2 < r\}.$$

Definition 2 (The Overlap Gap Property) *Let $r > 0$ and $0 < \zeta_1 < \zeta_2 < 1$. We say that the high-dimensional linear regression problem defined by (X, W, β^*) satisfies the Overlap Gap Property with parameters (r, ζ_1, ζ_2) if the following holds.*

(a) For every $\beta \in S_r$,

$$\frac{1}{k}\text{Overlap}(\beta) < \zeta_1 \text{ or } \frac{1}{k}\text{Overlap}(\beta) > \zeta_2.$$

(b) Both the sets

$$S_r \cap \{\beta : \frac{1}{k}\text{Overlap}(\beta) < \zeta_1\} \text{ and } S_r \cap \{\beta : \frac{1}{k}\text{Overlap}(\beta) > \zeta_2\}$$

are non-empty.

Intuitively, this means that the set of $\beta' s$ with closed to optimum objective value in (Φ_2) **"shatters" in two components**, one with **low** overlap size with the ground truth β^* and one with **high** overlap size with β^*.

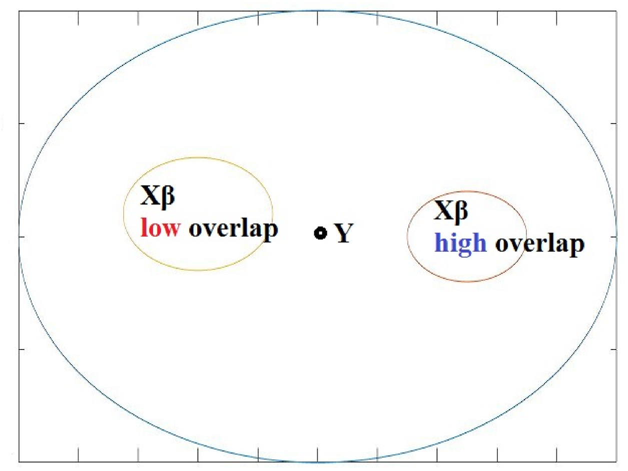


Figure 4: *The OGP around Y*

Theorem 2 *Suppose the assumptions of Theorem 1 hold. There exists $C > c > 0$ with the following properties.*

- *If $n^* < n < ck \log p$ then there exists $0 < \zeta_1 < \zeta_2 < 1$ and a sequence $r_k > 0$ such that w.h.p. as k increases the high-dimensional problem defined by our model* ***satisfies*** *the Overlap Gap Property with parameters (r, ζ_1, ζ_2).*
- *If $n > Ck \log p$ then for any $0 < \zeta_1 < \zeta_2 < 1$ and any sequence $r_k > 0$ w.h.p. as k increases the high-dimensional problem defined by our model* ***does not satisfy*** *the Overlap Gap Property with parameters (r, ζ_1, ζ_2).*

Corollary 1 *(Informal) If $n < ck \log p$ then any "successful" local search algorithm needs in the worst case to increase the distance from Y in at least one step.*

Summary

- We **positively answer** the question of whether information for recovering β^* exists when $n > n^*$.
- We establish a certain Overlap Gap Property(OGP) in the space of binary k-sparse vectors when $n < ck \log p$. We conjecture that OGP is **the source of algorithmic hardness** of the problem when $n^* < n < 2k \log p$.

Bibliography

1. D. Achlioptas and A. Coja-Oghlan. Algorithmic barriers from phase transi- tions. In Foundations of Computer Science, 2008.

3. D. L Donoho. Compressed sensing. IEEE Transactions on information theory, 2006

4. F. Krzakała, A. Montanari, F. Ricci-Tersenghi, G. Semerjian, and L. Zdeborová, Gibbs states and the set of solutions of random constraint satisfaction problems. PNAS, 2007

4. M. J Wainwright. Sharp thresholds for high-dimensional and noisy sparsity recovery using constrained quadratic programming (lasso), 2009.

5. W. Wang, M. J Wainwright, and Kannan Ramchandran. Information-theoretic limits on sparse signal recovery: Dense versus sparse measurement matrices, 2010.