Computational and Statistical Challenges in High Dimensional Statistical Models

Ilias Zadik

Operations Research Center (ORC), MIT

PhD Thesis Defense

Commitee: David Gamarnik (PhD advisor), Guy Bresler, Lester Mackey

June 12, 2019

Introduction- Big Data Challenges

Over the recent years, the **number** and **magnitude** of available **datasets** have been growing **enormously**.

Introduction- Big Data Challenges

Over the recent years, the **number** and **magnitude** of available **datasets** have been growing **enormously**.

Big impact across science:
From **artificial intelligence** to **economics** to **medicine** and many others.

Introduction- Big Data Challenges

Over the recent years, the **number** and **magnitude** of available **datasets** have been growing **enormously**.

Big impact across science:
From **artificial intelligence** to **economics** to **medicine** and many others.

Required novel **statistical and computational tools** on dealing with issues such as high dimensionality.

Introduction- Big Data Challenges

Over the recent years, the **number** and **magnitude** of available **datasets** have been growing **enormously**.

Big impact across science:
From **artificial intelligence** to **economics** to **medicine** and many others.

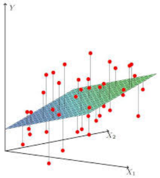
Required novel **statistical and computational tools** on dealing with issues such as high dimensionality.

Many open challenging theoretical questions
even for *simple high dimensional statistical models!*

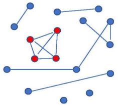
Thesis Overview: The Models

Two Long-Studied Stylized High Dimensional Models:

(1) *High Dimensional Linear Regression Model (HDLR), [Tibshirani '96]*
Recover vector of coefficients from few noisy linear samples.
Motivation: Fit linear models in high dimensional data.

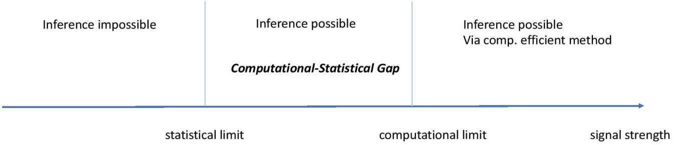


(2) *Planted Clique Model (PC) [Jerrum '92]*
Recover planted clique from a large observed network.
Motivation: Community detection in large networks.



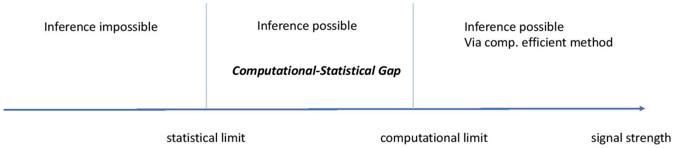
Thesis Overview: Contributions

HDLR (signal strength = sample size), PC (signal strength = clique size):



Thesis Overview: Contributions

HDLR (signal strength = sample size), PC (signal strength = clique size):



Under assumptions,

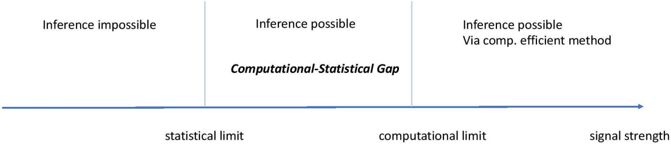
- Compute the exact **statistical limit** of the HDLR model *("All-to-Nothing Phase Transition")*
- Explain **computational-statistical gaps** of HDLR and PC models, through *statistical-physics* based methods. *("Overlap Gap Property")*
- Improved **computational limit** for **noiseless** HDLR model using *lattice basis reduction ("One Sample Suffices")*

Papers:

(Gamarnik, **Z.** *COLT* '17, *AOS* (major rev.) '18+)
(Gamarnik, **Z.** *AOS* (major rev.) '18+), (Gamarnik, **Z.** *NeurIPS* '18)
(Reeves, Xu, **Z.** COLT '19), (Gamarnik, **Z.** '19+)

Thesis Overview: Contributions

HDLR (signal strength $=$ sample size), PC (signal strength $=$ clique size):



Under assumptions,

- Compute the exact **statistical limit** of the HDLR model *("All-to-Nothing Phase Transition")*
- Explain **computational-statistical gaps** of HDLR and PC models, through *statistical-physics* based methods. *("Overlap Gap Property")*
- Improved **computational limit** for **noiseless** HDLR model using *lattice basis reduction ("One Sample Suffices")*

Papers:

(Gamarnik, **Z.** *COLT* '17, *AOS* (major rev.) '18+)
(Gamarnik, **Z.** *AOS* (major rev.) '18+), (Gamarnik, **Z.** *NeurIPS* '18)
(Reeves, Xu, **Z.** COLT '19), (Gamarnik, **Z.** '19+)

Outline of the Talk

(1) Introduction and Thesis Overview

(2) High Dimensional Linear Regression Model

- Background
- Statistical Limit: All-or-Nothing Phenomenon
- Computational-Statistical Gap and Overlap Gap Property

(3) Planted Clique Model and Overlap Gap Property

Outline of the Talk

(1) Introduction and Thesis Overview

(2) **High Dimensional Linear Regression Model**

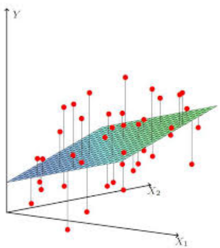
- **Background**
- Statistical Limit: All-or-Nothing Phenomenon
- Computational-Statistical Gap and Overlap Gap Property

(3) Planted Clique Model and Overlap Gap Property

High Dimensional Linear Regression

Let (unknown) $\beta^* \in \mathbb{R}^p$. p number of features.
For **data matrix** $X \in \mathbb{R}^{n \times p}$, and **noise** $W \in \mathbb{R}^n$,
observe n noisy linear samples of β^*, $Y = X\beta^* + W$.

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

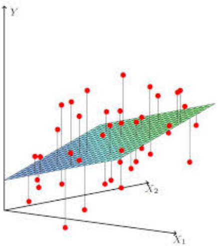


High-dimensional regime: $n \ll p, p \to +\infty$.

High Dimensional Linear Regression

Let (unknown) $\beta^* \in \mathbb{R}^p$. p number of features.
For **data matrix** $X \in \mathbb{R}^{n \times p}$, and **noise** $W \in \mathbb{R}^n$,
observe n noisy linear samples of β^*, $Y = X\beta^* + W$.

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



High-dimensional regime: $n \ll p, p \to +\infty$.
$n < p$ implies assumptions on β^* are necessary.
Reason: even if $W = 0$, $Y = X\beta^*$ underdetermined.

Assumptions on β^* and X, W

Assumptions on β^*:

(1) β^* is k-**sparse**: k non-zero coordinates, $k/p \to 0$, as $p \to +\infty$.
(A lot of research, e.g. *Compressed Sensing, Genomics, MRI*.)

(2) β^* is **binary valued**: $\beta^* \in \{0, 1\}^p$. (†)

(†) (non-trivial) *simplification* of **well-studied** $\beta^*_{\min} := \min_{\beta^*_i \neq 0} |\beta^*_i| = \Theta(1) > 0$ and *support recovery task.*

Assumptions on β^* and X, W

Assumptions on β^*:

(1) β^* is k-**sparse**: k non-zero coordinates, $k/p \to 0$, as $p \to +\infty$.
(A lot of research, e.g. *Compressed Sensing, Genomics, MRI*.)

(2) β^* is **binary valued**: $\beta^* \in \{0,1\}^p$. (†)

Distributional Assumptions on X, W:

(1) $X \in \mathbb{R}^{n \times p}$ has i.i.d. $\mathcal{N}(0,1)$ entries.

(2) $W \in \mathbb{R}^n$ has i.i.d. $\mathcal{N}(0,\sigma^2)$ entries.

(†) (non-trivial) *simplification* of **well-studied** $\beta^*_{\min} := \min_{\beta^*_i \neq 0} |\beta^*_i| = \Theta(1) > 0$ and *support recovery task.*

The Model

Setup

Let $\beta^* \in \{0,1\}^p$ be a **binary** k-**sparse** vector, $k/p \to 0$, as $p \to +\infty$. For

- $X \in \mathbb{R}^{n \times p}$ consisting of i.i.d $\mathcal{N}(0,1)$ entries
- $W \in \mathbb{R}^n$ consisting of i.i.d. $\mathcal{N}(0,\sigma^2)$ entries

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

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

The Model

Setup

Let $\beta^* \in \{0,1\}^p$ be a **binary** k-**sparse** vector, $k/p \to 0$, as $p \to +\infty$. For

- $X \in \mathbb{R}^{n \times p}$ consisting of i.i.d $\mathcal{N}(0,1)$ entries
- $W \in \mathbb{R}^n$ consisting of i.i.d. $\mathcal{N}(0,\sigma^2)$ entries

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

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

Goal: Statistical and Computational Limit

Minimum n so that given (Y, X), β^* is **(efficiently) recoverable** with probability tending to 1 as $n, k, p \to +\infty$ **(w.h.p.)**.

Rise of a Computational-Statistical Gap

Computational Results ([Wainwright '09],[Fletcher et al '11])

Set $n_{\text{alg}} = 2k \log p$. Assume $\text{SNR} = \frac{k}{\sigma^2} \to +\infty$.
If

$$n > (1+\epsilon) n_{\text{alg}}$$

LASSO *(convex relaxation)* and OMP *(greedy algorithm)* succeed w.h.p.

Rise of a Computational-Statistical Gap

Computational Results ([Wainwright '09],[Fletcher et al '11])

Set $n_{alg} = 2k \log p$. Assume $\mathrm{SNR} = \frac{k}{\sigma^2} \to +\infty$.
If

$$n > (1+\epsilon) n_{alg}$$

LASSO *(convex relaxation)* and OMP *(greedy algorithm)* succeed w.h.p.

Statistical Results

Let $n^* := 2k \log \frac{p}{k} / \log \left(\frac{k}{\sigma^2} + 1 \right)$. Assume $\mathrm{SNR} = \frac{k}{\sigma^2} \to +\infty$.

Rise of a Computational-Statistical Gap

Computational Results ([Wainwright '09],[Fletcher et al '11])

Set $n_{alg} = 2k \log p$. Assume $SNR = \frac{k}{\sigma^2} \to +\infty$.
If

$$n > (1+\epsilon) n_{alg}$$

LASSO *(convex relaxation)* and OMP *(greedy algorithm)* succeed w.h.p.

Statistical Results

Let $n^* := 2k \log \frac{p}{k} / \log\left(\frac{k}{\sigma^2} + 1\right)$. Assume $SNR = \frac{k}{\sigma^2} \to +\infty$.

- If $n < (1-\epsilon) n^*$ no algorithm can succeed w.h.p. [Wang et al '10]

Rise of a Computational-Statistical Gap

Computational Results ([Wainwright '09],[Fletcher et al '11])

Set $n_{alg} = 2k \log p$. Assume $\text{SNR} = \frac{k}{\sigma^2} \to +\infty$.
If

$$n > (1+\epsilon) n_{alg}$$

LASSO *(convex relaxation)* and OMP *(greedy algorithm)* succeed w.h.p.

Statistical Results

Let $n^* := 2k \log \frac{p}{k} / \log \left(\frac{k}{\sigma^2} + 1 \right)$. Assume $\text{SNR} = \frac{k}{\sigma^2} \to +\infty$.

- If $n < (1-\epsilon) n^*$ no algorithm can succeed w.h.p. [Wang et al '10]
- For some large $C > 0$, if $n \geq C n^*$, MLE succeeds [Rad' 11].

Rise of a Computational-Statistical Gap

Computational Results ([Wainwright '09],[Fletcher et al '11])

Set $n_{alg} = 2k \log p$. Assume $\mathrm{SNR} = \frac{k}{\sigma^2} \to +\infty$.
If

$$n > (1+\epsilon) n_{alg}$$

LASSO *(convex relaxation)* and OMP *(greedy algorithm)* succeed w.h.p.

Statistical Results

Let $n^* := 2k \log \frac{p}{k} / \log \left(\frac{k}{\sigma^2} + 1 \right)$. Assume $\mathrm{SNR} = \frac{k}{\sigma^2} \to +\infty$.

- If $n < (1-\epsilon) n^*$ no algorithm can succeed w.h.p. [Wang et al '10]
- For some large $C > 0$, if $n \geq C n^*$, MLE succeeds [Rad' 11].

Pictorial Representation

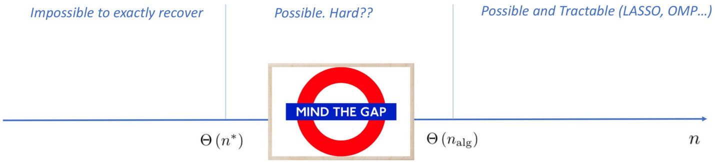


Figure: Computational-Statistical Gap

Pictorial Representation

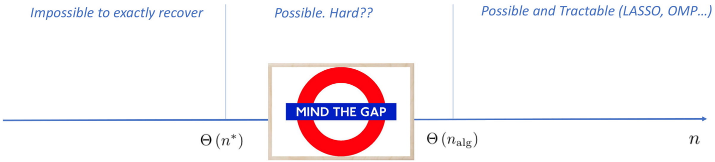


Figure: Computational-Statistical Gap

Contributions

(1) $n^* = 2k \log \frac{p}{k} / \log\left(\frac{k}{\sigma^2} + 1\right)$ is the **exact statistical limit** *(All-or-Nothing Phase Transition)*.

(2) $n_{\text{alg}} = 2k \log p$ is the **phase transition point** for (landscape) hardness *(Overlap Gap Property Phase Transition)*.

Outline of the Talk

(1) Introduction and Thesis Overview

(2) **High Dimensional Linear Regression Model**

- Background
- **Statistical Limit: All-or-Nothing Phenomenon**
- Computational-Statistical Gap and Overlap Gap Property

(3) Planted Clique Model and Overlap Gap Property

Maximum Likelihood Estimator (MLE)

$Y = X\beta^* + W$ with W iid $N(0, \sigma^2)$ entries.

The MLE

$\hat{\beta}_{\text{MLE}}$ is the optimal solution of least-squares

$$(\text{LS}): \min_{\beta \in \{0,1\}^p, \|\beta\|_0 = k} \|Y - X\beta\|_2$$

[Rad '11]: *success* with Cn^* samples.

All or Nothing Phenomenon- Result

Definition

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

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

All or Nothing Phenomenon- Result

Definition

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

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

Theorem ("All or Nothing Phase Transition" (G**Z** '17), (RX**Z** '19))

Let $\epsilon > 0$ *be arbitrary. Assume* $k \ll p$ *and* $k/\sigma^2 \geq C(\epsilon) > 0$,

- *If* $n > (1+\epsilon)\, n^*$, *then*

$$\frac{1}{k}\text{overlap}\left(\hat{\beta}_{\text{MLE}}\right) \to 1, \textit{ whp, as } n, p, k \to +\infty.$$

- *If* $n < (1-\epsilon)\, n^*$, *and* $k \ll \sqrt{p}$, *then* $\forall \hat{\beta} = \hat{\beta}\,(Y, X)$

$$\frac{1}{k}\text{overlap}\left(\hat{\beta}\right) \to 0, \textit{ whp, as } n, p, k \to +\infty.$$

All or Nothing Phenomenon - Comments

An "All or Nothing" phase transition!

- With $n \geq (1+\epsilon)n^*$,
  MLE recovers **all but** $o(1)$-fraction of the support.

All or Nothing Phenomenon - Comments

An "All or Nothing" phase transition!

- With $n \geq (1+\epsilon)n^*$,
  MLE recovers **all but** $o(1)$-fraction of the support.
- With $n \leq (1-\epsilon)n^*$,
  every estimator recovers **at most** $o(1)$-fraction of the support.

All or Nothing Theorem - Proof Sketch

Negative Result for $n \leq (1-\epsilon)n^*$:

- *Step 1:*
  "Impossibility of Testing": Data Look Like Pure Noise.

All or Nothing Theorem - Proof Sketch

Negative Result for $n \leq (1-\epsilon)n^*$:

- *Step 1:*
  "Impossibility of Testing": Data Look Like Pure Noise.
  Let P the law of $(Y = X\beta^* + W, X)$,
  and Q the law of $(Y = \lambda W, X)$, for some $\lambda > 0$.

All or Nothing Theorem - Proof Sketch

Negative Result for $n \leq (1-\epsilon)n^*$:

- *Step 1:*
  "Impossibility of Testing": Data Look Like Pure Noise.
  Let P the law of $(Y = X\beta^* + W, X)$,
  and Q the law of $(Y = \lambda W, X)$, for some $\lambda > 0$.
  We show,
  $$D_{KL}(P||Q) \to 0, \text{ as } p \to +\infty.$$

All or Nothing Theorem - Proof Sketch

Negative Result for $n \leq (1-\epsilon)n^*$:

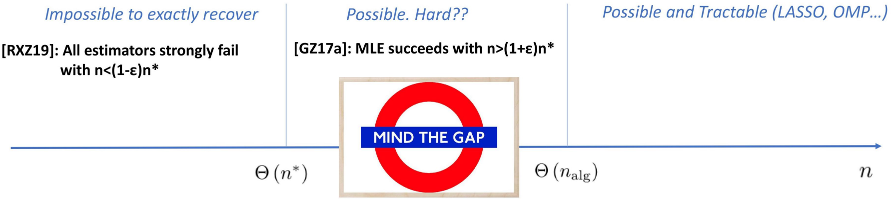
- *Step 1:*
  "Impossibility of Testing": Data Look Like Pure Noise.
  Let P the law of $(Y = X\beta^* + W, X)$,
  and Q the law of $(Y = \lambda W, X)$, for some $\lambda > 0$.
  We show,
  $$D_{KL}(P||Q) \to 0, \text{ as } p \to +\infty.$$
- *Step 2:*
  "Impossibility of Testing" implies "Impossibility of Estimation".

All or Nothing Theorem - Proof Sketch

Negative Result for $n \leq (1-\epsilon)n^*$:

- *Step 1:*
  "Impossibility of Testing": Data Look Like Pure Noise.
  Let P the law of $(Y = X\beta^* + W, X)$,
  and Q the law of $(Y = \lambda W, X)$, for some $\lambda > 0$.
  We show,
  $$D_{KL}(P||Q) \to 0, \text{ as } p \to +\infty.$$
- *Step 2:*
  "Impossibility of Testing" implies "Impossibility of Estimation".
  We show that for any estimator $\hat{\beta} = \hat{\beta}(Y, X)$:
  $$\text{overlap}\left(\hat{\beta}\right)/k \leq \left(1 + \sigma^2/k\right) D_{KL}(P||Q).$$

Summary for n* contribution



Sharp Information-Theoretic Limit n*

$(1+\epsilon)$n* samples MLE (asymptotically) succeeds.
$(1-\epsilon)$n* samples all estimators (asymptotically) strongly fail.

Outline of the Talk

(1) Introduction and Thesis Overview

(2) **High Dimensional Linear Regression Model**

- Background
- Statistical Limit: All-or-Nothing Phenomenon
- **Computational-Statistical Gap and Overlap Gap Property**

(3) Planted Clique Model and Overlap Gap Property

Computational-Statistical Gap

Impossible to exactly recover

[RXZ19]: All estimators strongly fail with n<(1-ε)n*

Possible. Hard??

[GZ17a]: MLE succeeds with n>(1+ε)n*

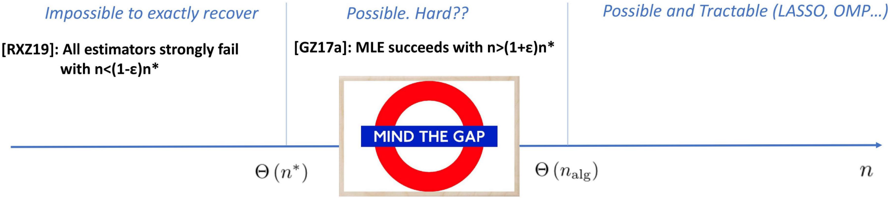
Possible and Tractable (LASSO, OMP...)

$\Theta(n^*)$

$\Theta(n_{\text{alg}})$

n

Computational-Statistical Gap



Contribution through Landscape Analysis

n_{alg} is a **phase transition point** for certain Overlap Gap Property (OGP) on the space of binary k-sparse vectors (origin in *spin glass theory*).
Conjecture computational hardness!

Computational Hardness: A Spin Glass Perspective

Computational gaps appear frequently in random environments

(1) *randoms CSPs*,
such as random-k-SAT (e.g. [MMZ '05], [ACORT '11])

(2) *average-case combinatorial opt problems*
such as max-independent set in ER graphs (e.g. [GS '17], [RV '17])

Computational Hardness: A Spin Glass Perspective

Computational gaps appear frequently in random environments

(1) *randoms CSPs*,
such as random-k-SAT (e.g. [MMZ '05], [ACORT '11])

(2) *average-case combinatorial opt problems*
such as max-independent set in ER graphs (e.g. [GS '17], [RV '17])

Between easy and hard regime there is an **"abrupt change in the geometry of the space of (near-optimal) solutions"** [ACO '08].

Computational Hardness: A Spin Glass Perspective

Computational gaps appear frequently in random environments

(1) *randoms CSPs*,
such as random-k-SAT (e.g. [MMZ '05], [ACORT '11])

(2) *average-case combinatorial opt problems*
such as max-independent set in ER graphs (e.g. [GS '17], [RV '17])

Between easy and hard regime there is an **"abrupt change in the geometry of the space of (near-optimal) solutions"** [ACO '08].

(Vague) Strategy of Studying the Geometry

Study **realizable overlap sizes** between "near-optimal" solutions.
Algorithms appear to work as long as there are **no gaps** in the overlaps.

Computational Hardness: A Spin Glass Perspective

Computational gaps appear frequently in random environments

(1) *randoms CSPs*,
such as random-k-SAT (e.g. [MMZ '05], [ACORT '11])

(2) *average-case combinatorial opt problems*
such as max-independent set in ER graphs (e.g. [GS '17], [RV '17])

Between easy and hard regime there is an **"abrupt change in the geometry of the space of (near-optimal) solutions"** [ACO '08].

(Vague) Strategy of Studying the Geometry

Study **realizable overlap sizes** between "near-optimal" solutions.
Algorithms appear to work as long as there are **no gaps** in the overlaps.

Overlap Gap Property, Shattering, Clustering, Free Energy Wells etc

The Overlap Gap Property (OGP) for Linear Regression

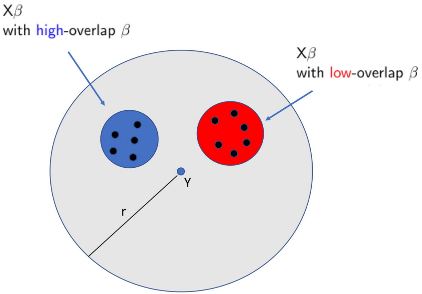
"Near-optimal solutions" $\{\beta \in \{0,1\}^p : \|\beta\|_0 = k,$ "small" $\|Y - X\beta\|_2\}$.

The Overlap Gap Property (OGP) for Linear Regression

"Near-optimal solutions" $\{\beta \in \{0,1\}^p : \|\beta\|_0 = k, \text{ "small" } \|Y - X\beta\|_2\}$.
Idea: Study overlaps between β and β^*.
$\text{overlap}(\beta) = |\text{Support}(\beta) \cap \text{Support}(\beta^*)|$.

The OGP (informally)

The set of β's with "small" $\|Y - X\beta\|_2$ partitions in
one group where β have **low** overlap with the ground truth β^* and
the other group where β have **high** overlap with the ground truth β^*.



The Overlap Gap Property for Linear Regression-definition

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

Definition (The Overlap Gap Property)

The linear regression problem satisfies OGP if there exists $r > 0$ and $0 < \zeta_1 < \zeta_2 < 1$ such that

(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.

OGP Phase Transition at $\Theta(n_{alg})$

Theorem (GZ '17a), (GZ '17b)

Suppose $k \leq \exp(\sqrt{\log p})$. *There exists* $C > 1 > c > 0$ *such that,*

- *If* $n^* < n < cn_{alg}$ *then w.h.p. OGP holds.*
- *If* $n > Cn_{alg}$ *then w.h.p. OGP does* **not** *hold.*

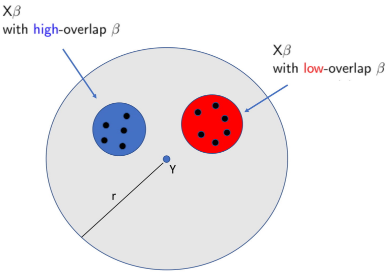


Figure: $n < cn_{alg}$

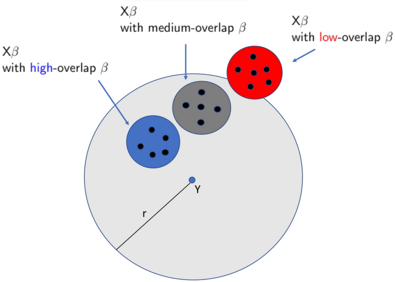


Figure: $n > Cn_{alg}$

OGP Phase Transition at $\Theta(n_{alg})$

Theorem (GZ '17a), (GZ '17b)

Suppose $k \leq \exp(\sqrt{\log p})$. *There exists* $C > 1 > c > 0$ *such that,*

- *If* $n^* < n < cn_{alg}$ *then w.h.p. OGP holds.*
- *If* $n > Cn_{alg}$ *then w.h.p. OGP does* **not** *hold.*

OGP coincides with the failure of
convex relaxation and **compressed sensing** methods!

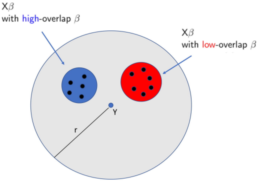


Figure: $n^* < n < cn_{alg}$

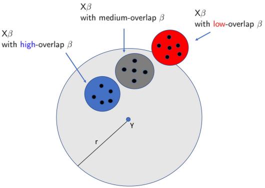
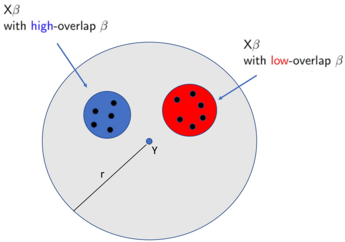


Figure: $n > Cn_{alg}$

OGP and Local Search

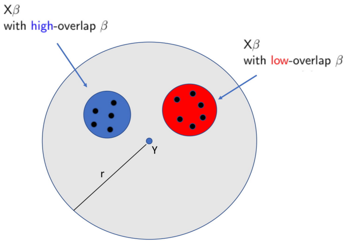
Local Step: $\beta \to \beta'$ if $d_H(\beta, \beta') = 2$. E.g. $\begin{bmatrix} * \\ 0 \\ 1 \\ * \end{bmatrix} \to \begin{bmatrix} * \\ 1 \\ 0 \\ * \end{bmatrix}$



OGP and Local Search

Local Step: $\beta \to \beta'$ if $d_H(\beta, \beta') = 2$. E.g. $\begin{bmatrix} * \\ 0 \\ 1 \\ * \end{bmatrix} \to \begin{bmatrix} * \\ 1 \\ 0 \\ * \end{bmatrix}$

(LS): $\min_{\beta \in \{0,1\}^p, \|\beta\|_0 = k} \|Y - X\beta\|_2$.



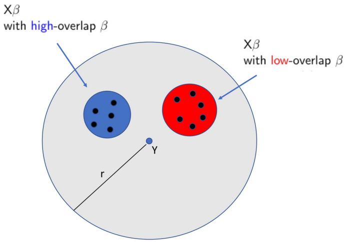
OGP and Local Search

Local Step: $\beta \to \beta'$ if $d_H(\beta, \beta') = 2$. E.g. $\begin{bmatrix} * \\ 0 \\ 1 \\ * \end{bmatrix} \to \begin{bmatrix} * \\ 1 \\ 0 \\ * \end{bmatrix}$

(LS): $\min_{\beta \in \{0,1\}^p, \|\beta\|_0 = k} \|Y - X\beta\|_2$.

Corollary: Local Search Barrier [GZ'17a]

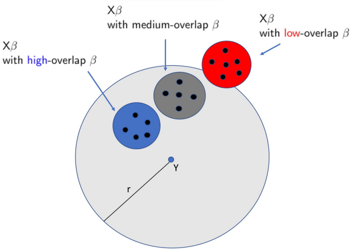
Under OGP, there are **low-overlap local minima** in (LS).
If $n < cn_{alg}$, greedy local-search algorithm **fails** (worst-case) w.h.p.



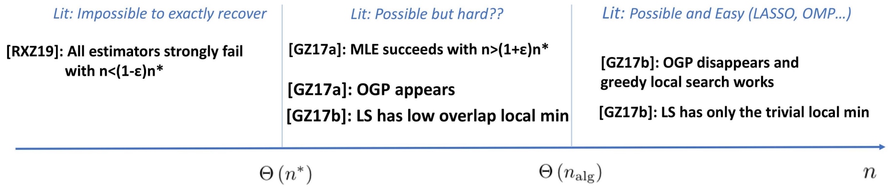
OGP and Local Search

Theorem (GZ '17b)

If $n > Cn_{alg}$, *the* **only local minimum** *in (LS) is* β^* *whp and greedy local search algorithm* **succeeds** *in* $O(k/\sigma^2)$ *iterations whp.*



Summary of Contribution



Sharp Information-Theoretic Limit n^*

$(1+\epsilon)n^*$ samples MLE (asymptotically) succeeds.
$(1-\epsilon)n^*$ samples all estimators (asymptotically) strongly fail.

OGP Phase Transition at n_{alg}

$n < cn_{\mathrm{alg}}$ OGP holds and $n > Cn_{\mathrm{alg}}$ OGP does not hold.
Computational Hardness conjectured!

Outline of the Talk

(1) Introduction and Thesis Overview

(2) High Dimensional Linear Regression Model

- Background
- Statistical Limit: All-or-Nothing Phenomenon
- Computational-Statistical Gap and Overlap Gap Property

(3) **Planted Clique Model and Overlap Gap Property**

The Planted Clique Model

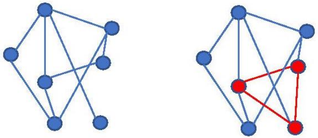
The Planted Clique Model [Jerrum '92]

Graph Generating Assumptions:

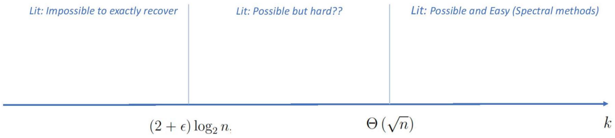
- Stage 1: $\mathcal{G}_0$ is an Erdos-Renyi $G(n, 1/2)$:
  n-vertex undirected graph, each edge appears w.p. $1/2$.
- Stage 2: k out of the n vertices of $\mathcal{G}_0$ are chosen u.a.r. to form a k-vertex clique, $\mathcal{PC}$. Call $\mathcal{G}$ the final graph.

Goal: Recover $\mathcal{PC}$ from observing $\mathcal{G}$.
Question: For how small $k = k_n$ can we recover?
Statistical limit + Computational limit.

$n = 7, k = 3$, $\mathcal{G}_0$ (left) and $\mathcal{G}$ (right) :

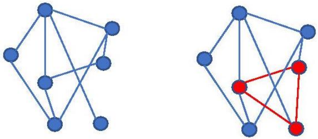


The Planted Clique Model-Literature

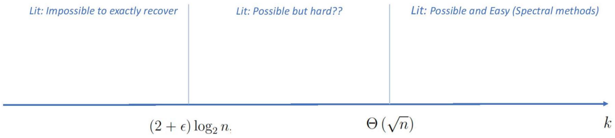


Literature:

- *Statistical Limit* (unique k-clique): $k = (2+\epsilon)\log_2 n$, for any $\epsilon > 0$.
- *(Apparent) Computational Limit*: $k = c\sqrt{n}$, for any $c > 0$. [AKS'98],[FR'10],[DM'13],[DGGP'14]



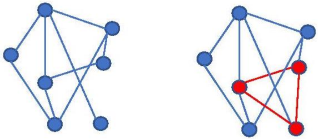
The Planted Clique Model-Literature



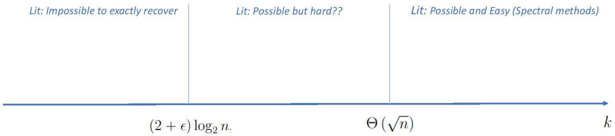
Literature:

- *Statistical Limit* (unique k-clique): $k = (2+\epsilon)\log_2 n$, for any $\epsilon > 0$.
- *(Apparent) Computational Limit*: $k = c\sqrt{n}$, for any $c > 0$. [AKS'98],[FR'10],[DM'13],[DGGP'14]

  Long-studied comp-stats gap [BR'13], [BHK+'16], [BBH'18]



The Planted Clique Model-Literature

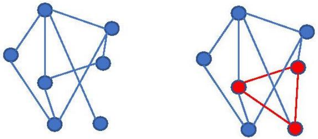


Literature:

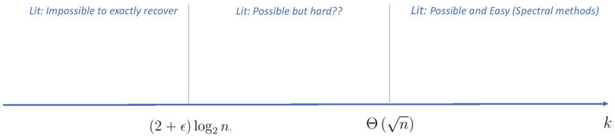
- *Statistical Limit* (unique k-clique): $k = (2+\epsilon)\log_2 n$, for any $\epsilon > 0$.
- *(Apparent) Computational Limit*: $k = c\sqrt{n}$, for any $c > 0$. [AKS'98],[FR'10],[DM'13],[DGGP'14]

  Long-studied comp-stats gap [BR'13], [BHK+'16], [BBH'18]

  Question: Is there an OGP phase transition around $k = \sqrt{n}$*?*

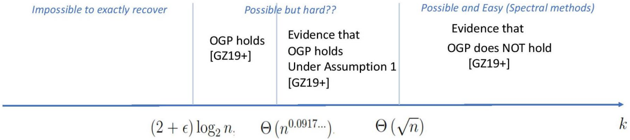


The Overlap Gap Property for Planted Clique: Results



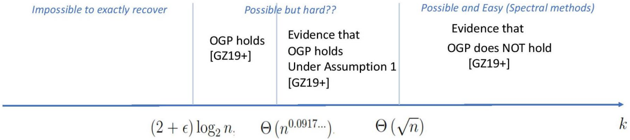
- Focus on *subgraphs of $\mathcal{G}$ of fixed vertex size* ("k-sparse binary β") with *many edges*, dense, ("small error $\|Y - X\beta\|_2$") and study *their overlap with $\mathcal{PC}$* ("overlap with β^*") .
  OGP: dense subgraphs have either high or low overlap with $\mathcal{PC}$.

The Overlap Gap Property for Planted Clique: Results



- Focus on *subgraphs of $\mathcal{G}$ of fixed vertex size* ("k-sparse binary β") with *many edges*, dense, ("small error $\|Y - X\beta\|_2$") and study *their overlap with $\mathcal{PC}$* ("overlap with β^*") .
  OGP: dense subgraphs have either high or low overlap with $\mathcal{PC}$.
- *Strong evidence* for OGP **phase transition** at $k = \sqrt{n}$.
  (Possible explanation for a long-studied hardness!).
- *Proof OGP appears* if $k \leq n^{0.0917}$.

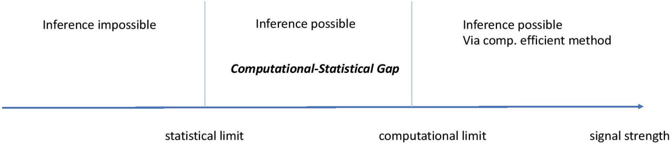
The Overlap Gap Property for Planted Clique: Results



- Focus on *subgraphs of $\mathcal{G}$ of fixed vertex size* ("k-sparse binary β")
  with *many edges*, dense, ("small error $\|Y - X\beta\|_2$")
  and study *their overlap with $\mathcal{PC}$* ("overlap with β^*") .
  OGP: dense subgraphs have either high or low overlap with $\mathcal{PC}$.
- *Strong evidence* for OGP **phase transition** at $k = \sqrt{n}$.
  (Possible explanation for a long-studied hardness!).
- *Proof OGP appears* if $k \leq n^{0.0917}$.
- **Assumption 1:**
  Concentration of the value of k-densest subgraph problem of $G(n, \frac{1}{2})$.
  Known for $k = \Theta(\log n)$ [BBSV'18], *proven* for $k \leq n^{0.0917..}$ [GZ'19],
  conjectured for all $k = o\left(\sqrt{n}\right)$.

Thesis Overview: Contributions

HDLR (signal strength = sample size), PC (signal strength = clique size):



Under assumptions,

- Compute the exact **statistical limit** of the HDLR model *("All-to-Nothing Phase Transition")*
- Explain **computational-statistical gaps** of HDLR and PC models, through *statistical-physics* based methods. *("Overlap Gap Property")*
- Improved **computational limit** for **noiseless** HDLR model using *lattice basis reduction ("One Sample Suffices")*

Papers:
(Gamarnik, **Z.** *COLT* '17, *AOS* (major rev.) '18+)
(Gamarnik, **Z.** *AOS* (major rev.) '18+), (Gamarnik, **Z.** *NeurIPS* '18)
(Reeves, Xu, **Z.** COLT '19), (Gamarnik, **Z.** '19+)