High Temperature Structure Detection in Ferromagnets

Matey Neykov

Department of Statistics & Data Science
Carnegie Mellon University

Ferromagnetic Zero-field Ising Model

$$\mathbb{P}(\boldsymbol{X} = \mathbf{x}) \propto \exp\left(\theta \sum_{1 \le u < v \le d} \xi_{uv} x_u x_v\right)$$

- $\mathbf{x} \in \{\pm 1\}^d$
- $\theta = \frac{1}{T} \geq 0$ *inverse temperature*
- If all $\xi_{uv} \geq 0$ the model is called *ferromagnetic*
- The model is overparametrized; call $\theta_{uv} = \theta \xi_{uv}$

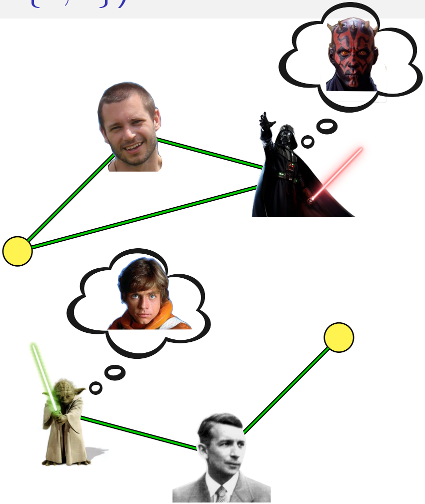
Corollary (to the Griffith's inequliaty)

"If you decrease any ξ_{uv}, no correlation $\mathbb{E} X_k X_\ell$ will increase"

A Simple Ferromagnet (all $\xi_{uv} \in \{0,1\}$)

$$\boldsymbol{X} = \underbrace{(X_1, X_2, \ldots, X_d)}_{d \text{ spins}} \in \{\pm 1\}^d$$

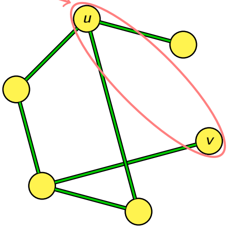
$$\mathbb{P}_{\theta,G}(\boldsymbol{X} = \boldsymbol{x}) \propto \exp\left(\theta \sum_{u \sim v} x_u x_v\right)$$



Graphical Model Interpretation

$$\boldsymbol{X} = \underbrace{(X_1, X_2, \ldots, X_d)}_{d \text{ spins}}$$

$G = (V, E)$



$X_u \perp\!\!\!\perp X_v | \textbf{rest}$

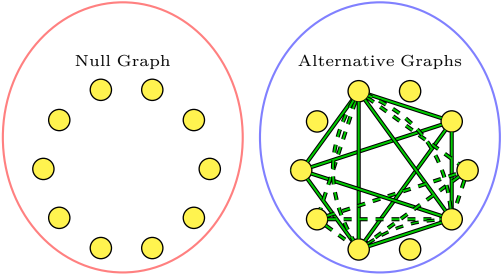
$\xi_{uv} = 0$ (or equivalently $\theta_{uv} = 0$)

Outline of the Talk

1. Fundamental Limit 
2. Scan Test For Clique Detection 
3. Computational Lower Bound 
4. Open Questions 

Worst Case Formalization

$\mathbf{H}_0 : \theta = 0$ vs $\mathbf{H}_1 : \theta \neq 0$, graph is a hidden $s-$clique



$\theta = 0$ $\qquad$ $\theta \neq 0$, G is an $s-$clique

$\boldsymbol{X} \sim \mathbb{P}_0$ $\qquad$ $\boldsymbol{X} \sim \mathbb{P}_{\theta,G}$

Minimax Risk:

$$R_n(\theta) = \inf_{\psi} \Big[\overbrace{\mathbb{P}_0^{\otimes n}(\psi = 1)}^{\text{Type I}} + \max_G \overbrace{\mathbb{P}_{\theta,G}^{\otimes n}(\psi = 0)}^{\text{Type II}} \Big]$$

$$\geq \inf_{\psi} [\mathbb{P}_0^{\otimes n}(\psi = 1) + \frac{1}{m} \sum_{i=1}^m \mathbb{P}_{\theta,G_i}^{\otimes n}(\psi = 0)]$$

Fundamental Limit 

Theorem (Cao, Neykov and Liu, 2018)

If $s = o(\sqrt{d})$

$$\theta \leq \sqrt{\frac{\log(d/s^2)}{6ns}} \wedge \frac{1}{32s}$$

it holds that

$$\liminf_{n \to \infty} R_n(\theta) = 1.$$

Roadmap

1. Fundamental Limit 
2. Scan Test For Clique Detection  
3. Computational Lower Bound
4. Open Questions 

A Simple Scan Test

For any s-clique graph G define the statistic:

$$\widehat{W}_G := \frac{1}{n}\sum_{l=1}^{n}\left(\frac{1}{|E(G)|}\sum_{(i,j)\in E(G)} X_{l,i}X_{l,j}\right), \text{ here } |E(G)| = \binom{s}{2}.$$

For a large absolute constant κ, define the test ψ:

$$\psi := \mathbb{1}\left(\max_G \widehat{W}_G > \frac{\kappa}{4}\sqrt{\frac{\log(ed/s)}{sn}}\right).$$

Performance Guarantee

Theorem (Cao, Neykov and Liu, 2018)

If $\theta < \frac{1}{2s}, s\log(ed/s) = o(n), (d/s)^s > 2/\alpha$, *and*

$$\theta \geq \kappa\sqrt{\frac{\log(ed/s)}{6ns}}$$

it holds that

$$\mathbb{P}_0(\psi = 1) + \max_G \mathbb{P}_{\theta,G}(\psi = 0) \leq \alpha$$

Roadmap

1. Fundamental Limit 
2. Scan Test For Clique Detection  
3. Computational Lower Bound

---

4. Open Questions 

Computational Lower Bound

Q: Is there a test which is as good as the scan test, only faster?

Computational Lower Bound

Q: Is there a test which is as good as the scan test, only faster?

A: Probably not ...

A PCA Conjecture

Consider the PCA testing problem

$$H_0^{\text{PCA}} : \boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n \sim N(0, \mathbf{I}),$$

vs

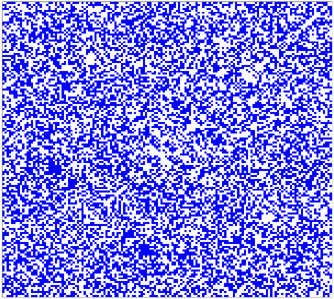
$$H_1^{\text{PCA}} : \boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n \sim N(0, \Sigma), \Sigma = \mathbf{I} + \sigma \mathbb{1}_S \mathbb{1}_S^{\intercal}$$

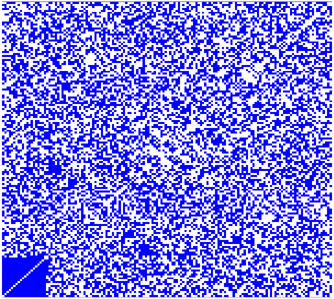
Conjecture (Cao, Neykov and Liu, 2018)

If $\sigma \lesssim n^{-1/2} \wedge s^{-1}$, then for any polynomial time test ψ we have

$$\liminf_{n \to \infty} \left[\mathbb{P}_I^{\otimes n}(\psi = 1) + \max_{\Sigma} \mathbb{P}_{\Sigma}^{\otimes n}(\psi = 0) \right] \geq \frac{1}{4}$$

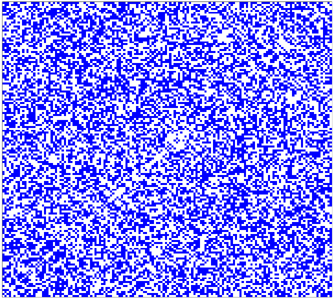
The PCA Conjecture and the Planted Clique Conjecture

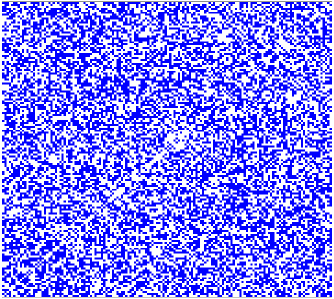




The PCA conjecture is established under the Planted Clique Conjecture by [Gao, Ma and Zhou, 2014], [Brennan and Bresler, 2019] and [Brennan Bresler and Huleihel, 2018] under various restrictions on n, s, d.

The PCA Conjecture and the Planted Clique Conjecture





The PCA conjecture is established under the Planted Clique Conjecture by [Gao, Ma and Zhou, 2014], [Brennan and Bresler, 2019] and [Brennan, Bresler and Huleihel, 2018] under various restrictions on n, s, d.

Computational Lower Bound

Theorem (Cao, Neykov and Liu, 2018)

Suppose that the PCA conjecture holds. If $\theta \lesssim n^{-1/2} \wedge s^{-1}$*, then for any polynomial time test* ψ *we have*

$$\liminf_{n\to\infty} \left[\mathbb{P}_0^{\otimes n}(\psi = 1) + \max_{G, G: s-clique} \mathbb{P}_{\theta,G}^{\otimes n}(\psi = 0) \right] \geq \frac{1}{4}$$

Computational Lower Bound Proof Sketch

- Need to show that there exists a polynomial time transformation between the PCA testing problem and the Ising model up to small total variation.
- This will imply that if we are able to test for a hidden s-clique in the Ising model, we will be able to detect whether the covariance in the PCA model has been corrupted.
- The transformation is remarkably simple: Take a Gaussian vector $\boldsymbol{Z}$ and transform it as $\mathrm{sign}(\boldsymbol{Z})$ (where sign is applied entry-wise).

Roadmap

1. Fundamental Limit 
2. Scan Test For Clique Detection 
3. Computational Lower Bound 
4. Open Questions 

- Can we remove the $\theta \lesssim s^{-1}$ assumption from the analysis of the scan test
- How do the upper and lower bounds change when $s \gg \sqrt{d}$

Acknowledgments

Yuan Cao **Han Liu**

Thanks!

Thank You!