Lecture 4.

Today

(i) Evidence for computational hardness

(ii) Some directions for future study

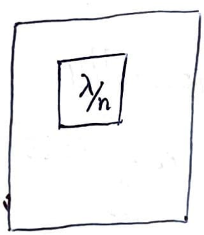
(i) Evidence for computational hardness.

$$M = \frac{\lambda}{n} vv^T + W \qquad v_i \overset{iid}{\sim} \begin{cases} 1 & \text{w.p. } \rho \\ 0 & \text{w.p. } 1-\rho \end{cases}$$

$W = W^T$

$\{W_{ij} : i<j\} \overset{iid}{\sim} N(0, \frac{1}{n})$

$\overset{ind}{\sim}$ $\{W_{ii} : 1 \le i \le n\} \overset{iid}{\sim} N(0, \frac{2}{n})$



Obs ① Detection of the hidden submatrix is easy!

$$\frac{1}{n} 1^T M 1 = \underbrace{\frac{\lambda}{n^2}(v^T 1)^2}_{\xrightarrow{P} \lambda\rho^2 > 0} + \underbrace{\frac{1^T W 1}{n}}_{\xrightarrow{P} 0}$$

(this test consistently differentiates between $H_0 : \lambda = 0$ vs. $H_1 : \lambda > 0$)

We now turn to ~~de~~ recovery.

Target (Stable recovery): We will say that the support can be recovered stably if $\exists\ \hat{v} := \hat{v}(M)$ & a universal constant $c > 0$ (ind of n, λ, ρ) s.t. $\frac{\langle \hat{v}, v\rangle}{n} > c\rho$ whp.

Q1. When does the top eigenvector perform stable recovery?

$$M = \lambda\rho \frac{v}{\sqrt{n\rho}} \cdot \left(\frac{v}{\sqrt{n\rho}}\right)^T + W \Rightarrow \text{top eigenvector works if } \lambda\rho > 1 \Leftrightarrow \lambda > \frac{1}{\rho}.$$

Q2. What about the information theoretic threshold?

Thm: If $\lambda > C\sqrt{\frac{1}{p}\log\frac{1}{p}}$ (for some universal const $C>0$ large enough) then the MLE recovers the support stably.
On the other hand, if $\lambda = o\left(\sqrt{\frac{1}{p}\log\frac{1}{p}}\right)$, no estimator recovers the support in a stable manner.

NB: The MLE in this setting

$$\hat{v} = \underset{\substack{v \in \{0,1\}^n \\ \sum v_i = np}}{\operatorname{argmax}} \; v^T M v$$ is computationally inefficient!

Obs: There is a wide gap between $\lambda = C\sqrt{\frac{1}{p}\log\frac{1}{p}}$ & $\lambda = 1/p$!

Q. What happens in the middle?

Conjecture: there does not exist any computationally efficient method for signal recovery in this intermediate SNR regime.

Q. How does ~~no~~ one provide evidence in favor of this conjecture?

A: Analyze restricted classes of algorithms (MCMC, spectral, convex relaxations etc).

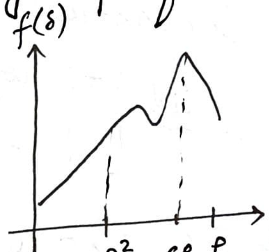
If none of the algorithms work in this regime, this is more evidence in favor of the conjecture!

<u>Today</u>: MCMC algorithms.

To this end, consider the restricted maximum likelihood

$$f_n(\delta) = \frac{1}{n} \max_{\substack{u \in \{0,1\}^n \\ \sum u_i = n\rho \\ \sum u_i v_i = n\delta}} u^T M u$$

<u>NB</u>: This is <u>not</u> observable, & is merely a proof device!



<u>Thm</u>: As $n \to \infty$, $f_n(\delta) \longrightarrow f(\delta)$.

Suppose $\sqrt{\frac{1}{\rho}\log\frac{1}{\rho}} \ll \lambda \ll \frac{1}{\rho^{1/2+\varepsilon}}$

(i) As $\lambda \gg \sqrt{\frac{1}{\rho}\log\frac{1}{\rho}}$, MLE performs stable recovery $\Rightarrow$ the global maximum occurs at $c\rho$!

(ii) If we guess $\hat{u}$ at random, $\frac{1}{n}\langle \hat{u}, v\rangle \approx \rho^2$. There exists a local maximum near ρ^2

Together, (i) + (ii) $\Rightarrow$ $f(\delta)$ is non-monotonic as a function of δ!

<u>Cor</u>: local MCs with stationary dist $\mu(u) \propto e^{\beta u^T M u}$ (β large) will get stuck at the first local max!

NB: For $\beta > 0$ large, the stationary dist approximates the MLE.

Refs: This discussion is based on Gamarnik, Jagannath & Sen '19.

Ben-Arous, Wein & Zadik '20
Barbier, Macris, Rush '20 } also study this problem with $\rho_n \to 0$ as $n \to \infty$.

Future directions

(i) Contextual Block models

$$\sigma_i \overset{iid}{\sim} \begin{cases} +1 & \text{wp } 1/2 \\ -1 & \text{wp } 1/2 \end{cases} \qquad u \sim N(0, I_p/p) \qquad \frac{p}{n} \to \frac{1}{\gamma}$$

$$\mathbb{P}[\{i,j\} \in E \mid \sigma] = \begin{cases} a/n & \text{if } \sigma_i = \sigma_j \\ b/n & \text{o.w.} \end{cases}$$

$$b_i = \sqrt{\frac{\mu}{n}}\, \sigma_i u + \frac{z_i}{\sqrt{p}}.$$

Q1. Spectral algorithms for weak recovery? (Analogue of the non-backtracking result for the block model)

Q2. What if we do not know λ, μ?

Q3. Community detection with $k > 2$ communities?

(ii) Universality of algorithms.

Information theoretic thresholds are model specific
Algorithmic performance less so.

(i) Spectral alg — Largest eigenvalue is universal as long as $\{W_{ij} : i<j\}$ are iid with suff light tail.

(ii) AMP — Similar universality results known.

(iii) TAP based approach — Not known. interesting to prove some universality.

Ambitous question: What if the noise has heavy tails? (Not much is known)

(iii) $\mathbb{Z}_2$-synchronization: $M = \frac{\lambda}{n} \sigma\sigma^T + M$, $\sigma_i \overset{iid}{\sim} \begin{cases} +1 & wp \ 1/2 \\ -1 \end{cases}$

$$SDP(\lambda) = \frac{1}{n} \max_{\substack{X \succcurlyeq 0 \\ X_{ii}=1 \ \forall i \in [n]}} \langle M, X \rangle$$

Thm: (i) If $\lambda \le 1$, $SDP(\lambda) \to 2$ whp

(ii) If $\lambda > 1+\varepsilon$, $SDP(\lambda) > 2 + \delta(\varepsilon) > 2$ whp.

Q. What is $\lim_{n\to\infty} SDP(\lambda)$ when $\lambda>1$?

Known: $2 < SDP(\lambda) \leq \lambda + \frac{1}{\lambda}$ if $\lambda > 1$.

Q. Is $\lim_{n\to\infty} SDP(\lambda) < \lambda + \frac{1}{\lambda}$?

Motivation: Certification of the ground state value $\max_{v\in\{\pm1\}^n} v^T M v$!