Information-theoretic bounds and phase transitions in clustering, sparse PCA, and submatrix localization

Jiaming Xu

Krannert School of Management
Purdue University

Joint work with Florent Krzakala (ENS) and Lenka Zdeborová (IPhT)
Jess Banks (SFI/Berkeley), Cris Moore (SFI), and Roman Vershynin (Michigan)

October 10, 2016

Outline of the talk

1. Sparse, spiked Wigner model
2. Extensions
3. Conclusions and remarks

Sparse, Spiked Wigner Model

$$Y = \frac{\lambda}{\sqrt{n}} v^* (v^*)^\top + W$$

- λ is a fixed constant

Sparse, Spiked Wigner Model

$$Y = \frac{\lambda}{\sqrt{n}} v^* (v^*)^\top + W$$

- λ is a fixed constant
- $W_{ii} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,2)$ and $W_{ij} = W_{ji} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$ for $i < j$

Sparse, Spiked Wigner Model

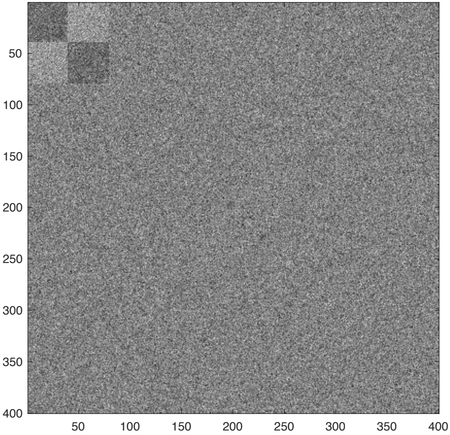
$$Y = \frac{\lambda}{\sqrt{n}} v^*(v^*)^\top + W$$

- λ is a fixed constant
- $W_{ii} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,2)$ and $W_{ij} = W_{ji} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$ for $i < j$
- $v_i^* \overset{\text{i.i.d.}}{\sim} (1-p)\delta_0 + \frac{p}{2}\delta_1 + \frac{p}{2}\delta_{-1}$ for a fixed $p \in [0,1]$

Sparse, Spiked Wigner Model

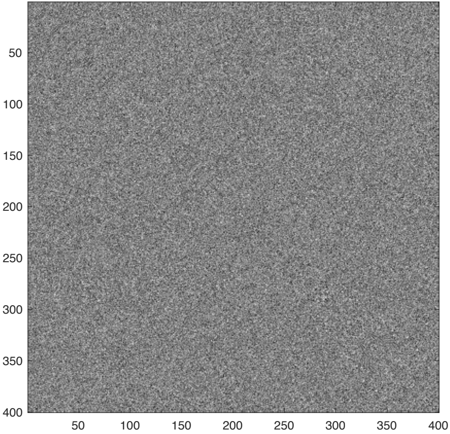
$$Y = \frac{\lambda}{\sqrt{n}} v^* (v^*)^\top + W$$

- λ is a fixed constant
- $W_{ii} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,2)$ and $W_{ij} = W_{ji} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$ for $i < j$
- $v_i^* \overset{\text{i.i.d.}}{\sim} (1-p)\delta_0 + \frac{p}{2}\delta_1 + \frac{p}{2}\delta_{-1}$ for a fixed $p \in [0,1]$
- $\|v^*\|_0 \approx np$: sparse level p



$$n = 400,\ p = 0.25,\ \lambda = 10$$

Of course not ordered



$$n = 400,\ p = 0.25,\ \lambda = 10$$

Motivation 1: Sparse PCA

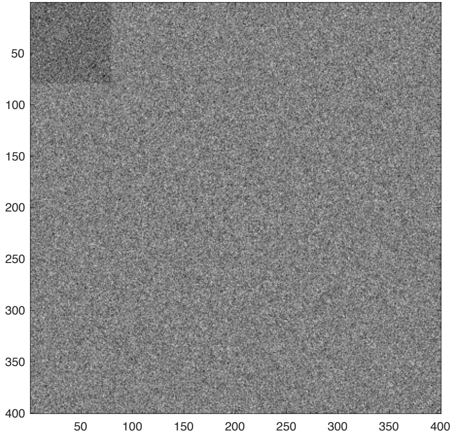
Sparse principal component analysis: [Johnstone-Lu 09], [Amini-Wainwright 09]...

$$Y = \frac{\lambda}{\sqrt{n}} u(v^*)^\top + W,$$

- $v^* \in \mathbb{R}^d$: sparse principal component
- W: Gaussian random matrix with i.i.d. entres

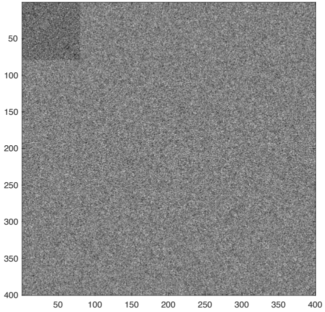
Motivation 2: Submatrix localization

Submatrix localization [Kolar-Balakrishnan-Rinaldo-Singh 11] [Butucea-Ingster 13] ...

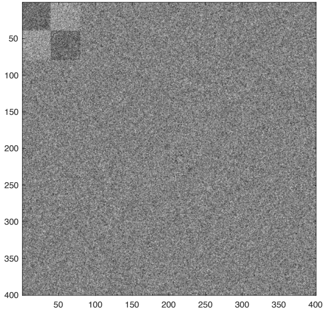


Motivation 2: Submatrix localization

Submatrix localization [Kolar-Balakrishnan-Rinaldo-Singh 11] [Butucea-Ingster 13] ...



Row sum statistic is informative



Row sum statistic is uninformative

Two statistical tasks

Focus on $\lambda = \Theta(1)$ and $p = \Theta(1)$

Two statistical tasks

Focus on $\lambda = \Theta(1)$ and $p = \Theta(1)$

- Estimation: $\mathbb{E}\left[\left(\frac{1}{n}\langle \hat{v}, v^* \rangle\right)^2\right] \geq \epsilon$
- Detection:

$$\mathcal{H}_0 : Y = W \quad \text{v.s.} \quad \mathcal{H}_1 : Y = \frac{\lambda}{\sqrt{n}} v^* (v^*)^\top + W$$

Type-I + Type-II error probabilities $\to 0$

Two statistical tasks

Focus on $\lambda = \Theta(1)$ and $p = \Theta(1)$

- Estimation: $\mathbb{E}\left[\left(\frac{1}{n}\langle \hat{v}, v^* \rangle\right)^2\right] \geq \epsilon$
- Detection:

$$\mathcal{H}_0 : Y = W \quad \text{v.s.} \quad \mathcal{H}_1 : Y = \frac{\lambda}{\sqrt{n}} v^* (v^*)^\top + W$$

Type-I + Type-II error probabilities $\to 0$

Main Questions

- When is estimation or detection informationally possible?
- Is IT-limit achievable in polynomial-time?

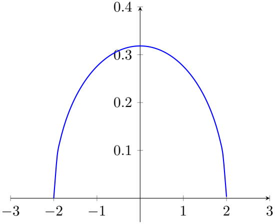
Prior work: Spectral phase transition [Péché 06]

$$Y = \frac{\lambda}{\sqrt{n}} v^* (v^*)^\top + W$$

Prior work: Spectral phase transition [Péché 06]

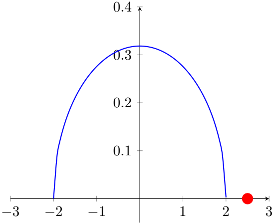
$$Y = \frac{\lambda}{\sqrt{n}} v^* (v^*)^\top + W$$

$\lambda p \leq 1$



$$\langle v_1(Y), v^*/\|v^*\|_2 \rangle^2 \to 0$$

$\lambda p > 1$



$$\langle v_1(Y), v^*/\|v^*\|_2 \rangle^2 \to 1 - (\lambda p)^{-2}$$

Prior work: Approximate message passing

Approximate message passing algorithm: [Thouless-Anderson-Palmer 77] [Rangan-Fletcher 12], [Deshpande-Montanari 14], [Lesieur-Krzakala-Zdeborová 15]

Conjecture (Lesieur-Krzakala-Zdeborová 15)

There exists $p^* \in (0,1)$ such that

1. *If $p \geq p^*$, then the IT limit is $\lambda p = 1$.*
2. *If $p < p^*$, then the computational limit is $\lambda p = 1$, but the IT-limit is strictly lower.*

Upper bound to the IT-limit

Theorem (Banks-Moore-Vershynin-X. 16)

Detection and estimation are information-theoretically possible, if

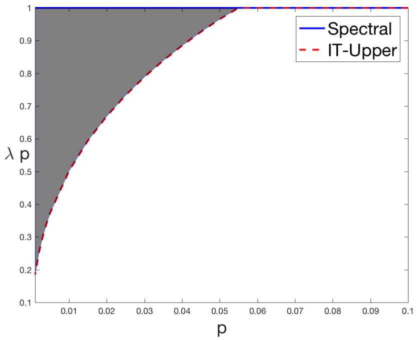
$$\lambda p > 2\sqrt{h(p) + p \log 2}$$

Upper bound to the IT-limit

Theorem (Banks-Moore-Vershynin-X. 16)

Detection and estimation are information-theoretically possible, if

$$\lambda p > 2\sqrt{h(p) + p \log 2}$$



IT-limit is below the spectral limit if $p \leq 0.054$.

Proof of upper bound: First moment method

Maximum likelihood estimation (MLE):

$$\begin{aligned} \hat{v} \in \arg\max_{v} \; & v^\top Y v \\ \text{s.t.} \quad & \|v\|_0 \le np(1+\epsilon_n) \\ & v \in \{0, \pm 1\}^n \end{aligned}$$

Proof of upper bound: First moment method

Maximum likelihood estimation (MLE):

$$\hat{v} \in \arg\max_{v} v^\top Y v$$
$$\text{s.t.} \quad \|v\|_0 \le np(1+\epsilon_n)$$
$$v \in \{0, \pm 1\}^n$$

- Estimation: show $\frac{1}{n} |\langle \hat{v}, v^* \rangle| \ge \delta$; it suffices to show

$$\max_{v : |\langle v, v^* \rangle| \le n\delta} v^\top Y v < (v^*)^\top Y v^*$$

Proof of upper bound: First moment method

Maximum likelihood estimation (MLE):

$$\hat{v} \in \arg\max_{v} v^\top Y v$$
$$\text{s.t.} \quad \|v\|_0 \le np(1+\epsilon_n)$$
$$v \in \{0, \pm 1\}^n$$

- Estimation: show $\frac{1}{n} |\langle \hat{v}, v^* \rangle| \ge \delta$; it suffices to show

$$\max_{v: |\langle v, v^* \rangle| \le n\delta} v^\top Y v < (v^*)^\top Y v^*$$

- Detection: show

$$\max_{v} v^\top Y v \ \text{ under } \mathcal{H}_0 \quad < \quad (v^*)^\top Y v^* \ \text{ under } \mathcal{H}_1$$

Lower bound to the IT-limit

Theorem (Banks-Moore-Vershynin-X. 16)

Detection and estimation are information-theoretically impossible, if

$$\lambda p < \sqrt{2p\,\mathcal{W}\left(\frac{1-p}{2\sqrt{ep}}\right)}$$

where $\mathcal{W}(y)$ *is the root* x *of* $xe^x = y$.

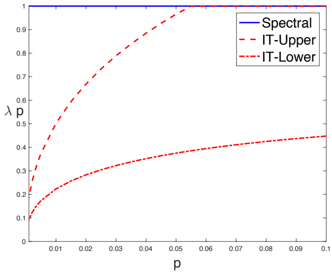
Lower bound to the IT-limit

Theorem (Banks-Moore-Vershynin-X. 16)

Detection and estimation are information-theoretically impossible, if

$$\lambda p < \sqrt{2p\,\mathcal{W}\left(\frac{1-p}{2\sqrt{ep}}\right)}$$

where $\mathcal{W}(y)$ is the root x of $xe^x = y$.



When $p \to 0$

- IT-upper limit: $\lambda p > 2\sqrt{-p \log p + O_p(p)}$
- IT-lower limit: $\lambda p < \sqrt{-2p \log p - O_p(p)}$
- Closed the gap of $\sqrt{2}$ [Verzelen 16]

Proof of lower bound: Second moment method

Theorem (Banks-Moore-Vershynin-X. 16)

Detection and estimation are information-theoretically impossible, if

$$\mathbb{E}_{Y \sim Q}\left[\left(\frac{P(Y)}{Q(Y)}\right)^2\right] = O(1).$$

Proof of lower bound: Second moment method

Theorem (Banks-Moore-Vershynin-X. 16)

Detection and estimation are information-theoretically impossible, if

$$\mathbb{E}_{Y\sim Q}\left[\left(\frac{P(Y)}{Q(Y)}\right)^2\right] = O(1).$$

- For estimation, it holds for Gaussian noise model (I-MMSE formula)

Proof of lower bound: Second moment method

Theorem (Banks-Moore-Vershynin-X. 16)

Detection and estimation are information-theoretically impossible, if

$$\mathbb{E}_{Y\sim Q}\left[\left(\frac{P(Y)}{Q(Y)}\right)^2\right] = O(1).$$

- For estimation, it holds for Gaussian noise model (I-MMSE formula)
- In sparse, spiked Wigner model,

$$\mathbb{E}_{Y\sim Q}\left[\left(\frac{P(Y)}{Q(Y)}\right)^2\right] \approx \mathbb{E}\left[\exp\left(\frac{\lambda^2 R^2}{2n}\right)\right]$$

where R is a T-step, symmetric random walk, where $T \sim \mathrm{Hyp}(n, np, p)$ [Cai-Ma-Wu 15].

Conjectured IT-limit

Conjecture (Lesieur-Krzakala-Zdeborová 15)

$$\lim_{n\to\infty} \frac{1}{n} I(v^*; Y) = \min_{\alpha\in[0,p]} i_{\mathrm{RS}}(\alpha; \lambda, p)$$

Moreover, let $\alpha^*(\lambda, p)$ *denote the smallest minimizer. Then the IT-limit for estimation is* $\lambda^*(p) = \inf\{\lambda : \alpha^*(\lambda, p) > 0\}$.

Conjectured IT-limit

Conjecture (Lesieur-Krzakala-Zdeborová 15)

$$\lim_{n\to\infty} \frac{1}{n} I(v^*; Y) = \min_{\alpha\in[0,p]} i_{\mathrm{RS}}(\alpha; \lambda, p)$$

Moreover, let $\alpha^*(\lambda, p)$ denote the smallest minimizer. Then the IT-limit for estimation is $\lambda^*(p) = \inf\{\lambda : \alpha^*(\lambda, p) > 0\}$.

Bethe mutual information i_{RS} (replica methods [Sherrington-Kirkpatrick 75] or cavity methods [Mezard-Parisi-Virasoro 86])

$$i_{\mathrm{RS}}(\alpha) = \frac{\lambda^2(p^2+\alpha^2)}{4} - \mathbb{E}\log\left(1-p+pe^{-\alpha\lambda^2/2}\cosh\left(\alpha\lambda^2\eta+\sqrt{\alpha}\lambda z\right)\right)$$

where $\eta \sim \mathrm{Bern}(p)$ and $z \sim \mathcal{N}(0,1)$

Conjectured IT-limit

Conjecture (Lesieur-Krzakala-Zdeborová 15)

$$\lim_{n\to\infty} \frac{1}{n} I(v^*; Y) = \min_{\alpha\in[0,p]} i_{\mathrm{RS}}(\alpha; \lambda, p)$$

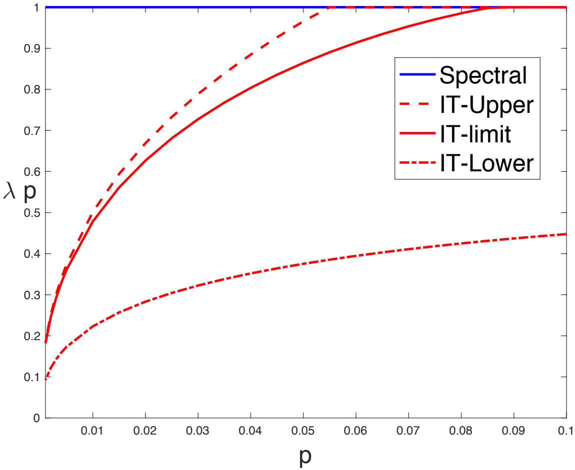
Moreover, let $\alpha^*(\lambda, p)$ denote the smallest minimizer. Then the IT-limit for estimation is $\lambda^*(p) = \inf\{\lambda : \alpha^*(\lambda, p) > 0\}$.

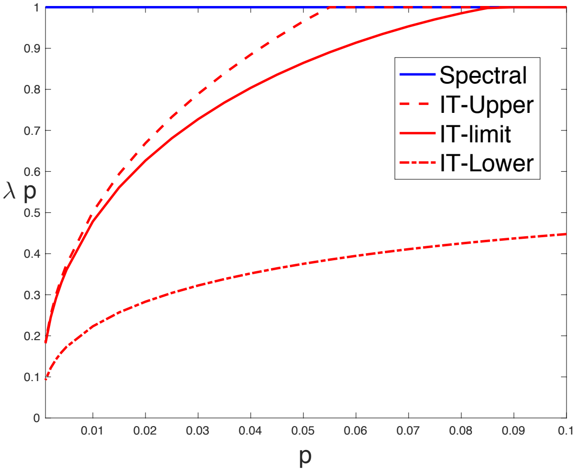
Bethe mutual information i_{RS} (replica methods [Sherrington-Kirkpatrick 75] or cavity methods [Mezard-Parisi-Virasoro 86])

$$i_{\mathrm{RS}}(\alpha) = \frac{\lambda^2(p^2+\alpha^2)}{4} - \mathbb{E}\log\left(1 - p + pe^{-\alpha\lambda^2/2}\cosh\left(\alpha\lambda^2\eta + \sqrt{\alpha}\lambda z\right)\right)$$

where $\eta \sim \mathrm{Bern}(p)$ and $z \sim \mathcal{N}(0,1)$

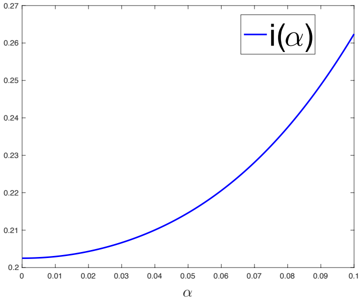
The corner case $p = 1$ is proved in [Deshpande-Abbe-Montanari 15]


Spectral
IT-Upper
IT-limit
IT-Lower
λ p
p
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0.01
0.02
0.03
0.04
0.05
0.06
0.07
0.08
0.09
0.1



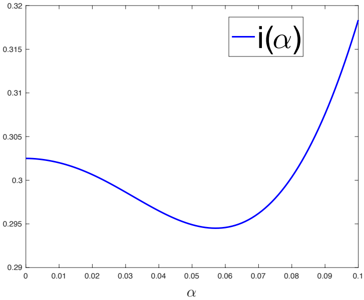
IT limit falls below spectral limit if $p \leq 0.085$. Why?

$\lambda p = 0.9$ (Below spectral limit)



$\alpha^* = 0$

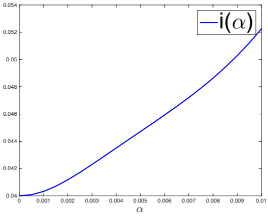
$\lambda p = 1.1$ (Above spectral limit)



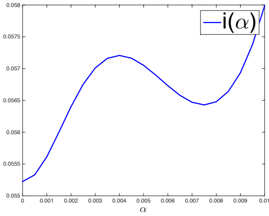
$\alpha^* > 0$

Statistical physics picture: sparse regime ($p = 0.01$)

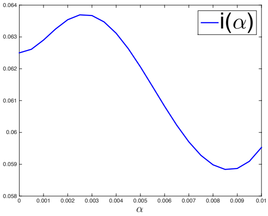
$\lambda p = 0.4$ (Uninformative)



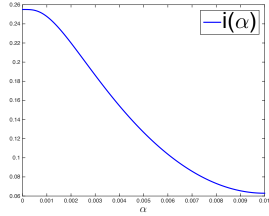
$\lambda p = 0.47$ (Uninformative: spinodal)



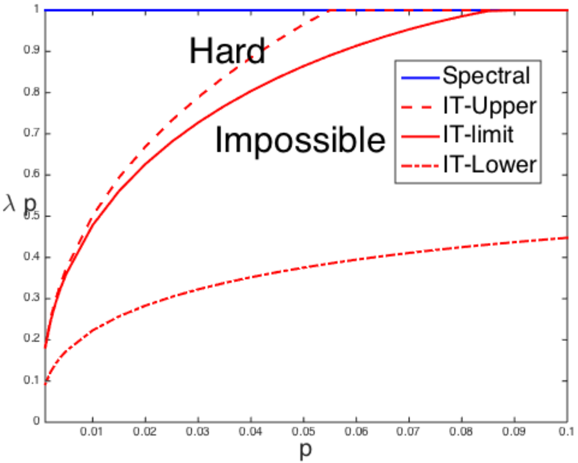
$\lambda p = 0.5$ (Informative: spinodal)



$\lambda p = 1.01$ (Above spectral limit)



Conjectured "possible but hard" regime [Lesieur-Krzakala-Zdeborová 15]



Proof of conjectured upper bound to mutual information

Theorem (Krzakala-X.-Zdeborová 16)

$$\frac{1}{n} I(v^*; Y) \leq \min_{\alpha \in [0,p]} i_{\mathrm{RS}}(\alpha; \lambda, p)$$

Proof of conjectured upper bound to mutual information

Theorem (Krzakala-X.-Zdeborová 16)

$$\frac{1}{n} I(v^*; Y) \leq \min_{\alpha \in [0,p]} i_{\mathrm{RS}}(\alpha; \lambda, p)$$

- Upper bound holds for any finite n

Proof of conjectured upper bound to mutual information

Theorem (Krzakala-X.-Zdeborová 16)

$$\frac{1}{n} I(v^*; Y) \le \min_{\alpha \in [0,p]} i_{\mathrm{RS}}(\alpha; \lambda, p)$$

- Upper bound holds for any finite n
- Asymptotic, matching lower bound is proved in [Barbier-Dia-Macris-Krzakala-Lesieur-Zdeborová 16] under the assumption that $i_{\mathrm{RS}}(\alpha)$ has at most three stationary points

Proof ideas: Interpolation method [Guerra 03]

- A simple denoising model: $y = \sqrt{\alpha}\lambda v^* + w$, where $w \sim \mathcal{N}(0, \mathbf{I}_{n\times n})$

$$\frac{1}{n} I(v^*; y) = I(v_1^*; y_1) = i_{\rm RS}(\alpha; \lambda, p) - \frac{(p-\alpha)^2 \lambda^2}{4}$$

Proof ideas: Interpolation method [Guerra 03]

- A simple denoising model: $y = \sqrt{\alpha}\lambda v^* + w$, where $w \sim \mathcal{N}(0, \mathbf{I}_{n \times n})$

$$\frac{1}{n} I(v^*; y) = I(v_1^*; y_1) = i_{\mathrm{RS}}(\alpha; \lambda, p) - \frac{(p-\alpha)^2 \lambda^2}{4}$$

- Interpolating between the denoising model and the Wigner model

$$Y_t = \frac{\sqrt{t}\lambda}{\sqrt{n}} v^* (v^*)^\top + W$$
$$y_t = \sqrt{\alpha(1-t)}\lambda v^* + w$$

Let $I_t = I(v^*; Y_t, y_t)$. Then $I_0 = I(v^*; y)$ and $I_1 = I(v^*; Y)$.

Proof ideas: Interpolation method [Guerra 03]

- A simple denoising model: $y = \sqrt{\alpha}\lambda v^* + w$, where $w \sim \mathcal{N}(0, \mathbf{I}_{n\times n})$

$$\frac{1}{n} I(v^*; y) = I(v_1^*; y_1) = i_{\rm RS}(\alpha; \lambda, p) - \frac{(p-\alpha)^2\lambda^2}{4}$$

- Interpolating between the denoising model and the Wigner model

$$Y_t = \frac{\sqrt{t}\lambda}{\sqrt{n}} v^*(v^*)^\top + W$$
$$y_t = \sqrt{\alpha(1-t)}\lambda v^* + w$$

Let $I_t = I(v^*; Y_t, y_t)$. Then $I_0 = I(v^*; y)$ and $I_1 = I(v^*; Y)$.

- Show that

$$\frac{1}{n}\frac{\mathrm{d}I_t}{\mathrm{d}t} \le \frac{(p-\alpha)^2\lambda^2}{4} \Longrightarrow \frac{1}{n} I_1 \le i_{\rm RS}(\alpha; \lambda, p)$$

Outline of the talk

1. Sparse, spiked Wigner model
2. Extensions
3. Conclusions and remarks

Extension 1: general channel output and prior

$$X = \frac{\lambda}{\sqrt{n}} v^* (v^*)^\top \longrightarrow \boxed{p_{\text{out}}(y|x)} \longrightarrow Y$$

- $v_i \overset{\text{i.i.d.}}{\sim} p_{\text{prior}}$
- $Y_{ij} \overset{\text{i.i.d.}}{\sim} p_{\text{out}}(\cdot|X_{ij})$ for $i \le j$
- p_{prior} and p_{out} are assumed to be independent of n

Channel universality

Theorem (Krzakala-X.-Zdeborová 16)

Suppose p_{prior} *has a finite support and* $\log p_{\text{out}}(y|x)$ *satisfies some mild regularity conditions. Then*

$$I(X;Y) = I(X; X + \sqrt{\Delta} W) + O(\sqrt{n})$$

Channel universality

Theorem (Krzakala-X.-Zdeborová 16)

Suppose p_{prior} has a finite support and $\log p_{\text{out}}(y|x)$ satisfies some mild regularity conditions. Then

$$I(X;Y) = I(X; X + \sqrt{\Delta} W) + O(\sqrt{n})$$

- Δ is the inverse Fisher information

$$\frac{1}{\Delta} \triangleq \mathbb{E}_{p_{\text{out}}(y|0)} \left[\left(\left. \frac{\partial \log p_{\text{out}}(y|x)}{\partial x} \right|_{y,0} \right)^2 \right]$$

Channel universality

Theorem (Krzakala-X.-Zdeborová 16)

Suppose $p_{\rm prior}$ *has a finite support and* $\log p_{\rm out}(y|x)$ *satisfies some mild regularity conditions. Then*

$$I(X;Y) = I(X; X + \sqrt{\Delta} W) + O(\sqrt{n})$$

- Δ is the inverse Fisher information

$$\frac{1}{\Delta} \triangleq \mathbb{E}_{p_{\rm out}(y|0)} \left[\left(\left.\frac{\partial \log p_{\rm out}(y|x)}{\partial x}\right|_{y,0} \right)^2 \right]$$

- Originally conjectured in [Lesieur-Krzakala-Zdeborová 15]

Channel universality

Theorem (Krzakala-X.-Zdeborová 16)

Suppose p_{prior} *has a finite support and* $\log p_{\text{out}}(y|x)$ *satisfies some mild regularity conditions. Then*

$$I(X;Y) = I(X; X + \sqrt{\Delta} W) + O(\sqrt{n})$$

- Δ is the inverse Fisher information

$$\frac{1}{\Delta} \triangleq \mathbb{E}_{p_{\text{out}}(y|0)} \left[\left(\left. \frac{\partial \log p_{\text{out}}(y|x)}{\partial x} \right|_{y,0} \right)^2 \right]$$

- Originally conjectured in [Lesieur-Krzakala-Zdeborová 15]
- Equivalence between Bernoulli and Gaussian was established in [Deshpande-Abbe-Montanari 15]

Channel universality

Theorem (Krzakala-X.-Zdeborová 16)

Suppose p_{prior} *has a finite support and* $\log p_{\text{out}}(y|x)$ *satisfies some mild regularity conditions. Then*

$$I(X;Y) = I(X; X + \sqrt{\Delta} W) + O(\sqrt{n})$$

- Δ is the inverse Fisher information

$$\frac{1}{\Delta} \triangleq \mathbb{E}_{p_{\text{out}}(y|0)}\left[\left(\left.\frac{\partial \log p_{\text{out}}(y|x)}{\partial x}\right|_{y,0}\right)^2\right]$$

- Originally conjectured in [Lesieur-Krzakala-Zdeborová 15]
- Equivalence between Bernoulli and Gaussian was established in [Deshpande-Abbe-Montanari 15]
- The proof relies on Lindeberg's principle

$$X = \frac{\mathrm{snr}}{\sqrt{n}} \begin{bmatrix} U_1^\top \\ U_2^\top \\ \vdots \\ U_m^\top \end{bmatrix} \begin{bmatrix} V_1 & V_2 & \cdots & V_n \end{bmatrix} \longrightarrow \boxed{p_{\mathrm{out}}(y|x)} \longrightarrow Y$$

- $U_i \in \mathbb{R}^k \overset{\text{i.i.d.}}{\sim} p_{\mathrm{prior}}$ and $V_j \in \mathbb{R}^k \overset{\text{i.i.d.}}{\sim} q_{\mathrm{prior}}$

Extension 2: High rank and asymmetric case

$$X = \frac{\text{snr}}{\sqrt{n}} \begin{bmatrix} U_1^\top \\ U_2^\top \\ \vdots \\ U_m^\top \end{bmatrix} \begin{bmatrix} V_1 & V_2 & \cdots & V_n \end{bmatrix} \longrightarrow \boxed{p_{\text{out}}(y|x)} \longrightarrow Y$$

- $U_i \in \mathbb{R}^k \overset{\text{i.i.d.}}{\sim} p_{\text{prior}}$ and $V_j \in \mathbb{R}^k \overset{\text{i.i.d.}}{\sim} q_{\text{prior}}$
- $Y_{ij} \overset{\text{i.i.d.}}{\sim} p_{\text{out}}(\cdot|X_{ij})$

Extension 2: High rank and asymmetric case

$$X = \frac{\mathrm{snr}}{\sqrt{n}} \begin{bmatrix} U_1^\top \\ U_2^\top \\ \vdots \\ U_m^\top \end{bmatrix} \begin{bmatrix} V_1 & V_2 & \cdots & V_n \end{bmatrix} \longrightarrow \boxed{p_{\mathrm{out}}(y|x)} \longrightarrow Y$$

- $U_i \in \mathbb{R}^k \overset{\text{i.i.d.}}{\sim} p_{\mathrm{prior}}$ and $V_j \in \mathbb{R}^k \overset{\text{i.i.d.}}{\sim} q_{\mathrm{prior}}$
- $Y_{ij} \overset{\text{i.i.d.}}{\sim} p_{\mathrm{out}}(\cdot|X_{ij})$
- p_{prior}, q_{prior}, and p_{out} are assumed to be independent of n

$$X = \frac{\mathrm{snr}}{\sqrt{n}} \begin{bmatrix} U_1^\top \\ U_2^\top \\ \vdots \\ U_m^\top \end{bmatrix} \begin{bmatrix} V_1 & V_2 & \cdots & V_n \end{bmatrix} \longrightarrow \boxed{p_{\mathrm{out}}(y|x)} \longrightarrow Y$$

- $U_i \in \mathbb{R}^k \overset{\text{i.i.d.}}{\sim} p_{\mathrm{prior}}$ and $V_j \in \mathbb{R}^k \overset{\text{i.i.d.}}{\sim} q_{\mathrm{prior}}$
- $Y_{ij} \overset{\text{i.i.d.}}{\sim} p_{\mathrm{out}}(\cdot|X_{ij})$
- p_{prior}, q_{prior}, and p_{out} are assumed to be independent of n
- Proposed in [Lesieur-Krzakala-Zdeborová 15]

Special case 1: Submatrix localization with k blocks

$$Y = \frac{\mu}{\sqrt{n}} \begin{bmatrix} 1 & & 0 \\ & 1 & \\ 0 & & 1 \end{bmatrix} + W$$

Special case 1: Submatrix localization with k blocks

$$Y = \frac{\mu}{\sqrt{n}} \begin{bmatrix} 1 & & 0 \\ & 1 & \\ 0 & & 1 \end{bmatrix} + W$$

$$= \frac{\mu}{\sqrt{n}} \begin{bmatrix} U_1^\top \\ U_2^\top \\ \vdots \\ U_n^\top \end{bmatrix} \begin{bmatrix} U_1 & U_2 & \cdots & U_n \end{bmatrix} + W$$

U_i is i.i.d. uniformly drawn from $\{e_1, \ldots, e_k\}$

Upper and lower bounds for submatrix localization

Theorem

Let

$$\mu^{\mathrm{up}} = 2k\sqrt{\frac{\log k}{k-1}}$$
$$\mu^{\mathrm{low}} = k\sqrt{\frac{2\log(k-1)}{k-1}}$$

Then detection and reconstruction are information-theoretically possible when $\mu > \mu^{\mathrm{up}}$ and impossible when $\mu < \mu^{\mathrm{low}}$.

Upper and lower bounds for submatrix localization

Theorem

Let

$$\mu^{\mathrm{up}} = 2k\sqrt{\frac{\log k}{k-1}}$$

$$\mu^{\mathrm{low}} = k\sqrt{\frac{2\log(k-1)}{k-1}}$$

Then detection and reconstruction are information-theoretically possible when $\mu > \mu^{\mathrm{up}}$ and impossible when $\mu < \mu^{\mathrm{low}}$.

- When k is large, upper and lower bounds differ by a factor of $\sqrt{2}$
- When $k \geq 11$, IT limit is below the spectral limit $\mu^{\mathrm{spectral}} = k$
- When $k = 2$, $\mu^{\mathrm{low}} = \mu^{\mathrm{spectral}} = \mu^{\mathrm{IT}} = 2$

Special case 2: Gaussian mixture clustering with k clusters

$$Y_{m\times n} = \sqrt{\frac{\rho}{n}} \begin{array}{|c|} \hline v_1 - \bar{v} \\ \hline v_2 - \bar{v} \\ \hline \vdots \\ \hline v_k - \bar{v} \\ \hline \end{array} + W_{m\times n}$$

- $v_1, \ldots, v_k \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \mathbf{I}_{n\times n})$ and $\bar{v} = (1/k)\sum_s v_s$

Special case 2: Gaussian mixture clustering with k clusters

$$Y_{m\times n} = \sqrt{\frac{\rho}{n}} \begin{bmatrix} v_1 - \bar{v} \\ v_2 - \bar{v} \\ \vdots \\ v_k - \bar{v} \end{bmatrix} + W_{m\times n}$$

- $v_1, \ldots, v_k \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \mathbf{I}_{n\times n})$ and $\bar{v} = (1/k)\sum_s v_s$
- Cluster center separation $\approx \sqrt{2n\rho}$
- Let $m = \alpha n$ for a fixed $\alpha > 0$

Upper and lower bounds for Gaussian mixture clustering

Theorem

Let

$$\rho^{\mathrm{up}} = \frac{2k}{k-1}\left(\sqrt{\frac{k\log k}{\alpha}} + \log k\right)$$

$$\rho^{\mathrm{low}} = k\sqrt{\frac{2\log(k-1)}{(k-1)\alpha}}$$

Then detection and reconstruction are possible when $\rho > \rho^{\mathrm{up}}$ *and impossible when* $\rho < \rho^{\mathrm{low}}$*.*

Upper and lower bounds for Gaussian mixture clustering

Theorem

Let

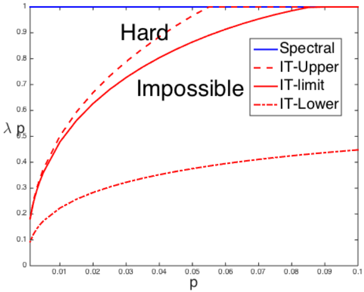
$$\rho^{\mathrm{up}} = \frac{2k}{k-1}\left(\sqrt{\frac{k\log k}{\alpha}} + \log k\right)$$

$$\rho^{\mathrm{low}} = k\sqrt{\frac{2\log(k-1)}{(k-1)\alpha}}$$

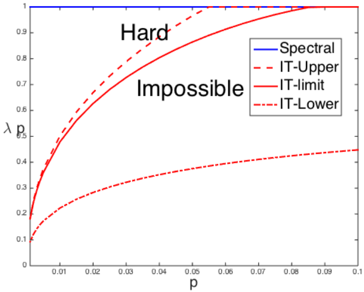
Then detection and reconstruction are possible when $\rho > \rho^{\mathrm{up}}$ and impossible when $\rho < \rho^{\mathrm{low}}$.

- When k is large, upper and lower bounds differ by a factor of $\sqrt{2}$
- When $k \geq 26$, IT limit is below the spectral limit $\rho^{\mathrm{spectral}} = \frac{k}{\sqrt{\alpha}}$
- When $k = 2$, $\rho^{\mathrm{low}} = \rho^{\mathrm{spectral}} = \rho^{\mathrm{IT}} = \frac{2}{\sqrt{\alpha}}$

Conclusion and remarks

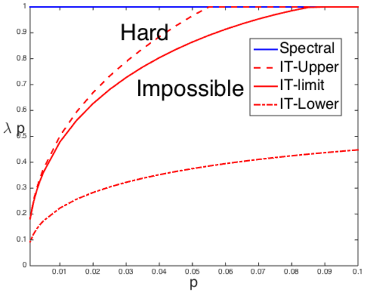


Conclusion and remarks



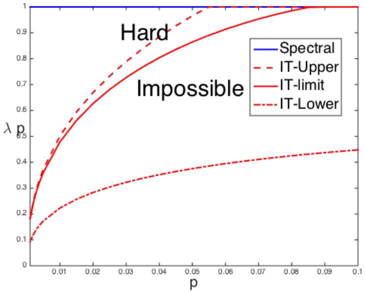
- First and second moment method: powerful tools to locate IT-limit [Abbe-Sandon 15] [Banks-Moore-Neeman-Netrapalli 16]

Conclusion and remarks



- First and second moment method: powerful tools to locate IT-limit [Abbe-Sandon 15] [Banks-Moore-Neeman-Netrapalli 16]
- Channel universality and Interpolation method

Conclusion and remarks



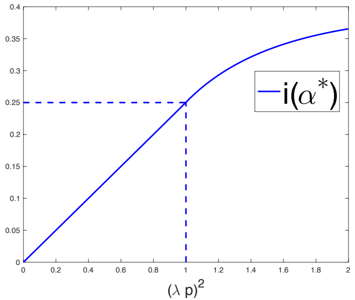
- First and second moment method: powerful tools to locate IT-limit [Abbe-Sandon 15] [Banks-Moore-Neeman-Netrapalli 16]
- Channel universality and Interpolation method
- Open question: computational limit?

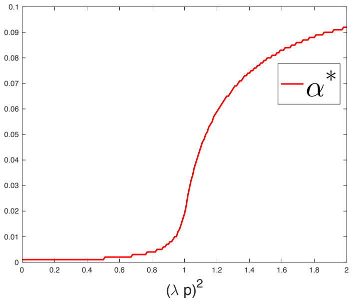
References

References

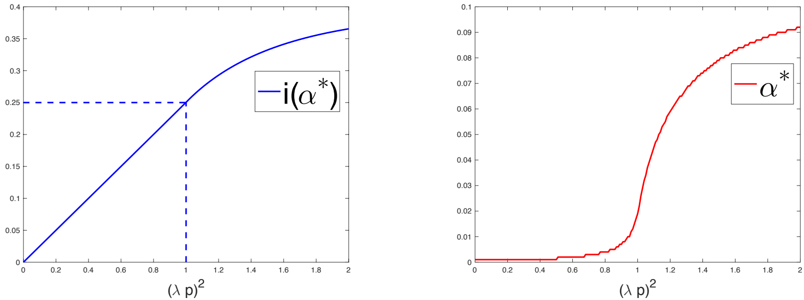
- J. Banks & C. Moore & R. Vershynin & J. X. (2016). *Information-theoretic bounds and phase transitions in clustering, sparse PCA, and submatrix localization*. arXiv:1607.05222
- F. Krzakala & J. X. & L. Zdeborová (2016). *Mutual Information in Rank-One Matrix Estimation*. arXiv:1603.08447 (ITW '16)

Dense regime: $p = 0.1$





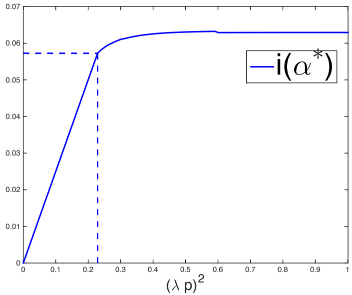
Dense regime: $p = 0.1$

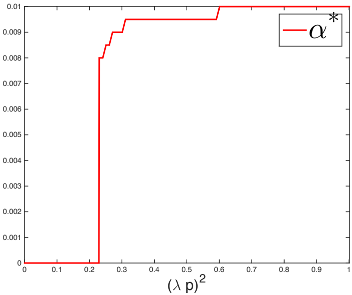


Second-order phase transition: discontinuity of 2nd derivative of $i_{RS}(\alpha^*)$

IT limit coincides with spectral limit; similar to binary symmetric SBM

Sparse regime: $p = 0.01$





First-order phase transition: discontinuity of 1st derivative of $i_{RS}(\alpha^*)$

IT limit falls below spectral limit; similar to SBM with $k > 4$ communities