Statistical mechanics, graph estimation and semidefinite programming

Andrea Montanari

Stanford University

September 29, 2015

Plan

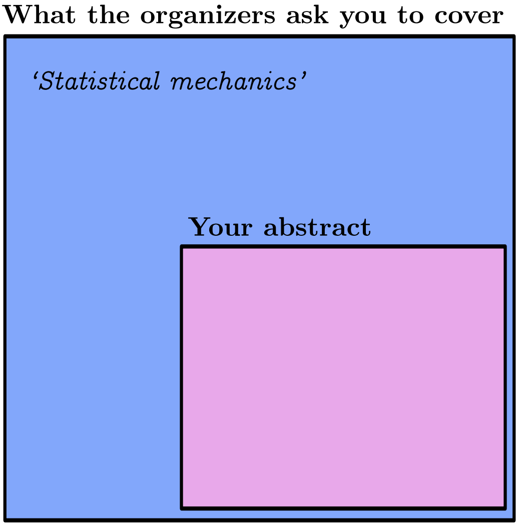
Planning for a tutorial

Planning for a tutorial

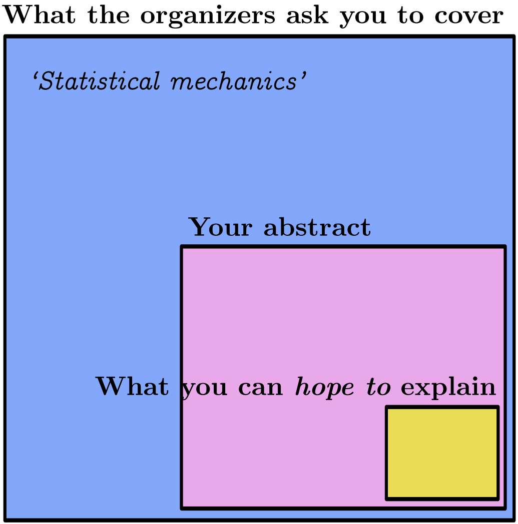
What the organizers ask you to cover

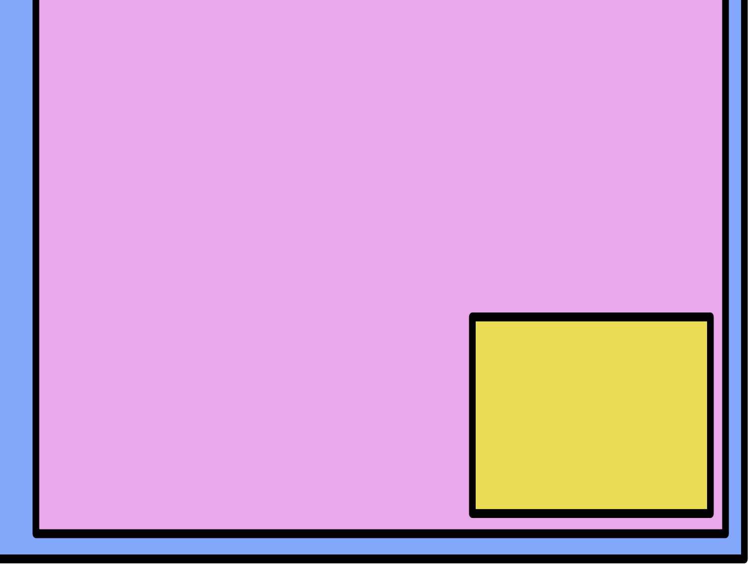
'Statistical mechanics'

Planning for a tutorial



Planning for a tutorial





Universality and the Lindeberg method

What is this tutorial about

Universality:

- The macroscopic properties of a statistical physics system are insensitive to the microscopic details of the model.

Lindeberg method:

- A technique to prove some universality properties.

Outline

References

Applications to Computer Science/Information Theory:

- A. Dembo, A. Montanari, S. Sen (2015), *Extremal cuts of sparse random graphs*, arXiv:1503.03923
- A. Montanari, S. Sen (2015), *Semidefinite Programs on Sparse Random Graphs*, arXiv:1504.05910
- Y. Deshpande, E. Abbe, A. Montanari (2015), *Asymptotic mutual information for the two-groups stochastic block model*, arXiv:1507.08685
- S.B. Korada, A. Montanari (2011), *Applications of the Lindeberg principle in communications and statistical learning*, IEEE Transactions on Information Theory, 57 (4), 2440-2450

References

A nice application to physics:

- P. Carmona, Y. Hu (2006), *Universality in Sherrington-Kirkpatrick's spin glass model*,
  Annales de l'IHP Probabilités et statistiques, 42, 2006

A nice 'pre-packaged' tool:

- S. Chatterjee (2005), *A simple invariance theorem*, arXiv:0508213

[Many other applications!!]

Caveat: This is going to be technical!

Warm up: CLT via Lindeberg method

CLT

- $X_1, X_2, \dots X_N$ independent, $\mathbb{E}[X_i] = 0$, $\mathbb{E}[X_i^2] = 1$.

We want to prove

$$\frac{1}{\sqrt{N}} \sum_{i=1}^{N} X_i \stackrel{\mathrm{d}}{\Rightarrow} \mathrm{N}(0,1)\,.$$

Idea:

- Easy for $Z_1, Z_2, \dots Z_N \sim_{i.i.d.} \mathrm{N}(0,1)$.
- Prove, for h smooth

$$\left|\mathbb{E}h\Big(\frac{1}{\sqrt{n}} \sum_{i=1}^{N} X_i\Big) - \mathbb{E}h\Big(\frac{1}{\sqrt{N}} \sum_{i=1}^{N} Z_i\Big)\right| \le o_n(1)$$

CLT

- $X_1, X_2, \dots X_N$ independent, $\mathbb{E}[X_i] = 0$, $\mathbb{E}[X_i^2] = 1$.

We want to prove

$$\frac{1}{\sqrt{N}} \sum_{i=1}^{N} X_i \overset{\mathrm{d}}{\Rightarrow} \mathrm{N}(0,1)\,.$$

Idea:

- Easy for $Z_1, Z_2, \dots Z_N \sim_{i.i.d.} \mathrm{N}(0,1)$.
- Prove, for h smooth

$$\left|\mathbb{E}h\Big(\frac{1}{\sqrt{n}} \sum_{i=1}^{N} X_i\Big) - \mathbb{E}h\Big(\frac{1}{\sqrt{N}} \sum_{i=1}^{N} Z_i\Big)\right| \le o_n(1)$$

CLT

- $X_1, X_2, \dots X_N$ independent, $\mathbb{E}[X_i] = 0$, $\mathbb{E}[X_i^2] = 1$.

We want to prove

$$\frac{1}{\sqrt{N}} \sum_{i=1}^{N} X_i \stackrel{\mathrm{d}}{\Rightarrow} \mathrm{N}(0,1)\,.$$

Idea:

- Easy for $Z_1, Z_2, \dots Z_N \sim_{i.i.d.} \mathrm{N}(0,1)$.
- Prove, for h smooth

$$\left|\mathbb{E}h\Big(\frac{1}{\sqrt{n}}\sum_{i=1}^{N} X_i\Big) - \mathbb{E}h\Big(\frac{1}{\sqrt{N}}\sum_{i=1}^{N} Z_i\Big)\right| \le o_n(1)$$

Side note

- Classically, a whole class of test functions h.

- Here: Work with a specific h.

Side note

- Classically, a whole class of test functions h.

- Here: Work with a specific h.

Notation

$$\mathbf{x} = (x_1, x_2, \dots, x_N) \in \mathbb{R}^N$$

$$f(\mathbf{x}) \equiv h\Big(\frac{1}{\sqrt{N}} \sum_{i=1}^{N} x_i\Big)$$

We want to prove

$$\big|\mathbb{E}f(\mathbf{X}) - \mathbb{E}f(\mathbf{Z})\big| \le o_n(1)$$

Notation

$\mathbf{x} = (x_1, x_2, \dots, x_N) \in \mathbb{R}^N$

$$f(\mathbf{x}) \equiv h\Big(\frac{1}{\sqrt{N}} \sum_{i=1}^{N} x_i\Big)$$

We want to prove

$$\big|\mathbb{E} f(\mathbf{X}) - \mathbb{E} f(\mathbf{Z})\big| \le o_n(1)$$

Key idea ('smart path')

$$\mathbf{X}^{(i)} = (X_1, X_2, \dots, X_{i-1}, X_i, Z_{i+1}, \dots, Z_n)$$

$$\big|\mathbb{E}f(\mathbf{X}) - \mathbb{E}f(\mathbf{Z})\big| \le \sum_{i=1}^{N} \big|\mathbb{E}f(\mathbf{X}^{(i)}) - \mathbb{E}f(\mathbf{X}^{(i-1)})\big|$$

This is brilliant:

- $\mathbf{X}$ and $\mathbf{Z}$ are difficult to relate to each other.
- $\mathbf{X}^{(i)}$ and $\mathbf{X}^{(i-1)}$ are easy to connect! (Taylor expansion!)

Key idea ('smart path')

$$\mathbf{X}^{(i)} = (X_1, X_2, \dots, X_{i-1}, X_i, Z_{i+1}, \dots, Z_n)$$

$$\big|\mathbb{E}f(\mathbf{X}) - \mathbb{E}f(\mathbf{Z})\big| \le \sum_{i=1}^{N} \big|\mathbb{E}f(\mathbf{X}^{(i)}) - \mathbb{E}f(\mathbf{X}^{(i-1)})\big|$$

This is brilliant:

- $\mathbf{X}$ and $\mathbf{Z}$ are difficult to relate to each other.
- $\mathbf{X}^{(i)}$ and $\mathbf{X}^{(i-1)}$ are easy to connect! (Taylor expansion!)

Key idea ('smart path')

$$\mathbf{X}^{(i)} = (X_1, X_2, \dots, X_{i-1}, X_i, Z_{i+1}, \dots, Z_n)$$

$$\left|\mathbb{E}f(\mathbf{X}) - \mathbb{E}f(\mathbf{Z})\right| \le \sum_{i=1}^{N} \left|\mathbb{E}f(\mathbf{X}^{(i)}) - \mathbb{E}f(\mathbf{X}^{(i-1)})\right|$$

This is brilliant:

- $\mathbf{X}$ and $\mathbf{Z}$ are difficult to relate to each other.
- $\mathbf{X}^{(i)}$ and $\mathbf{X}^{(i-1)}$ are easy to connect! (Taylor expansion!)

Taylor expansion ($x \geq 0$)

$$g(x) = g(0) + xg'(0) + \int_0^x (x - t)\, g''(t)\, \mathrm{d}t$$

$$\left| g(x) - \left\{ g(0) + g'(0)\, x + \frac{1}{2}\, g''(0)x^2 \right\} \right| \leq \int_0^x (x - t)\, |g''(t) - g''(0)|\, \mathrm{d}t$$

$$\leq (\|g''\|_\infty x^2) \vee (\frac{1}{6} \|g'''\|_\infty x^3)$$

$$\left| g(x) - \left\{ g(0) + g'(0)\, x + \frac{1}{2}\, g''(0)x^2 \right\} \right| \leq \mathcal{R}$$

$$\mathcal{R} \equiv \|g''\|_\infty x^2 \mathbf{1}_{\{|x|>K\}} + \frac{1}{6} \|g'''\|_\infty x^3 \mathbf{1}_{\{|x|\leq K\}}$$

Taylor expansion ($x \geq 0$)

$$g(x) = g(0) + xg'(0) + \int_0^x (x-t)\, g''(t)\, \mathrm{d}t$$

$$\Big| g(x) - \{g(0) + g'(0)\, x + \frac{1}{2}\, g''(0) x^2\} \Big| \leq \int_0^x (x-t)\, |g''(t) - g''(0)|\, \mathrm{d}t$$
$$\leq (\|g''\|_\infty x^2) \vee (\frac{1}{6} \|g'''\|_\infty x^3)$$

$$\Big| g(x) - \{g(0) + g'(0)\, x + \frac{1}{2}\, g''(0) x^2\} \Big| \leq \mathcal{R}$$
$$\mathcal{R} \equiv \|g''\|_\infty x^2 \mathbf{1}_{\{|x| > K\}} + \frac{1}{6} \|g'''\|_\infty x^3 \mathbf{1}_{\{|x| \leq K\}}$$

Taylor expansion ($x \geq 0$)

$$g(x) = g(0) + xg'(0) + \int_0^x (x-t)\, g''(t)\,\mathrm{d}t$$

$$\left| g(x) - \{ g(0) + g'(0)\,x + \frac{1}{2}\,g''(0)x^2 \} \right| \leq \int_0^x (x-t)\,|g''(t) - g''(0)|\,\mathrm{d}t$$
$$\leq (\|g''\|_\infty x^2) \vee (\frac{1}{6}\|g'''\|_\infty x^3)$$

$$\left| g(x) - \{ g(0) + g'(0)\,x + \frac{1}{2}\,g''(0)x^2 \} \right| \leq \mathcal{R}$$
$$\mathcal{R} \equiv \|g''\|_\infty x^2 \mathbf{1}_{\{|x|>K\}} + \frac{1}{6}\|g'''\|_\infty x^3 \mathbf{1}_{\{|x|\leq K\}}$$

Consider now a single step

$$\mathbf{X}^{(i)} = (X_1, X_2, \dots, X_{i-1}, X_i, Z_{i+1}, \dots, Z_n)$$

$$\mathbf{X}^{(i-1)} = (X_1, X_2, \dots, X_{i-1}, Z_i, Z_{i+1}, \dots, Z_n)$$

$$\mathbf{X}_0^{(i)} = (X_1, X_2, \dots, X_{i-1}, 0, Z_{i+1}i, \dots, Z_n)$$

Taylor expansion around $\mathbf{X}_0^{(i)}$

$$f(\mathbf{X}^{(i)}) = f(\mathbf{X}_0^{(i)}) + X_i \partial_i f(\mathbf{X}_0^{(i)}) + \frac{1}{2} X_i^2 \partial_i^2 f(\mathbf{X}_0^{(i)}) \pm \mathcal{R}_i$$

$$f(\mathbf{X}^{(i-1)}) = f(\mathbf{X}_0^{(i)}) + Z_i \partial_i f(\mathbf{X}_0^{(i)}) + \frac{1}{2} Z_i^2 \partial_i^2 f(\mathbf{X}_0^{(i)}) \pm \widetilde{\mathcal{R}}_i$$

$$\mathcal{R}_i \equiv \|\partial_i^2 f\|_\infty X_i^2 \mathbf{1}_{\{|X_i|>K\}} + \frac{1}{6} \|\partial_i^3 f\|_\infty X_i^3 \mathbf{1}_{\{|X_i|\le K\}},$$

$$\widetilde{\mathcal{R}}_i \equiv \|\partial_i^2 f\|_\infty Z_i^2 \mathbf{1}_{\{|Z_i|>K\}} + \frac{1}{6} \|\partial_i^3 f\|_\infty Z_i^3 \mathbf{1}_{\{|Z_i|\le K\}}.$$

Consider now a single step

$$\mathbf{X}^{(i)} = (X_1, X_2, \dots, X_{i-1}, X_i, Z_{i+1}, \dots, Z_n)$$
$$\mathbf{X}^{(i-1)} = (X_1, X_2, \dots, X_{i-1}, Z_i, Z_{i+1}, \dots, Z_n)$$
$$\mathbf{X}_0^{(i)} = (X_1, X_2, \dots, X_{i-1}, 0, Z_{i+1} i, \dots, Z_n)$$

Taylor expansion around $\mathbf{X}_0^{(i)}$

$$f(\mathbf{X}^{(i)}) = f(\mathbf{X}_0^{(i)}) + X_i \partial_i f(\mathbf{X}_0^{(i)}) + \frac{1}{2} X_i^2 \partial_i^2 f(\mathbf{X}_0^{(i)}) \pm \mathcal{R}_i$$
$$f(\mathbf{X}^{(i-1)}) = f(\mathbf{X}_0^{(i)}) + Z_i \partial_i f(\mathbf{X}_0^{(i)}) + \frac{1}{2} Z_i^2 \partial_i^2 f(\mathbf{X}_0^{(i)}) \pm \widetilde{\mathcal{R}}_i$$

$$\mathcal{R}_i \equiv \|\partial_i^2 f\|_\infty X_i^2 \mathbf{1}_{\{|X_i|>K\}} + \frac{1}{6} \|\partial_i^3 f\|_\infty X_i^3 \mathbf{1}_{\{|X_i|\le K\}},$$
$$\widetilde{\mathcal{R}}_i \equiv \|\partial_i^2 f\|_\infty Z_i^2 \mathbf{1}_{\{|Z_i|>K\}} + \frac{1}{6} \|\partial_i^3 f\|_\infty Z_i^3 \mathbf{1}_{\{|Z_i|\le K\}}.$$

Consider now a single step

$$f(\mathbf{X}^{(i)}) = f(\mathbf{X}_0^{(i)}) + X_i \partial_i f(\mathbf{X}_0^{(i)}) + \frac{1}{2} X_i^2 \partial_i^2 f(\mathbf{X}_0^{(i)}) \pm \mathcal{R}_i$$

$$f(\mathbf{X}^{(i-1)}) = f(\mathbf{X}_0^{(i)}) + Z_i \partial_i f(\mathbf{X}_0^{(i)}) + \frac{1}{2} Z_i^2 \partial_i^2 f(\mathbf{X}_0^{(i)}) \pm \widetilde{\mathcal{R}}_i$$

Take expectation with respect to X_i, Z_i

$$\mathbb{E}_i\{f(\mathbf{X}^{(i)}) - f(\mathbf{X}^{(i-1)})\} \leq$$

$$\leq \partial_i f(\mathbf{X}_0^{(i)}) \mathbb{E}_i(X_i - Z_i) + \frac{1}{2} \partial_i^2 f(\mathbf{X}_0^{(i)}) \mathbb{E}_i(X_i^2 - Z_i^2) + \mathbb{E}_i(\mathcal{R}_i + \widetilde{\mathcal{R}}_i)$$

$$= \mathbb{E}_i(\mathcal{R}_i + \widetilde{\mathcal{R}}_i)$$

Consider now a single step

$$f(\mathbf{X}^{(i)}) = f(\mathbf{X}_0^{(i)}) + X_i \partial_i f(\mathbf{X}_0^{(i)}) + \frac{1}{2} X_i^2 \partial_i^2 f(\mathbf{X}_0^{(i)}) \pm \mathcal{R}_i$$

$$f(\mathbf{X}^{(i-1)}) = f(\mathbf{X}_0^{(i)}) + Z_i \partial_i f(\mathbf{X}_0^{(i)}) + \frac{1}{2} Z_i^2 \partial_i^2 f(\mathbf{X}_0^{(i)}) \pm \widetilde{\mathcal{R}}_i$$

Take expectation with respect to X_i, Z_i

$$\mathbb{E}_i\{f(\mathbf{X}^{(i)}) - f(\mathbf{X}^{(i-1)})\} \le$$

$$\le \partial_i f(\mathbf{X}_0^{(i)}) \mathbb{E}_i(X_i - Z_i) + \frac{1}{2} \partial_i^2 f(\mathbf{X}_0^{(i)}) \mathbb{E}_i(X_i^2 - Z_i^2) + \mathbb{E}_i\big(\mathcal{R}_i + \widetilde{\mathcal{R}}_i\big)$$

$$= \mathbb{E}_i\big(\mathcal{R}_i + \widetilde{\mathcal{R}}_i\big)$$

Summing the terms back together

Lemma

$$\mathbb{E}\{f(\mathbf{X}) - f(\mathbf{Z})\} \le T_2(K) \max_{i\in n} \|\partial_i^2 f\|_\infty + T_3(K) \max_{i\in n} \|\partial_i^3 f\|_\infty \,,$$

$$T_2(K) \equiv \sum_{i=1}^{N} \Big\{\mathbb{E}[X_i^2 \mathbf{1}_{\{|X_i|>K\}}] + \mathbb{E}[Z_i^2 \mathbf{1}_{\{|Z_i|>K\}}]\Big\}$$

$$T_3(K) \equiv \frac{1}{6} \sum_{i=1}^{N} \Big\{\mathbb{E}[|X_i|^3 \mathbf{1}_{\{|X_i|\le K\}}] + \mathbb{E}[|Z_i|^3 \mathbf{1}_{\{|Z_i|\le K\}}]\Big\}$$

Morally: I can replace the X_i's by Gaussians if

- $f(\,\cdot\,)$ does not depend to strongly on any of the coordinates.
- X_i's have light tails

Summing the terms back together

Lemma

$$\mathbb{E}\{f(\mathbf{X}) - f(\mathbf{Z})\} \le T_2(K) \max_{i\in n} \|\partial_i^2 f\|_\infty + T_3(K) \max_{i\in n} \|\partial_i^3 f\|_\infty ,$$

$$T_2(K) \equiv \sum_{i=1}^{N} \Big\{ \mathbb{E}[X_i^2 \mathbf{1}_{\{|X_i|>K\}}] + \mathbb{E}[Z_i^2 \mathbf{1}_{\{|Z_i|>K\}}] \Big\}$$

$$T_3(K) \equiv \frac{1}{6} \sum_{i=1}^{N} \Big\{ \mathbb{E}[|X_i|^3 \mathbf{1}_{\{|X_i|\le K\}}] + \mathbb{E}[|Z_i|^3 \mathbf{1}_{\{|Z_i|\le K\}}] \Big\}$$

Morally: I can replace the X_i's by Gaussians if

- $f(\,\cdot\,)$ does not depend to strongly on any of the coordinates.
- X_i's have light tails

Getting the CLT $|h''(x)|, |h'''(x)| \le 1$, $K = \varepsilon\sqrt{N}$

$$f(\mathbf{x}) = h\Big(\frac{1}{\sqrt{N}}\sum_{i=1}^N x_i\Big)\,,$$
$$\|\partial_i^2 f\|_\infty = \frac{1}{N}\|h''\|_\infty\,,$$
$$\|\partial_i^3 f\|_\infty = \frac{1}{N^{3/2}}\|h'''\|_\infty\,.$$

$$\text{Bound} = \frac{1}{N}\sum_{i=1}^N \mathbb{E}[X_i^2 \mathbf{1}_{\{|X_i|>\varepsilon\sqrt{n}\}}] + \frac{1}{N^{3/2}}\sum_{i=1}^N \mathbb{E}[|X_i|^3 \mathbf{1}_{\{|X_i|\le\varepsilon\sqrt{N}\}}] + Z\text{'s}$$
$$\le \frac{1}{N}\sum_{i=1}^N \mathbb{E}[X_i^2 \mathbf{1}_{\{|X_i|>\varepsilon\sqrt{N}\}}] + \frac{\varepsilon}{N}\sum_{i=1}^N \mathbb{E}[|X_i|^2] + \dots$$
$$\le \frac{1}{N}\sum_{i=1}^N \mathbb{E}[X_i^2 \mathbf{1}_{\{|X_i|>\varepsilon\sqrt{N}\}}] + \varepsilon + \dots$$

Getting the CLT $|h''(x)|, |h'''(x)| \le 1$, $K = \varepsilon\sqrt{N}$

$$
\begin{aligned}
f(\mathbf{x}) &= h\Big(\frac{1}{\sqrt{N}}\sum_{i=1}^{N} x_i\Big)\,,\\
\|\partial_i^2 f\|_\infty &= \frac{1}{N}\|h''\|_\infty\,,\\
\|\partial_i^3 f\|_\infty &= \frac{1}{N^{3/2}}\|h'''\|_\infty\,.
\end{aligned}
$$

$$
\begin{aligned}
\text{Bound} &= \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}[X_i^2\mathbf{1}_{\{|X_i|>\varepsilon\sqrt{n}\}}] + \frac{1}{N^{3/2}}\sum_{i=1}^{N}\mathbb{E}[|X_i|^3\mathbf{1}_{\{|X_i|\le\varepsilon\sqrt{N}\}}] + Z\text{'s}\\
&\le \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}[X_i^2\mathbf{1}_{\{|X_i|>\varepsilon\sqrt{N}\}}] + \frac{\varepsilon}{N}\sum_{i=1}^{N}\mathbb{E}[|X_i|^2] + \dots\\
&\le \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}[X_i^2\mathbf{1}_{\{|X_i|>\varepsilon\sqrt{N}\}}] + \varepsilon + \dots
\end{aligned}
$$

Getting the CLT

Theorem (Lindeberg, 1922)

Assume that $N^{-1}\sum_{i=1}^{n}\mathbb{E}[X_i^2\mathbf{1}_{\{|X_i|>\varepsilon\sqrt{N}\}}]\to 0$ *for any* $\varepsilon>0$. *Then*

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}X_i\overset{\mathrm{d}}{\Rightarrow}\mathrm{N}(0,1)\,.$$

[To conclude...]

Remarks

- Introduced by Lindeberg (1922)
- Used for non-sum functions in mathematical physics, e.g. Guerra-Toninelli (2004)
- Chatterjee (2005) points out that Lemma holds for any $f : \mathbb{R}^N \to \mathbb{R}$
- Several variants. Example, $\theta \in [0, \pi/2]$

$$\mathbf{X}(\theta) \equiv \mathbf{Z} \cos\theta + \mathbf{X} \sin\theta$$

('smart path method', 'interpolation', ...)

Remarks

- Introduced by Lindeberg (1922)
- Used for non-sum functions in mathematical physics, e.g. Guerra-Toninelli (2004)
- Chatterjee (2005) points out that Lemma holds for any $f : \mathbb{R}^N \to \mathbb{R}$
- Several variants. Example, $\theta \in [0, \pi/2]$

$$\mathbf{X}(\theta) \equiv \mathbf{Z}\,\cos\theta + \mathbf{X}\,\sin\theta$$

('smart path method', 'interpolation', ...)

Remarks

- Introduced by Lindeberg (1922)
- Used for non-sum functions in mathematical physics, e.g. Guerra-Toninelli (2004)
- Chatterjee (2005) points out that Lemma holds for any $f : \mathbb{R}^N \to \mathbb{R}$
- Several variants. Example, $\theta \in [0, \pi/2]$

$$\mathbf{X}(\theta) \equiv \mathbf{Z} \cos\theta + \mathbf{X} \sin\theta$$

('smart path method', 'interpolation', ...)

Remarks

- Introduced by Lindeberg (1922)
- Used for non-sum functions in mathematical physics, e.g. Guerra-Toninelli (2004)
- Chatterjee (2005) points out that Lemma holds for any $f : \mathbb{R}^N \to \mathbb{R}$
- Several variants. Example, $\theta \in [0, \pi/2]$

$$\mathbf{X}(\theta) \equiv \mathbf{Z}\,\cos\theta + \mathbf{X}\,\sin\theta$$

('smart path method', 'interpolation', ...)

Remarks

- Introduced by Lindeberg (1922)
- Used for non-sum functions in mathematical physics, e.g. Guerra-Toninelli (2004)
- Chatterjee (2005) points out that Lemma holds for any $f : \mathbb{R}^N \to \mathbb{R}$
- Several variants. Example, $\theta \in [0, \pi/2]$

$$\mathbf{X}(\theta) \equiv \mathbf{Z} \cos\theta + \mathbf{X} \sin\theta$$

('smart path method', 'interpolation', ...)

Extremal cuts of random graphs

Background: Random graph models

Erdös-Renyi Random graph $G = (V, E) \sim \mathcal{G}(n, \gamma/n)$

- $|V| = n$ vertices
- Each edge present with probability γ/n

Random regular graph $G = (V, E) \sim \mathcal{G}^{\text{reg}}(n, \gamma)$

- $|V| = n$ vertices
- Uniformly random among all graphs with $\deg(i) = \gamma$ for all $i \in V$.

Average degree $\gamma = O(1)$

Background: Random graph models

Erdös-Renyi Random graph $G = (V, E) \sim \mathcal{G}(n, \gamma/n)$

- $|V| = n$ vertices
- Each edge present with probability γ/n

Random regular graph $G = (V, E) \sim \mathcal{G}^{\rm reg}(n, \gamma)$

- $|V| = n$ vertices
- Uniformly random among all graphs with $\deg(i) = \gamma$ for all $i \in V$.

Average degree $\gamma = O(1)$

Background: Random graph models

Erdös-Renyi Random graph $G = (V, E) \sim \mathcal{G}(n, \gamma/n)$

- $|V| = n$ vertices
- Each edge present with probability γ/n

Random regular graph $G = (V, E) \sim \mathcal{G}^{\rm reg}(n, \gamma)$

- $|V| = n$ vertices
- Uniformly random among all graphs with $\deg(i) = \gamma$ for all $i \in V$.

Average degree $\gamma = O(1)$

Background: Random graph models

Erdös-Renyi Random graph $G = (V, E) \sim \mathcal{G}(n, \gamma/n)$

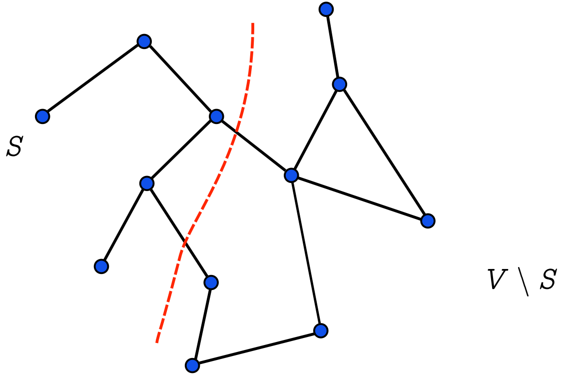
- $|V| = n$ vertices
- Each edge present with probability γ/n

Random regular graph $G = (V, E) \sim \mathcal{G}^{\text{reg}}(n, \gamma)$

- $|V| = n$ vertices
- Uniformly random among all graphs with $\deg(i) = \gamma$ for all $i \in V$.

Average degree $\gamma = O(1)$

Cuts



$$\mathrm{cut}_G(S) \equiv \Big|\{(i,j) \in E :\; i \in S,\, j \in V \setminus S\}\Big|$$

Interesting for many reasons

- Clustering similarity matrices
- Graph layout
- Community structure in social networks
- . . .

Extremal cuts

Minimum bisection

$$\mathsf{mcut}(G) = \min\left\{\mathrm{cut}_G(S)\ :\ \ S \subseteq V\,,\ |S| = n/2\ \right\}$$

Maximum bisection

$$\mathsf{MCUT}(G) = \max\left\{\mathrm{cut}_G(S)\ :\ \ S \subseteq V\,,\ |S| = n/2\ \right\}$$

Maximum Cut

$$\mathsf{MaxCut}(G) = \max\left\{\mathrm{cut}_G(S)\ :\ \ S \subseteq V\ \right\}$$

Extremal cuts

Minimum bisection

$$\text{mcut}(G) = \min\left\{\text{cut}_G(S)\ :\ S \subseteq V\,,\ |S| = n/2\right\}$$

Maximum bisection

$$\text{MCUT}(G) = \max\left\{\text{cut}_G(S)\ :\ S \subseteq V\,,\ |S| = n/2\right\}$$

Maximum Cut

$$\text{MaxCut}(G) = \max\left\{\text{cut}_G(S)\ :\ S \subseteq V\right\}$$

Extremal cuts

Minimum bisection

$$\mathsf{mcut}(G) = \min\left\{\mathrm{cut}_G(S)\ :\ S \subseteq V\,,\ |S| = n/2\right\}$$

Maximum bisection

$$\mathsf{MCUT}(G) = \max\left\{\mathrm{cut}_G(S)\ :\ S \subseteq V\,,\ |S| = n/2\right\}$$

Maximum Cut

$$\mathsf{MaxCut}(G) = \max\left\{\mathrm{cut}_G(S)\ :\ S \subseteq V\right\}$$

A long history

- Bollobas, 1984
- Alon, 1997
- Coppersmith, Gamarnik, Hajiaghayi, Sorkin, 2004
- Díaz, Serna, and Wormald, 2007
- Daudé, Martínez, Rasendrahasina, Ravelomanana, 2012
- Gamarnik, Li, 2014
- ...

Typical result:
If $G \sim \mathcal{G}(n, \gamma/n)$ then, with high probability

$$\frac{n\gamma}{4} + C_1 n\sqrt{\gamma} \le \mathsf{MCUT}(G) \le \frac{n\gamma}{4} + C_2 n\sqrt{\gamma}$$

A long history

- Bollobas, 1984
- Alon, 1997
- Coppersmith, Gamarnik, Hajiaghayi, Sorkin, 2004
- Díaz, Serna, and Wormald, 2007
- Daudé, Martínez, Rasendrahasina, Ravelomanana, 2012
- Gamarnik, Li, 2014
- ...

Typical result:
If $G \sim \mathcal{G}(n, \gamma/n)$ then, with high probability

$$\frac{n\gamma}{4} + C_1 n\sqrt{\gamma} \le \mathsf{MCUT}(G) \le \frac{n\gamma}{4} + C_2 n\sqrt{\gamma}$$

Insights from statistical physics

Conjectures

- Fu, Anderson, 1986
- Mézard, Parisi, 2001
- Zdéborova, Boettcher, 2010

Theorem (Franz, Leone, Toninelli, 2003, Bayati, Gamarnik, Tetali, 2009)

The following limit exists, for $G_n \sim \mathcal{G}(n, \gamma/n), \mathcal{G}^{\rm reg}(n, \gamma)$

$$\lim_{n\to\infty} \frac{1}{n} \mathsf{MaxCut}(G_n)$$

Subadditivity: no limit value.

Insights from statistical physics

Conjectures

- Fu, Anderson, 1986
- Mézard, Parisi, 2001
- Zdéborova, Boettcher, 2010

Theorem (Franz, Leone, Toninelli, 2003, Bayati, Gamarnik, Tetali, 2009)

The following limit exists, for $G_n \sim \mathcal{G}(n, \gamma/n), \mathcal{G}^{\rm reg}(n, \gamma)$

$$\lim_{n\to\infty} \frac{1}{n} \mathsf{MaxCut}(G_n)$$

Subadditivity: no limit value.

Insights from statistical physics

Conjectures

- Fu, Anderson, 1986
- Mézard, Parisi, 2001
- Zdéborova, Boettcher, 2010

Theorem (Franz, Leone, Toninelli, 2003, Bayati, Gamarnik, Tetali, 2009)

The following limit exists, for $G_n \sim \mathcal{G}(n, \gamma/n), \mathcal{G}^{\rm reg}(n,\gamma)$

$$\lim_{n\to\infty} \frac{1}{n} \mathsf{MaxCut}(G_n)$$

Subadditivity: no limit value.

A classical argument (Bollobas 1984)

Fix $S \subseteq V$, $S = n/2$

- Each edge is cut with probability $1/2$

$$\mathbb{E}\text{cut}(S) = \frac{n\gamma}{4}\,.$$

- Azuma-Höeffding argument

$$\mathbb{P}\Big\{\text{cut}(S) \geq \mathbb{E}\text{cut}(S) + \Delta\Big\} \leq \exp\Big(-\frac{\Delta^2}{4n\gamma}\Big)$$

- Union bound

$$\mathbb{P}\Big\{\max_{S,|S|=n/2} \text{cut}(S) \geq \frac{n\gamma}{4} + \delta n\sqrt{\gamma}\Big\} \leq 2^n\, e^{-n\delta^2/4}\,.$$

A classical argument (Bollobas 1984)

Fix $S \subseteq V$, $S = n/2$

- Each edge is cut with probability $1/2$

$$\mathbb{E}\text{cut}(S) = \frac{n\gamma}{4}\,.$$

- Azuma-Höeffding argument

$$\mathbb{P}\Big\{\text{cut}(S) \geq \mathbb{E}\text{cut}(S) + \Delta\Big\} \leq \exp\Big(-\frac{\Delta^2}{4n\gamma}\Big)$$

- Union bound

$$\mathbb{P}\Big\{\max_{S,|S|=n/2} \text{cut}(S) \geq \frac{n\gamma}{4} + \delta n\sqrt{\gamma}\Big\} \leq 2^n\, e^{-n\delta^2/4}\,.$$

A classical argument (Bollobas 1984)

Fix $S \subseteq V$, $S = n/2$

- Each edge is cut with probability $1/2$

$$\mathbb{E}\text{cut}(S) = \frac{n\gamma}{4}\,.$$

- Azuma-Höeffding argument

$$\mathbb{P}\Big\{\text{cut}(S) \ge \mathbb{E}\text{cut}(S) + \Delta\Big\} \le \exp\Big(-\frac{\Delta^2}{4n\gamma}\Big)$$

- Union bound

$$\mathbb{P}\Big\{\max_{S,|S|=n/2} \text{cut}(S) \ge \frac{n\gamma}{4} + \delta n\sqrt{\gamma}\Big\} \le 2^n\, e^{-n\delta^2/4}\,.$$

A classical argument (Bollobas 1984)

Fix $S \subseteq V$, $S = n/2$

- Each edge is cut with probability $1/2$

$$\mathbb{E}\text{cut}(S) = \frac{n\gamma}{4}\,.$$

- Azuma-Höeffding argument

$$\mathbb{P}\Big\{\text{cut}(S) \ge \mathbb{E}\text{cut}(S) + \Delta\Big\} \le \exp\Big(-\frac{\Delta^2}{4n\gamma}\Big)$$

- Union bound

$$\mathbb{P}\Big\{\max_{S,|S|=n/2} \text{cut}(S) \ge \frac{n\gamma}{4} + \delta n\sqrt{\gamma}\Big\} \le 2^n\, e^{-n\delta^2/4}\,.$$

A classical argument (Bollobas 1984)

Hence, with high probability

$$\mathsf{MCUT}(G) \le \frac{n\gamma}{4} + C_2 n\sqrt{\gamma}$$

In other words

The $n\gamma/4$ term is ‘trivial’...

A new result

A new result

Theorem (Dembo, Montanari, Sen, 2015)

Assume $G \sim \mathcal{G}(n, \gamma/n)$ *or* $G \sim \mathcal{G}^{\rm reg}(n, \gamma)$. *Then, with high probability,*

$$\frac{1}{n}\mathsf{mcut}(G) = \frac{\gamma}{4} - \mathsf{P}_*\sqrt{\frac{\gamma}{4}} + o_\gamma(\sqrt{\gamma})\,,$$

$$\frac{1}{n}\mathsf{MCUT}(G) = \frac{\gamma}{4} + \mathsf{P}_*\sqrt{\frac{\gamma}{4}} + o_\gamma(\sqrt{\gamma})\,,$$

$$\frac{1}{n}\mathsf{MaxCut}(G) = \frac{\gamma}{4} + \mathsf{P}_*\sqrt{\frac{\gamma}{4}} + o_\gamma(\sqrt{\gamma})\,.$$

where $\mathsf{P}_* = \dots$ *(wait a minute).*

Remarks

- The $\sqrt{\gamma}$ term has a well defined coefficient.
- The coefficient has a formula.
- It is the same for 3 problems, 2 graph models.
- $\mathsf{P}_* \approx 0.7632$

What is P_*?

GOE random matrix:

$$J \in \mathbb{R}^{n\times n}, \quad J = J^{\mathsf{T}}, \quad (J_{ij})_{i<j} \sim \mathsf{N}(0,1), \quad J_{ii} = 0\,.$$

Sherrington-Kirkpatrick spin-glass model

$$\mathcal{H}_J(\sigma) \equiv \frac{1}{2\sqrt{n}}\langle \sigma, J\sigma\rangle\,.$$

Finally

$$\mathsf{P}_* \equiv \lim_{n\to\infty}\frac{1}{n}\mathbb{E}\max_{\sigma\in\{+1,-1\}^n}\mathcal{H}_J(\sigma)\,.$$

What is P_*?

GOE random matrix:

$$J \in \mathbb{R}^{n\times n}\,, \quad J = J^{\mathsf{T}}\,, \quad (J_{ij})_{i<j} \sim \mathrm{N}(0,1), \quad J_{ii} = 0\,.$$

Sherrington-Kirkpatrick spin-glass model

$$\mathcal{H}_J(\boldsymbol{\sigma}) \equiv \frac{1}{2\sqrt{n}}\,\langle \boldsymbol{\sigma}, J\boldsymbol{\sigma}\rangle\,.$$

Finally

$$\mathrm{P}_* \equiv \lim_{n\to\infty} \frac{1}{n}\mathbb{E} \max_{\boldsymbol{\sigma}\in\{+1,-1\}^n} \mathcal{H}_J(\boldsymbol{\sigma})\,.$$

What is P_*?

GOE random matrix:

$$J \in \mathbb{R}^{n\times n}, \quad J = J^{\mathsf{T}}, \quad (J_{ij})_{i<j} \sim \mathrm{N}(0,1), \quad J_{ii} = 0\,.$$

Sherrington-Kirkpatrick spin-glass model

$$\mathcal{H}_J(\boldsymbol{\sigma}) \equiv \frac{1}{2\sqrt{n}}\,\langle \boldsymbol{\sigma}, J\boldsymbol{\sigma}\rangle\,.$$

Finally

$$\mathsf{P}_* \equiv \lim_{n\to\infty} \frac{1}{n}\mathbb{E} \max_{\boldsymbol{\sigma}\in\{+1,-1\}^n} \mathcal{H}_J(\boldsymbol{\sigma})\,.$$

What is P_*?

$$\mathsf{P}_* \equiv \lim_{n\to\infty} \frac{1}{n} \mathbb{E} \max_{\boldsymbol{\sigma}\in\{+1,-1\}^n} \mathcal{H}_J(\boldsymbol{\sigma})\,.$$

- The limit exists (consequence of Guerra, Toninelli, 2002)
- Given by *'Parisi's formula'* (Talagrand, 2006)
- Clarifies why 'standard' combinatorial methods were unsuccessful

What is P_*?

$$\mathsf{P}_* \equiv \lim_{n\to\infty} \frac{1}{n} \mathbb{E} \max_{\boldsymbol{\sigma}\in\{+1,-1\}^n} \mathcal{H}_J(\boldsymbol{\sigma})\,.$$

- The limit exists (consequence of Guerra, Toninelli, 2002)
- Given by *'Parisi's formula'* (Talagrand, 2006)
- Clarifies why 'standard' combinatorial methods were unsuccessful

(Informal) Implication

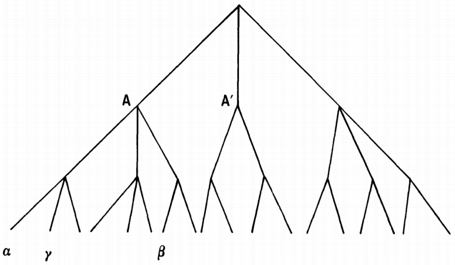


Fig. 1. — The tree of the states. The different states α, β, γ... are the extremities of the branches of the tree. The distance between two states is a monotonic function of the number of steps one has to climb along the tree to find a common ancestor.

MaxCut, max-bisection, min-bisection have ∞-RSB structure

Proof strategy: mcut

Two graph models

$$\text{SK model} \underset{\text{Lindeberg}}{\longleftrightarrow} \mathcal{G}(n, \gamma/n) \underset{\text{coupling}}{\leftrightarrow} \mathcal{G}^{\text{reg}}(n, \gamma)$$

Two graph models

$$\text{SK model} \underset{\text{Lindeberg}}{\longleftrightarrow} \mathcal{G}(n, \gamma/n) \underset{\text{coupling}}{\leftrightarrow} \mathcal{G}^{\text{reg}}(n, \gamma)$$

min-bisection

Hamiltonian $\mathcal{H}_{\mathsf{mcut}} : \Omega_n \to \mathbb{R}$

$$\mathcal{H}_{\mathsf{mcut}}(\boldsymbol{\sigma}) = \frac{1}{\sqrt{\gamma}} \sum_{(i,j)\in E} \sigma_i \sigma_j \,,$$

$$\Omega_n = \Big\{ \boldsymbol{\sigma} \in \{+1,-1\}^n : \; \sum_{i=1}^{n} \sigma_i = 0 \Big\} \,.$$

$$\begin{aligned} \mathsf{mcut}(G) &= \min_{\sigma\in\Omega_n} \sum_{(i,j)\in E} \Big(\frac{1-\sigma_i\sigma_j}{2}\Big) \\ &= \frac{1}{2}|E| - \frac{1}{2} \max_{\sigma\in\Omega_n} \sum_{(i,j)\in E} \sigma_i\sigma_j \\ &= \frac{n\gamma}{4} - \sqrt{\frac{\gamma}{4}} \max_{\sigma\in\Omega_n} \mathcal{H}_{\mathsf{mcut}}(\sigma) \,. \end{aligned}$$

min-bisection

Hamiltonian $\mathcal{H}_{\text{mcut}} : \Omega_n \to \mathbb{R}$

$$\mathcal{H}_{\text{mcut}}(\boldsymbol{\sigma}) = \frac{1}{\sqrt{\gamma}} \sum_{(i,j)\in E} \sigma_i \sigma_j \, ,$$

$$\Omega_n = \Big\{ \boldsymbol{\sigma} \in \{+1,-1\}^n : \; \sum_{i=1}^{n} \sigma_i = 0 \Big\} .$$

$$\begin{aligned} \text{mcut}(G) &= \min_{\sigma\in\Omega_n} \sum_{(i,j)\in E} \Big(\frac{1-\sigma_i\sigma_j}{2}\Big) \\ &= \frac{1}{2}|E| - \frac{1}{2} \max_{\sigma\in\Omega_n} \sum_{(i,j)\in E} \sigma_i\sigma_j \\ &= \frac{n\gamma}{4} - \sqrt{\frac{\gamma}{4}} \max_{\sigma\in\Omega_n} \mathcal{H}_{\text{mcut}}(\sigma) . \end{aligned}$$

We want to show

$$\left|\frac{1}{n}\mathbb{E}\max_{\sigma\in\Omega_n}\mathcal{H}_{\rm mcut}(\boldsymbol{\sigma})-\frac{1}{n}\mathbb{E}\max_{\sigma\in\{+1,-1\}^n}\mathcal{H}_J(\boldsymbol{\sigma})\right|\le o_\gamma(1)+o_n(1)\,.$$

$\Rightarrow$ Universality result

We want to show

$$\left|\frac{1}{n}\mathbb{E}\max_{\sigma\in\Omega_n}\mathcal{H}_{\sf mcut}(\boldsymbol{\sigma})-\frac{1}{n}\mathbb{E}\max_{\sigma\in\{+1,-1\}^n}\mathcal{H}_J(\boldsymbol{\sigma})\right|\le o_\gamma(1)+o_n(1)\,.$$

$\Rightarrow$ Universality result

Construct $\mathcal{H}_{\mathsf{mcut}}(\boldsymbol{\sigma})$ using iid random variables

For $\boldsymbol{\sigma} \in \Omega_n$

$$\mathcal{H}_{\mathsf{mcut}}(\boldsymbol{\sigma}) = \frac{1}{\sqrt{\gamma}} \sum_{i<j} \Big(\mathbf{1}_{(i,j)\in E} - \frac{\gamma}{n} \Big) \sigma_i \sigma_j$$

$$= \frac{1}{2\sqrt{n}} \langle \boldsymbol{\sigma}, A\boldsymbol{\sigma} \rangle = \mathcal{H}_A(\boldsymbol{\sigma})$$

where

$$A_{ij} = \begin{cases} \sqrt{\frac{n}{\gamma}}\Big(1 - \frac{\gamma}{n}\Big) & \text{with probability } \gamma/n, \\ -\sqrt{\frac{\gamma}{n}} & \text{with probability } 1 - \gamma/n. \end{cases}$$

$$\mathbb{E}\{A_{ij}\} = 0 , \quad \mathbb{E}\{A_{ij}^2\} = 1 - O(1/n) .$$

Construct $\mathcal{H}_{\text{mcut}}(\boldsymbol{\sigma})$ using iid random variables

For $\boldsymbol{\sigma} \in \Omega_n$

$$\mathcal{H}_{\text{mcut}}(\boldsymbol{\sigma}) = \frac{1}{\sqrt{\gamma}} \sum_{i<j} \Big(\mathbf{1}_{(i,j)\in E} - \frac{\gamma}{n} \Big) \sigma_i \sigma_j$$

$$= \frac{1}{2\sqrt{n}} \langle \boldsymbol{\sigma}, A\boldsymbol{\sigma} \rangle = \mathcal{H}_A(\boldsymbol{\sigma})$$

where

$$A_{ij} = \begin{cases} \sqrt{\frac{n}{\gamma}}\Big(1 - \frac{\gamma}{n}\Big) & \text{with probability } \gamma/n, \\ -\sqrt{\frac{\gamma}{n}} & \text{with probability } 1 - \gamma/n. \end{cases}$$

$$\mathbb{E}\{A_{ij}\} = 0, \quad \mathbb{E}\{A_{ij}^2\} = 1 - O(1/n).$$

Construct $\mathcal{H}_{\mathsf{mcut}}(\boldsymbol{\sigma})$ using iid random variables

For $\boldsymbol{\sigma} \in \Omega_n$

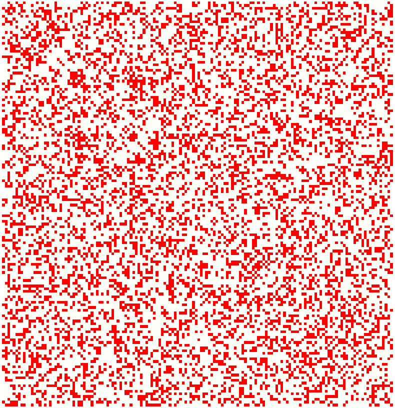
$$\mathcal{H}_{\mathsf{mcut}}(\boldsymbol{\sigma}) = \frac{1}{\sqrt{\gamma}} \sum_{i<j} \Big(\mathbf{1}_{(i,j)\in E} - \frac{\gamma}{n}\Big)\sigma_i\sigma_j$$
$$= \frac{1}{2\sqrt{n}} \langle \boldsymbol{\sigma}, A\boldsymbol{\sigma} \rangle = \mathcal{H}_A(\boldsymbol{\sigma})$$

where

$$A_{ij} = \begin{cases} \sqrt{\frac{n}{\gamma}}\Big(1 - \frac{\gamma}{n}\Big) & \text{with probability } \gamma/n, \\ -\sqrt{\frac{\gamma}{n}} & \text{with probability } 1 - \gamma/n. \end{cases}$$

$$\mathbb{E}\{A_{ij}\} = 0\,, \;\; \mathbb{E}\{A_{ij}^2\} = 1 - O(1/n)\,.$$

Centered, normalized adjacency matrix



$$A_{ij} = \begin{cases} \sqrt{\frac{n}{\gamma}}\left(1 - \frac{\gamma}{n}\right) & \text{with probability } \gamma/n, \\ -\sqrt{\frac{\gamma}{n}} & \text{with probability } 1 - \gamma/n. \end{cases}$$

Universality

We want to show

$$\left|\frac{1}{n}\mathbb{E}\max_{\sigma\in\Omega_n}\mathcal{H}_A(\boldsymbol{\sigma}) - \frac{1}{n}\mathbb{E}\max_{\sigma\in\{+1,-1\}^n}\mathcal{H}_J(\boldsymbol{\sigma})\right| \le o_\gamma(1) + o_n(1)\,.$$

Challenges:

- $\Omega_n \neq \{+1,-1\}^n$
- Define

$$\phi(A) \equiv \frac{1}{n}\max_{\sigma\in\Omega_n}\mathcal{H}_A(\boldsymbol{\sigma})$$

Then $A \mapsto \phi(A)$ is not differentiable.

Fixing problem #1

$$\Big|\mathbb{E}\max_{\sigma\in\Omega_n}\mathcal{H}_A(\boldsymbol{\sigma}) - \mathbb{E}\max_{\sigma\in\{\pm1\}^n}\mathcal{H}_J(\boldsymbol{\sigma})\Big| \le$$

$$\le \Big|\mathbb{E}\max_{\sigma\in\Omega_n}\mathcal{H}_A(\boldsymbol{\sigma}) - \mathbb{E}\max_{\sigma\in\Omega_n}\mathcal{H}_J(\boldsymbol{\sigma})\Big| + \Big|\mathbb{E}\max_{\sigma\in\Omega_n}\mathcal{H}_J(\boldsymbol{\sigma}) - \mathbb{E}\max_{\sigma\in\{\pm1\}^n}\mathcal{H}_J(\boldsymbol{\sigma})\Big|$$

- $\boldsymbol{\sigma}^* = \arg\max_{\sigma\in\{+1,-1\}^n}\mathcal{H}_J(\boldsymbol{\sigma})$ is uniformly random:

$$\langle\boldsymbol{\sigma}^*, \mathbf{1}\rangle = \Theta(\sqrt{n})$$

- Prove that $\max_{\sigma\in\{+1,-1\}^n}\mathcal{H}_J(\boldsymbol{\sigma})$ is 'almost' achieved on $\Omega_n = \{\boldsymbol{\sigma} : \langle\boldsymbol{\sigma}, \mathbf{1}\rangle = 0\}$.

Fixing problem #1

$$\left|\mathbb{E}\max_{\sigma\in\Omega_n}\mathcal{H}_A(\boldsymbol{\sigma}) - \mathbb{E}\max_{\sigma\in\{\pm 1\}^n}\mathcal{H}_J(\boldsymbol{\sigma})\right| \leq$$

$$\leq \left|\mathbb{E}\max_{\sigma\in\Omega_n}\mathcal{H}_A(\boldsymbol{\sigma}) - \mathbb{E}\max_{\sigma\in\Omega_n}\mathcal{H}_J(\boldsymbol{\sigma})\right| + \left|\mathbb{E}\max_{\sigma\in\Omega_n}\mathcal{H}_J(\boldsymbol{\sigma}) - \mathbb{E}\max_{\sigma\in\{\pm 1\}^n}\mathcal{H}_J(\boldsymbol{\sigma})\right|$$

- $\boldsymbol{\sigma}^* = \arg\max_{\sigma\in\{+1,-1\}^n}\mathcal{H}_J(\boldsymbol{\sigma})$ is uniformly random:

$$\langle\boldsymbol{\sigma}^*, \mathbf{1}\rangle = \Theta(\sqrt{n})$$

- Prove that $\max_{\sigma\in\{+1,-1\}^n}\mathcal{H}_J(\boldsymbol{\sigma})$ is 'almost' achieved on $\Omega_n = \{\boldsymbol{\sigma} : \langle\boldsymbol{\sigma}, \mathbf{1}\rangle = 0\}$.

Fixing problem #1

$$\left|\mathbb{E}\max_{\sigma\in\Omega_n}\mathcal{H}_A(\boldsymbol{\sigma}) - \mathbb{E}\max_{\sigma\in\{\pm1\}^n}\mathcal{H}_J(\boldsymbol{\sigma})\right| \le$$

$$\le \left|\mathbb{E}\max_{\sigma\in\Omega_n}\mathcal{H}_A(\boldsymbol{\sigma}) - \mathbb{E}\max_{\sigma\in\Omega_n}\mathcal{H}_J(\boldsymbol{\sigma})\right| + \left|\mathbb{E}\max_{\sigma\in\Omega_n}\mathcal{H}_J(\boldsymbol{\sigma}) - \mathbb{E}\max_{\sigma\in\{\pm1\}^n}\mathcal{H}_J(\boldsymbol{\sigma})\right|$$

- $\boldsymbol{\sigma}^* = \arg\max_{\sigma\in\{+1,-1\}^n}\mathcal{H}_J(\boldsymbol{\sigma})$ is uniformly random:

$$\langle\boldsymbol{\sigma}^*, \mathbf{1}\rangle = \Theta(\sqrt{n})$$

- Prove that $\max_{\sigma\in\{+1,-1\}^n}\mathcal{H}_J(\boldsymbol{\sigma})$ is 'almost' achieved on $\Omega_n = \{\boldsymbol{\sigma} : \langle\boldsymbol{\sigma}, \mathbf{1}\rangle = 0\}$.

Fixing problem #2

$$\phi(A) \equiv \frac{1}{n} \max_{\sigma \in \Omega_n} \mathcal{H}_A(\boldsymbol{\sigma})$$

Idea: Smooth max

$$\phi(\beta; A) \equiv \frac{1}{n\beta} \log \Big\{ \sum_{\sigma \in \Omega_n} e^{\beta \mathcal{H}_A(\boldsymbol{\sigma})} \Big\},$$

$$\phi(A) = \lim_{\beta \to \infty} \phi(\beta; A).$$

Need to:

- Bound $\partial^r_{A_{ij}} \phi(A; \beta)$
- Bound $|\phi(\beta; A) - \phi(\infty; A)|$

Fixing problem #2

$$\phi(A) \equiv \frac{1}{n} \max_{\sigma \in \Omega_n} \mathcal{H}_A(\boldsymbol{\sigma})$$

Idea: Smooth max

$$\phi(\beta; A) \equiv \frac{1}{n\beta} \log \Big\{ \sum_{\boldsymbol{\sigma} \in \Omega_n} e^{\beta \mathcal{H}_A(\boldsymbol{\sigma})} \Big\},$$

$$\phi(A) = \lim_{\beta \to \infty} \phi(\beta; A).$$

Need to:

- Bound $\partial^r_{A_{ij}} \phi(A; \beta)$
- Bound $|\phi(\beta; A) - \phi(\infty; A)|$

Fixing problem #2

$$\phi(A) \equiv \frac{1}{n} \max_{\sigma \in \Omega_n} \mathcal{H}_A(\boldsymbol{\sigma})$$

Idea: Smooth max

$$\phi(\beta; A) \equiv \frac{1}{n\beta} \log \Big\{ \sum_{\boldsymbol{\sigma} \in \Omega_n} e^{\beta \mathcal{H}_A(\boldsymbol{\sigma})} \Big\} ,$$

$$\phi(A) = \lim_{\beta \to \infty} \phi(\beta; A) .$$

Need to:

- Bound $\partial^r_{A_{ij}} \phi(A; \beta)$
- Bound $|\phi(\beta; A) - \phi(\infty; A)|$

It is a free energy!

$$\phi(\beta; A) \equiv \frac{1}{n\beta} \log \Big\{ \sum_{\boldsymbol{\sigma} \in \Omega_n} e^{\beta \mathcal{H}_A(\boldsymbol{\sigma})} \Big\}$$

Associated Boltzmann measure on Ω_n

$$\mu_{\beta,A}(\boldsymbol{\sigma}) \equiv \frac{1}{Z(\beta; A)} e^{\beta \mathcal{H}_A(\boldsymbol{\sigma})}$$

$$\mathsf{E}_\mu \equiv \text{expectation wrt } \mu$$

Bounding $\partial^r_{A_{ij}} \phi(A;\beta)$

$$\phi(\beta; A) \equiv \frac{1}{n\beta} \log \Big\{ \sum_{\boldsymbol{\sigma} \in \Omega_n} \exp\Big(\frac{\beta}{\sqrt{n}} \sum_{i<j} A_{ij} \boldsymbol{\sigma}_i \boldsymbol{\sigma}_j \Big) \Big\}$$

$$\partial_{A_{ij}} \phi(\beta; A) = \frac{1}{n^{3/2}} \mathsf{E}_\mu(\sigma_i \sigma_j) = O(n^{-3/2}) \,,$$

$$\partial^2_{A_{ij}} \phi(\beta; A) = \frac{\beta}{n^2} [\mathsf{E}_\mu(\sigma_i^2 \sigma_j^2) - \mathsf{E}_\mu(\sigma_i \sigma_j)^2] = O(n^{-2}) \,,$$

$$\partial^3_{A_{ij}} \phi(\beta; A) = O(n^{-5/2}) \,.$$

Bounding $\partial^r_{A_{ij}}\phi(A;\beta)$

$$\phi(\beta; A) \equiv \frac{1}{n\beta}\log\Big\{\sum_{\sigma\in\Omega_n}\exp\Big(\frac{\beta}{\sqrt{n}}\sum_{i<j}A_{ij}\sigma_i\sigma_j\Big)\Big\}$$

$$\partial_{A_{ij}}\phi(\beta;A) = \frac{1}{n^{3/2}}\mathsf{E}_\mu(\sigma_i\sigma_j) = O(n^{-3/2}),$$

$$\partial^2_{A_{ij}}\phi(\beta;A) = \frac{\beta}{n^2}[\mathsf{E}_\mu(\sigma_i^2\sigma_j^2) - \mathsf{E}_\mu(\sigma_i\sigma_j)^2] = O(n^{-2}),$$

$$\partial^3_{A_{ij}}\phi(\beta;A) = O(n^{-5/2}).$$

Bounding $\partial^r_{A_{ij}}\phi(A;\beta)$

$$\phi(\beta;A) \equiv \frac{1}{n\beta}\log\Big\{\sum_{\sigma\in\Omega_n}\exp\Big(\frac{\beta}{\sqrt{n}}\sum_{i<j}A_{ij}\sigma_i\sigma_j\Big)\Big\}$$

$$\partial_{A_{ij}}\phi(\beta;A) = \frac{1}{n^{3/2}}\mathsf{E}_\mu(\sigma_i\sigma_j) = O(n^{-3/2}),$$
$$\partial^2_{A_{ij}}\phi(\beta;A) = \frac{\beta}{n^2}[\mathsf{E}_\mu(\sigma_i^2\sigma_j^2) - \mathsf{E}_\mu(\sigma_i\sigma_j)^2] = O(n^{-2}),$$
$$\partial^3_{A_{ij}}\phi(\beta;A) = O(n^{-5/2}).$$

Bounding $\partial^r_{A_{ij}} \phi(A;\beta)$

$$\phi(\beta; A) \equiv \frac{1}{n\beta} \log \Big\{ \sum_{\sigma \in \Omega_n} \exp\Big(\frac{\beta}{\sqrt{n}} \sum_{i<j} A_{ij} \sigma_i \sigma_j \Big) \Big\}$$

$$\begin{aligned}
\partial_{A_{ij}} \phi(\beta; A) &= \frac{1}{n^{3/2}} \mathsf{E}_\mu(\sigma_i\sigma_j) = O(n^{-3/2})\,,\\
\partial^2_{A_{ij}} \phi(\beta; A) &= \frac{\beta}{n^2} [\mathsf{E}_\mu(\sigma_i^2\sigma_j^2) - \mathsf{E}_\mu(\sigma_i\sigma_j)^2] = O(n^{-2})\,,\\
\partial^3_{A_{ij}} \phi(\beta; A) &= O(n^{-5/2})\,.
\end{aligned}$$

Bound $|\phi(\beta; A) - \phi(\infty; A)|$

$$\phi(\beta; A) \equiv \frac{1}{n\beta} \log \Big\{ \sum_{\boldsymbol{\sigma}\in\Omega_n} e^{\beta \mathcal{H}_A(\boldsymbol{\sigma})} \Big\} = \frac{1}{n\beta} \log Z(\beta; A)$$

$$\frac{\partial \phi}{\partial \beta}(\beta; A) \equiv -\frac{1}{n\beta^2} \log Z(\beta; A) + \frac{1}{n\beta} \mathsf{E}_\mu \mathcal{H}_A(\boldsymbol{\sigma})$$

$$= \frac{1}{n\beta^2} \mathsf{E}_\mu \{\beta \mathcal{H}_A(\boldsymbol{\sigma}) - \log Z(\beta; A)\}$$

$$= \frac{1}{n\beta^2} \mathsf{E}_\mu \log \mu_{\beta,A}(\boldsymbol{\sigma})$$

$$\frac{\partial \phi}{\partial (\beta^{-1})}(\beta; A) = \frac{1}{n} H(\mu_{\beta,A}) \in [0, \log 2]$$

...put together the pieces

Bound $|\phi(\beta; A) - \phi(\infty; A)|$

$$\phi(\beta; A) \equiv \frac{1}{n\beta} \log \Big\{ \sum_{\boldsymbol{\sigma} \in \Omega_n} e^{\beta \mathcal{H}_A(\boldsymbol{\sigma})} \Big\} = \frac{1}{n\beta} \log Z(\beta; A)$$

$$\frac{\partial \phi}{\partial \beta}(\beta; A) \equiv -\frac{1}{n\beta^2} \log Z(\beta; A) + \frac{1}{n\beta} \mathsf{E}_\mu \mathcal{H}_A(\boldsymbol{\sigma})$$

$$= \frac{1}{n\beta^2} \mathsf{E}_\mu \{\beta \mathcal{H}_A(\boldsymbol{\sigma}) - \log Z(\beta; A)\}$$

$$= \frac{1}{n\beta^2} \mathsf{E}_\mu \log \mu_{\beta,A}(\boldsymbol{\sigma})$$

$$\frac{\partial \phi}{\partial(\beta^{-1})}(\beta; A) = \frac{1}{n} H(\mu_{\beta,A}) \in [0, \log 2]$$

...put together the pieces

Bound $|\phi(\beta;A) - \phi(\infty;A)|$

$$\phi(\beta;A) \equiv \frac{1}{n\beta}\log\Big\{\sum_{\boldsymbol{\sigma}\in\Omega_n} e^{\beta\mathcal{H}_A(\boldsymbol{\sigma})}\Big\} = \frac{1}{n\beta}\log Z(\beta;A)$$

$$\begin{aligned}\frac{\partial\phi}{\partial\beta}(\beta;A) &\equiv -\frac{1}{n\beta^2}\log Z(\beta;A) + \frac{1}{n\beta}\mathsf{E}_\mu\mathcal{H}_A(\boldsymbol{\sigma})\\ &= \frac{1}{n\beta^2}\mathsf{E}_\mu\{\beta\mathcal{H}_A(\boldsymbol{\sigma}) - \log Z(\beta;A)\}\\ &= \frac{1}{n\beta^2}\mathsf{E}_\mu\log\mu_{\beta,A}(\boldsymbol{\sigma})\end{aligned}$$

$$\frac{\partial\phi}{\partial(\beta^{-1})}(\beta;A) = \frac{1}{n}H(\mu_{\beta,A}) \in [0,\log 2]$$

. . . put together the pieces

Bound $|\phi(\beta; A) - \phi(\infty; A)|$

$$\phi(\beta; A) \equiv \frac{1}{n\beta} \log\Big\{ \sum_{\boldsymbol{\sigma}\in\Omega_n} e^{\beta \mathcal{H}_A(\boldsymbol{\sigma})} \Big\} = \frac{1}{n\beta} \log Z(\beta; A)$$

$$\begin{aligned} \frac{\partial \phi}{\partial \beta}(\beta; A) &\equiv -\frac{1}{n\beta^2} \log Z(\beta; A) + \frac{1}{n\beta} \mathsf{E}_\mu \mathcal{H}_A(\boldsymbol{\sigma}) \\ &= \frac{1}{n\beta^2} \mathsf{E}_\mu \{ \beta \mathcal{H}_A(\boldsymbol{\sigma}) - \log Z(\beta; A) \} \\ &= \frac{1}{n\beta^2} \mathsf{E}_\mu \log \mu_{\beta,A}(\boldsymbol{\sigma}) \end{aligned}$$

$$\frac{\partial \phi}{\partial (\beta^{-1})}(\beta; A) = \frac{1}{n} H(\mu_{\beta,A}) \in [0, \log 2]$$

. . . put together the pieces

Bound $|\phi(\beta;A) - \phi(\infty;A)|$

$$\phi(\beta;A) \equiv \frac{1}{n\beta}\log\Big\{\sum_{\boldsymbol{\sigma}\in\Omega_n} e^{\beta\mathcal{H}_A(\boldsymbol{\sigma})}\Big\} = \frac{1}{n\beta}\log Z(\beta;A)$$

$$\begin{aligned}\frac{\partial\phi}{\partial\beta}(\beta;A) &\equiv -\frac{1}{n\beta^2}\log Z(\beta;A) + \frac{1}{n\beta}\mathsf{E}_\mu\mathcal{H}_A(\boldsymbol{\sigma})\\ &= \frac{1}{n\beta^2}\mathsf{E}_\mu\{\beta\mathcal{H}_A(\boldsymbol{\sigma}) - \log Z(\beta;A)\}\\ &= \frac{1}{n\beta^2}\mathsf{E}_\mu\log\mu_{\beta,A}(\boldsymbol{\sigma})\end{aligned}$$

$$\frac{\partial\phi}{\partial(\beta^{-1})}(\beta;A) = \frac{1}{n}H(\mu_{\beta,A}) \in [0,\log 2]$$

. . . put together the pieces

Semidefinite programs on random graphs

Basic remark

$$\left.\begin{array}{r}\mathsf{mcut}\\ \mathsf{MCUT}\\ \mathsf{MaxCut}\end{array}\right\}\text{ are NP-hard}$$

(and hard to approximate)

(. . . very rich theory)

Example: min bisection

Theorem (Khot 2004)

Assume SAT cannot be solved in (randomized) time below 2^{n^ε}*. Then min-bisection cannot be approximated in polynomial time within a factor* $1 + \delta(\varepsilon)$*.*

Theorem (Räcke 2008)

There exists a polynomial-time algorithm that approximates the min-bisection within a factor $O(\log n)$*.*

Stronger assumption

Theorem (Khot, Kindler, Mossel, O'Donnel 2005)

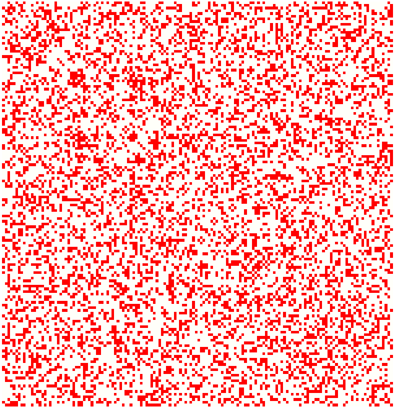
Under the Unique Games Conjecture, there is no polynomial-time algorithm that approximates MaxCut *better than within a factor* 0.878568*.*

What about random instances?

(relevant for machine learning, statistics, signal processing,. . .)

$$G \sim \mathcal{G}(n, \gamma/n)$$

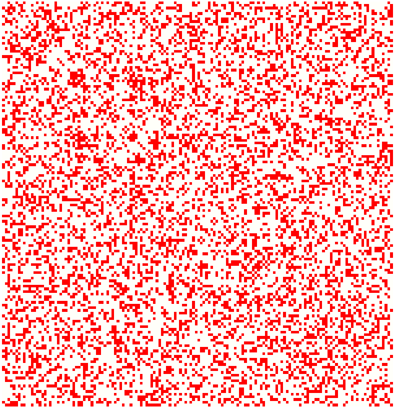
Centered adjacency matrix



$$(A_G)_{ij} = \begin{cases} 1 - \frac{\gamma}{n} & \text{if } (i,j) \in E, \\ -\frac{\gamma}{n} & \text{otherwise.} \end{cases}$$

Not normalized $\mathbb{E}\{(A_G)_{ij}^2\} = (\gamma/n)(1 + O(n^{-1}))$

Centered adjacency matrix



$$(A_G)_{ij} = \begin{cases} 1 - \frac{\gamma}{n} & \text{if } (i,j) \in E, \\ -\frac{\gamma}{n} & \text{otherwise.} \end{cases}$$

Not normalized $\mathbb{E}\{(A_G)_{ij}^2\} = (\gamma/n)(1 + O(n^{-1}))$

min bisection

$$\begin{aligned} \text{maximize} \quad & \langle \boldsymbol{\sigma}, A_G \boldsymbol{\sigma} \rangle \,, \\ \text{subject to} \quad & \langle \mathbf{1}, \boldsymbol{\sigma} \rangle = 0 \,, \\ & \sigma_i \in \{+1, -1\} \,. \end{aligned}$$

Note: $\langle \mathbf{1}, A_G \mathbf{1} \rangle = O(\sqrt{n}) \ll \max_{\boldsymbol{\sigma}} \langle \boldsymbol{\sigma}, A_G \boldsymbol{\sigma} \rangle$

$\Rightarrow$ can drop constraint $\langle \mathbf{1}, \boldsymbol{\sigma} \rangle = 0$

min bisection

$$\begin{aligned} \text{maximize} \quad & \langle \boldsymbol{\sigma}, A_G \boldsymbol{\sigma} \rangle \,, \\ \text{subject to} \quad & \langle \mathbf{1}, \boldsymbol{\sigma} \rangle = 0 \,, \\ & \sigma_i \in \{+1, -1\} \,. \end{aligned}$$

Note: $\langle \mathbf{1}, A_G \mathbf{1} \rangle = O(\sqrt{n}) \ll \max_{\boldsymbol{\sigma}} \langle \boldsymbol{\sigma}, A_G \boldsymbol{\sigma} \rangle$
$\Rightarrow$ can drop constraint $\langle \mathbf{1}, \boldsymbol{\sigma} \rangle = 0$

Another way to look at it: Lagrangian

$$\begin{aligned} \text{maximize} \quad & 2 \sum_{(i,j)\in E} \sigma_i \sigma_j - \lambda \Big(\sum_{i=1}^{n} \sigma_i \Big)^2 , \\ \text{subject to} \quad & \sigma_i \in \{+1, -1\} . \end{aligned}$$

$\lambda =$ regularization parameter

As $\lambda \to \infty$, forces $\sum_{i=1}^{n} \sigma_i = 0$

Lagrangian

$$\begin{aligned} \text{maximize} \quad & \left\langle \boldsymbol{\sigma}, \Big(A_G - (\lambda - (\gamma/n))\, \mathbf{1}\,\mathbf{1}^{\mathsf{T}}\Big)\boldsymbol{\sigma}\right\rangle , \\ \text{subject to} \quad & \boldsymbol{\sigma}_i \in \{+1, -1\} . \end{aligned}$$

$\lambda =$ regularization parameter

Setting $\lambda = \gamma/n$ is sufficient

Typical value

$$\begin{aligned} \text{maximize} \quad & \langle \boldsymbol{\sigma}, A_G \boldsymbol{\sigma} \rangle\,, \\ \text{subject to} \quad & \sigma_i \in \{+1, -1\}\,. \end{aligned}$$

Value (last part of the talk)

$$\begin{aligned} \mathsf{OPT}(G) &= 2n\mathsf{P}_* \sqrt{\gamma} + n\, o_\gamma(\sqrt{\gamma}) \\ &\approx 1.5264\, n\sqrt{\gamma} \end{aligned}$$

Typical value

$$\begin{aligned} \text{maximize} \quad & \langle \boldsymbol{\sigma}, A_G \boldsymbol{\sigma} \rangle \,, \\ \text{subject to} \quad & \sigma_i \in \{+1, -1\} \,. \end{aligned}$$

Value (last part of the talk)

$$\begin{aligned} \mathsf{OPT}(G) &= 2n\mathsf{P}_* \sqrt{\gamma} + n\, o_\gamma(\sqrt{\gamma}) \\ &\approx 1.5264\, n\sqrt{\gamma} \end{aligned}$$

Idea: Convex relaxations

Spectral relaxation

$$\begin{aligned} \text{maximize} \quad & \langle \boldsymbol{\sigma}, A_G \boldsymbol{\sigma} \rangle \,, \\ \text{subject to} \quad & \sigma_i \in \{+1, -1\} \end{aligned} \quad .$$

Spectral relaxation

$$\begin{aligned} \text{maximize} \quad & \langle \boldsymbol{\sigma}, A_G \boldsymbol{\sigma} \rangle \,, \\ \text{subject to} \quad & \cancel{\sigma_i \in \{+1, -1\}} \;\; \|\sigma\|_2^2 = n \,. \end{aligned}$$

Spectral relaxation

$$
\begin{aligned}
\text{maximize} \quad & \langle \boldsymbol{\sigma}, A_G \boldsymbol{\sigma} \rangle \,, \\
\text{subject to} \quad & \cancel{\sigma_i \in \{+1, -1\}} \;\; \|\boldsymbol{\sigma}\|_2^2 = n \,.
\end{aligned}
$$

Value

$$
\lambda_{\max}(A_G)\, n
$$

Spectral relaxation very bad in the sparse regime!

Theorem (Krivelevich, Sudakov 2003+Vu 2005)

If $G \sim \mathcal{G}(n, \gamma/n)$*, then, with high probability,*

$$\lambda_{\max}(A_G) = \begin{cases} 2\sqrt{\gamma}\,(1+o(1)) & \text{if } \gamma \gg (\log n)^4, \\ \sqrt{\log n/(\log\log n)}(1+o(1)) & \text{if } \gamma = O(1). \end{cases}$$

Compare with $\mathrm{OPT}(G)/n \approx 1.5264\,\sqrt{\gamma}$

Spectral relaxation very bad in the sparse regime!

Theorem (Krivelevich, Sudakov 2003+Vu 2005)

If $G \sim \mathcal{G}(n, \gamma/n)$, then, with high probability,

$$\lambda_{\max}(A_G) = \begin{cases} 2\sqrt{\gamma}\,(1+o(1)) & \text{if } \gamma \gg (\log n)^4, \\ \sqrt{\log n/(\log\log n)}(1+o(1)) & \text{if } \gamma = O(1). \end{cases}$$

Compare with $\mathrm{OPT}(G)/n \approx 1.5264\,\sqrt{\gamma}$

Where does this come from? 1. Dense graphs

Wigner heuristics:

Assume $A = A^{\mathsf{T}}$ with $(A_{i,j})_{i<j}$ independent, $\mathbb{E}\{A_{ij}\} = 0$, $\mathbb{E}\{A_{ij}^2\} = 1/n$. Then

$$\mathrm{ESD}_n \equiv \frac{1}{n}\sum_{i=1}^{n}\delta_{\lambda_i} \overset{\mathrm{w}}{\Rightarrow} \mu_{\rm sc}(\mathrm{d}x) \equiv \frac{\sqrt{4-x^2}}{2\pi}\,\mathbf{1}_{|x|\le 2}\,\mathrm{d}x$$

Further $\lambda_{\max}(A) \to \sup\{x : x \in \mathrm{supp}\mu_{\rm sc}\}$, $\lambda_{\min}(A) \to \inf\{x : x \in \mathrm{supp}\mu_{\rm sc}\}$.

Requires tail conditions on A_{ij}
e.g. $\mathbb{E}\{|A_{ij}|^k\} \le C^k$.

Where does this come from? 1. Dense graphs

Wigner heuristics:

Assume $A = A^{\mathsf{T}}$ with $(A_{i,j})_{i<j}$ independent, $\mathbb{E}\{A_{ij}\} = 0$, $\mathbb{E}\{A_{ij}^2\} = 1/n$. Then

$$\text{ESD}_n \equiv \frac{1}{n}\sum_{i=1}^{n} \delta_{\lambda_i} \overset{\text{w}}{\Rightarrow} \mu_{\text{sc}}(\mathrm{d}x) \equiv \frac{\sqrt{4-x^2}}{2\pi}\,\mathbf{1}_{|x|\le 2}\,\mathrm{d}x$$

Further $\lambda_{\max}(A) \to \sup\{x : x \in \text{supp}\mu_{\text{sc}}\}$, $\lambda_{\min}(A) \to \inf\{x : x \in \text{supp}\mu_{\text{sc}}\}$.

Requires tail conditions on A_{ij}
e.g. $\mathbb{E}\{|A_{ij}|^k\} \le C^k$.

Where does this come from? 1. Dense graphs

$\mathbb{E}\{A_{ij}^2\} = \gamma(1 + O(1/n))$

Wigner heuristics:

$$\lim_{n\to\infty} \frac{1}{\sqrt{\gamma}} \lambda_{\max}(A_G) = 2$$

Where does this come from? 2. Sparse graphs

$$\lambda_{\max}(A_G)^2 \geq \langle \mathbf{e}_i, A_G^2 \mathbf{e}_i \rangle = \sum_{j=1}^{n} A_{ij}^2$$

$$\geq \sum_{j \in \partial i} A_{ij}^2 = \Big(1 - \frac{\gamma}{n}\Big)^2 \mathrm{deg}_G(i)$$

Hence

$$\lambda_{\max}(A_G) \geq \sqrt{\max \deg_G(i)}\,(1 + O(1/n)) = \sqrt{\frac{\log n}{\log \log n}}(1 + O(1/n))$$

Where does this come from? 2. Sparse graphs

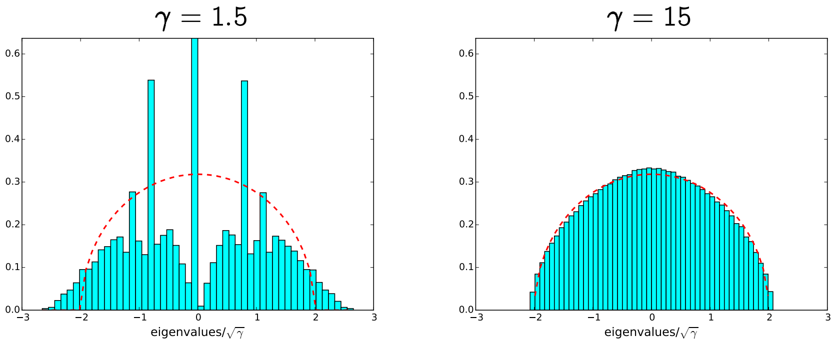
$$\lambda_{\max}(A_G)^2 \geq \langle \mathbf{e}_i, A_G^2 \mathbf{e}_i \rangle = \sum_{j=1}^{n} A_{ij}^2$$

$$\geq \sum_{j \in \partial i} A_{ij}^2 = \left(1 - \frac{\gamma}{n}\right)^2 \mathrm{deg}_G(i)$$

Hence

$$\lambda_{\max}(A_G) \geq \sqrt{\max \mathrm{deg}_G(i)}\,(1 + O(1/n)) = \sqrt{\frac{\log n}{\log \log n}}(1 + O(1/n))$$

Illustration: $n = 10,000$



What we learned

- Spectral relaxations are sensitive to degree heterogeneities
- Leading eigenvectors are highly localized/uninformative.

Idea: Enforce that solution has all entries of rougly same size!

What we learned

- Spectral relaxations are sensitive to degree heterogeneities
- Leading eigenvectors are highly localized/uninformative.

Idea: Enforce that solution has all entries of rougly same size!

Semidefinite Programming (SDP)

$$\begin{aligned} \text{maximize} \quad & \langle A_G, \boldsymbol{\sigma}\boldsymbol{\sigma}^{\mathsf{T}} \rangle\,, \\ \text{subject to} \quad & \sigma_i \in \{+1, -1\}\,. \end{aligned}$$

SDP(A_G)

$$\begin{aligned} \text{maximize} \quad & \langle A_G, X \rangle\,, \\ \text{subject to} \quad & X_{ii} = 1\,, \\ & X \succeq 0\,. \end{aligned}$$

Semidefinite Programming (SDP)

$$\begin{aligned} \text{maximize} \quad & \langle A_G, \boldsymbol{\sigma}\boldsymbol{\sigma}^{\mathsf{T}} \rangle\,, \\ \text{subject to} \quad & \sigma_i \in \{+1, -1\}\,. \end{aligned}$$

$\mathsf{SDP}(A_G)$

$$\begin{aligned} \text{maximize} \quad & \langle A_G, X \rangle\,, \\ \text{subject to} \quad & X_{ii} = 1\,, \\ & X \succeq 0\,. \end{aligned}$$

Does it work?

SDP is universal (sparse $\approx$ dense)

Theorem (Montanari, Sen 2015)

If $G \sim \mathcal{G}(n;\gamma/n)$, $\gamma = O(1)$, then, with high probability,

$$\frac{1}{n}\mathsf{SDP}(A_G) = 2\sqrt{\gamma} + o(\sqrt{\gamma})$$

- $\mathsf{SDP}(A_G)$ behaves much better than the principal eigenvalue

SDP is universal (sparse $\approx$ dense)

Theorem (Montanari, Sen 2015)

If $G \sim \mathcal{G}(n;\gamma/n)$, $\gamma = O(1)$, then, with high probability,

$$\frac{1}{n}\mathsf{SDP}(A_G) = 2\sqrt{\gamma} + o(\sqrt{\gamma})$$

- $\mathsf{SDP}(A_G)$ behaves much better than the principal eigenvalue

Proof idea

This is a universality result!

$$\frac{1}{\sqrt{\gamma}}\mathsf{SDP}(A_G) \underset{\text{Lindeberg}}{\longleftrightarrow} \mathsf{SDP}(W)$$

$W \sim \text{GOE}(n)$, i.e. $(W_{ij})_{i<j} \sim_{iid} \text{N}(0, 1/n)$

Following same strategy as before

- Prove $\lim_{n\to\infty} \mathsf{SDP}(W) = 2$.
- Replace $\mathsf{SDP}(A)$ by a suitable 'smooth-max' $\phi(\beta; A)$
- Bound $\partial^r_{A_{ij}} \phi(\beta; A)$.
- Bound $|\mathsf{SDP}(A) - \phi(\beta; A)|$

Following same strategy as before

- Prove $\lim_{n\to\infty} \mathsf{SDP}(W) = 2$.
- Replace $\mathsf{SDP}(A)$ by a suitable 'smooth-max' $\phi(\beta; A)$
- Bound $\partial^r_{A_{ij}} \phi(\beta; A)$.
- Bound $|\mathsf{SDP}(A) - \phi(\beta; A)|$

Does not quite work!

Where is the problem? Introducing the smooth max

SDP(A):

$$\begin{aligned} \text{maximize} \quad & \langle A, X \rangle \,, \\ \text{subject to} \quad & X_{ii} = 1 \,, \\ & X \succeq 0 \,. \end{aligned}$$

Equivalent (non-convex) formulation

$$\begin{aligned} \text{maximize} \quad & \sum_{i,j=1}^{n} A_{i,j} \langle \boldsymbol{\sigma}_i, \boldsymbol{\sigma}_j \rangle \,, \\ \text{subject to} \quad & \boldsymbol{\sigma} \in \mathbb{R}^n \,, \quad \|\boldsymbol{\sigma}_i\|_2 = 1 \,. \end{aligned}$$

Introducing the smooth max

$$\phi(\beta; A) \equiv \frac{1}{n\beta} \log \left\{ \int \exp\left(\beta \sum_{i,j=1}^{n} A_{i,j} \langle \boldsymbol{\sigma}_i, \boldsymbol{\sigma}_j \rangle\right) \nu_0(\mathrm{d}\boldsymbol{\sigma}) \right\}$$

- $\nu_0(\mathrm{d}\boldsymbol{\sigma}) \equiv$ uniform measure on $S^{n-1} \times \cdots \times S^{n-1}$
- $S^{n-1} \equiv \{\mathbf{x} \in \mathbb{R}^n : \ \|\mathbf{x}\|_2 = 1\}$
- Easy to show $\lim_{\beta\to\infty} \phi(\beta; A) = \mathsf{SDP}(A)$
- Associated Boltzmann measure

$$\mu_{\beta,A}(\mathrm{d}\boldsymbol{\sigma}) = \frac{1}{Z(\beta; A)} \exp\left\{\beta \sum_{i,j=1}^{n} A_{ij} \langle \boldsymbol{\sigma}_i, \boldsymbol{\sigma}_j \rangle\right\} \nu_0(\mathrm{d}\boldsymbol{\sigma})$$

Where is the problem?

We want to prove, *uniformly in n*, and for $A \in \{A_G, W\}$

$$\left|\mathbb{E}\phi(\beta; A, n) - \mathbb{E}\phi(\infty; A, n)\right| \le \varepsilon(\beta)$$

with $\lim_{\beta\to\infty} \varepsilon(\beta) = 0$.

Let us try as before

$$\frac{\partial\phi(\beta; A)}{\partial(\beta^{-1})} = -\frac{1}{n} D(\mu_{\beta,A}||\nu_0)$$

- $\mu_{\beta,A}$ distribution on a continuous space
- Dimension $= n(n-1)$

Where is the problem?

We want to prove, *uniformly in n*, and for $A \in \{A_G, W\}$

$$\Big|\mathbb{E}\phi(\beta; A, n) - \mathbb{E}\phi(\infty; A, n)\Big| \le \varepsilon(\beta)$$

with $\lim_{\beta\to\infty} \varepsilon(\beta) = 0$.

Let us try as before

$$\frac{\partial\phi(\beta; A)}{\partial(\beta^{-1})} = -\frac{1}{n} D(\mu_{\beta,A}||\nu_0)$$

- $\mu_{\beta,A}$ distribution on a continuous space
- Dimension $= n(n-1)$

Idea

- Constrain $\text{rank}(X) = k = O_n(1)$
- Bound the error due to 'rank truncation'

Concretely

Define $\mathrm{OPT}_k(A)$

$$\begin{aligned} \text{maximize} \quad & \langle A, X \rangle \,, \\ \text{subject to} \quad & X_{ii} = 1 \,, \\ & X \succeq 0 \,, \\ & \mathrm{rank}(X) = k \end{aligned}$$

For $A = A_G$

- $k = 1$: Minimum bisection
- $k = n$: SDP

Concretely

Define $\mathrm{OPT}_k(A)$

$$\begin{aligned} \text{maximize} \quad & \langle A, X \rangle\,, \\ \text{subject to} \quad & X_{ii} = 1\,, \\ & X \succeq 0\,, \\ & \mathrm{rank}(X) = k \end{aligned}$$

For $A = A_G$

- $k = 1$: Minimum bisection
- $k = n$: SDP

Equivalent (non-convex) definition

$\mathsf{OPT}_k(A)$:

$$\begin{aligned} \text{maximize} \quad & \sum_{i,j=1}^{n} A_{ij} \langle \boldsymbol{\sigma}_i, \boldsymbol{\sigma}_j \rangle \,, \\ \text{subject to} \quad & \boldsymbol{\sigma} = (\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2, \dots, \boldsymbol{\sigma}_n)^{\mathsf{T}} \in \mathbb{R}^{n \times k} \,, \\ & \boldsymbol{\sigma}_i \in \mathbb{R}^k \,, \quad \|\boldsymbol{\sigma}_i\|_2 = 1 \,. \end{aligned}$$

Free energy

$$\phi_k(\beta; A) \equiv \frac{1}{n\beta} \log \left\{ \int \exp\left(\beta \sum_{i,j=1}^{n} A_{i,j} \langle \boldsymbol{\sigma}_i, \boldsymbol{\sigma}_j \rangle \right) \nu_{0,k}(\mathrm{d}\boldsymbol{\sigma}) \right\}$$

- $\nu_{0,k}(\mathrm{d}\boldsymbol{\sigma}) \equiv$ uniform measure on $S^{k-1} \times \cdots \times S^{k-1}$

Equivalent (non-convex) definition

$\mathsf{OPT}_k(A)$:

$$\begin{aligned} \text{maximize} \quad & \sum_{i,j=1}^{n} A_{ij} \langle \boldsymbol{\sigma}_i, \boldsymbol{\sigma}_j \rangle \,, \\ \text{subject to} \quad & \boldsymbol{\sigma} = (\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2, \dots, \boldsymbol{\sigma}_n)^{\mathsf{T}} \in \mathbb{R}^{n\times k} \,, \\ & \boldsymbol{\sigma}_i \in \mathbb{R}^k \,, \quad \|\boldsymbol{\sigma}_i\|_2 = 1 \,. \end{aligned}$$

Free energy

$$\phi_k(\beta; A) \equiv \frac{1}{n\beta} \log \left\{ \int \exp \left(\beta \sum_{i,j=1}^{n} A_{i,j} \langle \boldsymbol{\sigma}_i, \boldsymbol{\sigma}_j \rangle \right) \nu_{0,k}(\mathrm{d}\boldsymbol{\sigma}) \right\}$$

- $\nu_{0,k}(\mathrm{d}\boldsymbol{\sigma}) \equiv$ uniform measure on $S^{k-1} \times \cdots \times S^{k-1}$

Proof strategy

- Prove $\lim_{n\to\infty} \mathsf{SDP}(W) = 2$.
- Replace SDP by the rank-constrained problem OPT_k.
- Bound $|\mathsf{SDP}(A) - \mathsf{OPT}_k(A)| \le o_k(1)$, uniformly in n.
- Replace $\mathsf{OPT}_k(A)$ by the free energy $\phi_k(\beta; A)$
- Bound $|\mathsf{OPT}_k(A) - \phi_k(\beta; A)|$
- Bound $\partial^r_{A_{ij}} \phi(\beta; A)$.

Now this works! (Details omitted)

Proof strategy

- Prove $\lim_{n\to\infty} \mathsf{SDP}(W) = 2$.
- Replace SDP by the rank-constrained problem OPT_k.
- Bound $|\mathsf{SDP}(A) - \mathsf{OPT}_k(A)| \le o_k(1)$, uniformly in n.
- Replace $\mathsf{OPT}_k(A)$ by the free energy $\phi_k(\beta; A)$
- Bound $|\mathsf{OPT}_k(A) - \phi_k(\beta; A)|$
- Bound $\partial^r_{A_{ij}} \phi(\beta; A)$.

This works because the spins are bounded. (Details omitted)

Proof strategy

- Prove $\lim_{n\to\infty} \mathsf{SDP}(W) = 2$.
- Replace SDP by the rank-constrained problem OPT_k.
- Bound $|\mathsf{SDP}(A) - \mathsf{OPT}_k(A)| \leq o_k(1)$, uniformly in n.
- Replace $\mathsf{OPT}_k(A)$ by the free energy $\phi_k(\beta; A)$
- Bound $|\mathsf{OPT}_k(A) - \phi_k(\beta; A)|$
- Bound $\partial^r_{A_{ij}} \phi(\beta; A)$

What about this? A new Grothendieck inequality

Grothendieck inequalities

- First proved by Alexander Grothendieck in 1956
- Connect non-convex quadratic programs to SDPs
- Very popular in CS Theory since Alon, Naor 2006
- We cannot use ealier inequalities:
  - Symmetric matrices
  - (Rank $= k$) $\to$ (Rank $= n$)

Higher-rank Grothendieck inequality

Theorem (Montanari, Sen, 2015)

For $k \geq 1$, let $\mathbf{g} \sim \mathrm{N}(0, \mathbf{I}_{k\times k}/k)$ be a vector with i.i.d. centered normal entries with variance $1/k$, and define

$$\alpha_k \equiv (\mathbb{E}\|\mathbf{g}\|_2)^2 \,.$$

Then, for any symmetric matrix B, we have the inequalities

$$\mathsf{OPT}_k(B) \geq \Big(2 - \frac{1}{\alpha_k}\Big)\mathsf{SDP}(B) + \Big(\frac{1}{\alpha_k} - 1\Big)\,\mathsf{OPT}_k(-B)\,.$$

Remarks

- $\alpha_k = 1 - O(k^{-1})$
- Obviously $\mathsf{OPT}_k(B) \leq \mathsf{SDP}(B)$
- The additive error can be bounded crudely: $\mathsf{OPT}_k(-B) \geq 0$ with high probabilty.

Higher-rank Grothendieck inequality

Theorem (Montanari, Sen, 2015)

For $k \geq 1$, let $\mathbf{g} \sim \mathrm{N}(0, \mathbf{I}_{k\times k}/k)$ be a vector with i.i.d. centered normal entries with variance $1/k$, and define

$$\alpha_k \equiv (\mathbb{E}\|\mathbf{g}\|_2)^2 \,.$$

Then, for any symmetric matrix B, we have the inequalities

$$\mathsf{OPT}_k(B) \geq \Big(2 - \frac{1}{\alpha_k}\Big)\mathsf{SDP}(B) + \Big(\frac{1}{\alpha_k} - 1\Big)\,\mathsf{OPT}_k(-B)\,.$$

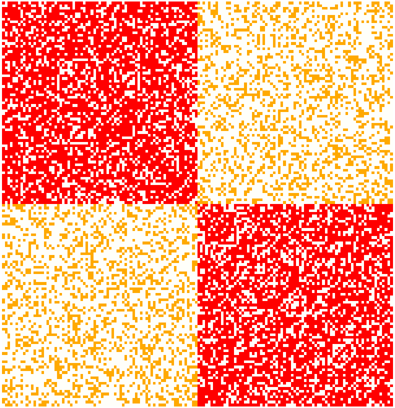
Remarks

- $\alpha_k = 1 - O(k^{-1})$
- Obviously $\mathsf{OPT}_k(B) \leq \mathsf{SDP}(B)$
- The additive error can be bounded crudely: $\mathsf{OPT}_k(-B) \geq 0$ with high probabilty.

One application

The hidden partition problem

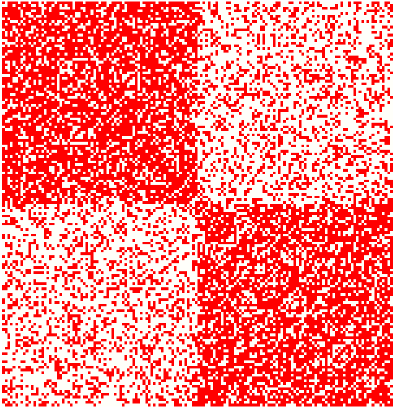
(a.k.a. community detection)



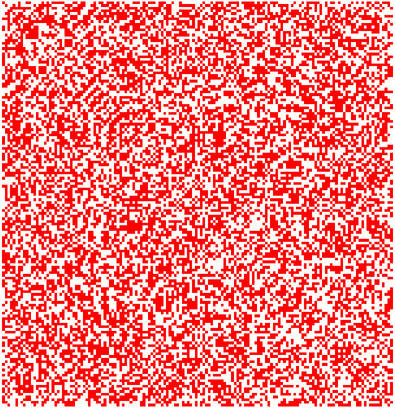
Vertices V, $|V| = n$, $V = S_1 \cup S_2$, $|S_1| = |S_2| = n/2$

$$\mathbb{P}\{(i,j) \in E\} = \begin{cases} p & \text{if } \{i,j\} \subseteq S_1 \text{ or } \{i,j\} \subseteq S_2, \\ q < p & \text{otherwise.} \end{cases}$$

Of course entries are not colored. . .



. . . and rows/columns are not ordered



Background: Hypothesis testing (sparse graph $p = a/n$, $q = b/n$)

Hypothesis H_0:

$$\mathbb{P}\{(i,j) \in E\} = \frac{a+b}{2n}$$

Hypothesis H_1: $V = S_1 \cup S_2$, $|S_1| = |S_2| = n/2$

$$\mathbb{P}\{(i,j) \in E\} = \begin{cases} a/n & \text{if } \{i,j\} \subseteq S_1 \text{ or } \{i,j\} \subseteq S_2, \\ b/n & \text{otherwise.} \end{cases}$$

Detection thresholds

Classical Statistics / Information Theory:
Is there *any algorithm* that detects the hidden structure?

Computational:
Is there *a poly-time algorithm* that detects the hidden structure?

Detection thresholds

Classical Statistics / Information Theory:
Is there *any algorithm* that detects the hidden structure?

Computational:
Is there *a poly-time algorithm* that detects the hidden structure?

Information theory threshold

Theorem (Mossel, Neeman, Sly, 2012)

There is a test that succeed with high probability if and only if $a + b > 2$ *and*

$$\frac{a-b}{\sqrt{2(a+b)}} > 1\,.$$

[Proves conjecture by Decelle, Krzakala, Moore, Zdeborova, 2011]

Information theory threshold

Theorem (Mossel, Neeman, Sly, 2012)

There is a test that succeed with high probability if and only if $a + b > 2$ *and*

$$\frac{a-b}{\sqrt{2(a+b)}} > 1\,.$$

[Proves conjecture by Decelle, Krzakala, Moore, Zdeborova, 2011]

Computational threshold

- Dyer, Frieze 1989 $p = na > q = nb$ fixed.
- Condon, Karp 2001 $a - b \gg n^{1/2}$
- McSherry 2001 $a - b \gg \sqrt{b \log n}$
- Coja-Oghlan 2010 $a - b \gg \sqrt{b}$
- Massoulie 2013 and Mossel, Neeman, Sly, 2013

$$\frac{a-b}{\sqrt{2(a+b)}} > 1$$

Very ingenious algorithms!

Computational threshold

- Dyer, Frieze 1989 — $p = na > q = nb$ fixed.
- Condon, Karp 2001 — $a - b \gg n^{1/2}$
- McSherry 2001 — $a - b \gg \sqrt{b \log n}$
- Coja-Oghlan 2010 — $a - b \gg \sqrt{b}$
- Massoulie 2013 and Mossel, Neeman, Sly, 2013

$$\frac{a-b}{\sqrt{2(a+b)}} > 1$$

Very ingenious algorithms!

Computational threshold

- Dyer, Frieze 1989 $p = na > q = nb$ fixed.
- Condon, Karp 2001 $a - b \gg n^{1/2}$
- McSherry 2001 $a - b \gg \sqrt{b \log n}$
- Coja-Oghlan 2010 $a - b \gg \sqrt{b}$
- Massoulie 2013 and Mossel, Neeman, Sly, 2013

$$\frac{a - b}{\sqrt{2(a + b)}} > 1$$

Very ingenious algorithms!

What if I am not ingenious?

Maximum Likelihood testing

$\mathsf{OPT}(A_G)$:

$$\begin{aligned} \text{maximize} \quad & \langle \boldsymbol{\sigma}, A_G \boldsymbol{\sigma} \rangle \,, \\ \text{subject to} \quad & \sigma_i \in \{+1, -1\} \quad . \end{aligned}$$

$$G \sim \mathcal{G}(n, \gamma/n) \Rightarrow \mathsf{OPT}(A_G) \approx \big(2\mathsf{P}_* \sqrt{\gamma} + o(\sqrt{\gamma})\big)\, n.$$

$$T_{\rm ML}(G) = \begin{cases} 1 & \text{if } \mathsf{OPT}(A_G) \geq (2+\varepsilon)\mathsf{P}_* \sqrt{\gamma} n, \\ 0 & \text{otherwise.} \end{cases}$$

Maximum Likelihood testing

$\mathsf{OPT}(A_G)$:

$$\begin{aligned} \text{maximize} \quad & \langle \boldsymbol{\sigma}, A_G \boldsymbol{\sigma} \rangle\,, \\ \text{subject to} \quad & \sigma_i \in \{+1, -1\} \quad . \end{aligned}$$

$$G \sim \mathcal{G}(n, \gamma/n) \Rightarrow \mathsf{OPT}(A_G) \approx \big(2\mathsf{P}_* \sqrt{\gamma} + o(\sqrt{\gamma})\big)\, n.$$

$$T_{\rm ML}(G) = \begin{cases} 1 & \text{if } \mathsf{OPT}(A_G) \geq (2+\varepsilon)\mathsf{P}_* \sqrt{\gamma} n, \\ 0 & \text{otherwise.} \end{cases}$$

$T_{\mathrm{ML}}(\,\cdot\,)$ is hard to compute!

SDP(A_G):

$$\begin{aligned} \text{maximize} \quad & \langle \boldsymbol{\sigma}, A_G \boldsymbol{\sigma} \rangle\,, \\ \text{subject to} \quad & \sigma_i \in \{+1, -1\} \quad . \end{aligned}$$

$$G \sim \mathcal{G}(n, \gamma/n) \Rightarrow \mathsf{SDP}(A_G) \approx \big(2\sqrt{\gamma} + o(\sqrt{\gamma})\big)n.$$

$$T_{\mathsf{SDP}}(G) = \begin{cases} 1 & \text{if } \mathsf{SDP}(A_G) \geq (2+\varepsilon)\sqrt{\gamma}\, n, \\ 0 & \text{otherwise.} \end{cases}$$

$T_{\mathrm{ML}}(\,\cdot\,)$ is hard to compute!

SDP(A_G):

$$\begin{aligned} \text{maximize} \quad & \langle \boldsymbol{\sigma}, A_G \boldsymbol{\sigma} \rangle\,, \\ \text{subject to} \quad & \sigma_i \in \{+1, -1\} \quad . \end{aligned}$$

$G \sim \mathcal{G}(n, \gamma/n) \Rightarrow \mathsf{SDP}(A_G) \approx \big(2\sqrt{\gamma} + o(\sqrt{\gamma})\big)n.$

$$T_{\mathsf{SDP}}(G) = \begin{cases} 1 & \text{if } \mathsf{SDP}(A_G) \geq (2+\varepsilon)\sqrt{\gamma}n, \\ 0 & \text{otherwise.} \end{cases}$$

- This is really **off-the-shelf**
- How well does it work?

False positives

Under $H_0 : G \sim \mathcal{G}(n, \gamma/n)$, $\gamma = (a+b)/2$

$$\mathbb{P}_0(T_{\mathsf{SDP}}(G) = 1) = \mathbb{P}_0(\mathsf{SDP}(G) \geq (2+\varepsilon)\sqrt{\gamma}n) \to 0$$

provided $\gamma \geq \gamma_0$

False negatives

Under $H_1 : G \sim \mathcal{G}(n, a/n, b/n)$,

$$\mathbb{P}_1(T_{\mathsf{SDP}}(G) = 0) = \mathbb{P}_1(\mathsf{SDP}(G) < (2+\varepsilon)\sqrt{\gamma}n)$$

we need a lower bound on $\mathsf{SDP}(G)\dots$

Lower bound on SDP(A_G)

Ground truth

$$\sigma_{*,i} = \begin{cases} +1 & \text{if } i \in S_1, \\ -1 & \text{if } i \in S_2, \end{cases}$$

$$\begin{aligned} \mathrm{SDP}(A_G) &\geq \langle A_G, \boldsymbol{\sigma}_* \boldsymbol{\sigma}_*^{\mathsf{T}} \rangle \\ &= \sum_{i,j \in S_1} \mathbf{1}_{(i,j)\in E} + \sum_{i,j \in S_2} \mathbf{1}_{(i,j)\in E} - 2 \sum_{i \in S_1, j \in S_2} \mathbf{1}_{(i,j)\in E} \\ &= \Big(\frac{a-b}{2}\Big) + O(1/\sqrt{n}) \end{aligned}$$

$\mathbb{P}_1(T_{\mathrm{SDP}}(G) = 0) \to 0$ provided

$$\frac{a-b}{2} \geq (2+\varepsilon')\sqrt{\gamma}$$

Lower bound on $\mathsf{SDP}(A_G)$

Ground truth

$$\sigma_{*,i} = \begin{cases} +1 & \text{if } i \in S_1, \\ -1 & \text{if } i \in S_2, \end{cases}$$

$$\begin{aligned}\mathsf{SDP}(A_G) &\geq \langle A_G, \boldsymbol{\sigma}_* \boldsymbol{\sigma}_*^{\mathsf{T}} \rangle \\ &= \sum_{i,j \in S_1} \mathbf{1}_{(i,j)\in E} + \sum_{i,j \in S_2} \mathbf{1}_{(i,j)\in E} - 2 \sum_{i \in S_1, j \in S_2} \mathbf{1}_{(i,j)\in E} \\ &= \Big(\frac{a-b}{2}\Big) + O(1/\sqrt{n})\end{aligned}$$

$\mathbb{P}_1(T_{\mathsf{SDP}}(G) = 0) \to 0$ provided

$$\frac{a-b}{2} \geq (2+\varepsilon')\sqrt{\gamma}$$

Lower bound on $\mathsf{SDP}(A_G)$

Ground truth

$$\sigma_{*,i} = \begin{cases} +1 & \text{if } i \in S_1, \\ -1 & \text{if } i \in S_2, \end{cases}$$

$$\begin{aligned} \mathsf{SDP}(A_G) &\geq \langle A_G, \boldsymbol{\sigma}_* \boldsymbol{\sigma}_*^{\mathsf{T}} \rangle \\ &= \sum_{i,j \in S_1} \mathbf{1}_{(i,j) \in E} + \sum_{i,j \in S_2} \mathbf{1}_{(i,j) \in E} - 2 \sum_{i \in S_1, j \in S_2} \mathbf{1}_{(i,j) \in E} \\ &= \Big(\frac{a-b}{2} \Big) + O(1/\sqrt{n}) \end{aligned}$$

$\mathbb{P}_1(T_{\mathsf{SDP}}(G) = 0) \to 0$ provided

$$\frac{a-b}{2} \geq (2 + \varepsilon') \sqrt{\gamma}$$

Conclusion: It works well

Theorem (Montanari, Sen 2015)

The test based on SDP *works with high probability if*

$$\frac{a-b}{\sqrt{2(a+b)}} > 2 + o_{a+b}(1)\,.$$

- At most suboptimal by factor $2 + o_{a+b}(1)$
- Forthcoming: $2 \to 1$
- Earlier result by Guedon, Vershynin 2015: $(a-b) \geq 10^4\sqrt{a+b}$.
- (studied also in Abbe, Bandeira, Hall 2014+ Hajek, Wu, Xu 2015)

Conclusion: It works well

Theorem (Montanari, Sen 2015)

The test based on SDP *works with high probability if*

$$\frac{a-b}{\sqrt{2(a+b)}} > 2 + o_{a+b}(1)\,.$$

- At most suboptimal by factor $2 + o_{a+b}(1)$
- Forthcoming: $2 \to 1$
- Earlier result by Guedon, Vershynin 2015: $(a-b) \geq 10^4\sqrt{a+b}$.
- (studied also in Abbe, Bandeira, Hall 2014+ Hajek, Wu, Xu 2015)

Conclusion: It works well

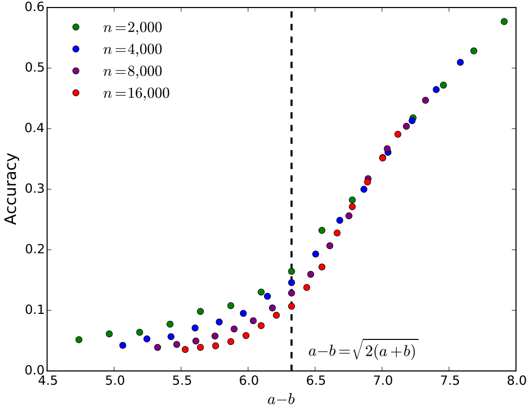
Theorem (Montanari, Sen 2015)

The test based on SDP *works with high probability if*

$$\frac{a-b}{\sqrt{2(a+b)}} > 2 + o_{a+b}(1)\,.$$

- At most suboptimal by factor $2 + o_{a+b}(1)$
- Forthcoming: $2 \to 1$
- Earlier result by Guedon, Vershynin 2015: $(a-b) \geq 10^4\sqrt{a+b}$.
- (studied also in Abbe, Bandeira, Hall 2014+ Hajek, Wu, Xu 2015)

It works very well in practice!



- Average degree $\gamma \equiv (a + b)/2 = 10$.

[Javanmard, Montanari, Ricci-Tersenghi, in preparation]

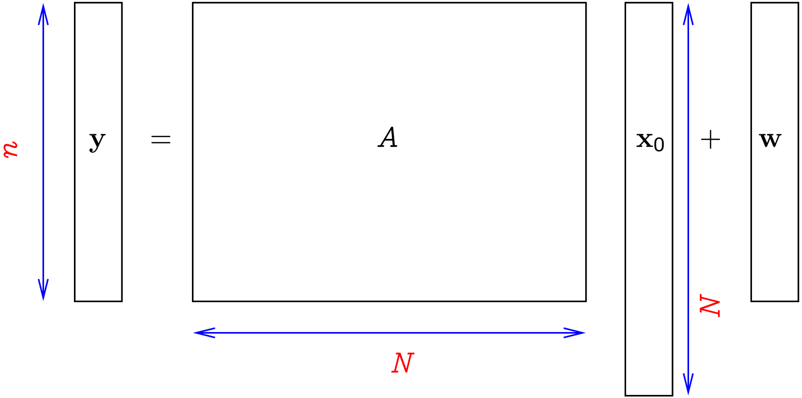
Compressed sensing

Linear observation model

$$\mathbf{y} = A\,\mathbf{x}_0 + \mathbf{w}$$

Estimate $x_0 \in \mathbb{R}^N$ given $y \in \mathbb{R}^n$, $A \in \mathbb{R}^{n\times N}$.
Noise: $\mathbf{w} \sim \mathrm{N}(0, \sigma^2 \mathbf{I}_n)$

Setting



The LASSO

$$\begin{aligned} \widehat{\mathbf{x}}(A, \mathbf{y}) &= \text{argmin}_{\mathbf{x} \in \mathbb{R}^N} \mathcal{L}_{A,\mathbf{y}}(\mathbf{x}) \\ \mathcal{L}_{A,\mathbf{y}}(\mathbf{x}) &= \lambda \|\mathbf{x}\|_1 + \frac{1}{2}\|\mathbf{y} - A\mathbf{x}\|_2^2 \end{aligned}$$

Finds sparse solutions

[Tibshirani 96; Chen, Donoho 95]

The LASSO

$$\begin{aligned}\hat{\mathbf{x}}(A,\mathbf{y}) &= \mathrm{argmin}_{\mathbf{x}\in\mathbb{R}^N}\mathcal{L}_{A,\mathbf{y}}(\mathbf{x})\\ \mathcal{L}_{A,\mathbf{y}}(\mathbf{x}) &= \lambda\,\|\mathbf{x}\|_1 + \frac{1}{2}\|\mathbf{y}-A\mathbf{x}\|_2^2\end{aligned}$$

Finds sparse solutions

[Tibshirani 96; Chen, Donoho 95]

Random sensing mechanism

- **IID model:** A_{ij} iid, $\mathbb{E}[A_{ij}] = 0$, $\mathbb{E}[A_{ij}^2] = 1/n$, $\mathbb{E}[A_{ij}^6] \le C/n^3$
- **Gaussian model:** $A_{ij} \sim_{iid} \mathrm{N}(0, 1/n)$

Why?

- 'Nearly ideal' sensing mechanism
- 'Designed' sensing matrices are pseudo-random
- Connections with random polytope geometry [Donoho, Tanner, Chandrasekaran, Hassibi, Stojnic, Tropp,...]

Random sensing mechanism

- **IID model:** A_{ij} iid, $\mathbb{E}[A_{ij}] = 0$, $\mathbb{E}[A_{ij}^2] = 1/n$, $\mathbb{E}[A_{ij}^6] \le C/n^3$
- **Gaussian model:** $A_{ij} \sim_{iid} \mathrm{N}(0, 1/n)$

Why?

- 'Nearly ideal' sensing mechanism
- 'Designed' sensing matrices are pseudo-random
- Connections with random polytope geometry

  [Donoho, Tanner, Chandrasekaran, Hassibi, Stojnic, Tropp,...]

Asymptotically exact characterization for Gaussian model

Gaussian model : Asymptotic characterization

Theorem (Bayati, Montanari, 2010)

Assume the Gaussian model with, $n/N \to \delta$ and

$$\text{EmpiricalLaw}(\mathbf{x}_0) \equiv \frac{1}{N}\sum_{i=1}^{N} \delta_{x_{0,i}} \stackrel{\mathrm{w}}{\Rightarrow} p_{X_0}\,.$$

Let $(\tau^2, \theta) \in \mathbb{R}^2_{\geq 0}$ be the solution of ($X_0 \sim p_{X_0}$, $Z \sim \mathrm{N}(0,1)$)

$$\tau^2 = \sigma^2 + \frac{1}{\delta}\mathbb{E}\{[\eta(X_0 + \tau\, Z;\theta) - X_0]^2\}\,,$$

$$\lambda = \theta\{1 - (1/\delta)\mathbb{P}\{|X_0 + \tau\, Z| \geq \theta\}\}\,.$$

Then,

$$\lim_{N\to\infty} \frac{1}{N}\|\widehat{\mathbf{x}}^{\mathrm{LASSO}}(\lambda) - \mathbf{x}_0\|^2 \stackrel{\text{a.s.}}{=} (\tau_\infty^2 - \sigma^2)\delta\,.$$

Gaussian model : Asymptotic characterization

Theorem (Bayati, Montanari, 2010)

Assume the Gaussian model with, $n/N \to \delta$ and

$$\mathrm{EmpiricalLaw}(\mathbf{x}_0) \equiv \frac{1}{N}\sum_{i=1}^{N}\delta_{x_{0,i}} \stackrel{\mathrm{w}}{\Rightarrow} p_{X_0}\,.$$

Let $(\tau^2, \theta) \in \mathbb{R}^2_{\geq 0}$ be the solution of ($X_0 \sim p_{X_0}$, $Z \sim \mathrm{N}(0,1)$)

$$\tau^2 = \sigma^2 + \frac{1}{\delta}\mathbb{E}\{[\eta(X_0 + \tau\, Z;\theta) - X_0]^2\}\,,$$

$$\lambda = \theta\{1 - (1/\delta)\mathbb{P}\{|X_0 + \tau\, Z| \geq \theta\}\}\,.$$

Then,

$$\lim_{N\to\infty} \frac{1}{N}\|\widehat{\mathbf{x}}^{\mathrm{LASSO}}(\lambda) - \mathbf{x}_0\|^2 \stackrel{\mathrm{a.s.}}{=} (\tau_\infty^2 - \sigma^2)\delta\,.$$

Gaussian model : Asymptotic characterization

Theorem (Bayati, Montanari, 2010)

Assume the Gaussian model with, $n/N \to \delta$ *and*

$$\text{EmpiricalLaw}(\mathbf{x}_0) \equiv \frac{1}{N}\sum_{i=1}^{N}\delta_{x_{0,i}} \stackrel{\text{w}}{\Rightarrow} p_{X_0}\,.$$

Let $(\tau^2, \theta) \in \mathbb{R}^2_{\geq 0}$ *be the solution of (*$X_0 \sim p_{X_0}$*,* $Z \sim \text{N}(0,1)$*)*

$$\tau^2 = \sigma^2 + \frac{1}{\delta}\mathbb{E}\{[\eta(X_0 + \tau\, Z; \theta) - X_0]^2\}\,,$$
$$\lambda = \theta\{1 - (1/\delta)\mathbb{P}\{|X_0 + \tau\, Z| \geq \theta\}\}\,.$$

Then,

$$\lim_{N\to\infty}\frac{1}{N}\|\widehat{\mathbf{x}}^{\text{LASSO}}(\lambda) - \mathbf{x}_0\|^2 \stackrel{\text{a.s.}}{=} (\tau_\infty^2 - \sigma^2)\delta\,.$$

More generally

Theorem (Bayati, Montanari, 2010)

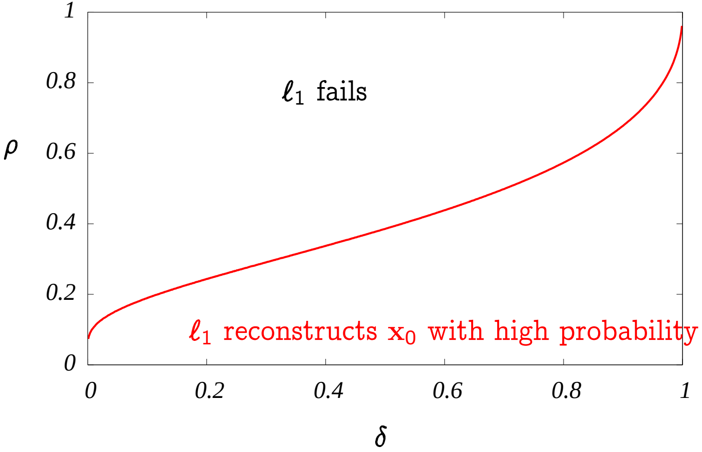
...same assumptions...

Let $\psi : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ *locally Lipschitz,* $\psi(a, b) \leq C(1 + a^2 + b^2)$. *Then*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \psi(\widehat{x}_i^{\rm LASSO}(\lambda), x_{0,i}) \stackrel{\rm a.s.}{=} \mathbb{E}\{\psi(\eta(X_0 + \tau\, Z; \theta), X_0)\}\,.$$

Setting $\sigma = 0$: 'ℓ_0–ℓ_1 equivalence'

$$N, n \to \infty,\ n/N = \delta, \qquad |\mathrm{supp}(x_0)|/N = \delta\rho$$



[Donoho 2006, Affentranger-Schneider 1992]

What about other sensing matrices?

Conjecture (Donoho, Taner, 2009)

The above predictions are universal for a broad range of random matrices.

Conjecture (Weak form)

Universality holds of the IID model.

Conjecture (Donoho, Taner, 2009)

The above predictions are universal for a broad range of random matrices.

Conjecture (Weak form)

Universality holds of the IID model.

Numerical evidence: Random Fourier measurements

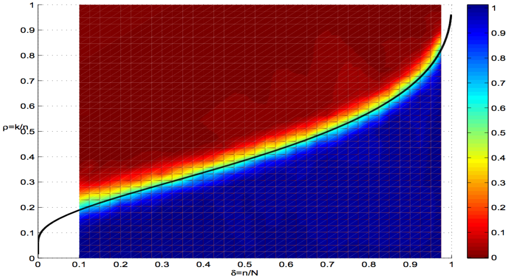
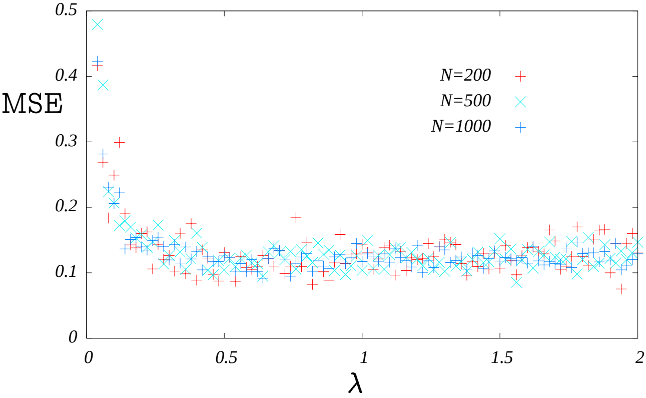


FIGURE 3. *Compressed Sensing from random Fourier measurements.* Shaded attribute: fraction of realizations in which $\ell_1$minimization (1.2) reconstructs an image accurate to within six digits. Horizontal axis: undersampling fraction $\delta = n/N$. Vertical axis: sparsity fraction $\rho = k/n$.

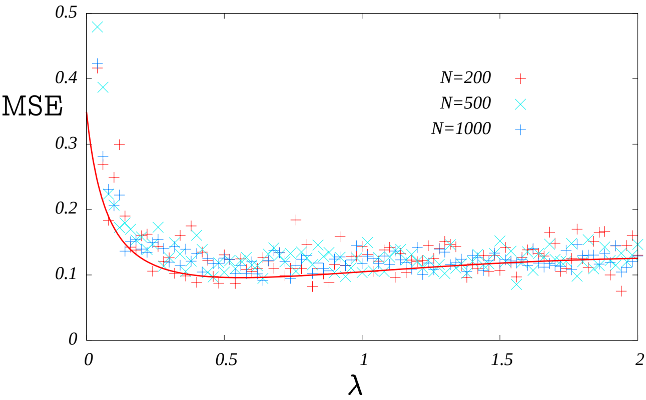
[Donoho, Tanner 2009]

Numerical evidence: Clinical data



A is $n \times N$, $n = 0.64N$

Numerical evidence: Clinical data



A is $n \times N$, $n = 0.64N$

Universality: Rigorous results

- Korada, Montanari, 2011:
  Universality of the cost $\min \mathcal{L}_{A,\mathbf{y}}(\mathbf{x})$

  (Lindeberg method)

- Bayati, Lelarge, Montanari 2012:
  Universality of the $\ell_0 - -\ell_1$ **phase transition**

  (Moment method)

  (A_{ij} subgaussian, with a density)

Universality: Rigorous results

- Korada, Montanari, 2011:
  Universality of the cost $\min \mathcal{L}_{A,\mathbf{y}}(\mathbf{x})$

  (Lindeberg method)

- Bayati, Lelarge, Montanari 2012:
  Universality of the $\ell_0 - -\ell_1$ **phase transition**

  (Moment method)

  (A_{ij} subgaussian, with a density)

Universality: Rigorous results

- Korada, Montanari, 2011:
  Universality of the cost $\min \mathcal{L}_{A,\mathbf{y}}(\mathbf{x})$

  (Lindeberg method)

- Bayati, Lelarge, Montanari 2012:
  Universality of the $\ell_0 - -\ell_1$ phase transition

  (Moment method)
  (A_{ij} subgaussian, with a density)

Universality: Rigorous results

- Korada, Montanari, 2011:
  Universality of the cost $\min \mathcal{L}_{A,\mathbf{y}}(\mathbf{x})$

  (Lindeberg method)

- Bayati, Lelarge, Montanari 2012:
  Universality of the $\ell_0 - -\ell_1$ **phase transition**

  (Moment method)

  (A_{ij} subgaussian, with a density)

Statement

$$\mathcal{L}_{A,\mathbf{y}}(\mathbf{x}) = \lambda \|\mathbf{x}\|_1 + \frac{1}{2}\|\mathbf{y} - A\mathbf{x}\|_2^2$$

$$\mathcal{L}^*_M(A, \mathbf{y}) = \frac{1}{n} \min \left\{ \mathcal{L}_{A,\mathbf{y}}(\mathbf{x}) : \quad \mathbf{x} \in \mathbb{R}^N, \|\mathbf{x}\|_\infty \le M \right\}$$

Theorem (Korada, Montanari, 2011)

Let $A^{(n)} = (A_{ij})_{1\le i\le n, 1\le j\le N}$ *and* $B^{(n)} = (B_{ij})_{1\le i\le n, 1\le j\le n}$ *denote two sensing matrices from the IID model and the Gaussian model, and assume* $\|\mathbf{x}_0\|_\infty \le M = O_n(1)$*.*
Consider the limit $n, N \to \infty$ *with* $n/N \to \delta$ *fixed. Then*

$$\lim_{n,N\to\infty} \left\{ \mathbb{E}\mathcal{L}^*_M(A^{(n)}, \mathbf{y}) - \mathbb{E}\mathcal{L}^*_M(B^{(n)}, \mathbf{y}) \right\} = 0\,.$$

Statement

$$\mathcal{L}_{A,\mathbf{y}}(\mathbf{x}) = \lambda\,\|\mathbf{x}\|_1 + \frac{1}{2}\|\mathbf{y} - A\mathbf{x}\|_2^2$$

$$\mathcal{L}^*_M(A,\mathbf{y}) = \frac{1}{n}\min\left\{\mathcal{L}_{A,\mathbf{y}}(\mathbf{x}) : \;\; \mathbf{x}\in\mathbb{R}^N, \|\mathbf{x}\|_\infty \le M\right\}$$

Theorem (Korada, Montanari, 2011)

Let $A^{(n)} = (A_{ij})_{1\le i\le n, 1\le j\le N}$ *and* $B^{(n)} = (B_{ij})_{1\le i\le n, 1\le j\le n}$ *denote two sensing matrices from the IID model and the Gaussian model, and assume* $\|\mathbf{x}_0\|_\infty \le M = O_n(1)$*.*
Consider the limit $n, N\to\infty$ *with* $n/N\to\delta$ *fixed. Then*

$$\lim_{n,N\to\infty}\left\{\mathbb{E}\mathcal{L}^*_M(A^{(n)},\mathbf{y}) - \mathbb{E}\mathcal{L}^*_M(B^{(n)},\mathbf{y})\right\} = 0\,.$$

Remarks

$$\lim_{n,N\to\infty}\Big\{\mathbb{E}\mathcal{L}^*_M(A^{(n)},\mathbf{y})-\mathbb{E}\mathcal{L}^*_M(B^{(n)},\mathbf{y})\Big\}=0\,.$$

- Not hard: $\mathcal{L}^*_M(A^{(n)},\mathbf{y})$ concentrates around its expectation.
- From [BayatiMontanari2010]: $\lim_{n,N\to\infty}\mathbb{E}\mathcal{L}^*_M(A^{(n)},\mathbf{y})$ exists and depends non-trivially on $\delta, p_{X_0}, \sigma^2$

As in previous examples

- Replace $\mathcal{L}^*_M(A)$ by free energy $\phi(\beta; A)$.
- Bound $\left|\mathcal{L}^*_M(A) - \phi(\beta; A)\right|$
- Bound $\left|\phi(\beta; A^{(n)}) - \phi(\beta; B^{(n)})\right|$ (Lindeberg)

Free energy

$$\phi(\beta; A) = \frac{1}{N\beta} \log \left\{ \int_{B_M} e^{-\beta \mathcal{L}_{A,\mathbf{y}}(\mathbf{x})} \mathrm{d}\mathbf{x} \right\}$$

$$B_M \equiv [-M, M]^N$$

$$\mathcal{L}_{A,\mathbf{y}}(\mathbf{x}) = \lambda \|\mathbf{x}\|_1 + \frac{1}{2} \|\mathbf{w} - A(\mathbf{x} - \mathbf{x}_0)\|_2^2$$

Associated Boltzmann measure

$$\mu_{\beta,A}(\mathrm{d}\mathbf{x}) = \frac{1}{Z(\beta, A)} e^{-\beta \mathcal{L}_{A,\mathbf{y}}(\mathbf{x})} \mathrm{d}\mathbf{x}$$

Free energy

$$\phi(\beta; A) = \frac{1}{N\beta} \log \left\{ \int_{B_M} e^{-\beta \mathcal{L}_{A,\mathbf{y}}(\mathbf{x})} \mathrm{d}\mathbf{x} \right\}$$

$$B_M \equiv [-M, M]^N$$

$$\mathcal{L}_{A,\mathbf{y}}(\mathbf{x}) = \lambda \|\mathbf{x}\|_1 + \frac{1}{2}\|\mathbf{w} - A(\mathbf{x} - \mathbf{x}_0)\|_2^2$$

Associated Boltzmann measure

$$\mu_{\beta,A}(\mathrm{d}\mathbf{x}) = \frac{1}{Z(\beta, A)}\, e^{-\beta \mathcal{L}_{A,\mathbf{y}}(\mathbf{x})} \mathrm{d}\mathbf{x}$$

Taking derivatives

$$\partial_{A_{ij}}\phi(\beta;A) = -\frac{1}{N}\mathsf{E}_\mu\{\partial_{A_{ij}}\mathcal{L}_{A,\mathbf{y}}(\mathbf{x})\}$$

$$\mathcal{L}_{A,\mathbf{y}}(\mathbf{x}) = \lambda\,\|\mathbf{x}\|_1 + \frac{1}{2}\|\mathbf{w} - A(\mathbf{x}-\mathbf{x}_0)\|_2^2$$

$$\partial_{A_{ij}}\mathcal{L}_{A,\mathbf{y}}(\mathbf{x}) = -[\mathbf{w} - A(\mathbf{x}-\mathbf{x}_0)]_i(x_j - x_{0,j})$$

Naive bound:

$$\begin{aligned}\left|\partial_{A_{ij}}\mathcal{L}_{A,\mathbf{y}}(\mathbf{x})\right| &\le 2M\left|[A(\mathbf{x}-\mathbf{x}_0)]_i\right| + \dots \\ &\le 4M^2\|A_{i,\cdot}\|_1 + \dots \le C\,N^{1/2}\end{aligned}$$

Taking derivatives

$$\partial_{A_{ij}} \phi(\beta; A) = -\frac{1}{N} \mathsf{E}_\mu \{\partial_{A_{ij}} \mathcal{L}_{A,\mathbf{y}}(\mathbf{x})\}$$

$$\mathcal{L}_{A,\mathbf{y}}(\mathbf{x}) = \lambda \|\mathbf{x}\|_1 + \frac{1}{2}\|\mathbf{w} - A(\mathbf{x} - \mathbf{x}_0)\|_2^2$$

$$\partial_{A_{ij}} \mathcal{L}_{A,\mathbf{y}}(\mathbf{x}) = -[\mathbf{w} - A(\mathbf{x} - \mathbf{x}_0)]_i (x_j - x_{0,j})$$

Naive bound:

$$\begin{aligned} \left|\partial_{A_{ij}} \mathcal{L}_{A,\mathbf{y}}(\mathbf{x})\right| &\le 2M \left|[A(\mathbf{x} - \mathbf{x}_0)]_i\right| + \dots \\ &\le 4M^2 \|A_{i,\cdot}\|_1 + \dots \le C\, N^{1/2} \end{aligned}$$

Second/third derivatives

$$\partial^2_{A_{ij}} \phi(\beta; A) = \cdots + \frac{\beta}{N} \mathsf{E}_\mu \{ \partial_{A_{ij}} \mathcal{L}_{A,\mathbf{y}}(\mathbf{x}) \}^2 \,,$$

$$\partial^3_{A_{ij}} \phi(\beta; A) = \cdots + \frac{\beta^2}{N} \mathsf{E}_\mu \{ \partial_{A_{ij}} \mathcal{L}_{A,\mathbf{y}}(\mathbf{x}) \}^3 \,.$$

Naive bound:

$$\|\partial^2_{A_{ij}} \phi(\beta; A)\|_\infty \le C \,, \qquad \|\partial^3_{A_{ij}} \phi(\beta; A)\|_\infty \le C\, N^{1/2} \,.$$

Second/third derivatives

$$\partial^2_{A_{ij}} \phi(\beta; A) = \cdots + \frac{\beta}{N} \mathsf{E}_\mu \{\partial_{A_{ij}} \mathcal{L}_{A,\mathbf{y}}(\mathbf{x})\}^2 ,$$
$$\partial^3_{A_{ij}} \phi(\beta; A) = \cdots + \frac{\beta^2}{N} \mathsf{E}_\mu \{\partial_{A_{ij}} \mathcal{L}_{A,\mathbf{y}}(\mathbf{x})\}^3 .$$

Naive bound:

$$\|\partial^2_{A_{ij}} \phi(\beta; A)\|_\infty \le C \, , \qquad \|\partial^3_{A_{ij}} \phi(\beta; A)\|_\infty \le C \, N^{1/2} .$$

Lindeberg Bound

$$
\begin{aligned}
&\left|\mathbb{E}\phi(\beta; A^{(n)}) - \mathbb{E}\phi(\beta; B^{(n)})\right| \leq \\
&\quad \leq \left\|\partial^2_{A_{ij}}\phi(\beta; A)\right\|_\infty \sum_{i,j} \mathbb{E}\{|A_{ij}|^2 \mathbf{1}_{|A_{ij}|>K/\sqrt{n}}\} \\
&\qquad + \left\|\partial^3_{A_{ij}}\phi(\beta; A)\right\|_\infty \sum_{i,j} \mathbb{E}\{|A_{ij}|^3 \mathbf{1}_{|A_{ij}|\leq K/\sqrt{n}}\} + \dots \\
&\quad \leq C \cdot Nn \cdot \frac{K^2}{n} \, e^{-K^2/2} + CN^{1/2} \cdot Nn \cdot \frac{K^3}{n^{3/2}}
\end{aligned}
$$

Where is the issue?

$$\partial^2_{A_{ij}} \phi(\beta; A) = \cdots + \frac{\beta}{N} \mathsf{E}_\mu \{\partial_{A_{ij}} \mathcal{L}_{A,\mathbf{y}}(\mathbf{x})\}^2,$$

$$\partial_{A_{ij}} \mathcal{L}_{A,\mathbf{y}}(\mathbf{x}) = -[\mathbf{w} - A(\mathbf{x} - \mathbf{x}_0)]_i (x_j - x_{0,j})$$

For instance, consider $\mathsf{E}_\mu\{[A(\mathrm{x} - \mathrm{x}_0)]_i\}$

Naive bound:

$$\big|\mathsf{E}_\mu\{[A(\mathrm{x} - \mathrm{x}_0)]_i\}\big| \le \mathsf{E}_\mu\{\|A_{i,\cdot}\|_1\}\|\mathrm{x} - \mathrm{x}_0\|_\infty \le C\, N^{1/2}$$

Independent vectors heuristics

$$\big|\mathsf{E}_\mu\{[A(\mathrm{x} - \mathrm{x}_0)]_i\}\big| \approx \frac{1}{N^{1/2}} \mathsf{E}_\mu\{\|A_{i,\cdot}\|_2^2\}^{1/2} \mathsf{E}_\mu\{\|\mathrm{x} - \mathrm{x}_0\|_2^2\}^{1/2} \le C$$

Where is the issue?

$$\partial^2_{A_{ij}} \phi(\beta; A) = \cdots + \frac{\beta}{N} \mathsf{E}_\mu \{\partial_{A_{ij}} \mathcal{L}_{A,\mathbf{y}}(\mathbf{x})\}^2 ,$$
$$\partial_{A_{ij}} \mathcal{L}_{A,\mathbf{y}}(\mathbf{x}) = -[\mathbf{w} - A(\mathbf{x} - \mathbf{x}_0)]_i (x_j - x_{0,j})$$

For instance, consider $\mathsf{E}_\mu\{[A(\mathbf{x} - \mathbf{x}_0)]_i\}$

Naive bound:

$$\big|\mathsf{E}_\mu\{[A(\mathrm{x} - \mathrm{x}_0)]_i\}\big| \le \mathsf{E}_\mu\{\|A_{i,\cdot}\|_1\}\|\mathrm{x} - \mathrm{x}_0\|_\infty \le C\, N^{1/2}$$

Independent vectors heuristics

$$\big|\mathsf{E}_\mu\{[A(\mathrm{x} - \mathrm{x}_0)]_i\}\big| \approx \frac{1}{N^{1/2}} \mathsf{E}_\mu\{\|A_{i,\cdot}\|_2^2\}^{1/2} \mathsf{E}_\mu\{\|\mathrm{x} - \mathrm{x}_0\|_2^2\}^{1/2} \le C$$

Where is the issue?

$$\partial^2_{A_{ij}} \phi(\beta; A) = \cdots + \frac{\beta}{N} \mathsf{E}_\mu \{\partial_{A_{ij}} \mathcal{L}_{A,\mathbf{y}}(\mathbf{x})\}^2 \,,$$
$$\partial_{A_{ij}} \mathcal{L}_{A,\mathbf{y}}(\mathbf{x}) = -[\mathbf{w} - A(\mathbf{x} - \mathbf{x}_0)]_i (x_j - x_{0,j})$$

For instance, consider $\mathsf{E}_\mu\{[A(\mathbf{x}-\mathbf{x}_0)]_i\}$
Naive bound:

$$\big|\mathsf{E}_\mu\{[A(\mathbf{x}-\mathbf{x}_0)]_i\}\big| \le \mathsf{E}_\mu\{\|A_{i,\cdot}\|_1\}\|\mathbf{x}-\mathbf{x}_0\|_\infty \le C\, N^{1/2}$$

Independent vectors heuristics

$$\big|\mathsf{E}_\mu\{[A(\mathbf{x}-\mathbf{x}_0)]_i\}\big| \approx \frac{1}{N^{1/2}} \mathsf{E}_\mu\{\|A_{i,\cdot}\|_2^2\}^{1/2} \mathsf{E}_\mu\{\|\mathbf{x}-\mathbf{x}_0\|_2^2\}^{1/2} \le C$$

Where is the issue?

$$\partial^2_{A_{ij}} \phi(\beta; A) = \cdots + \frac{\beta}{N} \mathsf{E}_\mu\{\partial_{A_{ij}} \mathcal{L}_{A,\mathbf{y}}(\mathbf{x})\}^2,$$
$$\partial_{A_{ij}} \mathcal{L}_{A,\mathbf{y}}(\mathbf{x}) = -[\mathbf{w} - A(\mathbf{x} - \mathbf{x}_0)]_i (x_j - x_{0,j})$$

For instance, consider $\mathsf{E}_\mu\{[A(\mathbf{x} - \mathbf{x}_0)]_i\}$

Naive bound:

$$\left|\mathsf{E}_\mu\{[A(\mathbf{x} - \mathbf{x}_0)]_i\}\right| \le \mathsf{E}_\mu\{\|A_{i,\cdot}\|_1\}\|\mathbf{x} - \mathbf{x}_0\|_\infty \le C\, N^{1/2}$$

Independent vectors heuristics

$$\left|\mathsf{E}_\mu\{[A(\mathbf{x} - \mathbf{x}_0)]_i\}\right| \approx \frac{1}{N^{1/2}} \mathsf{E}_\mu\{\|A_{i,\cdot}\|_2^2\}^{1/2} \mathsf{E}_\mu\{\|\mathbf{x} - \mathbf{x}_0\|_2^2\}^{1/2} \le C$$

The independent vectors heuristics is kroughly correct!

Proof: Read the paper!

The independent vectors heuristics is kroughly correct!

Proof: Read the paper!

Conclusion

Conclusion

- **Universality!!**

  A model is not really predictive unless it has some degree of universality

- **Lindeberg!!**

  A tool you should put in your bagpack

Thanks!

Conclusion

- **Universality!!**

  A model is not really predictive unless it has some degree of universality

- **Lindeberg!!**

  A tool you should put in your bagpack

Thanks!

Conclusion

- **Universality!!**

  A model is not really predictive unless it has some degree of universality

- **Lindeberg!!**

  A tool you should put in your bagpack

Thanks!

Conclusion

- **Universality!!**

  A model is not really predictive unless it has some degree of universality

- **Lindeberg!!**

  A tool you should put in your bagpack

Thanks!

Conclusion

- **Universality!!**

  A model is not really predictive unless it has some degree of universality

- **Lindeberg!!**

  A tool you should put in your bagpack

Thanks!

Conclusion

- **Universality!!**

  A model is not really predictive unless it has some degree of universality

- **Lindeberg!!**

  A tool you should put in your bagpack

Thanks!