Computational aspects in Statistics:

Sparse PCA & Ising blockmodel

Quentin Berthet

Cargese Summer School - 2018

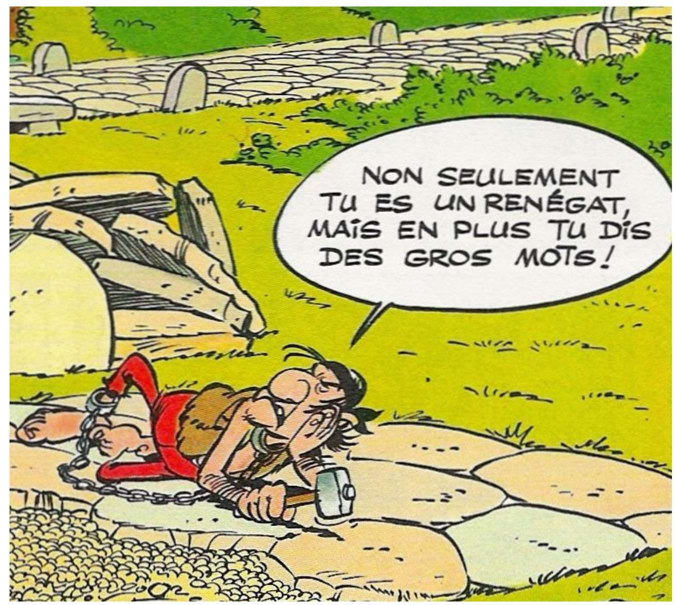
Objectives & Jargon

- Inference problems: detection/estimation, with theoretical guarantees.
- Statement about a **parameter** of size $\boldsymbol{p}$, based on an i.i.d. **sample** of size n.
- Finite sample results, with fixed probability δ, up to universal constants:

e.g. for $X_1, \dots, X_n$ i.i.d. from $\mathcal{N}(\mu, I_p)$,

$$\|\bar{X}_n - \mu\|_2^2 \leq C\, \frac{p + \log(1/\delta)}{n} \quad \text{w.p. } 1 - \delta.$$

- High-dimensional setting: large p.
- No prior, structural assumptions.
- Important notions:
    - Optimal statistical rates.
    - Algorithmic efficiency.

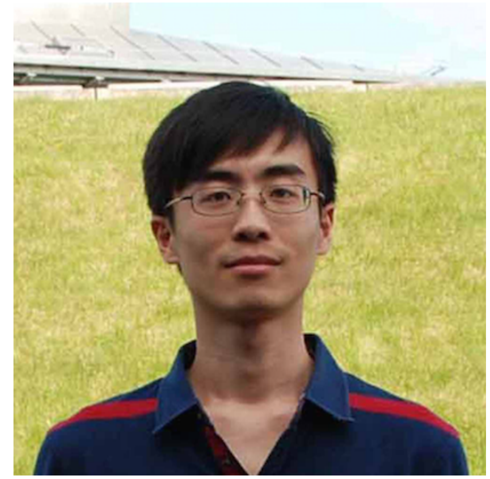


Astérix en Corse, Goscinny, Uderzo (73)

Sparse PCA

P. Rigollet (MIT)

R. Samworth (Cambridge)



T. Wang (Cambridge)

- **Optimal Detection of Sparse Principal Components in High Dimension**

  Q.B., P. Rigollet, *Ann. Statis. (2013)*

- **Computational Lower Bounds for Sparse PCA**

  Q.B., P. Rigollet, *COLT (2013)*

- **Statistical and Computational Trade-offs in Estimation of Sparse Principal Components**

  T. Wang, Q.B., R. Samworth, *Ann. Statis. (2016)*

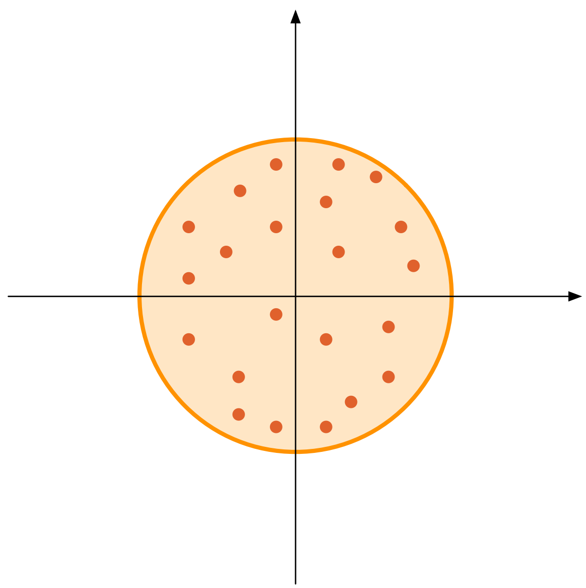
Sparse principal component detection

$X_1, \ldots, X_n \in \mathbf{R}^p$ independent, centered Gaussian with unknown covariance.

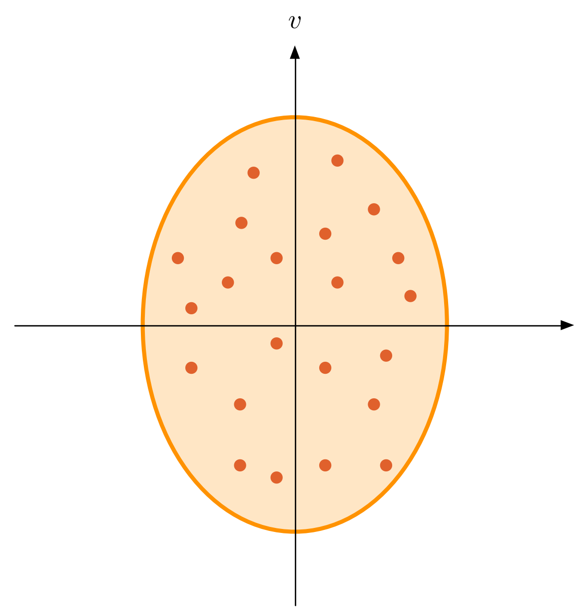
Testing problem between two hypotheses, Johnstone (01), & Lu (09).

$$\begin{cases} H_0 : I_p \\ H_1 : I_p + \theta v v^\top, \quad v \in \mathcal{B}_0(k) \end{cases}$$

v is a k-sparse unit vector. $\mathcal{B}_0(k) = \{v \in \mathbf{R}^p \,:\, |v|_2 = 1\,,\ |v|_0 \leq k\}$.



Isotropy: $\mathcal{N}(0, I_p)$



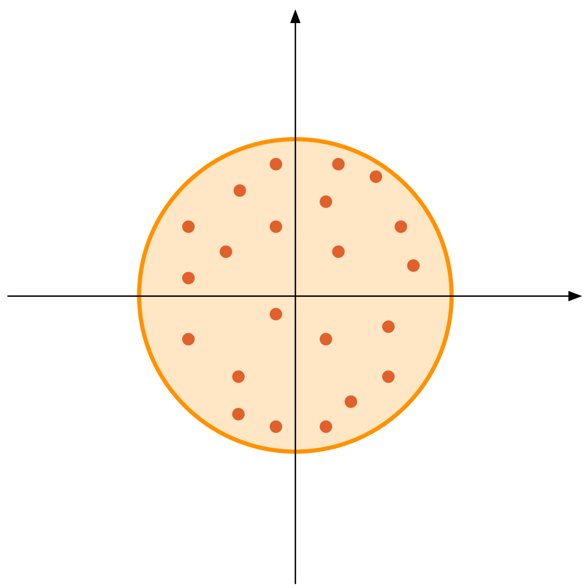
Sparse PC: $\mathcal{N}(0, I_p + \theta\, v v^\top)$

Importance of sparsity assumption

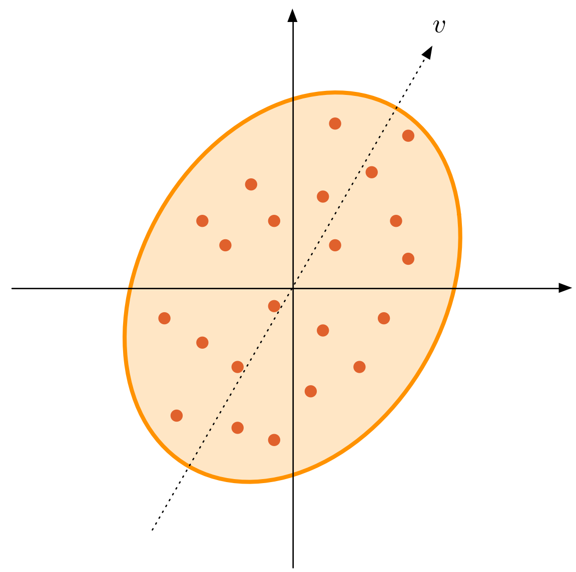
General problem of principal component detection, Onatski et al. (13)

$$\begin{cases} H_0 : I_p \\ H_1 : I_p + \theta v v^\top , \quad v \in \mathcal{S}^{p-1} \end{cases}$$

v is a unit vector. $\mathcal{S}^{p-1} = \{v \in \mathbf{R}^p \, : \, |v|_2 = 1\}$.



Isotropy: $\mathcal{N}(0, I_p)$



Principal Component: $\mathcal{N}(0, I_p + \theta\, v v^\top)$

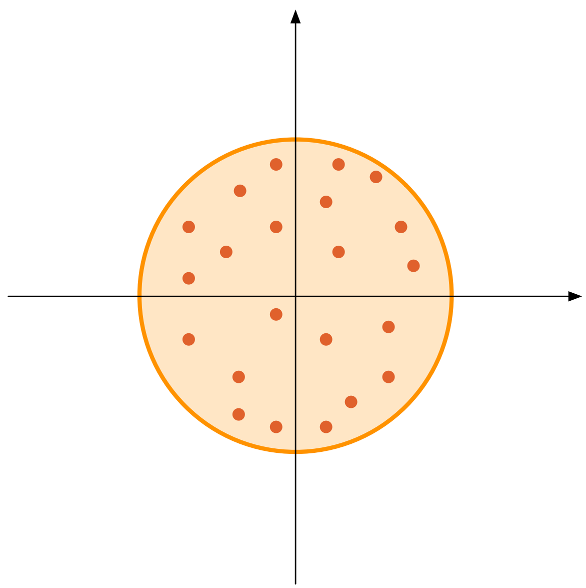
BBP transition instuition, Baik, Ben Arous, Péché (05): Strong signal: $\theta > \sqrt{\frac{p}{n}}$.

Sparse principal component detection

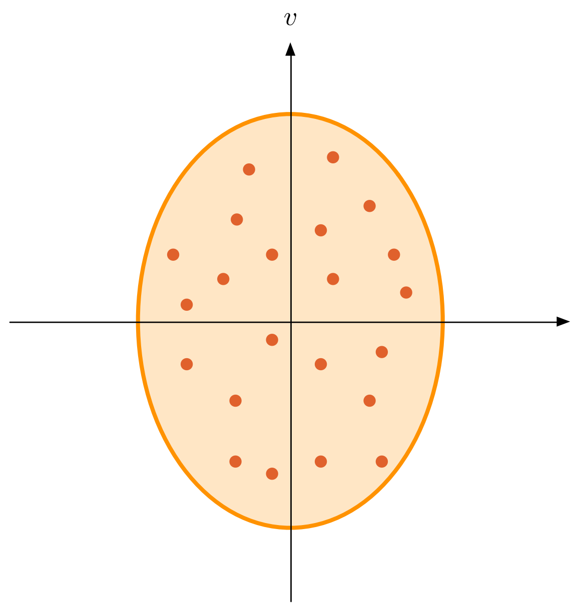
Testing problem between two hypotheses.

$$\begin{cases} H_0 : I_p \\ H_1 : I_p + \theta v v^\top, \quad v \in \mathcal{B}_0(k) \end{cases}$$

v is a k-sparse unit vector. $\mathcal{B}_0(k) = \{v \in \mathbf{R}^p \,:\, |v|_2 = 1\,,\, |v|_0 \leq k\}$.



Isotropy: $\mathcal{N}(0, I_p)$



Sparse PC: $\mathcal{N}(0, I_p + \theta\, v v^\top)$

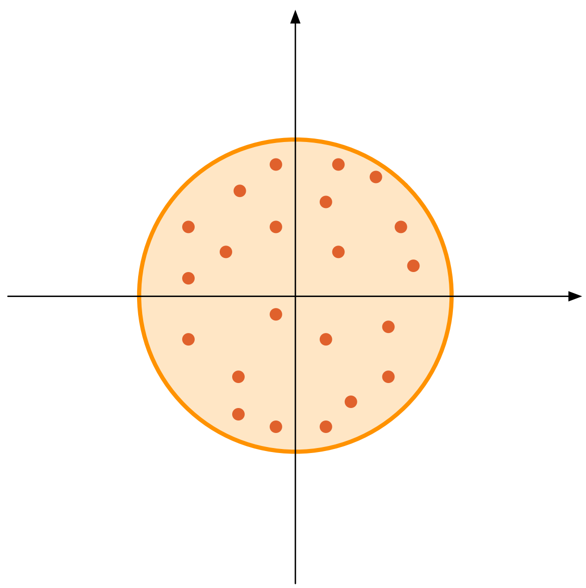
The **signal strength** is θ, quantifies the distance between the distributions.

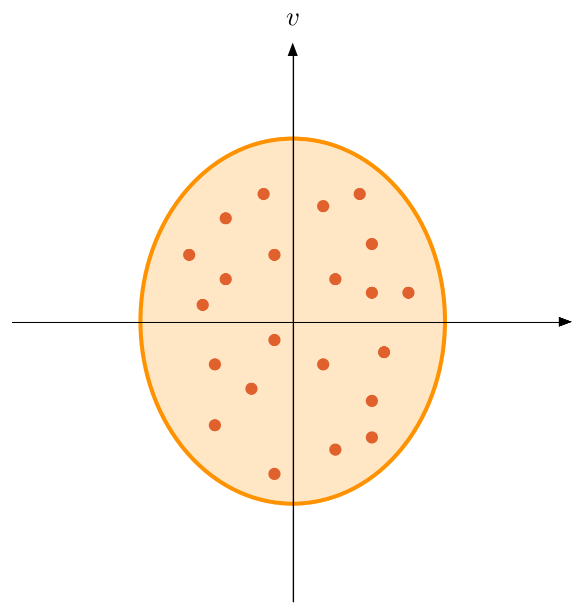
Sparse principal component detection

Testing problem between two hypotheses.

$$\begin{cases} H_0 : I_p \\ H_1 : I_p + \theta v v^\top, \quad v \in \mathcal{B}_0(k) \end{cases}$$

v is a k-sparse unit vector. $\mathcal{B}_0(k) = \{v \in \mathbf{R}^p \,:\, |v|_2 = 1\,,\, |v|_0 \leq k\}$.





Isotropy: $\mathcal{N}(0, I_p)$

Sparse PC: $\mathcal{N}(0, I_p + \theta\, v v^\top)$, small θ

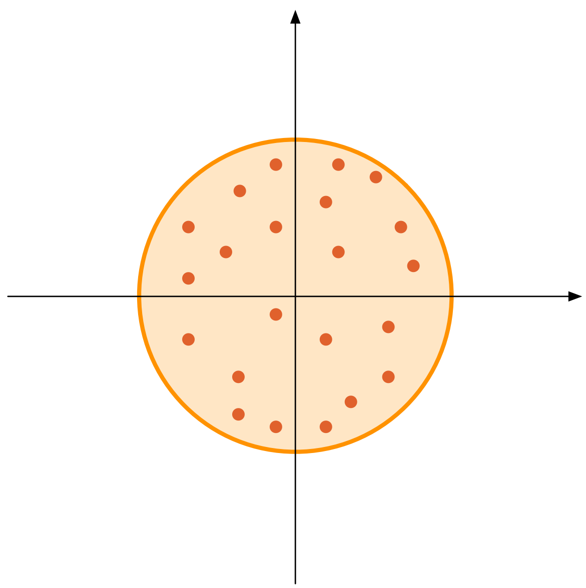
The **signal strength** is θ, quantifies the distance between the distributions.

Sparse principal component detection

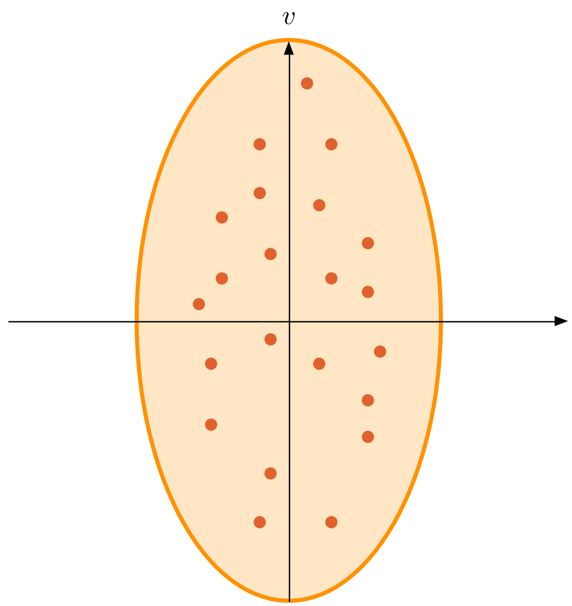
Testing problem between two hypotheses.

$$\begin{cases} H_0 : I_p \\ H_1 : I_p + \theta v v^\top, \quad v \in \mathcal{B}_0(k) \end{cases}$$

v is a k-sparse unit vector. $\mathcal{B}_0(k) = \{v \in \mathbf{R}^p : |v|_2 = 1, |v|_0 \leq k\}$.



Isotropy: $\mathcal{N}(0, I_p)$



Sparse PC: $\mathcal{N}(0, I_p + \theta\, v v^\top)$, large θ

The **signal strength** is θ, quantifies the distance between the distributions.

Results - Detection

Statistics

- Maximizing $x^\top \hat{\Sigma} x$ over unit sparse vectors, $\lambda^k_{\text{max}}(\hat{\Sigma})$. Test powerful in regime

$$\theta \gtrsim \sqrt{\frac{k \log(p/k)}{n}}.$$

- Lower bounds of the same order, optimal result.

Computations

- SDP relaxation of $\lambda^k_{\text{max}}(\hat{\Sigma})$ by d'Aspremont et al (07), requires

$$\theta \gtrsim \sqrt{\frac{k^2 \log(p)}{n}}.$$

- Can be better than $\lambda_{\text{max}}(\hat{\Sigma})$, but still a suboptimal result.

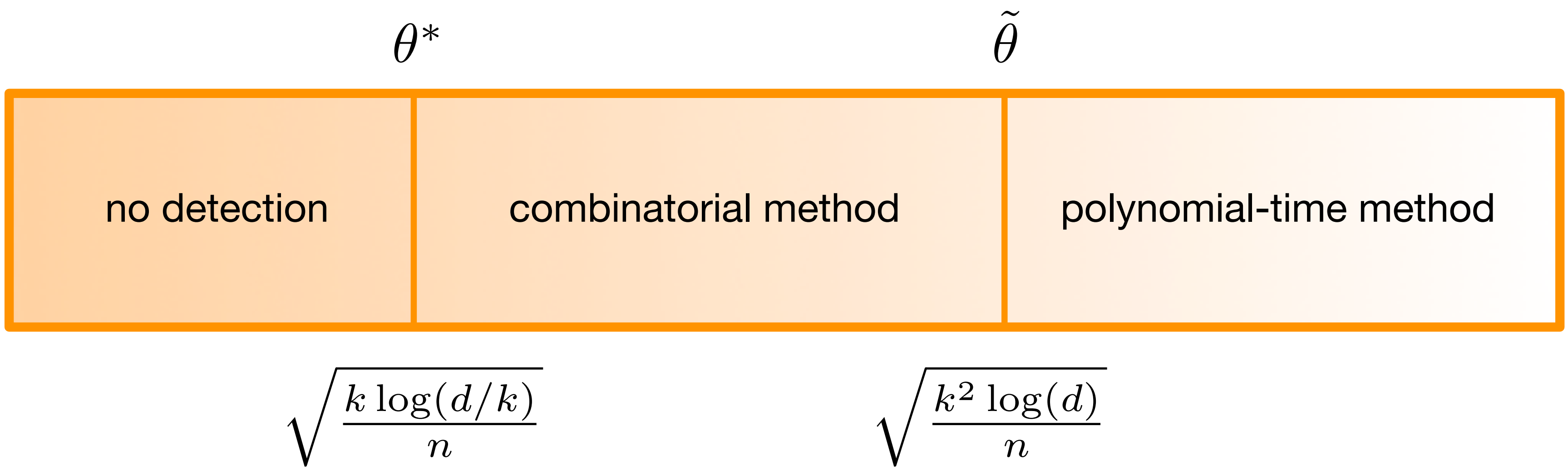
Overall picture - Testing

Computationally efficient tests seem to require

$$\theta \approx \sqrt{\frac{k^2 \log(p)}{n}}$$

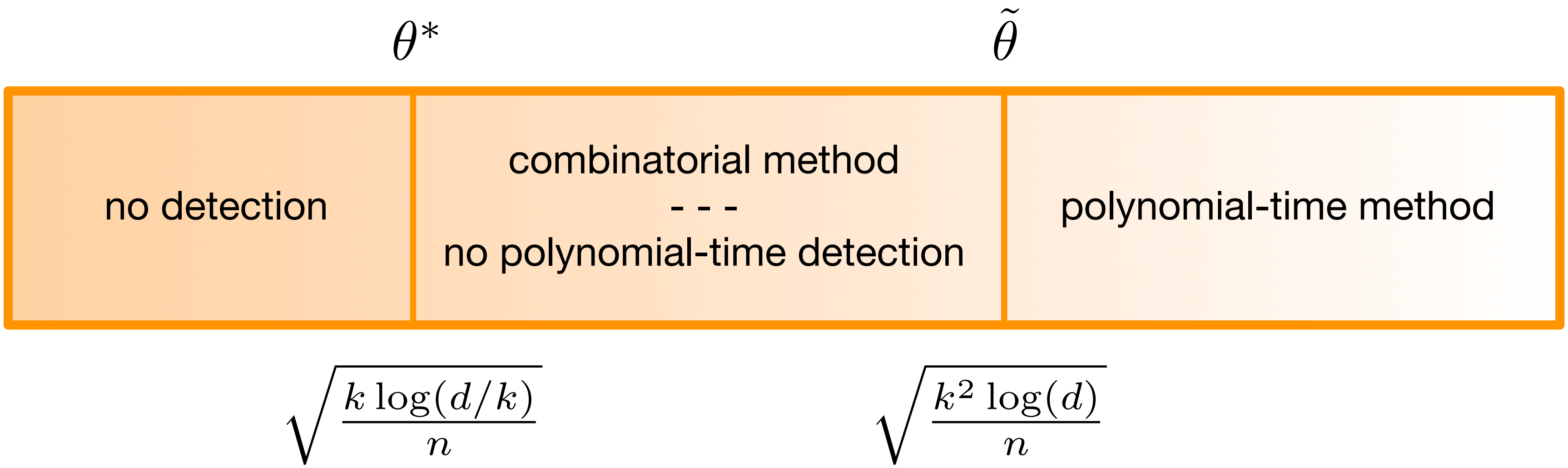
Rate observed for several explicit testing methods.

Tests: Diagonal method - Johnstone (01), SDP - d'Aspremont et al. (07), MDP - Berthet and Rigollet (13), other heuristics.



The actual situation?

Situation suggested by those results

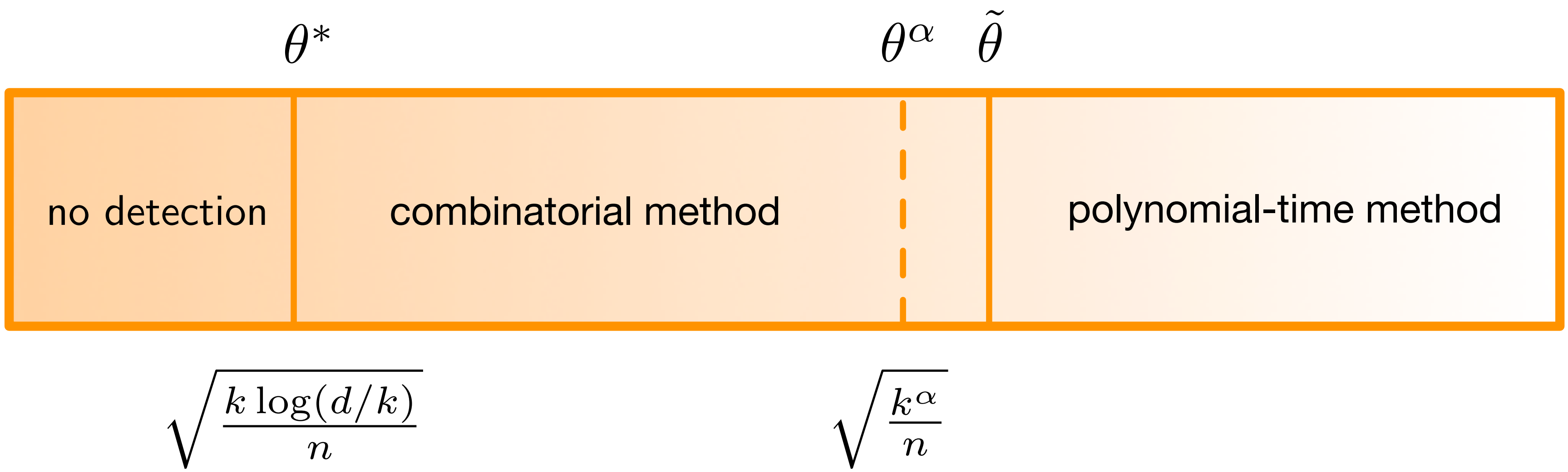


So far, only upper bounds, suggestions.

Need for **Complexity Theoretic Lower Bounds**

Reduction

We employ the following strategy:



If detection at $\theta \leq \theta^\alpha = \sqrt{\frac{k^\alpha}{n}}$, $\alpha \in [1, 2)$ with $\psi \in \mathrm{P}$ . . .

. . . then we can solve a computationally hard problem in polynomial time.

Reduction in statistics: Le Cam deficiency

- Notion of similarity between statistical problems on $\mathcal{P} = \{P_\gamma : \gamma \in \Gamma\}$ and $\mathcal{Q} = \{Q_\gamma : \gamma \in \Gamma\}$.

$$\delta(\mathcal{P}, \mathcal{Q}) = \inf_{T:\mathsf{Markov}} \sup_{\gamma \in \Gamma} d_{\mathsf{TV}}(P_\gamma, TQ_\gamma)$$

"if you can solve one problem, you can solve the other up to probability $\delta(\mathcal{P}, \mathcal{Q})$."

- Given an estimator $\hat{\gamma}_{\mathcal{P}}$, one can construct an estimator $\hat{\gamma}_{\mathcal{Q}}$ such that

$$Q_\gamma(\hat{\gamma}_{\mathcal{Q}} \neq \gamma) \leq P_\gamma(\hat{\gamma}_{\mathcal{P}} \neq \gamma) + \delta(\mathcal{P}, \mathcal{Q}) \,.$$

- Done by transforming the instances, using $\hat{\gamma}_{\mathcal{P}}$ as a blackbox.
- Similar to the notion of worst-case hardness in computer science.

Reduction in statistics: Le Cam deficiency

- Notion of **computational** similarity between statistical problems on $\mathcal{P} = \{P_\gamma \,:\, \gamma \in \Gamma\}$ and $\mathcal{Q} = \{Q_\gamma \,:\, \gamma \in \Gamma\}$.

$$\delta_{\textbf{comp}}(\mathcal{P}, \mathcal{Q}) = \inf_{T:\text{Markov } \textbf{comp.}} \sup_{\gamma \in \Gamma} d_{\mathsf{TV}}(P_\gamma, TQ_\gamma)$$

"if you can **computationally** solve one problem, you can solve the other up to probability $\delta_{\textbf{comp}}(\mathcal{P}, \mathcal{Q})$."

- Given a **computationally tractable** estimator $\hat{\gamma}_{\mathcal{P}}$, one can construct a **computationally tractable** estimator $\hat{\gamma}_{\mathcal{Q}}$ such that

$$Q_\gamma(\hat{\gamma}_{\mathcal{Q}} \neq \gamma) \leq P_\gamma(\hat{\gamma}_{\mathcal{P}} \neq \gamma) + \delta_{\textbf{comp}}(\mathcal{P}, \mathcal{Q})\,.$$

- Done by transforming the instances, using $\hat{\gamma}_{\mathcal{P}}$ as a blackbox.
- Similar to the notion of worst-case hardness in computer science.

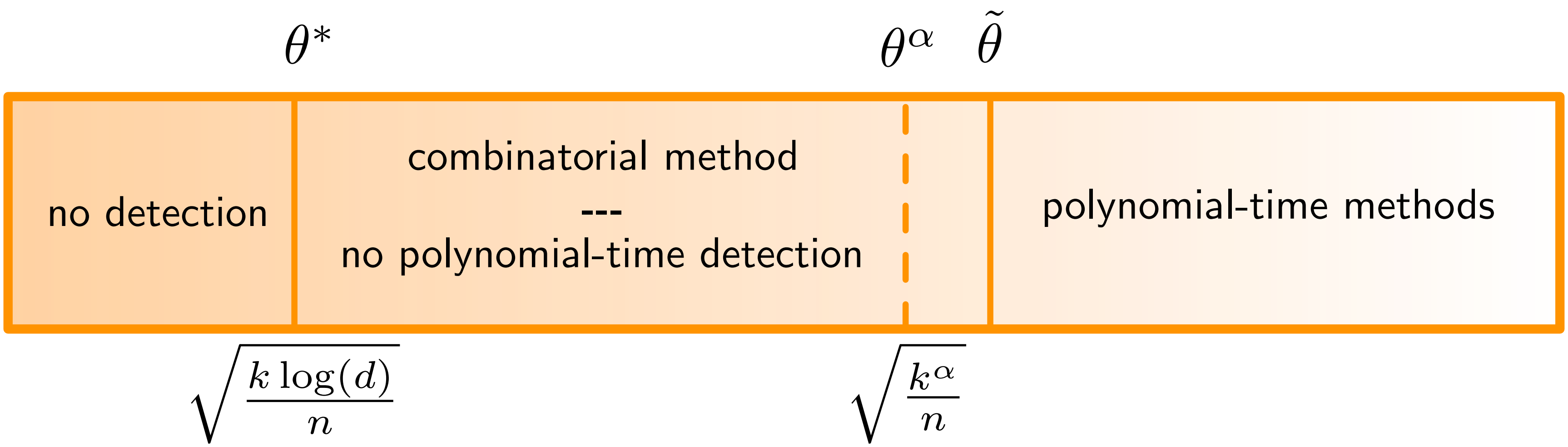
Reasoning by contradiction

- What is the consequence of a test that works for sparse PCA at regime $\theta^\alpha = \sqrt{\frac{k\alpha}{n}}$, $\alpha \in [1, 2)$? **(hypothetical blackbox)**
- **Theorem:** One can transform in polynomial time an instance of size m of planted clique of size $k = m^{\frac{1}{4-\alpha}}$ in an instance of sparse PCA at regime θ^α.

- **Corrollary:** Computational limit to solve Sparse PCA in regime θ^α, $\alpha \in [1, 2)$ if computational limit to solve planted clique in regime $k = m^c$, $c < 1/2$ **(conjectured)**.

Computational hardness in statistics

- Optimal signal level: information-theoretic barriers, combinatorial method.
- Linking sparse PCA and planted clique B., Rigollet (13), Wang et al. (16).



Take-home message: gap of $\sqrt{k}$ for computationally efficient methods.

- Line of work on establishing computational limits by reduction.
- Approach related to "algorithm-independence" point of view.

Ising blockmodel

P. Rigollet (MIT)

P. Srivastava (Caltech $\rightarrow$ Tata Inst.)

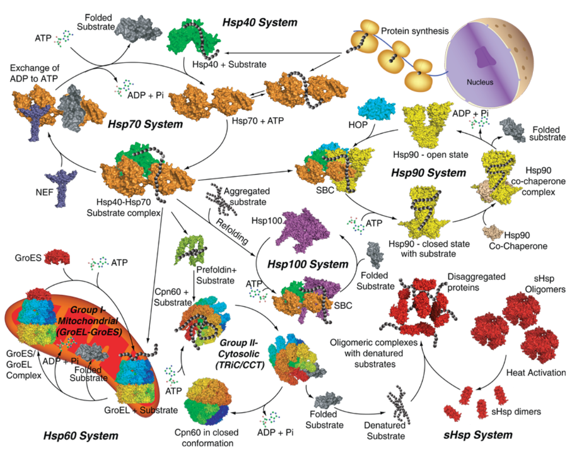
- **Exact recovery in the Ising blockmodel**

  Q. B., P.Rigollet, and P. Srivastava

  Ann. Statis. (to appear)

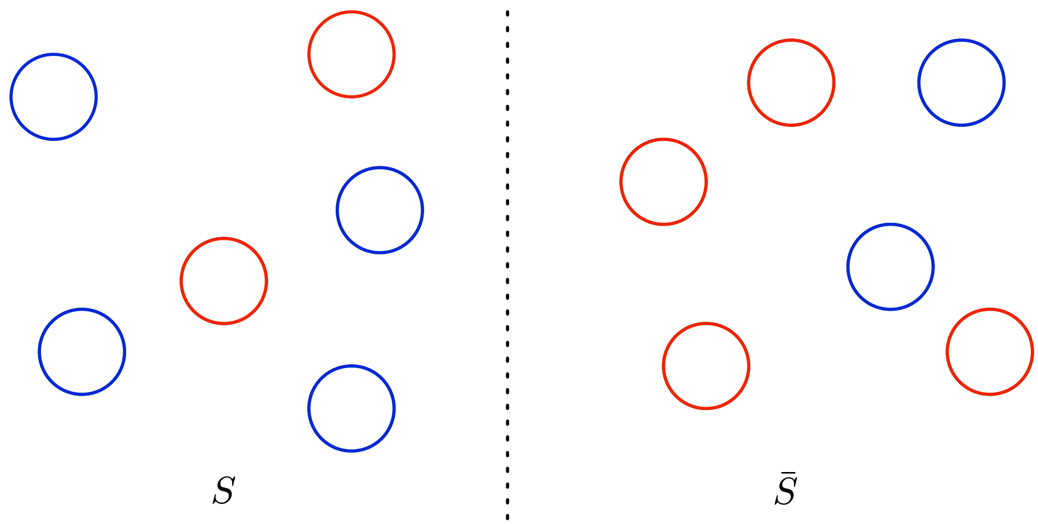
Motivation

- Finding communities in populations, based on similar **behavior** and **influence**.
- One of the justifications for **stochastic blockmodels**
- What if we observe the **behavior**, not the graph?



Motivation

Model with p individuals, observation $\sigma \in \{-1,1\}^p$, balanced communities $(S, \bar{S})$.



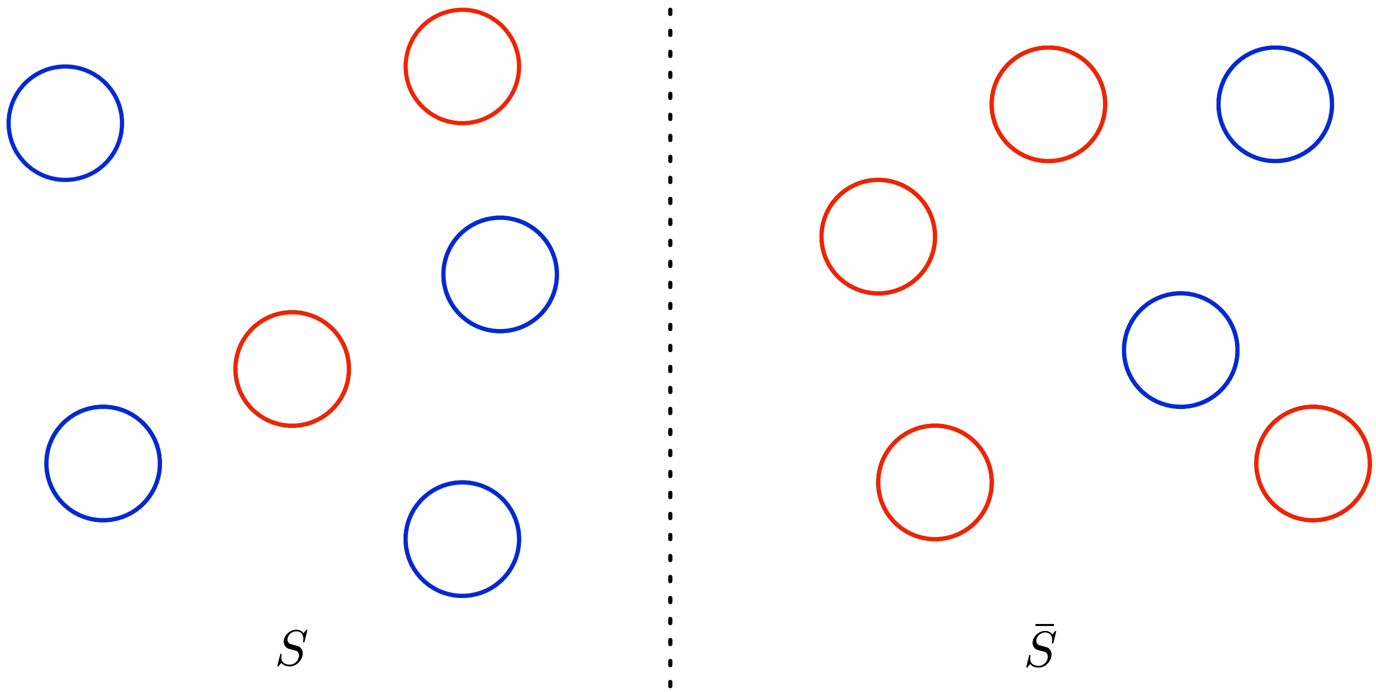
For a given question: members of the community influence an individual's answer.

Example: A metal sheet has temperature at point x, y given by $T(x,y) = T_0 + k(x^2+y^2)$. What is $T(r,\theta)$?

$+1$) $T(r,\theta) = T_0 + kr^2$

-1) $T(r,\theta) = T_0 + k(r^2+\theta^2)$

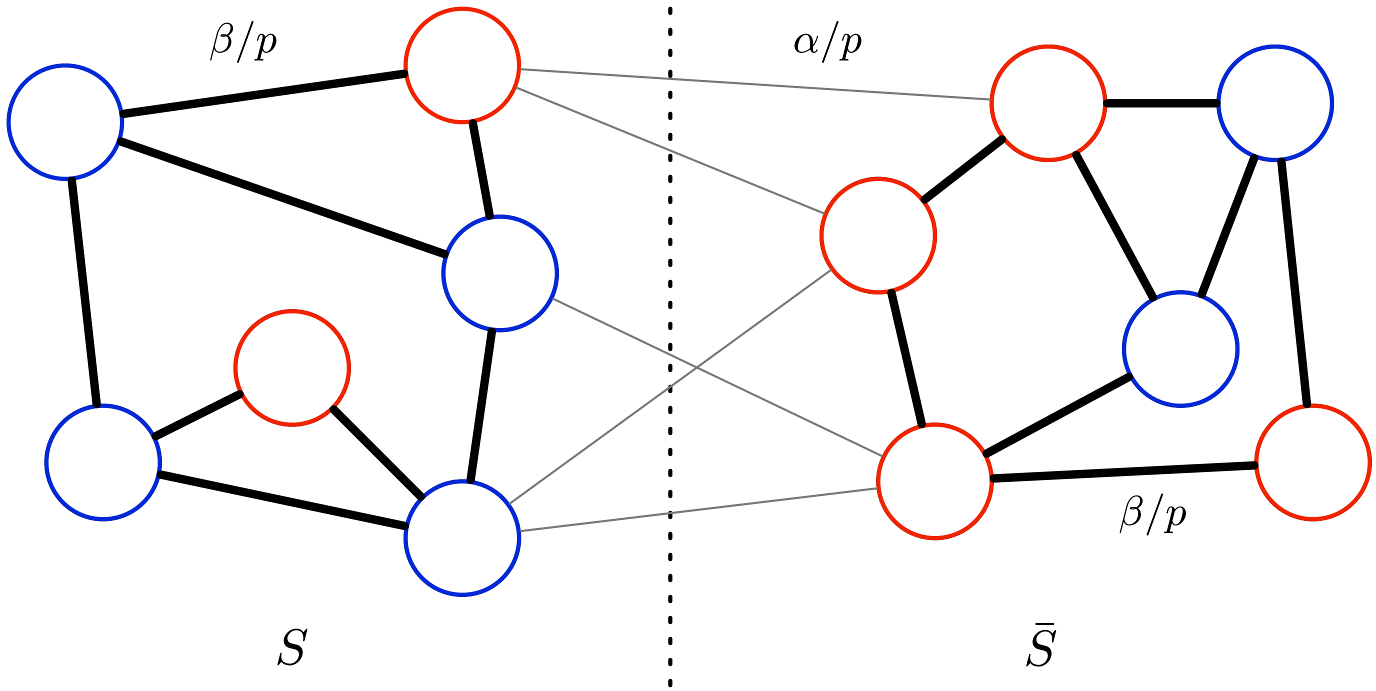
Motivation



Model with p individuals, $\sigma \in \{-1, 1\}^p$ and balanced communities $(S, \bar{S})$.

$$\mathbf{P}_S(\sigma) = \underbrace{\qquad\qquad\qquad\qquad\qquad\qquad}_{?} .$$

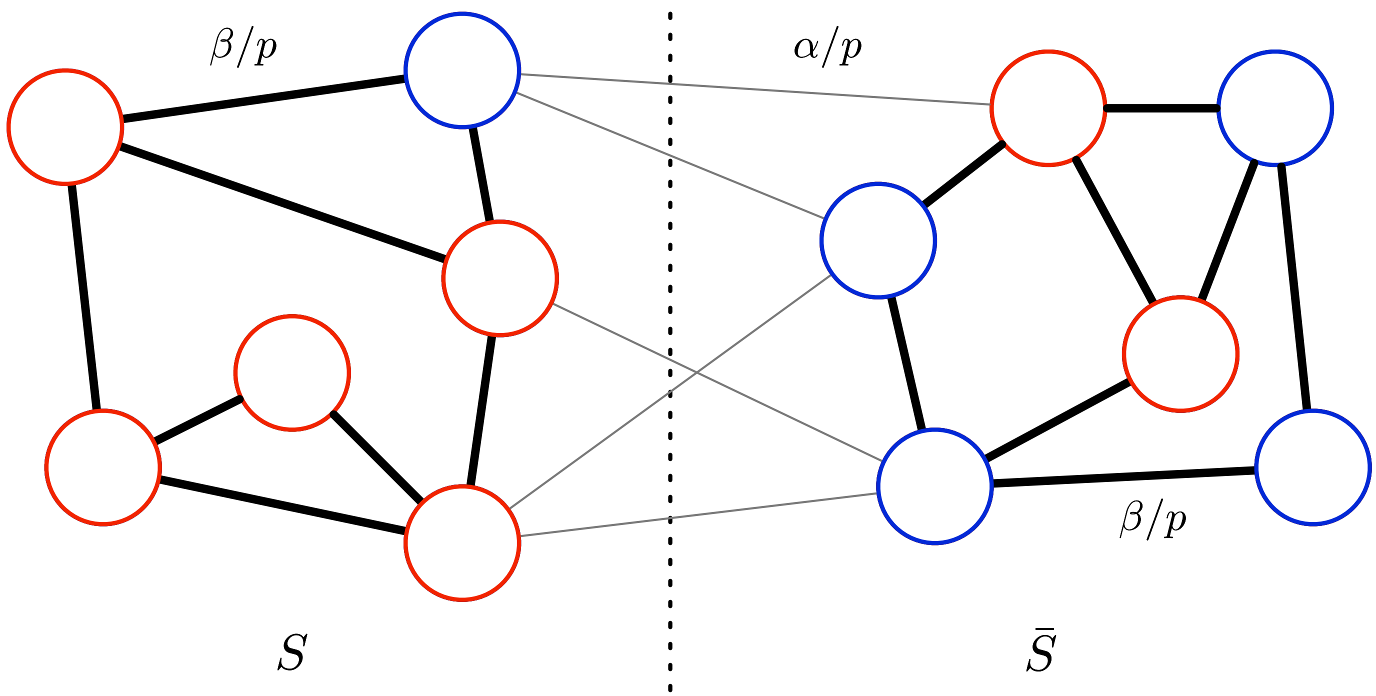
Motivation



Model with p individuals, $\sigma \in \{-1, 1\}^p$ and balanced communities $(S, \bar{S})$.

$$\mathbf{P}_S(\sigma) = \frac{1}{Z_{\alpha,\beta}} \exp\left[\frac{\beta}{2p}\sum_{i\sim j}\sigma_i\sigma_j + \frac{\alpha}{2p}\sum_{i\nsim j}\sigma_i\sigma_j\right].$$

Motivation



Model with p individuals, $\sigma \in \{-1, 1\}^p$ and balanced communities $(S, \bar{S})$.

$$\mathbf{P}_S(\sigma) = \frac{1}{Z_{\alpha,\beta}} \exp\left[\frac{\beta}{2p} \sum_{i \sim j} \sigma_i \sigma_j + \frac{\alpha}{2p} \sum_{i \nsim j} \sigma_i \sigma_j\right].$$

Problem description

Ising blockmodel:

$$\mathbf{P}_S(\sigma) = \frac{1}{Z_{\alpha,\beta}} \exp\Big[\frac{\beta}{2p}\sum_{i\sim j}\sigma_i\sigma_j + \frac{\alpha}{2p}\sum_{i\nsim j}\sigma_i\sigma_j\Big] = \frac{1}{Z_{\alpha,\beta}} \exp\big(-\mathcal{H}_{S,\alpha,\beta}(\sigma)\big)\,.$$

Energy decreases (probability increases) with more agreement inside each block.

- Blockmodel: $\mathbf{P}_S(\sigma_i = \sigma_j) = \begin{cases} b & \text{for all } i \sim j \\ a & \text{for all } i \nsim j \end{cases}$
- Balance: $|S| = |\bar{S}| = p/2$,
- Homophily: $\beta > 0 \Leftrightarrow b > 1/2$,
- Assortativity: $\beta > \alpha \Leftrightarrow b > a$.
- Relationship with Hopfield model, and posterior for the SBM.

Observations: $\sigma^{(1)}, \ldots, \sigma^{(n)} \in \{-1,1\}^p$ i.i.d. from $\mathbf{P}_S$.

Objective: recover the *balanced* partition $(S, \bar{S})$ from observations.

Stochastic blockmodels

- **one** observation of random graph on p vertices

$$\mathbf{P}(i \leftrightarrow j) = \begin{cases} b & \text{for all } i \sim j \\ a & \text{for all } i \nsim j \end{cases}$$

- Exact recovery using SDP iff

$$a = \mathsf{a}\frac{\log p}{p}, b = \mathsf{b}\frac{\log p}{p}$$

and $(\mathsf{a}+\mathsf{b})/2 > 1 + \sqrt{\mathsf{ab}}$

Abbe, Bandeira, Hall '14
Mossel, Neeman, Sly '14
Xu, Lelarge, Massoulieé '14
Hajek, Wu '16

Graphical models / MRF

- n observations $\sigma^{(1)}, \ldots, \sigma^{(n)}$ i.i.d.

$$\mathbf{P}(\sigma) \propto \exp\left[\frac{\beta}{2p}\sum_{i,j} J_{ij}\sigma_i\sigma_j\right]$$

- Goal estimating sparse $J = \{J_{ij}\}$ (max degree d)
- Sample complexity $n \gg 2^d \log p$

Chow-Liu '68
Bresler, Mossel, Sly '08
Santhanam, Wainwright '12
Bresler '15
Vuffray, Misra, Lokhov, Chertkov '16

Wigner matrices

Wishart matrices

Problem overview

- By symmetry $\mathbf{E}[\sigma_i] = 0$, what of the second moment?
- Structure of the problem visible in the **covariance matrix** Σ

$$\Sigma = \mathbf{E}[\sigma\sigma^\top] = \left(\begin{array}{c|c} \Delta & \Omega \\ \hline \Omega & \Delta \end{array}\right) + (1-\Delta)I_p\,.$$

- Difficulty of the problem related with the values of quantities $\Delta, \Omega \in (-1, 1)$

$$\Delta = 2b - 1\,, \quad \Omega = 2a - 1\,.$$

- Parallel with the **stochastic block model** on graphs with independent edges
- **Summary - Analysis**
  - **Deviations:** Behavior of $\hat{\Sigma}$ around Σ, sample size guarantees.
  - **Population:** Results function of $\Delta - \Omega$, scaling in p and α, β.

Maximum likelihood estimation

- Log-likelihood $\mathcal{L}_n(S) = -n \log Z_{\alpha,\beta} + \frac{n}{2}\mathbf{Tr}[\hat{\Sigma} Q_S]$
- Maximum likelihood estimator:

$$\hat{V} \in \operatorname*{argmax}_{V \in \mathcal{P}} \mathbf{Tr}[\hat{\Sigma} V]\,, \quad \text{where} \quad \mathcal{P} = \{vv^\top : v \in \{-1,1\}^p, v^\top \mathbf{1}_{[p]} = 0\}\,.$$

- Define $\Gamma = P\Sigma P$ and $\hat{\Gamma} = P\hat{\Sigma}P$, for a projector P on the orthogonal of $\mathbf{1}$:

$$\Gamma = (1-\Delta)P + p\,\frac{\Delta - \Omega}{2} u_S u_S^\top\,, \qquad u_S = \frac{1}{\sqrt{p}}(\mathbf{1}_S - \mathbf{1}_{\bar{S}})$$

- For all $V \in \mathcal{P}$, $\mathbf{Tr}[\hat{\Gamma} V] = \mathbf{Tr}[\hat{\Sigma} V]$, so equivalently

$$\hat{V} \in \operatorname*{argmax}_{V \in \mathcal{P}} \mathbf{Tr}[\hat{\Gamma} V]$$

SDP relaxation

$$\hat{V} \in \underset{V \in \mathcal{P}}{\operatorname{argmax}} \, \mathbf{Tr}[\hat{\Gamma} V] \,, \quad \text{where} \quad \mathcal{P} = \{vv^\top : v \in \{-1,1\}^p, v^\top \mathbf{1}_{[p]} = 0\} \,.$$

NP-Hard (Min bisection)

- Semidefinite convex relaxation of $\mathcal{P}$

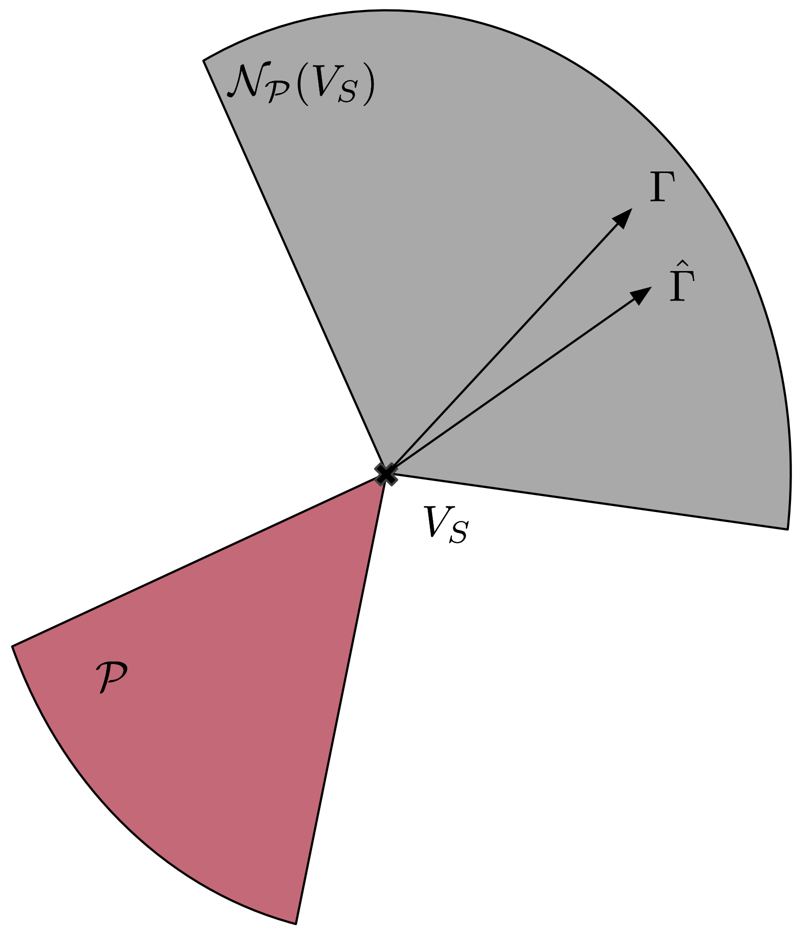
$$\mathcal{E} = \{V : \mathrm{diag}(V) = \mathbf{1}, V \succeq 0\} \,.$$

Change of variable $V = vv^\top$

MAXCUT Goemans-Williamson (95)

- Point V solution of $\max_{V \in \mathcal{C}} \mathbf{Tr}[\hat{\Gamma} V]$ equivalent to $\hat{\Gamma} \in \mathcal{N}_{\mathcal{C}}(V)$

- Relaxation is tight for population matrix Γ: $\hat{V} = V_S$ if $n = \infty$.



SDP relaxation

$$\hat{V} \in \operatorname*{argmax}_{V \in \mathcal{P}} \mathbf{Tr}[\hat{\Gamma} V]\,, \quad \text{where} \quad \mathcal{P} = \{vv^\top : v \in \{-1,1\}^p, v^\top \mathbf{1}_{[p]} = 0\}\,.$$

NP-Hard (Min bisection)

- Semidefinite convex relaxation of $\mathcal{P}$

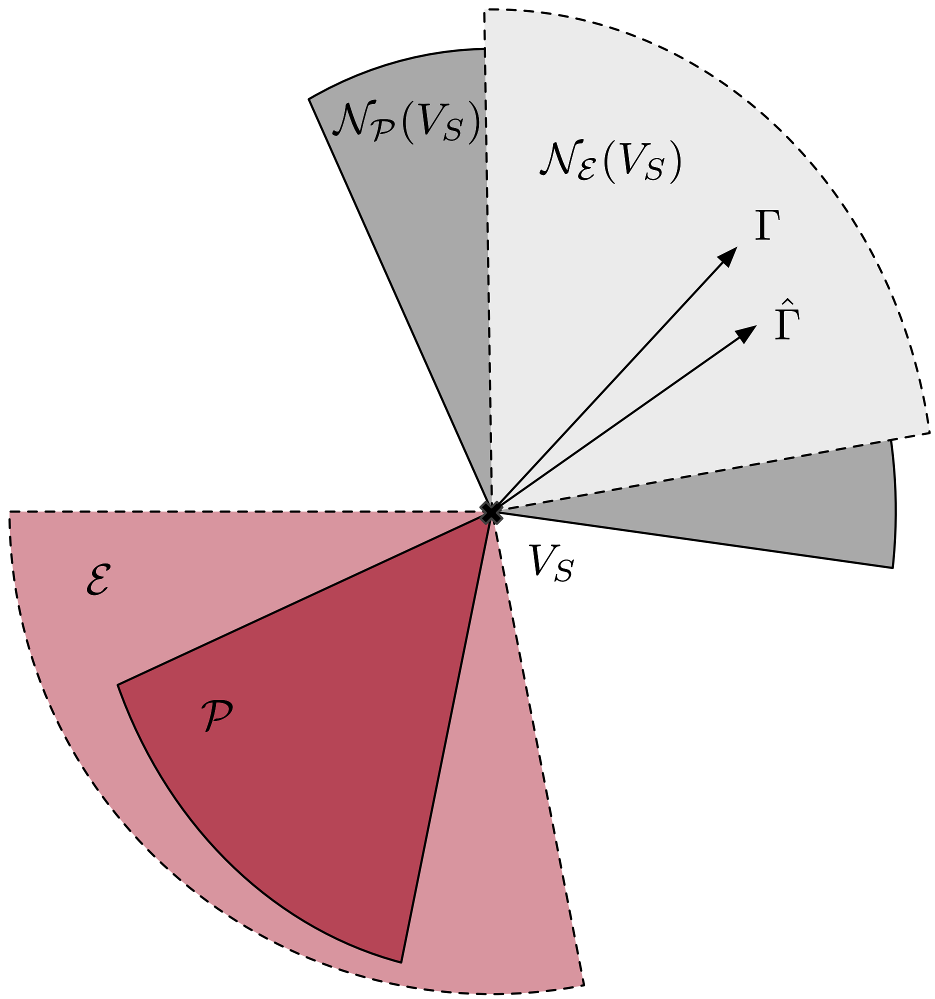
  $$\mathcal{E} = \{V : \text{diag}(V) = \mathbf{1}, V \succeq 0\}\,.$$

  Change of variable $V = vv^\top$

  MAXCUT Goemans-Williamson (95)

- Point V solution of $\max_{V \in \mathcal{C}} \mathbf{Tr}[\hat{\Gamma} V]$ equivalent to $\hat{\Gamma} \in \mathcal{N}_{\mathcal{C}}(V)$

- Relaxation is tight for population matrix Γ: $\hat{V} = V_S$ if $n = \infty$.



Exact recovery

- **Upper bound**: we have $\hat{V} = V_S$ with probability $1-\delta$ for

$$n \gtrsim \frac{1}{C_{\alpha,\beta}} \frac{\log(p/\delta)}{\Delta - \Omega},$$

by bounding function of $\Gamma - \hat{\Gamma}$, a sum of independent matrices. Tropp (12)

- **Matching lower bound**: Information theoretic argument yields

$$n \leq \frac{\gamma}{\beta - \alpha} \frac{\log(p/4)}{\Delta - \Omega} \implies \mathbf{P}(\text{recovery}) \lesssim \gamma$$

- Full understanding of the scaling of $\Delta - \Omega$ needed.

The Curie–Weiss model ($\alpha = \beta$)

$$\Sigma = \left(\begin{array}{c|c} \Delta & \Delta \\ \hline \Delta & \Delta \end{array}\right) + (1-\Delta) I_p$$

- **Mean magnetization**: $\mu = \frac{\mathbf{1}^\top \sigma}{p} \in [-1, 1]$. Observe that

$$\Delta = \frac{1}{p^2} \sum_{i,j=1}^{p} \mathbf{E}[\sigma_i \sigma_j] - \frac{1}{p} = \mathbf{E}[\mu^2] - \frac{1}{p} \approx \mathbf{E}[\mu^2]$$

- **Free energy**: μ is a sufficient statistic

$$\mathbf{P}_\beta(\mu) \approx \frac{1}{Z_\beta} \exp\Big(-\frac{p}{4}\, g_\beta^{\mathsf{CW}}(\mu)\Big)\,, \quad g_\beta^{\mathsf{CW}}(\mu) = -2\beta\mu^2 + 4h\Big(\frac{1+\mu}{2}\Big)$$

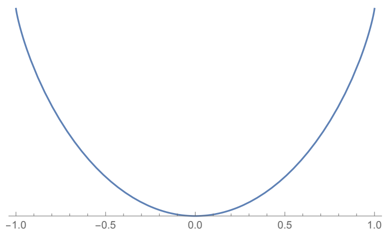
- **Ground states**: Minimizers $G \subset [-1, 1]$ of $g_\beta^{\mathsf{CW}}(\mu)$.

- **Concentration**: $\mu \approx$ ground state with exponentially large probability so

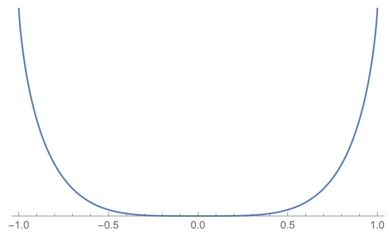
$$\Delta \approx \mathbf{E}[\mu^2] \approx \frac{1}{|G|} \sum_{\mathsf{s} \in G} \mathsf{s}^2\,.$$

Free energy of the Curie-Weiss model

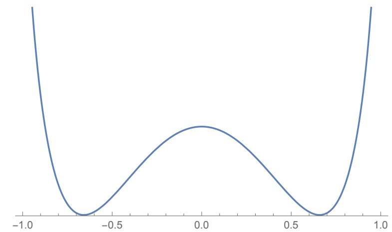
Ground states $\mathcal{G} = \{\tilde{\mu}(\beta), -\tilde{\mu}(\beta)\}, \tilde{\mu}(\beta) \geq 0$:



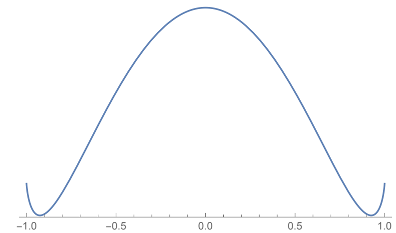
$\beta = 0$



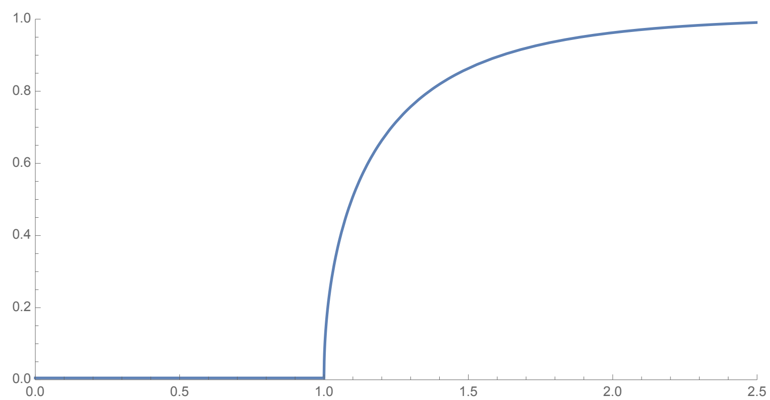
$\beta = 1$



$\beta = 1.2$



$\beta = 1.8$



$\beta \mapsto \tilde{\mu}(\beta)$

$$\Delta \approx \frac{1}{|G|} \sum_{\mathsf{s} \in G} \mathsf{s}^2 = \tilde{\mu}(\beta)^2$$

Free energy of the Ising blockmodel

- Energy function of the **block magnetizations**: $(\mu_S, \mu_{\bar{S}}) = \frac{2}{p}(\mathbf{1}_S^\top \sigma, \mathbf{1}_{\bar{S}}^\top \sigma)$

$$\mathbf{P}_S(\sigma) = \frac{1}{Z_{\alpha,\beta}} \exp\Big(-\frac{p}{8}\big(-\beta\mu_S^2 - \beta\mu_{\bar{S}}^2 - 2\alpha\,\mu_S\,\mu_{\bar{S}}\big)\Big)$$

- **Marginal:** number of configurations with magnetizations μ is $\binom{(p/2)}{\frac{1+\mu}{2}(p/2)}$

$$\mathbf{P}_S(\mu_S, \mu_{\bar{S}}) \approx \frac{1}{Z_{\alpha,\beta}} \exp\Big(-\frac{p}{8}\, g_{\alpha,\beta}(\mu_S, \mu_{\bar{S}})\Big)$$

where $g_{\alpha,\beta}$ is the free energy defined by

$$g_{\alpha,\beta}(\mu_S, \mu_{\bar{S}}) = -\beta\mu_S^2 - \beta\mu_{\bar{S}}^2 - 2\alpha\,\mu_S\,\mu_{\bar{S}} + 4h\Big(\frac{1+\mu_S}{2}\Big) + 4h\Big(\frac{1+\mu_{\bar{S}}}{2}\Big)\,.$$

The Ising blockmodel model ($\alpha < \beta$) $\Sigma = \left(\begin{array}{c|c} \Delta & \Omega \\ \hline \Omega & \Delta \end{array}\right) + (1-\Delta) I_p$

- **Block magnetizations**: $\mu_S = \frac{\mathbf{1_S}^\top \sigma}{p/2}, \mu_{\bar{S}} = \frac{\mathbf{1_{\bar{S}}}^\top \sigma}{p/2} \in [-1, 1]$. Observe that

$$\Delta \approx \frac{2}{p^2} \sum_{i \sim j} \mathbf{E}[\sigma_i \sigma_j] \approx \frac{1}{2} \mathbf{E}[\mu_S^2 + \mu_{\bar{S}}^2] \quad \text{and} \quad \Omega = \frac{2}{p^2} \sum_{i \nsim j} \mathbf{E}[\sigma_i \sigma_j] = \mathbf{E}[\mu_S \, \mu_{\bar{S}}]$$

- **Free energy**: $(\mu_S, \mu_{\bar{S}}) \in [-1, 1]^2$ is a sufficient statistic

$$\mathbf{P}_S(\mu_S, \mu_{\bar{S}}) \approx \frac{1}{Z_{\alpha,\beta}} \exp\Big(- \frac{p}{8} \, g_{\alpha,\beta}(\mu_S, \mu_{\bar{S}}) \Big)$$

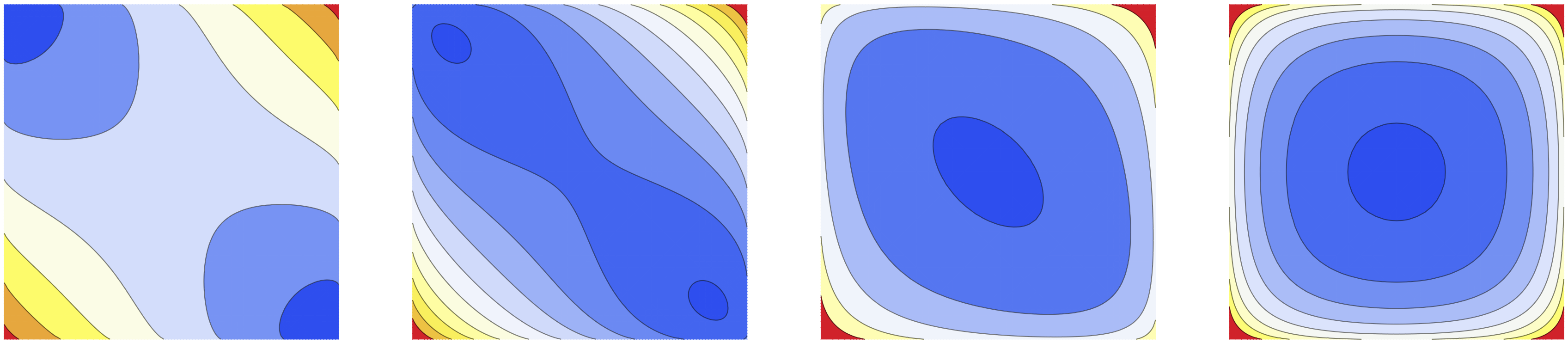
- **Ground states**: Minimizers $G \subset [-1, 1]^2$ of $g^{\mathsf{CW}}_{\alpha,\beta}(\mu_S, \mu_{\bar{S}})$.

- **Concentration**: $(\mu_S, \mu_{\bar{S}}) \approx$ ground states with exp. large probability so

$$\Delta - \Omega \approx \frac{1}{2} \mathbf{E}[(\mu_S - \mu_{\bar{S}})^2] \approx \frac{1}{|G|} \sum_{\mathsf{s} \in G} (\mathsf{s}_1 - \mathsf{s}_2)^2 \, .$$

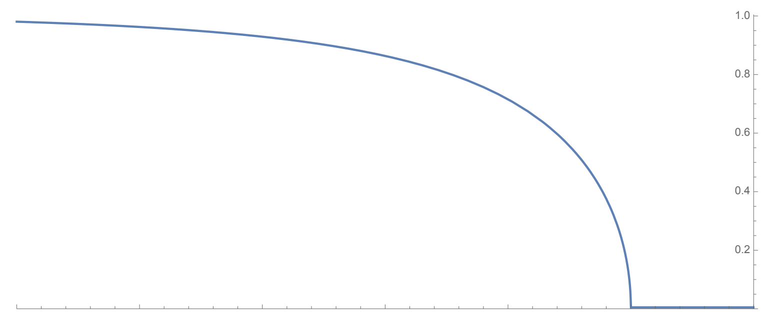
Ground states for the Ising blockmodel

$$g_{\alpha,\beta}(\mu_S, \mu_{\bar{S}}) = -\beta\mu_S^2 - \beta\mu_{\bar{S}}^2 - 2\alpha\,\mu_S\,\mu_{\bar{S}} + 4h\Big(\frac{1+\mu_S}{2}\Big) + 4h\Big(\frac{1+\mu_{\bar{S}}}{2}\Big)$$



$\alpha = -6$ $\qquad$ $\alpha = -2.5$ $\qquad$ $\alpha = -0.5$ $\qquad$ $\alpha = 0$

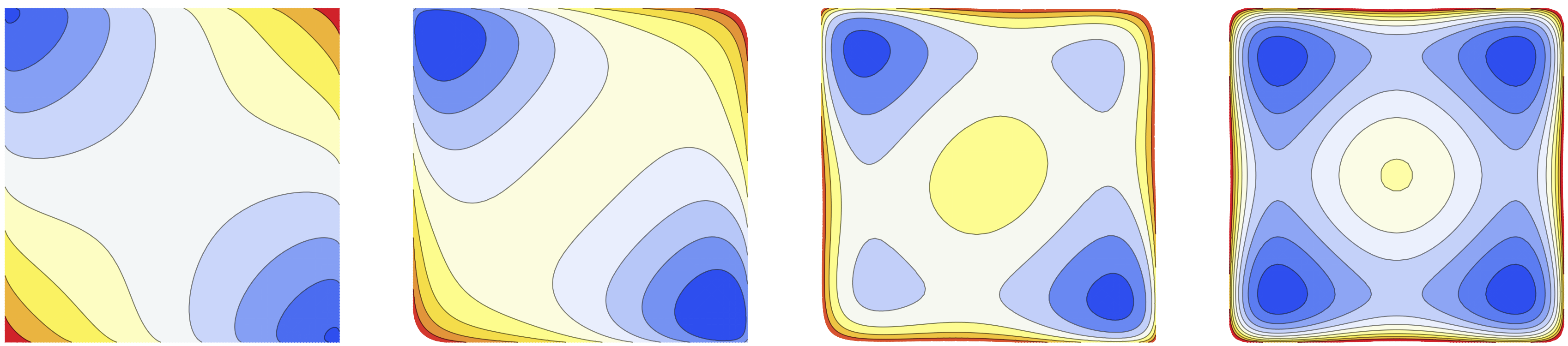
Ground states on the skew-diagonal ($\tilde{\mu}_S = -\tilde{\mu}_{\bar{S}}$) for $\alpha \leq 0$ and fixed $\beta = 1.5 < 2$



$\alpha \mapsto \tilde{\mu}_S(\alpha, \beta = 1.5)$

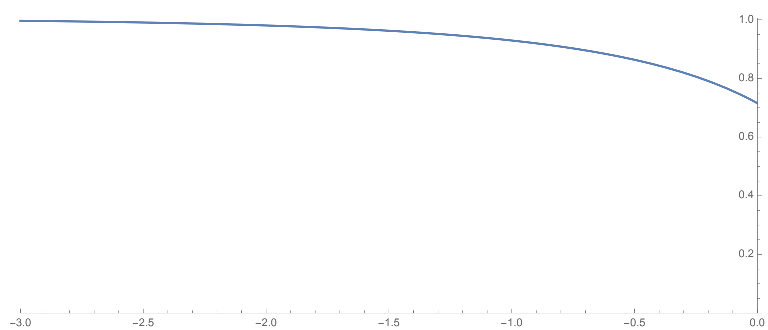
Ground states for the Ising blockmodel

$$g_{\alpha,\beta}(\mu_S, \mu_{\bar S}) = -\beta\mu_S^2 - \beta\mu_{\bar S}^2 - 2\alpha\, \mu_S\, \mu_{\bar S} + 4h\Big(\frac{1+\mu_S}{2}\Big) + 4h\Big(\frac{1+\mu_{\bar S}}{2}\Big)$$



$\alpha = -4$ $\qquad$ $\alpha = -0.9$ $\qquad$ $\alpha = -0.2$ $\qquad$ $\alpha = 0$

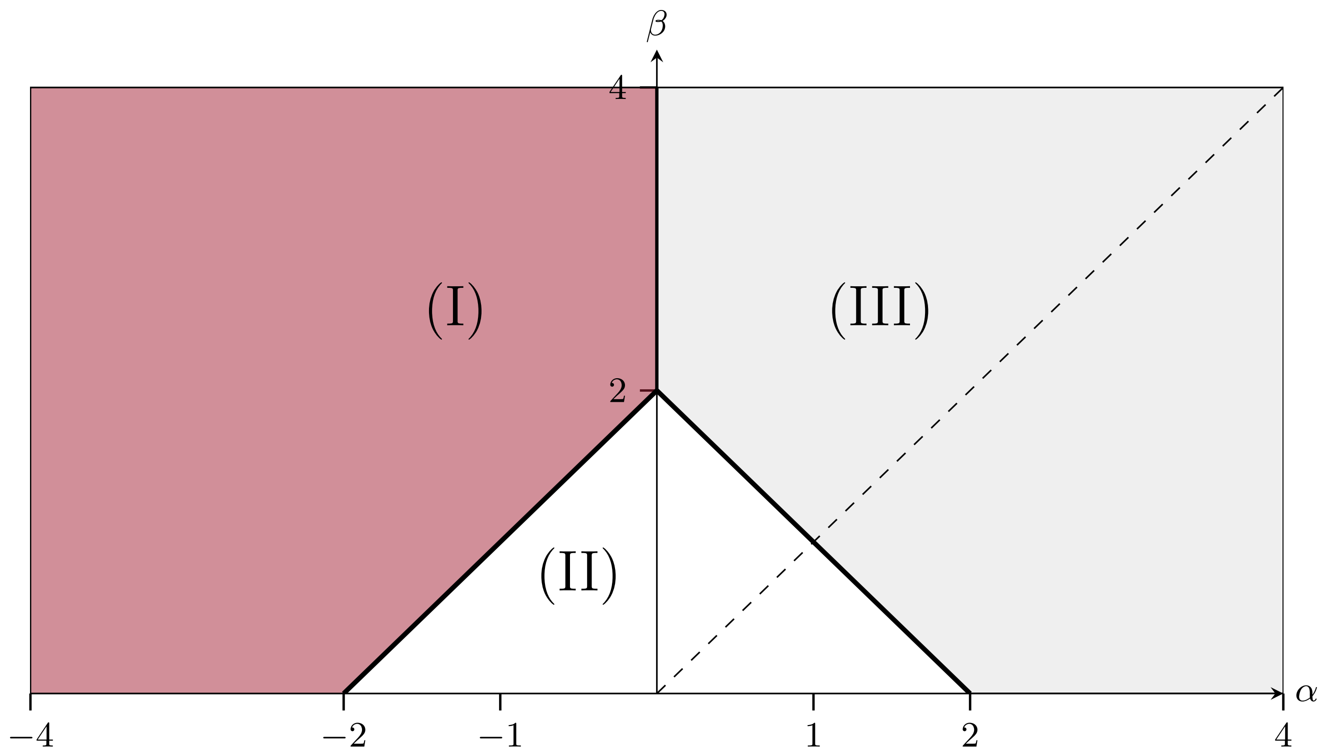
Ground states on the skew-diagonal $(\mu_S = -\mu_{\bar S})$ for $\alpha \le 0$ and fixed $\beta = 2.5 > 2$



$\alpha \mapsto \tilde\mu_S(\alpha, \beta) \qquad \beta = 2.5$

Phase diagram

Full understanding of the position of the ground states for $\beta > 0$, $\alpha < \beta$



- Phase diagram for all the parameter regions
  - Region (I): Two ground states $(\tilde{\mu}_S, \tilde{\mu}_{\bar{S}}) = \pm(\tilde{x}, -\tilde{x})$.
  - Region (II): One ground state at $(0, 0)$.
  - Region (III): Two ground states $(\tilde{\mu}_S, \tilde{\mu}_{\bar{S}}) = \pm(\tilde{x}, \tilde{x})$.

Concentration

- Quantities of interest as expectations of the mean block magnetizations

$$\Delta \approx \frac{1}{2}\mathbf{E}[\mu_S^2 + \mu_{\bar{S}}^2] \quad , \quad \Omega \approx \mathbf{E}[\mu_S \mu_{\bar{S}}] \quad \text{and} \quad \Delta - \Omega \approx \frac{1}{2}\mathbf{E}[(\mu_S - \mu_{\bar{S}})^2] \,.$$

- Gaussian approximation of the discrete distribution with $Z \sim \mathcal{N}(0, I_2)$.

$$\mathbf{E}_{\alpha,\beta}[\varphi(\mu)] \simeq_p \frac{1}{|G|} \sum_{\tilde{s} \in G} \mathbf{E}\Big[\varphi(\tilde{s} + 2\sqrt{\frac{2}{p}} H^{-1/2} Z)\Big] \qquad \forall \, \varphi \,.$$

- Approximation of the gap $\Delta - \Omega$:

$$\Delta - \Omega \simeq_p \begin{cases} 2\tilde{x}^2 & \text{in region (I)} \\ \dfrac{C_{\alpha,\beta}}{p} & \text{in region (II)} \\ \dfrac{C'_{\alpha,\beta}}{p} & \text{in region (III)} \end{cases}$$

Naive estimation

- Covariance matrix:

$$\Sigma = \mathbf{E}[\sigma\sigma^\top] = \left(\begin{array}{c|c} \Delta & \Omega \\ \hline \Omega & \Delta \end{array}\right) + (1-\Delta) I_p \,.$$

- Empirical covariance matrix:

$$\hat{\Sigma} = \frac{1}{n}\sum_{t=1}^{n} \sigma^{(t)}\sigma^{(t)\top} = \Sigma \pm \sqrt{\frac{\log p}{n}} \quad \text{entrywise}$$

- Threshold off-diagonal entries of $\hat{\Sigma}$ at $(\Delta + \Omega)/2$

- Exact recovery if

$$n \gtrsim \begin{cases} \log p & \text{in region (I)} \\ p^2 \log p & \text{in region (II)} \\ p^2 \log p & \text{in region (III)} \end{cases}$$

Exact recovery

- **Upper bound**: we have $\hat{V} = V_S$ with probability $1-\delta$ for

$$n \gtrsim \frac{1}{C_{\alpha,\beta}} \frac{\log(p/\delta)}{\Delta - \Omega},$$

by bounding function of $\Gamma - \hat{\Gamma}$, a sum of independent matrices. Tropp 12

- **Matching lower bound**: Fano's inequality yields

$$n \leq \frac{\gamma}{\beta - \alpha} \frac{\log(p/4)}{\Delta - \Omega} \implies \mathbf{P}(\text{recovery}) \lesssim \gamma$$

- Full understanding of the scaling of $\Delta - \Omega$ gives optimal rates.

$$n \gtrsim \begin{cases} \log p & \text{in region (I)} \\ p \log p & \text{in regions (II) and (III)} \end{cases}$$

with constant factors illustrating further these transitions.

Conclusion

- **Contributions**
  - Model for interactions between individuals in different communities.
  - Analysis from statistical physics to understand parameters of the problem.
  - Study of convex relaxations with an analysis on normal cones.

- **Open questions**
  - Exact recovery threshold, conjecture that $n^* = \frac{C^* \log(p)}{(\beta-\alpha)(\Delta-\Omega)}$.
  - Rates for partial recovery in Hamming distance.
  - Generalization to multiple blocks, more complex structures.

THANK YOU

Work supported by the Isaac Newton Trust and the Alan Turing Institute.