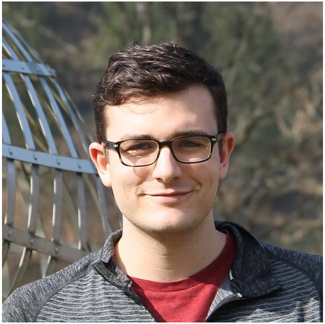
Exact recovery in graphical models

Piyush Srivastava

Tata Institute of Fundamental Research
Mumbai

ICTS, January 2018



Quentin Berthet

Philippe Rigollet

- **Exact recovery in the Ising blockmodel**
  Q. Berthet, P. Rigollet, and P. S.

arXiv:1612.03880v2

To appear in the *Annals of Statistics*

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?

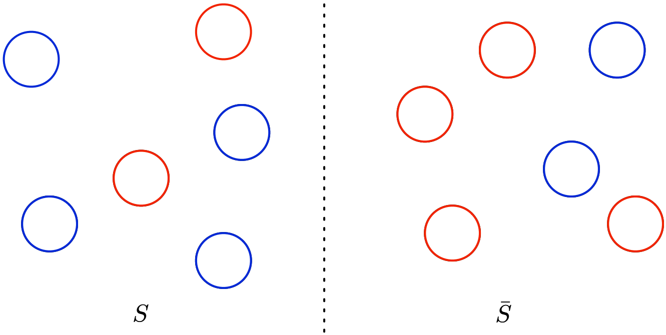
Motivation: Learning the model

- Population of interacting agents are often modeled as graphs
- **Setting:** The population has an (unknown) community structure
  - This is what the graph represents
- But one observes only **behavior** of the nodes, not the graph!

Behaviour ≡ "Sample from the 'model'"
Community structure ≡ "the underlying 'model'"

- Common framework in economics, biology (protein-protein interactions) ...
  - But one needs to be wary of application specific caveats

Motivation: the abstraction



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

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

Problem: Find S from observations of σ

Prior beliefs

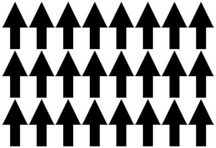
- To solve the problem, one needs to fix a search space of models
- Specifically, there needs to be some reasonably concrete model of the inter-community and intra-community interactions

Prior beliefs

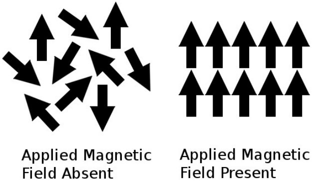
- To solve the problem, one needs to fix a search space of models
- Specifically, there needs to be some reasonably concrete model of the inter-community and intra-community interactions
- We choose to model them in a fashion inspired by model inspired from statistical mechanics

Magnets: the Ising model

Some magnetic materials lose their magnetism just above a critical temperature (Curie temperature)



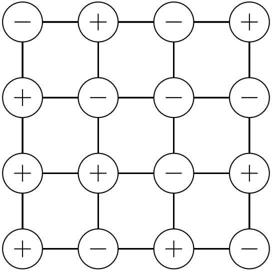
$T < T_c$



$T > T_c$

Perhaps the first phase transition to be abstracted: Ising model (1925)

Ising model



Gibbs distribution

$$\mu(\sigma) \propto \exp\left(\beta \sum_{u \sim v} \sigma_u \sigma_v\right)$$

- Nodes "represent" magnetic domains, β represents inverse temperature $1/T$
- Mean magentization: Average of $\sum_v \sigma_v$ according to μ

$$M(\beta) = \underset{\sigma \sim \mu}{E}[\sum_v \sigma_v]$$

Ising model: Beyond magnets

The Ising model⋆ makes an appearance in a rather wide variety of areas:

- Individual choice theory, as the logit response

  e.g. [McKelvey & Palfrey, *Games and Econ. Behaviour*, 1995]

- Spread of opinions in social networks

  e.g. [Montanari & Saberi, *PNAS*, 2010]

- Computer vision

  e.g. [Geman & Graffigne, *Proc. of the ICM*, 1986]

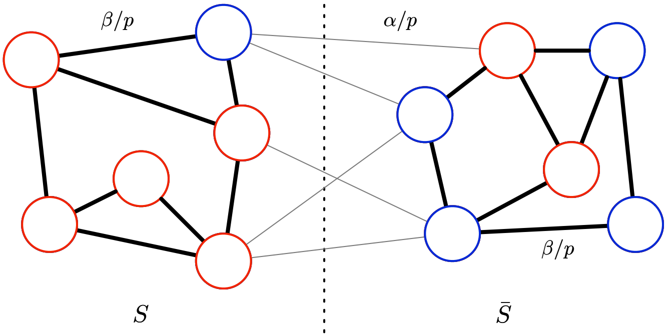
⋆ ...and other related models

Prior beliefs

- To solve the problem, one needs to fix a search space of models
- Specifically, there needs to be some reasonably concrete model of the inter-community and intra-community interactions
- We choose to model them in a fashion inspired by the Ising model

> Recall the earlier cited applications to networks
> e.g. [Montanari & Saberi, *PNAS*, 2010]

The Ising blockmodel



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

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

[Berthet, Rigollet, S., *Ann. Stat.*, to appear]

Problem description

Ising blockmodel:

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

- Balance: $|S| = |\bar{S}| = p/2$
- Homophily: $\beta > 0$
- Assortativity: $\beta > \alpha$

Problem description

Ising blockmodel:

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

- Balance: $|S| = |\bar{S}| = p/2$
- Homophily: $\beta > 0$
- Assortativity: $\beta > \alpha$

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

Problem overview

- Structure of the problem visible in the **covariance matrix** Σ

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

Problem overview

- Structure of the problem visible in the **covariance matrix** Σ

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

Main object of study

Scaling of $\Delta - \Omega$ with p?

Problem overview

- Structure of the problem visible in the **covariance matrix** Σ

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

Main object of study

Scaling of $\Delta - \Omega$ with p?

Is it clear that $\Delta - \Omega$ is positive?

The role of the correlation matrix: Lower bound

- Consider patterns S and T that differ on exactly one pair of nodes
  - Let $\mathbf{P}_S$ and $\mathbf{P}_T$ be the induced distributions on samples

The role of the correlation matrix: Lower bound

- Consider patterns S and T that differ on exactly one pair of nodes
  - Let $\mathbf{P}_S$ and $\mathbf{P}_T$ be the induced distributions on samples

For an algorithm to be able to distinguish between $\mathbf{P}_S$ and $\mathbf{P}_T$, they should be "far" in a statistical sense

The role of the correlation matrix: Lower bound

- Consider patterns S and T that differ on exactly one pair of nodes
  - Let $\mathbf{P}_S$ and $\mathbf{P}_T$ be the induced distributions on samples

Lemma

Relative entropy $D_{KL}(\mathbf{P}_T \| \mathbf{P}_S) = \frac{(\beta-\alpha)(p-2)}{p} \cdot (\Delta - \Omega)$

$$\Downarrow$$

Theorem

Sample complexity $n \gtrsim \dfrac{\log p}{\Delta - \Omega}$

The role of the correlation matrix: Lower bound

- Consider patterns S and T that differ on exactly one pair of nodes
  - Let $\mathbf{P}_S$ and $\mathbf{P}_T$ be the induced distributions on samples

Lemma

Relative entropy $D_{KL}(\mathbf{P}_T \| \mathbf{P}_S) = \frac{(\beta-\alpha)(p-2)}{p} \cdot (\Delta - \Omega)$

$$\Downarrow$$

Theorem

Sample complexity $n \gtrsim \dfrac{\log p}{\Delta - \Omega}$

- Note that the lemma implies that $\Delta - \Omega \geq 0$

The role of the covariance matrix: Upper bound

Empirical covariance matrix

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

The role of the covariance matrix: Upper bound

Empirical covariance matrix

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

Log likelihood

$$\mathcal{L}_{\sigma^{(1)},\sigma^{(2)},\ldots,\sigma^{(n)}}(S) := \log \mathbf{P}_S(\sigma^{(1)},\sigma^{(2)},\ldots,\sigma^{(n)}) = \text{const} + \frac{n(\beta-\alpha)}{2p}\mathbf{Tr}\left[\hat{\Sigma} v_S v_S^{\top}\right]$$

$$v_S := (\mathbb{1}_S - \mathbb{1}_{\overline{S}})$$

The role of the covariance matrix: Upper bound

Empirical covariance matrix

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

Log likelihood

$$\mathcal{L}_{\sigma^{(1)},\sigma^{(2)},\dots,\sigma^{(n)}}(S) := \log \mathbf{P}_S(\sigma^{(1)},\sigma^{(2)},\dots,\sigma^{(n)}) = \text{const} + \frac{n(\beta-\alpha)}{2p}\mathbf{Tr}\left[\hat{\Sigma} v_S v_S^\top\right]$$

$$v_S := (\mathbb{1}_S - \mathbb{1}_{\overline{S}})$$

Maximum likelihood estimator

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

$$\text{where} \quad \mathcal{P} = \{vv^\top : v \in \{-1,1\}^p, v^\top \mathbf{1}_{[p]} = 0\}$$

The role of the covariance matrix: Upper bound

Empirical covariance matrix

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

Log likelihood

$$\mathcal{L}_{\sigma^{(1)},\sigma^{(2)},\ldots,\sigma^{(n)}}(S) := \log \mathbf{P}_S(\sigma^{(1)},\sigma^{(2)},\ldots,\sigma^{(n)}) = \text{const} + \frac{n(\beta-\alpha)}{2p}\mathbf{Tr}\left[\hat{\Sigma} v_S v_S^\top\right]$$

$$v_S := (\mathbb{I}_S - \mathbb{I}_{\overline{S}})$$

Maximum likelihood estimator

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

$$\text{where} \quad \mathcal{P} = \{vv^\top : v \in \{-1,1\}^p, v^\top \mathbf{1}_{[p]} = 0\}$$

MINIMUM-BISECTION: NP-hard

The role of the covariance matrix: Upper bound

SDP relaxation

Maximum likelihood estimator

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

$$\text{where} \quad \mathcal{P} = \{vv^\top : v \in \{-1,1\}^p, v^\top \mathbf{1}_{[p]} = 0\}$$

The role of the covariance matrix: Upper bound

SDP relaxation

Maximum likelihood estimator

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

$$\text{where} \quad \mathcal{P} = \{vv^\top : v \in \{-1,1\}^p, v^\top \mathbf{1}_{[p]} = 0\}$$

- Easier to work with projections on space $\perp$ to $\mathbf{1}$:

$$\Gamma = P\Sigma P \qquad \hat{\Gamma} = P\hat{\Sigma}P,$$

  where

$$P = I - \frac{1}{p}\mathbf{1}\mathbf{1}^\top$$

- $\mathbf{Tr}[\hat{\Gamma}V] = \mathbf{Tr}[\hat{\Sigma}V]$ for $V \in \mathcal{P}$

The role of the covariance matrix: Upper bound

SDP relaxation

Maximum likelihood estimator

$$\hat{V} = \underset{V \in \mathcal{P}}{\operatorname{argmax}}\, \mathbf{Tr}[\hat{\Gamma} V]\,,$$

where $\mathcal{P} = \{vv^\top : v \in \{-1,1\}^p, v^\top \mathbf{1}_{[p]} = 0\}$

- Easier to work with projections on space $\perp$ to $\mathbf{1}$:

$$\Gamma = P\Sigma P \qquad \hat{\Gamma} = P\hat{\Sigma}P,$$

where

$$P = I - \frac{1}{p}\mathbf{1}\mathbf{1}^\top$$

- $\mathbf{Tr}[\hat{\Gamma}V] = \mathbf{Tr}[\hat{\Sigma}V]$ for $V \in \mathcal{P}$

The role of the covariance matrix: Upper bound

SDP relaxation

Maximum likelihood estimator

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

where $\quad \mathcal{P} = \{vv^\top : v \in \{-1,1\}^p, v^\top \mathbf{1}_{[p]} = 0\}$

SDP relaxation

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

where $\quad \mathcal{E} = \{V : \mathsf{diag}(V) = \mathbf{1}, V \succeq 0\}$

The role of the covariance matrix: Upper bound

SDP relaxation

Maximum likelihood estimator

$$\hat{V} = \underset{V \in \mathcal{P}}{\operatorname{argmax}} \, \mathbf{Tr}[\hat{\Gamma} V] \,,$$

$$\text{where} \quad \mathcal{P} = \{vv^\top : v \in \{-1,1\}^p, v^\top \mathbf{1}_{[p]} = 0\}$$

SDP relaxation

$$\hat{V} = \underset{V \in \mathcal{E}}{\operatorname{argmax}} \, \mathbf{Tr}[\hat{\Gamma} V] \,,$$

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

Lemma: Condition for zero integrality gap (from SDP duality)

$$\hat{V} = v_S v_S^\top \iff L_S(\hat{\Gamma}) := \mathbf{diag}(\hat{\Gamma} v_S v_S^\top) - \hat{\Gamma} \succeq 0$$

The role of the covariance matrix: Upper bound

SDP relaxation

Maximum likelihood estimator

$$\hat{V} = \underset{V \in \mathcal{P}}{\text{argmax}}\, \mathbf{Tr}[\hat{\Gamma} V]\,,$$

$$\text{where} \quad \mathcal{P} = \{vv^\top : v \in \{-1,1\}^p, v^\top \mathbf{1}_{[p]} = 0\}$$

SDP relaxation

$$\hat{V} = \underset{V \in \mathcal{E}}{\text{argmax}}\, \mathbf{Tr}[\hat{\Gamma} V]\,,$$

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

Lemma: Condition for zero integrality gap (from SDP duality)

$$\hat{V} = v_S v_S^\top \iff L_S(\hat{\Gamma}) := \mathbf{diag}(\hat{\Gamma} v_S v_S^\top) - \hat{\Gamma} \succeq 0$$

Satisfied when $\hat{\Gamma} = \Gamma$ ("infinitely many samples")

The role of the correlation matrix: Upper bound

SDP relaxation: Finitely many samples

Lemma: Condition for zero integrality gap (from SDP duality)

$$\hat{V} = v_S v_S^\top \iff L_S(\hat{\Gamma}) := \mathbf{diag}(\hat{\Gamma} v_S v_S^\top) - \hat{\Gamma} \succeq 0$$

Satisfied when $\hat{\Gamma} = \Gamma$ ("infinitely many samples")

The role of the correlation matrix: Upper bound

SDP relaxation: Finitely many samples

Lemma: Condition for zero integrality gap (from SDP duality)

$$\hat{V} = v_S v_S^\top \iff L_S(\hat{\Gamma}) := \mathbf{diag}(\hat{\Gamma} v_S v_S^\top) - \hat{\Gamma} \succeq 0$$

Satisfied when $\hat{\Gamma} = \Gamma$ ("infinitely many samples")

Spectrum of $L_S(\Gamma)$ ("infinite samples")

$$L_S(\Gamma) = \left(1 - \Delta + p\frac{\Delta - \Omega}{2}\right) \frac{\mathbf{1}_{[p]}}{\sqrt{p}} \frac{\mathbf{1}_{[p]}^\top}{\sqrt{p}} + p\frac{\Delta - \Omega}{2} \cdot \tilde{I}_{\perp(\mathbf{1}, v_S)} + 0 \cdot v_S v_S^\top$$

The role of the correlation matrix: Upper bound

SDP relaxation: Finitely many samples

Lemma: Condition for zero integrality gap (from SDP duality)

$$\hat{V} = v_S v_S^\top \iff L_S(\hat{\Gamma}) := \mathbf{diag}(\hat{\Gamma} v_S v_S^\top) - \hat{\Gamma} \succeq 0$$

Satisfied when $\hat{\Gamma} = \Gamma$ ("infinitely many samples")

Spectrum of $L_S(\Gamma)$ ("infinite samples")

$$L_S(\Gamma) = \left(1 - \Delta + p\frac{\Delta - \Omega}{2}\right) \frac{\mathbf{1}_{[p]}}{\sqrt{p}} \frac{\mathbf{1}_{[p]}^\top}{\sqrt{p}} + p\frac{\Delta - \Omega}{2} \cdot \tilde{I}_{\perp(\mathbf{1}, v_S)} + 0 \cdot v_S v_S^\top$$

Thus, the relaxation is exact in the finite sample case if

$$\|L_S(\Gamma) - L_S(\hat{\Gamma})\|_{\mathsf{op}} \leq p\,\frac{\Delta - \Omega}{2} \implies L_S(\hat{\Gamma}) \succeq 0 \implies \hat{V} = V_S$$

The role of the correlation matrix: Tight bounds

Upper and lower bounds

- The SDP relaxation is exact if $\|L_S(\Gamma) - L_S(\hat{\Gamma})\|_{\text{op}} \leq p(\Delta - \Sigma)/2$.
- From matrix concentration, this holds if

$$n \gtrsim \frac{\log p}{\Delta - \Omega}$$

- Previously, we saw $n \gtrsim \frac{\log p}{\Delta - \Omega}$ is required. Combining, we have:

The role of the correlation matrix: Tight bounds

Upper and lower bounds

- The SDP relaxation is exact if $\|L_S(\Gamma) - L_S(\hat{\Gamma})\|_{\text{op}} \leq p(\Delta - \Sigma)/2$.
- From matrix concentration, this holds if

$$n \gtrsim \frac{\log p}{\Delta - \Omega}$$

- Previously, we saw $n \gtrsim \frac{\log p}{\Delta - \Omega}$ is required. Combining, we have:

Theorem: Sample complexity and the correlation matrix

The sample complexity for exactly recovering the partition $(S, \overline{S})$ in the Ising blockmodel satisfies

$$n \simeq \frac{\log p}{\Delta - \Omega}$$

The role of the correlation matrix: Tight bounds

Upper and lower bounds

- The SDP relaxation is exact if $\|L_S(\Gamma) - L_S(\hat{\Gamma})\|_{\text{op}} \leq p(\Delta - \Sigma)/2$.
- From matrix concentration, this holds if

$$n \gtrsim \frac{\log p}{\Delta - \Omega}$$

- Previously, we saw $n \gtrsim \frac{\log p}{\Delta - \Omega}$ is required. Combining, we have:

Theorem: Sample complexity and the correlation matrix

The sample complexity for exactly recovering the partition $(S, \overline{S})$ in the Ising blockmodel satisfies

$$n \simeq \frac{\log p}{\Delta - \Omega}$$

Question

How does $\Delta - \Omega$ behave?

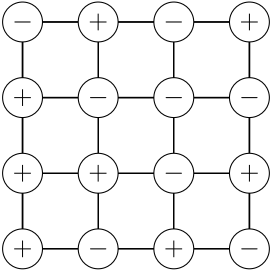
Analyzing $\Delta - \Omega$
Phase diagram of the Ising blockmodel

Ising blockmodel

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

$$\sigma \in \{-1,1\}^p$$

Ising model



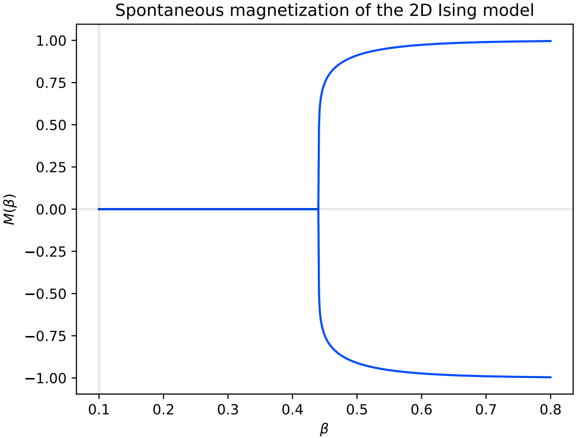
Gibbs distribution

$$\mu(\sigma) \propto \exp\left(\beta \sum_{u \sim v} \sigma_u \sigma_v\right)$$

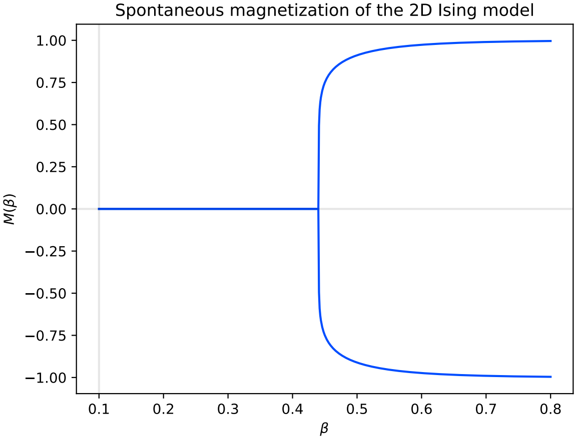
- Nodes "represent" magnetic domains, β represents inverse temperature $1/T$
- Mean magentization: Average of $\sum_v \sigma_v$ according to μ

$$M(\beta) = \underset{\sigma \sim \mu}{E}[\sum_v \sigma_v]$$

Phase transition in the Ising model [Onsager, 1944]



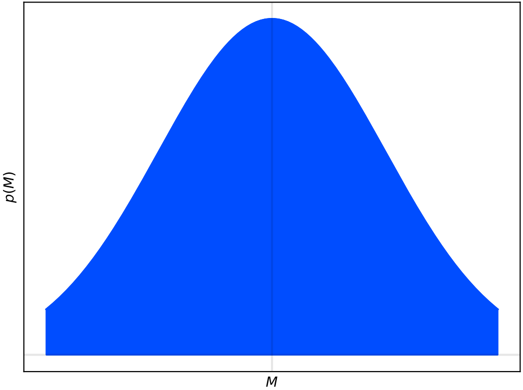
Phase transition in the Ising model [Onsager, 1944]



- Need to take an infinite volume limit to actually see this

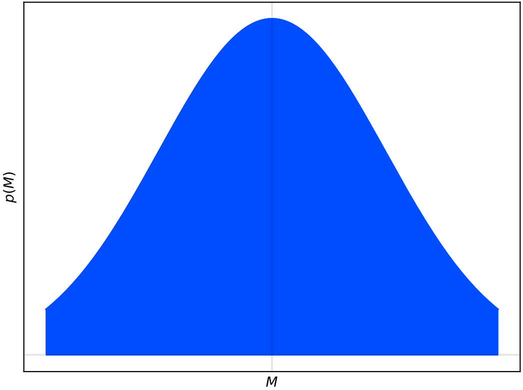
Phase transition in the Ising model: finite volume

- But one can see a “phase transition” in some finite settings as well



Phase transition in the Ising model: finite volume

- But one can see a "phase transition" in some finite settings as well
- As the size of the "lattice" increases, the bimodal nature of μ becomes more pronounced



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

"Mean field" Ising

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

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

"Mean field" Ising

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

- **Free energy**:

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

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

"Mean field" Ising

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

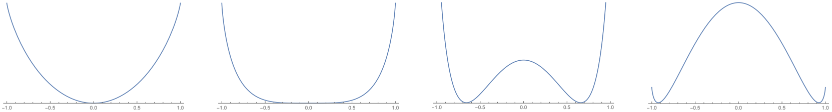
- **Free energy**:

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

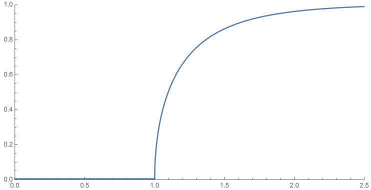
- **Ground states**: Minimizers $G \subset [-1, 1]$ of $g_\beta^{\mathsf{CW}}(\mu)$.

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 is determined by mean magnetizations: $(\mu_S, \mu_{\bar{S}}) = \frac{2}{p}\big(\mathbf{1}_S^\top \sigma, \mathbf{1}_{\bar{S}}^\top \sigma\big)$

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

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

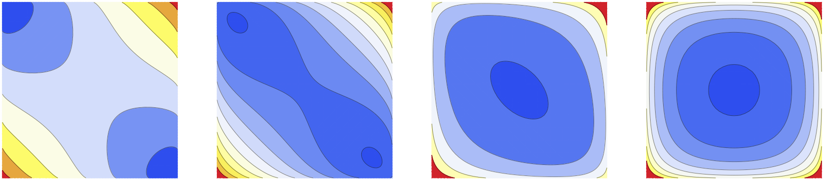
$$\mathbf{P}_S(\mu_S, \mu_{\bar{S}}) \approx \frac{1}{Z_{\alpha,\beta}} \exp\Big(-\frac{p}{8}\, g_{\alpha,\beta}(\mu_S, \mu_{\bar{S}})\Big)$$

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

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

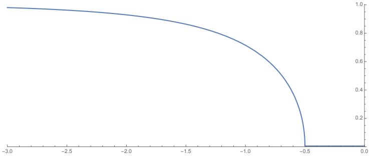
Ground states for the Ising blockmodel

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



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

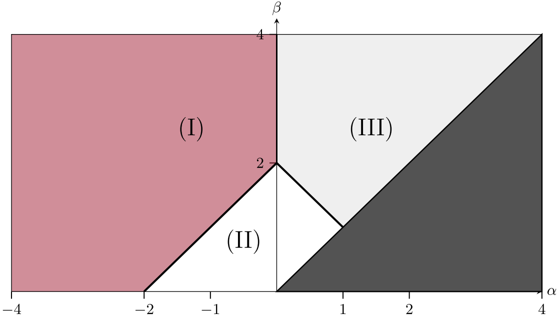
Ground states on the skew-diagonal ($\tilde{\mu}_S = -\tilde{\mu}_{\bar{S}}$) for $\alpha \leq 0$ and fixed $\beta = 1.5 < 2$



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

Phase diagram

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



Theorem

- *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

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

Theorem: Gaussian approximation of the Gibbs distribution

For "nice" $\varphi : [-1,1]^2 \to \mathbb{R}^+$,

$$\mathbf{E}_{\alpha,\beta}[\varphi(\mu)] \simeq_p \frac{1}{|G|}\sum_{\tilde{s}\in G} \mathbf{E}_{Z\sim\mathcal{N}(0,I)}\left[\varphi(\tilde{s} + 2\sqrt{\frac{2}{p}}H^{-1/2}Z)\right]$$

Concentration

- Quantities of interest as expectations

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

Theorem: Gaussian approximation of the Gibbs distribution

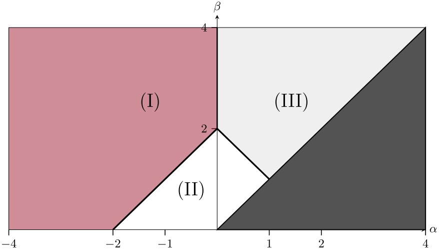
For "nice" $\varphi : [-1,1]^2 \to \mathbb{R}^+$,

$$\mathbf{E}_{\alpha,\beta}[\varphi(\mu)] \simeq_p \frac{1}{|G|} \sum_{\tilde{s} \in G} \mathbf{E}_{Z \sim \mathcal{N}(0,I)}\left[\varphi(\tilde{s} + 2\sqrt{\frac{2}{p}} H^{-1/2} Z)\right]$$

- Applied to $\Delta - \Omega$:

$$\Delta - \Omega \simeq_p \begin{cases} 2\tilde{x}^2 (= \Theta(1)) & \text{in region (I)} \\ \dfrac{C_{\alpha,\beta}}{p} & \text{in region (II)} \\ \dfrac{C'_{\alpha,\beta}}{p} & \text{in region (III)} \end{cases}$$

Exact recovery



Recall that

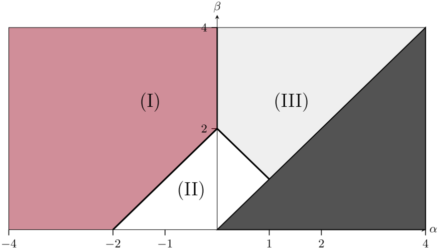
Theorem: Sample complexity and the correlation matrix

The sample complexity for exactly recovering the partition $(S, \overline{S})$ in the Ising blockmodel satisfies

$$n \simeq \frac{\log p}{\Delta - \Omega}$$

Combining the above estimates, we get...

Exact recovery



Main result: Optimal sample size for exact recovery

$$\text{(I)} : n^* \approx \log(p) \qquad \text{(II) and (III)} : n^* \approx p \log(p)\,.$$

Conclusion

- Open questions
  - **Non-independent samples:** Samples drawn from a logit response/Glauber dynamics?
  - **Algorithmic results:** Can the SDP be circumvented?
  - Generalization to multiple blocks, more complex structures

Conclusion

- Open questions
  - **Non-independent samples:** Samples drawn from a logit response/Glauber dynamics?
  - **Algorithmic results:** Can the SDP be circumvented?
  - Generalization to multiple blocks, more complex structures

- Phase transitions and algorithms
  - **Statistical/sample complexity**

Conclusion

- Open questions
  - **Non-independent samples:** Samples drawn from a logit response/Glauber dynamics?
  - **Algorithmic results:** Can the SDP be circumvented?
  - Generalization to multiple blocks, more complex structures

- Phase transitions and algorithms
  - **Statistical/sample complexity**
  - **Algorithmic complexity**

- Different views of phase transitions: **"Correlation decay"** vs. **"Analyticity of free energy"** vs. **"Glauber/heat bath dynamics"**
  - We essentially saw only the first in this talk
  - All three seem to have strong connections with algorithms,
  - ...but in weirdly different ways!

Conclusion

- Open questions
  - **Non-independent samples:** Samples drawn from a logit response/Glauber dynamics?
  - **Algorithmic results:** Can the SDP be circumvented?
  - Generalization to multiple blocks, more complex structures

- Phase transitions and algorithms
  - **Statistical/sample complexity**
  - **Algorithmic complexity**

- Different views of phase transitions: **"Correlation decay"** vs. **"Analyticity of free energy"** vs. **"Glauber/heat bath dynamics"**
  - We essentially saw only the first in this talk
  - All three seem to have strong connections with algorithms,
  - ...but in weirdly different ways!
  - *A story for another time...*

piyush.srivastava@tifr.res.in

Conclusion

- Open questions
  - **Non-independent samples:** Samples drawn from a logit response/Glauber dynamics?
  - **Algorithmic results:** Can the SDP be circumvented?
  - Generalization to multiple blocks, more complex structures

- Phase transitions and algorithms
  - **Statistical/sample complexity**
  - **Algorithmic complexity**

- Different views of phase transitions: **"Correlation decay"** vs. **"Analyticity of free energy"** vs. **"Glauber/heat bath dynamics"**
  - We essentially saw only the first in this talk
  - All three seem to have strong connections with algorithms,
  - ...but in weirdly different ways!
  - *A story for another time...*

piyush.srivastava@tifr.res.in

Conclusion

- Open questions
  - **Non-independent samples:** Samples drawn from a logit response/Glauber dynamics?
  - **Algorithmic results:** Can the SDP be circumvented?
  - Generalization to multiple blocks, more complex structures

- Phase transitions and algorithms
  - **Statistical/sample complexity**
  - **Algorithmic complexity**

- Different views of phase transitions: **"Correlation decay"** vs. **"Analyticity of free energy"** vs. **"Glauber/heat bath dynamics"**
  - We essentially saw only the first in this talk
  - All three seem to have strong connections with algorithms,
  - ...but in weirdly different ways!
  - *A story for another time...*

piyush.srivastava@tifr.res.in

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

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

Graphical models / MRF

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

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

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

Wigner matrices

Wishart matrices

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

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

Abbé, Bandeira, Hall '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 estimate sparse $J = \{J_{ij}\}_{ij}$ (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