SBM $\sum_{a=1}^{q} n_a = 1$ group sizes $\quad s_i = a$ with prob. n_a, $s_i \in \{1, \dots, q\}$

c_{ab} affinity matrix, $p_{ab} = \frac{c_{ab}}{N}$ sparse graphs

What is optimal? Depends on the dataset. Real datasets were not generated by SBM ... ~~so real problems~~ so worst case guarantees ~~for~~ would be needed. → most of statistics? Typical × worst.

~~Making our life ... In order~~ First step towards theoretical understanding: Generate the data from SBM with known n_a, c_{ab}, q!

$\boxed{\text{What is optimal inference then?}}$

What quantity to optimize? e.g. $\frac{1}{N}\sum_{i=1}^{N} \delta_{s_i^* \hat{s}_i} = Q$

s_i^*: true value; $\hat{s}_i$: estimator computed from $A_{ij}, \vec{n}, c, q$ (θ)

[IT & Algorithmic limitations in high-dimensional setting.]

$\hat{s}_i(A_{ij}, \vec{n}, c, q) = ?$ so that Q maximized?

$Q(s^*, A_{ij})$ so we can only maximize in expectation ~~of~~ over s^*

given A_{ij} & θ what is the probability of ~~...~~ what was s^*?

$$P(s \mid A, \theta) = \frac{P(s,\theta)\, P(A \mid s, \theta)}{P(A \mid \theta)} = \frac{1}{Z} \prod_{i=1}^{N} n_{s_i} \prod_{i<j} \left(1 - \frac{c_{s_i s_j}}{N}\right)^{1-A_{ij}} \left(\frac{c_{s_i s_j}}{N}\right)^{A_{ij}}$$

posterior distribution in Bayesian inference.

$$\max_{\hat{s}(A,\theta)} \left[\sum_{\{s_i\}_{i=1\dots N}} P(s \mid A, \theta) \frac{1}{N} \sum_{j=1}^{N} \delta_{s_j \hat{s}_j(A,\theta)} \right] =$$

$$= \max_{\hat{s}(A,\theta)} \left[\frac{1}{N} \sum_{i=1}^{N} \mu(s_j \mid A, \theta)\, \delta_{s_j, \hat{s}_j(A,\theta)} \right]$$

max term by term to get

$\hat{s}_j(A,\theta) = \operatorname{argmax} \mu(s_j)$

Bayes optimal estimation.

CENTRE DE RECHERCHES MATHÉMATIQUES

Where $\mu(s_i)$ is the marginal distribution of the posterior

$s_i \in \{1,2,\dots q\}$

$s \in \{1,2,\dots,q\}^N$ $A_{ij} \in \mathbb{R}^{N\times N}$

$$Z(A,\theta) = \sum_{\{s_i\}_{i=1\dots N}} \underbrace{\prod_{i=1}^{N} n_{s_i}}_{P(s|\theta)} \underbrace{\prod_{i<j} \left(1-\frac{c_{s_is_j}}{N}\right)^{1-A_{ij}} c_{s_is_j}^{A_{ij}}}_{P(A|s,\theta)}$$

$$\mu(s_i) = \sum_{\substack{\{s_j\}_{j=1\dots N}\\ j\neq i}} P(s|A,\theta)$$ definition of the marginal.

How to compute $\mu(s_i)$?

E.g. no graph known $P(s) = \frac{1}{Z}\prod_{i=1}^{N} n_{s_i}$

$Z=1$ $\mu(s_i) = n_{s_i}$ uninformative marginals!

With the likelihood term computing marginals exactly is hard. #P-hard.

~~But in these lectures I will teach you how~~

Tiago showed that you can do it with MCMC. Good option, but hard to analyze theoretically or prove something about.

This lecture will be about another class of algorithms and an analysis tool that permits deeper understanding of the behavior of the Bayes optimal estimator and its tractability. Methods & concepts rather than algorithms.

Summary of results (justifications in the rest of the lecture).

Focus on the challenging case where the average degree in every group is the same

$$\sum_{a=1}^{q} n_a c_{ab} = c_b = c \ \forall b$$

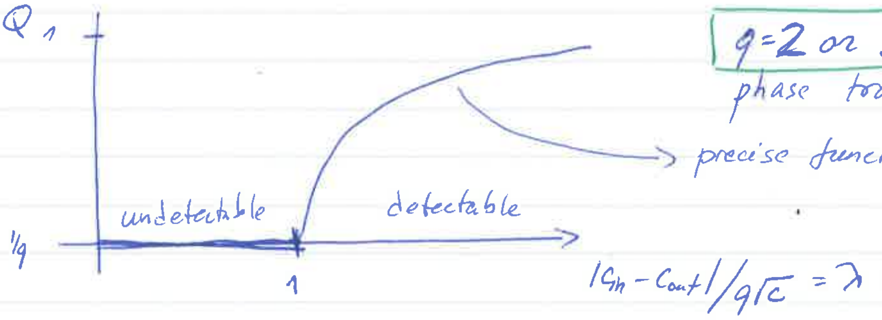
$\hookrightarrow$ average degree

(A) $n_a = \frac{1}{q}$ $\qquad c_{ab} = c_{out}$ if $a \neq b$
$\qquad\qquad\qquad = c_{in}$ if $a = b$

$$c = \frac{c_{in} + (q-1)c_{out}}{q}$$

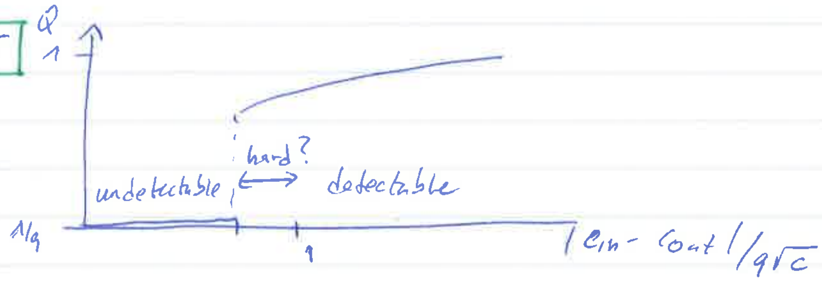
$$|c_{in} - c_{out}| = q\sqrt{c}$$

$\boxed{q = 2 \text{ or } 3}$ phase transition



precise function depends on q, c_{in}, c_{out}

$\boxed{q \geq 5}$



$q = 4$ depends on c_{in} & c_{out}
(large c → continuous
small c & $c_{in} > c_{out}$
discontinuous)

Comments: ① $q = 2$ established rigorously Mossel, Sly, Neeman
Bordenave, Lelarge, Massoulie...
$q = 3$ undetectability at $\lambda = 1$ still open

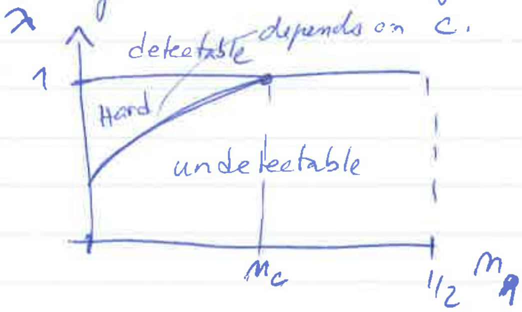
② contiguity between SBM & random ER graph
in the undetectable phase.
High prob. properties hold in both.
SBM graphs are ~~bas~~ for ~~all~~ practical purposes
undistinguishable from ER!

③ Abbe + collaborators, C. Moore & collaborators
gaps from 1 proven for $q \geq 4$ (not tight).

④ All known poly algorithms work only for $\lambda \geq 1$.
E.g. non-backtracking spectral (to come)
Belief propagation (to come)

⑤ Conjecture BP matches Bayes optimal out
of the hard region. (open in sparse, Theorem in dense)

(B) Two groups, one smaller n_1 and denser (to compensate for the degree). Average degree $c = n_0 c_{00} + n_1 c_{01} = n_1 c_{11} + n_0 c_{10}$



$$\lambda = \frac{n_1 (c_{11} - c_{10})}{\sqrt{c}}$$

$$n_c = \frac{1}{2} - \frac{1}{2\sqrt{3}}$$

Hard phase / discontinuous phase tr. exist even for $q = 2$.

Comments ① small n_1 related to planted clique problem. → the best known inference problem where IT cliques $> O(\log N)$ detectable, but algorithms fail for $O(\sqrt{N})$.

② To define un-detectability in the asymmetric case $\mu(S_i) = n_{S_i}$
(our definition of overlap not best)

Where do these results come from: Heuristic statistical physics + series of later proofs (still many open parts).

Why would people in physics study SBM? Long before interest in data / networks...
SBM = kind of a $\boxed{\text{mean-field spin glass}}$

Spin glass = gold with few % of iron
↳ not magnetic ↳ magnetic
Intriguing magnetic properties. Understand & predict

To introduce a spin glass to ~~math~~ non-physicist take the (planted) spin glass game.

Dense spin glass game.

N people, deal ± 1 cards to everyone (roughly the same # of $+1$ as -1). S_i^* card of person i.

Ask each pair (ij) of people a task to generate $W_{ij} \sim N(0,1)$, return

$$W_{ij} + \frac{\beta^*}{\sqrt{N}} S_i^* S_j^* = J_{ij}$$

Goal of the game: S^* is hidden, estimate it from J_{ij} (β known).

Bayes optimal inference (again) $\qquad S_i = \pm 1 \qquad \beta^* = \beta$

$$P(S|J,\mu) = \frac{1}{Z} \prod_{i<j} e^{-\frac{1}{2}\left(J_{ij} - \frac{\beta}{\sqrt{N}} S_i S_j\right)^2} =$$

$$= \frac{1}{Z} \prod_{i<j} e^{+\beta \frac{J_{ij}}{\sqrt{N}} S_i S_j} = \frac{1}{Z} e^{-\beta H}$$

$$H = - \sum_{i<j} \frac{J_{ij}}{\sqrt{N}} S_i S_j \qquad \text{spin glass Hamiltonian}$$

Z — partition function

$P(S|J,\mu)$ — Boltzmann measure

etc. β — inverse temperature

Substitution $\qquad \tilde{S}_i = S_i^* \tilde{S}_i$

$$H = - \sum_{i<j} \underbrace{\left(\frac{W_{ij}}{\sqrt{N}} S_i^* S_j^* + \frac{\beta^*}{N}\right)}_{} \tilde{S}_i \tilde{S}_j$$

J_{ij} random with mean $\frac{\beta^*}{N}$, variance $\frac{1}{N}$

[Bayes optimal inference = spin glass at a special value of temperature ($\beta = \beta^*$) related to bias in J_{ij} (Nishimori temperature). $\qquad \beta_c = 1$

Read the results off phase diagrams from 20-40 years old physics papers.

End of lecture 1. (1.5h)

CENTRE DE RECHERCHES MATHÉMATIQUES

Spin glass game and SBM ⑥ not so similar?

(symmetric)

Low rank matrix factorization:

A class of problems of statistical inference that includes SBM & planted spin glass.

$$P(S\mid Y) = \frac{1}{Z(Y)} \underbrace{\prod P_0(S_i)}_{\text{nodes}} \underbrace{\prod_{i<j}}_{\text{edges}} P_{out}(Y_{ij} \mid S_i K S_j)$$

(symmetric matrix)

$Y \in R^{N\times N}$ symmetric — asymmetric (graphons, tensors, hypergraphs...)

$S_i \in R^q$ $\quad i = 1 \dots N$

$K \in R^{q\times q}$ known, P_0, P_{out} known functions.

Examples:

SBM : $S_i = (0,1,0,\dots 0,0) = \vec{e}_a$ $\quad S_i(r) = 1$ if i belongs to group r, $=0$ otherwise

$K_{ab} = c_{ab}$

$P_0(S_i) = \sum_{a=1}^{q} n_a \vec{e}_a$

$P_{out}(Y=1\mid w) = \frac{w}{N}$
$\phantom{P_{out}(Y=}0 \quad = 1-\frac{w}{N}$ $\Big\}$ $\boxed{\text{sparse}}$ SBM

Planted spin glass: $\boxed{\text{dense}}$ $\quad P_0(S) = \frac{1}{2}\left(\delta(S+1)+\delta(S-1)\right)$

$P_{out}(Y\mid w) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(Y - w/\sqrt{N}\right)^2}$ $\quad \forall i<j$

Censored SBM / sparse and discrete J_{ij} $\quad (ij)\in E$ of ER graph

$P(J_{ij} = S_i^* S_j^*) = \rho$ $\qquad \rho = 1/2$ standard diluted spin glass

$P(J_{ij} = -S_i^* S_j^*) = 1-\rho$

Bethe-Peierls approximation or Belief Propagation (BP)

How to compute Z & $\mu(S_i)$ on a tree graph?

Derive BP for a generic P_{out} and (so far) discrete $P_0(S_i)$

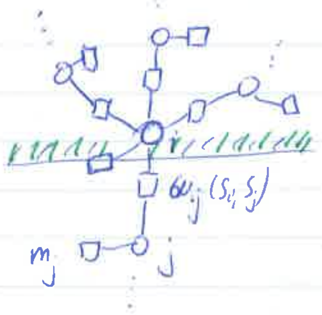
Also the weighted or layered SBM.

or latent space model

BP on a tree graph $G = (V, E)$ $\qquad s_i \in \mathcal{X}$ discrete set (for now)

$$P(s) = \frac{1}{Z} \prod_i n_i(s_i) \prod_{(ij)\in E} \omega_{ij}(s_i, s_j)$$

$$Z = \sum_{\{s_i\}_{i=1\ldots N}} \prod n_i(s_i) \prod_{(ij)\in E} \omega_{ij}(s_i, s_j)$$



Define $$Z^{i\to j}(s_i) \equiv \sum_{\{s_j\}_{\ell \text{ above } i}} \prod_{\ell \text{ above } j} n_\ell(s_\ell) \prod_{(\ell\ell')\in E_{\text{above } i}} \omega_{\ell\ell'}(s_\ell, s_{\ell'})$$

$$= n_i(s_i) \prod_{k\in\partial i\setminus j} \sum_{s_k} \omega_{ik}(s_i, s_k)\, Z^{k\to i}(s_k)$$

Conditionally on s_i the branches for different k are not connected $\Rightarrow$ independent.

Define messages $$\tilde{m}^{i\to j}(s_i) = \frac{Z^{i\to j}(s_i)}{\sum_s Z^{i\to j}(s)}$$

$$\tilde{m}^{i\to j}(s_i) = \frac{n_i(s_i) \prod_{k\in\partial i\setminus j} \sum_{s_k} \omega_{ik}(s_i, s_k) \frac{Z^{k\to i}(s_k)}{\sum_s Z^{k\to i}(s)}}{\sum_s n_i(s) \prod_{k\in\partial i\setminus j} \sum_{s_k} \omega_{ik}(s, s_k) \frac{Z^{k\to i}(s_k)}{\sum_r Z^{k\to i}(r)}} =$$

$$\tilde{m}^{i\to j}(s_i) = \frac{1}{\tilde{Z}^{i\to j}} n_i(s_i) \prod_{k\in\partial i\setminus j} \sum_{s_k} \omega_{ik}(s_i, s_k)\, \tilde{m}^{k\to i}(s_k)$$

where $$\tilde{Z}^{i\to j} = \sum_s n_i(s) \prod_{k\in\partial i\setminus j} \sum_{s_k} \omega_{ik}(s, s_k)\, \tilde{m}^{k\to i}(s_k)$$

BP on a tree

Marginals $$\mu_{BP}(s_i) = \frac{1}{\tilde{Z}^i} n_i(s_i) \prod_{k\in\partial i} \sum_{s_k} \omega_{ik}(s_i, s_k)\, \tilde{m}^{k\to i}(s_k)$$

Bethe free energy (exercise to derive this on a tree)
(exact on a tree)
$$\boxed{f = -\log Z = -\sum_i \log \tilde{Z}^i + \sum_{(ij)\in E} \log \sum_{s_i, s_j \in \mathcal{X}} \omega_{ij}(s_i, s_j)\, m^{i\to j}(s_i)\, m^{j\to i}(s_j)}$$

8

No magic so far (on trees).

BP is an iterative algorithm that often works well also on loopy graphs.

Specifically : [Conjecture : Out of the hard region BP matches the Bayes optimal estimator in the limit $N \to \infty$.

What is the overlap / error ? Sparse graphs not easy to analyze. Dense are simpler (but still highly non-trivial).

$\boxed{\text{Dense matrix factorization}}$ setting :

$$P(s \mid Y) = \frac{1}{Z} \prod_{i=1}^{N} P_0(s_i) \prod_{i<j} P_{out}\left(Y_{ij} \mid \frac{1}{\sqrt{N}} S_i S_j^T\right)$$

$s_i \in \mathbb{R}^q \qquad N \to \infty$

P_0, P_{out} do not depend on N.

Midane
P_0 finite support at least $\sigma^2 < \infty$
$P_{out}(Y \mid w)$ 3 times diff with bounded derivs.

The problem : generate S^* from $\prod_{i=1}^{N} P_0(s_i)$ then generate Y_{ij} from $P_{out}\left(Y_{ij} \mid \frac{1}{\sqrt{N}} S_i S_j^T\right)$, then estimate S^* from Y ?

SAY EXAMPLES HERE !

In this case results rather explicit. Statement of 3 theorems.

① Universality the MMSE in the large N limit depends of P_{out} only through the Fisher Information

$$\frac{1}{\Delta} = \mathbb{E}_{P_{out}(Y, w=0)}\left[\left(\frac{\partial \log P_{out}(Y \mid w)}{\partial w}\Big|_{w=0}\right)^2\right]$$

Examples Ⓐ $Y_{ij} = W_{ij} + \frac{\beta}{\sqrt{N}} s_i s_j$, $W_{ij} \sim N(0,1)$ $\qquad P_{out}(Y \mid w) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(Y - \beta w)^2}$

$$\frac{\partial \log P_{out}}{\partial w}\Big|_{w=0} = \beta Y \qquad \frac{1}{\Delta} = \beta^2$$

Ⓑ Symmetric SBM with q groups : $\qquad s_i \in \mathbb{R}^q$ $(0,1)$ or $(1,0)$.

$$P(Y_{ij} = 1) = p + \frac{\tilde{\mu}}{\sqrt{N}} s_i s_j \qquad p = O(1)$$

$$0 = 1 - p - \frac{\tilde{\mu}}{\sqrt{N}} s_i s_j$$

$$|p_{in} - p_{out}| = \frac{|\tilde{\mu}|}{\sqrt{N}}$$

→ Highlight proof of universality.

CENTRE DE RECHERCHES MATHÉMATIQUES
50 1968-1969 2018-2019

$$\frac{\partial \log(p+\tilde{\mu}w)}{\partial w}\Big|_{w=0} = \frac{\tilde{\mu}}{p} \qquad Y=1$$

$$\frac{\partial}{\partial w}\log(1-p-\tilde{\mu}w)\Big|_{w=0} = \frac{-\tilde{\mu}}{1-p} \qquad Y=0$$

$$\frac{1}{\Delta} = p\,\frac{\tilde{\mu}^2}{p^2} + (1-p)\frac{\tilde{\mu}^2}{(1-p)^2} = \tilde{\mu}^2\,\frac{1}{p(1-p)}$$

3 Spin-glass game
Censored SBM, only fraction of p of answers seen and for $S_i = \pm 1$ seen ones:

$Y_{ij} = +1$ with probability $\frac{1}{2} + \frac{\mu'}{\sqrt{N}} S_i S_j$

$\quad\;\; = -1$ —"— $\frac{1}{2} - \frac{\mu'}{\sqrt{N}} S_i S_j$

$Y_{ij} = 0$ with prob. $1-p$

$Y=+1 \qquad \frac{\partial}{\partial w}\log p(\frac{1}{2}+\mu' w)\Big|_{w=0} = \frac{2p\mu'}{p}$

$\quad -1 \qquad \frac{\partial \log p(\frac{1}{2}-\mu' w)}{\partial w}\Big|_{w=0} = -2\mu'$

$\quad 0 \qquad \frac{\partial \log p}{\partial w} = 0$

$$\frac{1}{\Delta} = \frac{p}{2}(2\mu')^2 + \frac{p}{2}(2\mu')^2 = 2p(\mu')^2$$

2 Expression for the free-energy and the MMSE $(q=1)$

MSE $\quad \frac{1}{N}\sum_{i=1}^{N}(S_i^* - \hat{S}_i)^2$

MMSE computed with $\hat{S}_i$ being mean of the marginal

$$\min_{\hat{S}} \int \prod_i dS_i\, P(S|Y)\, \frac{1}{N}\sum_{i=1}^{N}(S_i-\hat{S}_i)^2 \quad \to \text{minimized by marginal mean.}$$

(in physics we minimize energy)

$f = -\lim_{N\to\infty} \frac{1}{N}\,\mathbb{E}_{x^*,w} \log Z_N(Y)$ we have

Thanks to universality consider

$$Y_{ij} = \underbrace{w_{ij}}_{\mathcal{N}(0,1)} + \sqrt{\frac{\lambda}{N}}\, S_i^* S_j^*$$

$$\boxed{\Delta = \frac{1}{\lambda}}$$

$$f = \min_m \left[\frac{\lambda m^2}{4} - \mathbb{E}_{\substack{x^* \sim P_0(x^*),\; z\sim\mathcal{N}(0,1)}} \log \int dx\, P_0(x)\, e^{-\frac{m\lambda x^2}{2} + m\lambda x x^* + \sqrt{\lambda_m}\, x z}\right]$$

MMSE $= \arg\min f(m) + \mathbb{E}_{x^* \sim P_0}(x^{*2})$

www.crm.math.ca

$$Y_{ij} = G_{ij} + \frac{\beta}{\sqrt{N}} S_i S_j \qquad \lambda = \beta^2$$

Free energy $\tilde{F}_N(Y) = -\frac{1}{N}\log \tilde{Z}(Y)$ finite N, Y dependent

$$\tilde{Z}(Y) = \int \prod_{i=1}^{N} dS_i \underbrace{\qquad\qquad\qquad\qquad}$$

$$P(S|Y) = \frac{1}{\tilde{Z}(Y)} \prod_{i=1}^{N} P_0(S_i) \prod_{i<j} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(Y_{ij} - \sqrt{\frac{\lambda}{N}} S_i S_j\right)^2}$$

S-independent constanants can be "hidden" in Z:

$$P(S|Y) = \frac{1}{Z(Y)} \prod_{i=1}^{N} P_0(S_i) \prod_{i<j} e^{-\frac{1}{2}\frac{\lambda}{N} S_i^2 S_j^2 + Y_{ij}\sqrt{\frac{\lambda}{N}} S_i S_j}$$

$$Z_N(Y) = \int \prod_{i=1}^{N} dS_i \prod_{i=1}^{N} P_0(S_i) \prod_{i<j} e^{-\frac{\lambda}{2N} S_i^2 S_j^2 + Y_{ij}\sqrt{\frac{\lambda}{N}} S_i S_j}$$

Theorem: $\boxed{\text{rank } 1}$ $K=1$

$$f = -\lim_{N\to\infty} \mathbb{E}_{x^*, W} \frac{1}{N} \log Z_N(Y) \quad \text{is w.h.p.} \quad +(\text{self-averaging})$$

$$f = \min_m \left[\frac{\lambda m^2}{4} - \mathbb{E}_{x^*, z} \log \int dx\, P_0(x)\, e^{-\frac{\lambda m}{2} x^2 + \sqrt{\lambda m}\, x (\sqrt{\lambda m}\, x^* + z)} \right] = f_{RS}(m)$$

$x^* \sim P_0$, $z \sim N(0,1)$

Colorary: MMSE $= E_{x^*}(x^{*2}) - \arg\min f_{RS}(m)$

What just happened?

Original problem N-dimensional, depending on $N \times N$ matrix Y.
The expression in the theorem is scalar ! $(q=1)$

⇝ Originally obtained using non-rigorous replica formula. Proof (Barbier et al '16) uses the "knowledge" of the formula to prove it. Notice that the scalar formula is analogous to free energy in scalar denoising problem:

End of lecture 2. (1.5h)

SCALAR DENOISING

Generate $x^* \sim P_0(x^*)$, observe $y = \sqrt{\lambda}\, x^* + z$ all $\in \mathbb{R}$.

$$P(x|y) = \frac{1}{Z(y)} P_0(x) \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y - \sqrt{\lambda} x)^2} = \frac{1}{Z(y)} P_0(x)\, e^{-\frac{\lambda}{2} x^2 + \sqrt{\lambda}\, x (\sqrt{\lambda}\, x^* + z)}$$

$$E_{x^*, z} \log Z(y) = E_{x^*, z} \log \int dx\, P_0(x)\, e^{-\frac{\lambda}{2} x^2 + \sqrt{\lambda}\, x (\sqrt{\lambda}\, x^* + z)}$$

→ $\lambda_{eff} = \lambda m$

→ $\lambda m^2/4$ and minimization over m is missing.

Proof by Guerra interpolation :

Auxiliary inference problem at "time" t : $s_i^* \sim P_0(s^*)$

$$Y_{ij} = \sqrt{\frac{\lambda t}{N}}\, s_i^* s_j^* + w_{ij} \sim \mathcal{N}(0,1)$$

$$\tilde{y}_i = \sqrt{\lambda m (1-t)}\, s_i^* + \xi_i \sim \mathcal{N}(0,1)$$

$$P(S \mid Y, \tilde{y}) = \frac{1}{\tilde{Z}(Y,\tilde{y})} \prod_{i=1}^N P_0(s_i) \prod_{i=1}^N \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(\tilde{y}_i - s_i\sqrt{\lambda m(1-t)})^2} \cdot \prod_{i<j} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(Y_{ij} - \sqrt{\frac{\lambda t}{N}} s_i s_j)^2}$$

$\boxed{t=0}$ N decoupled scalar problems ✓

$\boxed{t=1}$ original N-dimensional problem

$$= \frac{1}{Z_t(Y,\tilde{y})} \prod_{i=1}^N P_0(s_i) \prod_{i=1}^N e^{\tilde{y}_i s_i \sqrt{\lambda m(1-t)} - \frac{1}{2} s_i^2 \lambda m(1-t)} \cdot \prod_{i<j} e^{-\frac{\lambda t}{N} s_i^2 s_j^2 + Y_{ij}\sqrt{\frac{\lambda t}{N}} s_i s_j}$$

$$f_t = -\lim_{N\to\infty} \mathbb{E}_{s^*,w,\xi} \frac{1}{N} \log Z_t(Y,\tilde{y})$$

$f_{t=1}$ = looking for

$$f_{t=0} = - \mathbb{E}_{x^*,z} \log \int dx\, P_0(x)\, e^{-\frac{1}{2}x^2\lambda m + x\sqrt{\lambda m}(x^*\sqrt{\lambda m} + z)}$$

$$f_{t=1} = f_{t=0} + \int_0^1 dt \frac{\partial f(t)}{\partial t}$$

skipping long yet straigtforward computation

$$= \underbrace{f_{t=0} + \frac{\lambda m^2}{2}}_{f_{RS}(m)} + \int_0^1 dt \left[\frac{\lambda}{4} \mathbb{E}_{s^*,w,\xi}\left(\left\langle \left(\frac{\vec{s}^{(1)}\cdot\vec{s}^{(2)}}{N} - m \right)^2 \right\rangle \right) - \frac{\lambda}{2} \mathbb{E}_{s^*,w,\xi} \left\langle \left(\frac{\vec{s}\cdot\vec{s}^*}{N} - m \right)^2 \right\rangle \right]$$

↳ What we need.

average w.r.t. the Boltzmann measure at t

→ quenched free energy ✓. ~~How does~~ self-averaging follows ?.

Nishimori properties of Bayes optimal inference

In word: Random sample from posterior has all the properties of the x^*. Under averages over posterior the ~~tho~~ two can be ~~replaced~~ exchanged.

~~Tho~~ Statetement of Nishimori: $\mathbb{E} f(x^a, x') = \mathbb{E} f(x, x^*)$

$x^* \sim P_0(X)$, then y generated from $P(y|x^*)$

$$\mathbb{E}\left[f(x,x^*)\right] = \overbrace{\int P_0(x^*) \int P(y|x^*)}^{E_{x^*, w_i \xi}} \int f(x,x^*) \overbrace{P(x|y)}^{\langle \cdot \rangle_{\text{Boltzman}}} \, dx\, dy\, dx^* =$$

$$= \underbrace{\int P(y)}_{E^n_{w_i \xi}} \iint f(x^a, x^*) \underbrace{P(x^a|y)\, P(x^*|y)}_{\langle\!\langle \ \rangle\!\rangle} \, dx^a dx^* dy =$$

$P(y) = \int dx^* \, p(y|x^*) p(x^*)$

$$= \mathbb{E}_{x^*, w_i \xi}\left[f(x^a, x')\right] \quad + 2 \text{ Boltzman}$$

So:

$$f_{t=1} = \underbrace{f_{t=0} + \frac{\lambda m^2}{2}}_{f_{RS}(m)} + \int_0^1 dt \, \frac{\lambda}{4} \, \mathbb{E}\left(\left\langle\left(\frac{S \cdot S^*}{N} - m\right)^2\right\rangle\right)$$

$\geq 0 \ \forall m$ (expectation of a square) $\forall m$

$$\boxed{f_{t=1} \leq \min_m f_{RS}(m)}$$

upper bound.

Lower bound that matches the upper bound: slightly different Guerra interpolation (no need for Nishimori hold also out of Nishimori ... but not always tight ?) RSB....

Non-analyticities in $f(\lambda)$ are phase transitions.

$$\text{MMSE} = \mathbb{E}_x(x^*) - \operatorname{argmin}_m f_{RS}(m)$$

Is the MMSE algor computationally achiable?

Key question.

3 BP in this dense setting (and for real-valued vars.)

BP on sparse graphs. Real-valued $\sum \to \int dx$.

$$m^{i\to j}(s_j) = \frac{1}{N^{*i\to j}} \int ds_i \, \omega_{ij}(s_i s_j) \, \tilde{m}^{i\to j}(s_i)$$

$$\tilde{m}^{i\to j}(s_i) = \frac{1}{\tilde{N}^{i\to j}} m_i(s_i) \prod_{k\in\partial i\setminus j} m^{k\to i}(s_i)$$

$\tilde{N}^{i\to j} = N^{*i\to j} \cdot \tilde{N}^{i\to j}$

↳ $s_{j|i} \to$ indep. constants.

Dense regime $\to \prod_{k\in\partial i\setminus j}$ product over many independent (assumed) probabilities.

$\rightarrow \omega_{ij}(s_i s_j) = P_{out}\left(Y_{ij} \,\middle|\, \frac{1}{\sqrt{N}} s_i s_j\right)$ $\quad w = \frac{s_i s_j}{\sqrt{N}}$

↳ depends only weakly on s_i, s_j!

Expand:

$$\omega_{ij}(s_i s_j) = P_{out}(Y_{ij}|0) \cdot \left[1 + S_{ij} \frac{s_i s_j}{\sqrt{N}} + R_{ij} \frac{s_i^2 s_j^2}{N} + O\left(\frac{1}{N^{3/2}}\right)\right]$$

where

$$S_{ij} \equiv \frac{\partial}{\partial w} \log P_{out}(Y_{ij}|w)\Big|_{w=0} \qquad \text{Fisher score matrix}$$

$$R_{ij} \equiv S_{ij}^2 + \frac{\partial^2}{\partial g^2} \log P_{out}(Y_{ij}|w)\Big|_{w=0}$$

Fisher information $\frac{1}{\Delta} = \langle S_{ij}^2 \rangle = \frac{1}{N^2}\sum_{ij} S_{ij}^2$

Examples: Gaus: $P_{out}(Y|w) = \frac{1}{\sqrt{2\pi\Delta}} e^{-\frac{(Y-w)^2}{2\Delta}}$ $\qquad S_{ij} = \frac{Y_{ij}}{\Delta} \qquad R_{ij} = \frac{Y_{ij}^2}{\Delta^2} - \frac{1}{\Delta}$

from page 9:

SBM: $\sqrt{N}|p_{in} - p_{out}| = \mu$

$S_{ij}(Y_{ij}=1) = \frac{\tilde{\mu}}{p} \qquad S_{ij}(Y_{ij}=0) = -\frac{\tilde{\mu}}{1-p} \qquad R_{ij} = 0$

Censored SBM:

$S_{ij}(Y_{ij}=1) = 2\mu' \qquad R_{ij} = 0$

$S_{ij}(Y_{ij}=-1) = -2\mu'$

$S_{ij}(Y_{ij}=0) = 0$

Exponential $\quad P_{out}(Y|w) = \frac{1}{2} e^{-|Y-w|} \qquad S_{ij} = \mathrm{sgn}(Y_{ij})$

Rewrite BP while keeping only leading term in N, one defines

$$\hat{s}_{k\to i} \equiv \int ds_k\ \tilde{m}^{k\to i}(s_k)\ s_k$$

$$\sigma_{k\to i} \equiv \int ds_k\ \tilde{m}^{k\to i}(s_k)\ s_k^2 - (\hat{s}_{k\to i})^2$$

BP becomes (relaxed - BP)

$$\begin{cases} B_{i\to j} = \frac{1}{\sqrt{N}} \sum_{k\neq j} S_{ki}\, \hat{s}_{k\to i} \\ A_{i\to j} = \frac{1}{N\Delta} \sum_{k\neq j} (\hat{s}_{k\to i})^2 \\ \hat{s}_{i\to j} = f_a(A_{i\to j}, B_{i\to j}) \\ \sigma_{i\to j} = \partial_B f_a(A_{i\to j}, B_{i\to j}) \end{cases}$$

(Bayes optimal, if not additional term)

where

$f_a(A,B)$ is mean of $P(x;A,B) = \frac{1}{Z(A,B)} P_0(x)\, e^{Bx - \frac{x^2 A}{2}}$

Does not look so simple, but we are sending 2 scalars, not probability distributions.

Still N^2 messages. Reduce to N.

Approximate message passing
AMP

$$\begin{cases} B_i^t = \frac{1}{\sqrt{N}} \sum_{k=1}^{N} S_{ki}\, \hat{s}_k^t - \hat{s}_i^{t-1} \frac{1}{\Delta N} \sum_{k=1}^{N} \sigma_k \\ A_i^t = \frac{1}{\Delta N} \sum_{k=1}^{N} (\hat{s}_k^t)^2 \\ \hat{s}_i^{t+1} = f_a(A_i^t, B_i^t) \\ \sigma_i^{t+1} = \partial_B f_a(A_i^t, B_i^t) \end{cases}$$

↳ Onsager term (previous time step)

$P_{out} \rightarrow S, \Delta$

$P_0 \rightarrow f_a$

Otherwise solves a huge class of problems.

new algo cool. But how is it related to the free energy, MMSE?

AMP compared to other Bayesian generic-purpose alg's:

variational mean field → as fast, but ~~and~~ less precise
MCMC → slow to be precise
$\boxed{\text{neither is analyzable}}$

State evolution of AMP:

$$m^t \equiv \frac{1}{N} \sum_{i=1}^{N} \hat{S}_i^t S_i^*$$

W.h.p. $$m^{t+1} = E_{x^*, z}\left[f_a\left(\frac{m^t}{\Delta}, \frac{m^t}{\Delta} x^* + \sqrt{\frac{m^t}{\Delta}}\, z \right) \cdot x^* \right]$$

$$f_a\left(\lambda m^t, \lambda m^t x^* + \sqrt{\lambda m^t}\, z \right) \cdot x^*$$

Fixed points of SE ≡ stationary points of the free energy.

MMSE given by global min of $f_{RS}(m)$
AMP-MSE given by local min reached by SE from $m=0$.

Shape of $f_{RS}(m)$ decides wheather MMSE = AMP-MSE or not & position of stationary points

Zoology of phase transitions:
* Zero-mean priors have a fixed point $m=0$ ⇒ undetectable phase exists.
* If more that one fixed point investigate basins of attraction.

⇝ Probably prepare slides to illustrate.

→ Connection to spectral algorithms (Spectrum of S, non-backtrack)
→ Back to sparse systems (no universality, AMP is not ~~bin~~ asymptotically correct, but the picture still holds and can be evaluated numerically).