# Bayesian Estimation

$X \sim P_X$ $\qquad$ $Y \sim P_{Y|X}$

Original Signal $\qquad$ Measurement/Observation

Graphical model: $X \to Y$

Bayes formula:

$$P_{X|Y}(x|y) = \frac{P_{Y|X}(y|x)\, P_X(x)}{P_Y(y)} = \frac{P_{Y|X}(y|x)\, P_X(x)}{\sum_{x'} P_{Y|X}(y|x')\, P_X(x')}$$

One could ask what is

1) $\arg\max_{x} P_{X|Y}(x|y)$ ? MAP estimator

2) $E[X|Y=y]$ ? MMSE estimator

3) $\mathrm{Var}[X|Y=y]$ ?

All these have to do eith with the distribution $P_{X|Y}$ or an optimization related to that.

Thinking about $P_{X|Y}$ relates it to statistical physics.

What are these $x, y$'s like

Example:

1) Corruption by Gaussian noise:

$$X \sim P_X \qquad Y = X + Z$$

$$X \sim \mathcal{N}(0, \Delta I)$$

$$P_{X|Y}(x|y) = \frac{e^{-\frac{\sum_\mu (y_\mu - x_\mu)^2}{2\Delta}} P_X(x)}{\int e^{-\frac{\sum_\mu (y_\mu - x'_\mu)^2}{2\Delta}} P_X(x')\, dx'}$$

$P_X$ could be a nontrivial distribution,

$$X_\mu \overset{iid}{\sim} P_{X_\mu}, \quad P_{X_\mu}(x) = p\,\delta(x) + (1-p)\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}}$$

2) Generalized linear model

$\Xi \sim P_\Xi,\ W \sim P_W,\ Y \sim P_{Y|W,\Xi}$

$$P_W(w) \propto e^{-\sum_i \beta r(w_i)}$$

$$P_{Y|W,\Xi}(y,w) \propto e^{-\beta \sum_\mu \ell(y_\mu, w^T \xi_\mu)}$$

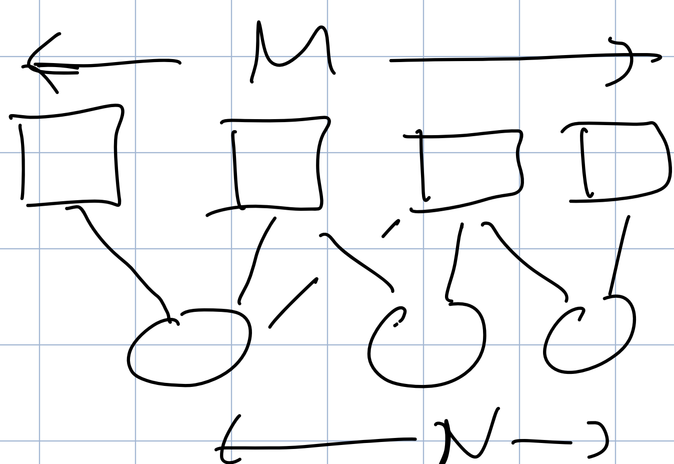
Here X is really w!

$$P_{W|Y,\Xi}(w) = \frac{1}{Z(\{(\xi_\mu, y_\mu)\}, \beta)} \prod_{i=1}^{d} e^{-\beta r(w_i)} \prod_{\mu=1}^{n} e^{-\beta \ell(y_\mu, w^T \xi_\mu)}$$

Many Bayesian Inference problems can be set a graphical models. They can often be represented via a factor graph

Squares

circles




Factor nodes

Variable nodes

Variables have index $i$, factors $a$

$\partial i = \{a \mid (i,a) \in E\}$, $\partial a = \{i \mid (i,a) \in E\}$,

$$P(\{s_i\}_{i=1}^{N}) = \frac{1}{Z_N} \prod_{a=1}^{M} f_a(\{s_i\}_{i \in \partial a})$$

$$Z_N = \sum_{\{s_i\}_{i=1}^N} \prod_{a=1}^{M} f_a(\{s_i\}_{i\in\partial a})$$

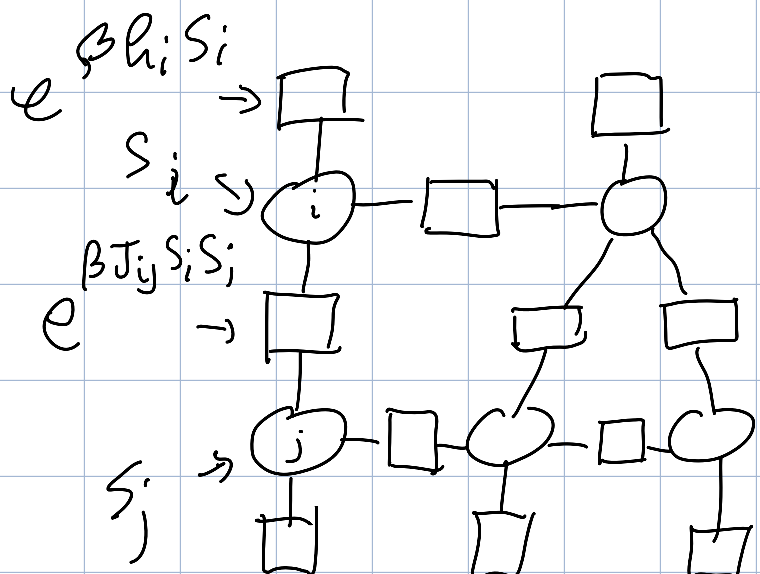
Examples:

1) Ising model

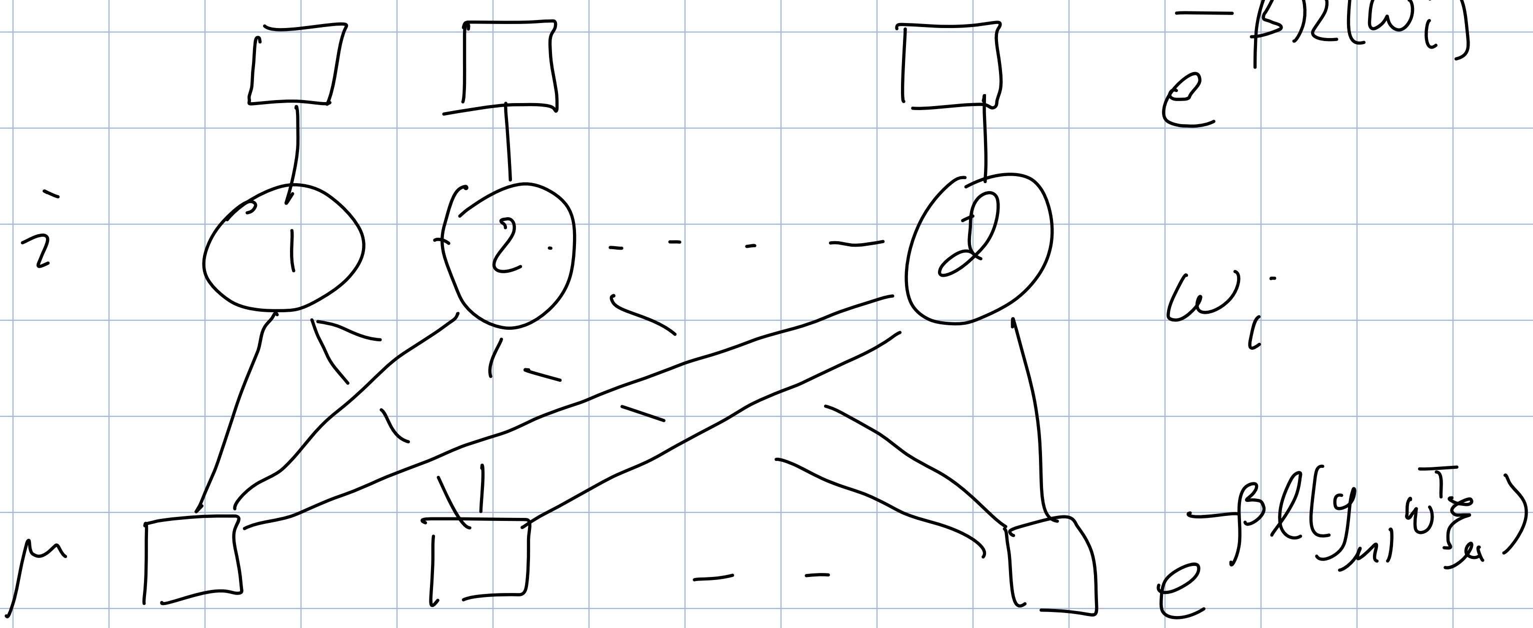
Graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$

$$\mathcal{H} = -\sum_{(ij)\in\mathcal{E}} J_{ij} s_i s_j - \sum_{i\in\mathcal{V}} h_i s_i$$

Factor graph version



## 2) Generalized Linear model



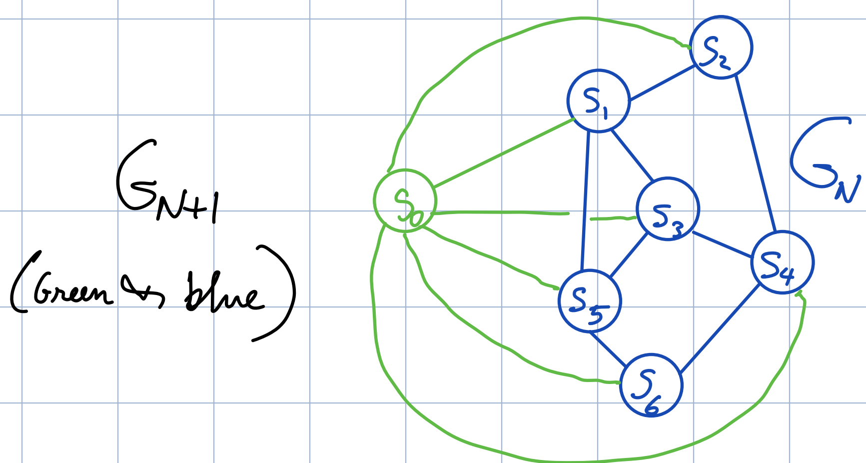

If the factor graph is a tree, we can solve it by Belief Propagation (BP). If it is sparse and tree-like, we can do loopy BP. For some dense problems, we can do Approximate Message Passing (AMP).

AMP has been a major contribution of statistical physics back to statistics, thanks to the work of Andrea Montanari and his collaborators.

However, to appreciate it one needs to understand the Onsager term. The best place to see it is in Spin Glass physics, the Thouless-Anderson-Palmer (TAP) equation.

# Cavity method and the Onsager term

Go from $G_N$ to $G_{N+1}$




$$\mathcal{H}_{N+1}(S_{0:N}) = \mathcal{H}_N(S_{1:N}) - \sum_{(0,i)\in\mathcal{E}_{N+1}} J_{0i} S_0 S_i - h_0 S_0$$

$$= \mathcal{H}_N - h_0^e(S_{1:N})\, S_0$$

where the effective field

$$h_0^e(S_{1:N}) := h_0 + \sum_{(0,i)\in\mathcal{E}_{N+1}} J_{0i} S_i$$

$$P_{N+1}(s_0, h_0^e) = \frac{e^{\beta h_0^e s_0} P_N(h_0^e)}{\sum_{s_0} \int e^{\beta h_0^e s_0} P_N(h_0^e)\, dh_0^e}$$

We would like to compute $E_{N+1}[S_0]$

$E_N[h_0^e]$ & $E_{N+1}[h_0^e]$.

What if we could postulate $P_N(h_0^e)$

$$\simeq \frac{1}{\sigma_N\sqrt{2\pi}} e^{-\frac{(h_0^e-\mu)^2}{2\sigma_N^2}}$$ when $N$ is large?

If so, $\mu = h_0 + \sum_{(0,i)\in \mathcal{E}_{N+1}} J_{0i} E_N[S_i]$

Note that $E_N[\ ]$ is the expectation in the system with a cavity, the 0-th site.

We insert a spin into that cavity and then compute the resulting changes, trying to understand what happens as $N \to \infty$.

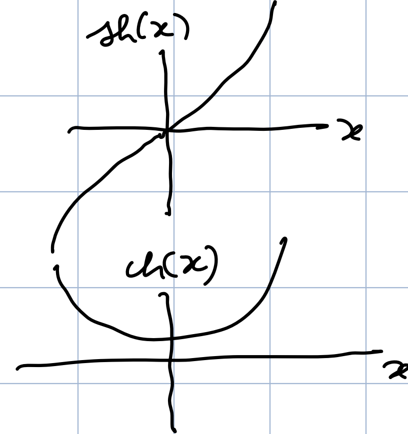
We will tackle $\sigma_N$ later (it is model-type dependent).

In the immediate discussion, I replace $h_0^e(S_{1:N})$ by just $h$.

$$E_{N+1}[S_0] = \frac{\int \sum_{S_0} S_0 e^{\beta h S_0} P_N(h)\, dh}{\int \sum_{S_0} e^{\beta h S_0} P_N(h)\, dh}$$

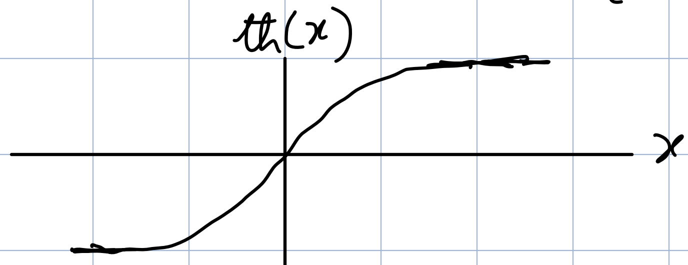
$$= \frac{\int \operatorname{sh}(\beta h)\, P_N(h)\, dh}{\int \operatorname{ch}(\beta h)\, P_N(h)\, dh}$$

$$sh(x) = \sinh(x) = \frac{1}{2}\left(e^{x} - e^{-x}\right)$$



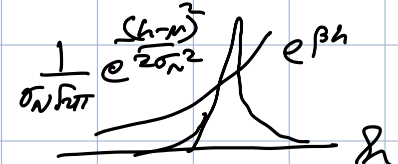

$$ch(x) = \cosh(x) = \frac{1}{2}\left(e^{x} + e^{-x}\right)$$

$$th(x) = \tanh(x) = \frac{\sinh(x)}{\cosh(x)} = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}}$$




Let us continue $\quad E_{N+1}[S_0] = \dfrac{\int sh(\beta h) \frac{1}{\sigma_N\sqrt{2\pi}} e^{-\frac{(h-\mu)^2}{2\sigma_N^2}} dh}{\int ch(\beta h) \frac{1}{\sigma_N\sqrt{2\pi}} e^{-\frac{(h-\mu)^2}{2\sigma_N^2}} dh}$

Now $\displaystyle\int e^{\pm\beta h} \frac{1}{\sigma_N\sqrt{2\pi}} e^{-\frac{(h-\mu)^2}{2\sigma_N^2}} dh$




$$= e^{\pm\beta\mu + \frac{\beta^2\sigma_N^2}{2}}$$

Using that, $\quad E_{N+1}[S_0] = \dfrac{\frac{1}{2}\left(e^{\beta\mu} - e^{-\beta\mu}\right) e^{\frac{\beta^2\sigma_N^2}{2}}}{\frac{1}{2}\left(e^{\beta\mu} + e^{-\beta\mu}\right) e^{\frac{\sigma_N^2\beta^2}{2}}}$

$$= th(\beta\mu)$$

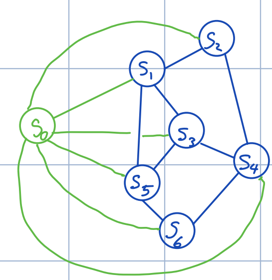
Thus

$$\boxed{E_{N+1}[S_0] = \tanh\left(\beta E_N[h_0^e]\right) = \tanh\left(\beta h_0 + \beta \sum_{(0,i)\in\mathcal{E}_{N+1}} J_{0i}\, E_N[S_i]\right)}$$

Expectation in the N+1 spin system

Expectation in the N spin system (w. cavity)

The eqn connecting $E_{N+1}[S_0]$ and $E_N[S_i]$ is similar in spirit to belief propagation.




We could have written this as

$$m_0 = \tanh\left(\beta h_0 + \beta \sum_{i\neq 0} J_{ij}\, m_{i\to 0}\right)$$

Considering any site in the graph as 0 we get a set of equations:

$$m_i = th\left(\beta h_i + \beta \sum_{j \neq i} J_{ij}\, m_{j\to i}\right).$$

We then need to get equations for $m_{j\to i}$ which is problematic.

For the time being, we try get a relation between $E_{N+1}[S_0]$ and $E_{N+1}[S_i]$ directly.

Perhaps we ask how $E_N[h_e^0]$ is related to $E_{N+1}[h_e^0]$.

$$P_{N+1}(h_0^e) = \sum_{S_0} P_{N+1}(S_0, h_0^e)$$

$$P_{N+1}(h) = \frac{ch(\beta h)\, P_N(h)}{\int ch(\beta h')\, P_N(h')\, dh'}$$

$$E_{N+1}[h] = \frac{\int h\, ch(\beta h)\, P_N(h)\, dh}{\int ch(\beta h)\, P_N(h)\, dh}$$

$$\int_{-\infty}^{\infty} h\, ch\, \beta h\, P_N(h)\, dh$$

$$= \frac{d}{d\beta} \int sh(\beta h)\, P_N(h)\, dh$$

(Using the Gaussian ansatz)

$$= \frac{d}{d\beta} \quad sh(\beta\mu)\, e^{\frac{\beta^2 \sigma_N^2}{2}}$$

$$= \left(\mu \cosh \beta\mu + \beta\sigma_N^2 \, sh(\beta\mu)\right) e^{\frac{\beta^2\sigma_N^2}{2}}$$

$$\int_{-\infty}^{\infty} ch(\beta h) \, P_N(h) \, dh$$

$$= ch(\beta\mu) \, e^{\frac{\beta^2\sigma_N^2}{2}}$$

So the ratio is $\mu + \beta\sigma_N^2 \, th(\beta\mu)$

Alternatively, we could use Stein's lemma $E[g(X)(X-\mu)] = \sigma^2 E[g'(X)]$

with differentiable $g$, when $X \sim \mathcal{N}(\mu, \sigma^2)$

$$\left[ \; E[g'(X)] = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} g'(x) \, e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \right.$$

$$= -\frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} g(x) \frac{d}{dx} e^{-\frac{(x-\mu)^2}{2\sigma^2}} = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} g(x) \frac{(x-\mu)}{\sigma^2} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

$$= \frac{1}{\sigma^2} E[g(X)(X-\mu)] \Big]$$

with $g(X) = \cosh \beta X$

$$\frac{E[g(X)X]}{E[g(X)]} = \frac{\mu E[g(X)] + \sigma^2 [g'(X)]}{E[g(X)]}$$

$$= \mu + \beta\sigma^2 \, \mathrm{th}\, \beta\mu$$

So $E_{N+1}[h_0^e] = \mu + \beta\sigma_N^2 \, \mathrm{th}(\beta\mu)$

$$\Rightarrow \boxed{E_{N+1}[h_0^e] = E_N[h_0^e] + \beta\sigma_N^2 \, E_{N+1}[S_0]}$$

See that $E_{N+1}[S_0] = \mathrm{th}(\beta E_N[h_0^e])$

and $E_N[h_0^e] = E_{N+1}[h_0^e] - \beta\sigma_N^2 E_{N+1}[S_0]$.

In the full system $S_i$'s feel $S_0$.

Hence we get

$$E_{N+1}[S_0] = \text{th}\Big(\beta h_0 + \beta \sum_{(0,i)\in\mathcal{E}_{N+1}} J_{0i} E_{N+1}[S_0] - \beta^2 \sigma_N^2 E_{N+1}[S_0]\Big)$$

Often written as

$$m_j = \text{th}\Big(\beta h_j + \beta \sum_{(j,i)\in\mathcal{E}} J_{ji} m_i - \beta^2 \sigma_N^2 m_j\Big)$$

Onsager reaction term

What $\sigma_N^2$? Well, that depends on the model!

Random Field Ising Model (RFIM)

$$J_{ij}^{(N)} = \frac{J}{N}$$

$$h_0^e = \sum_{i=1}^{N} J_{0i}\, S_i + h_0$$

$$= \frac{J}{N} \sum_i S_i + h_0$$

$$\mathrm{Var}_N[h_0^e] = \frac{J^2}{N^2} \sum_{i,j} \mathrm{cov}_N[S_i, S_j] = \frac{J^2}{N^2} \sum_i \mathrm{Var}_N[S_i]$$

$$+ \frac{2J^2}{N^2} \sum_{i<j} \mathrm{cov}_N[S_i, S_j]$$

$$\mathrm{Var}_N[S_i] = E_N[S_i^2] - (E_N[S_i])^2$$

$$\simeq 1 - m_i^2$$

$E_N[S_i]$ and $E_{N+1}[S_i]$ are different but

$S_0 \overset{-J/N}{\longrightarrow\!\!\!-\!\!\!-} S_i$ is a small influence and

$$E_{N+1}[S_i] - E_N[S_i] = O\left(\frac{1}{N}\right).$$

How to estimate $N$ dependence of

$\mathrm{cov}[S_i, S_j]$, when $i \neq j$? It is $O\left(\frac{1}{N}\right)$

(justification from high temp expn.

later).

So $\quad \sigma_N^2 = \frac{J^2}{N^2} O(N) + \frac{2J^2}{N^2} O(N^2 \frac{1}{N})$

$= J^2 O\left(\frac{1}{N}\right)$

Thus we get

$m_i \approx th\left(\frac{J}{N} \sum_i m_i + h_i\right)$

If we call $\frac{1}{N} \sum_i m_i = m$

$m_i \simeq th\,(Jm + h_i)$

and $\quad m = \frac{1}{N} \sum_i th(Jm + h_i)$

$\simeq E_h\, th(Jm + h)$

Now consider the problem where

$J_{ij}$ are random,

with $E[J_{ij}] = 0$, $\mathrm{Var}[J_{ij}] = J^2/N$

$$\mathrm{Var}_N[h_0^e] = \sum_i J_{0i}^2 \mathrm{Var}[S_i] + \sum_{i \neq j} J_{0i} J_{0j} \mathrm{Cov}[S_i, S_j]$$

$$\approx \frac{J^2}{N} \sum_i \mathrm{Var}[S_i] + O\left(\frac{1}{N^{1/2}} z\right)$$

random normal var

$$\approx J^2 \Big(1 - \underbrace{\frac{1}{N} \sum m_i^2}_{q}\Big)$$

$$m_i = \text{th}\left(\beta h_i + \beta \sum_j J_{ij} m_j - \beta^2 J^2 (1-q) m_i\right)$$

This is the Thouless - Anderson - Palmer (TAP) equation, with a non-trivial Onsager term.

How to solve it? In 2009, Bolthausen suggested that the iteration

$$m_i^{t+1} = \text{th}\left(\beta h_i + \beta \sum_j J_{ij} m_j^t - \beta^2 J^2 (1-q_t) m_i^{t-1}\right)$$

with $q_t = \frac{1}{N} \sum_i m_i^{t\,2}$, can be analyzed and $(h^e)^1, (h^e)^2, \cdots$ could be shown to be asymptotically jointy Gaussian.

Here we switch notation and breakup

$$h_i^e = h_i + x_i$$

total effective field ↗ $h_i^e$; applied field ↗ $h_i$; $x_i$ ↖ effective field from spins

Focus on the uniform field case $h_i = h$

Iteration for vectors $x^{(t)}, m^{(t)} \in \mathbb{R}^N$,

$$\underline{x}^{(t+1)} = J\underline{m}^{(t)} - \beta(1-q_t)\underline{m}^{(t-1)}$$

$$\underline{m}^{(t)} = \mathrm{th}(\beta \underline{x}^{(t)} + \beta h) = f(\underline{x}^{(t)})$$

Function applies component by component

Pretend $E\left[X_i^{(s)} X_j^{(t)}\right] = K_{s,t}\, \delta_{ij}$

with $K_{s,t} = \frac{1}{N}\langle m^{(s-1)}, m^{(t-1)}\rangle := q_{s-1,t-1}$

$q_{t,t} := q_t$ .

TAP: $q_t = \frac{1}{N}\|m^t\|^2 \simeq \frac{1}{N}\left\| th\left(\beta\sqrt{q_{t-1}}\,\underline{Z} + \beta h\right)\right\|^2$

$$= E_{Z\sim N(0,1)}\left[\left(th\left(\beta\sqrt{q_{t-1}}\,Z + h\right)\right)^2\right]$$

This is the $q_t$ iteration equation.

Fixed point: $q^* = E_{Z\sim N(0,1)}\left[\left(th\left(\beta\sqrt{q^*}\,Z + \beta h\right)\right)^2\right]$

Condition for a stable fixed point

$$E_{Z\sim N(0,1)}\left[\frac{1}{\left(ch\left(\beta\sqrt{q^*}\,Z + \beta h\right)\right)^4}\right] \leq \beta^2 .$$

Equality gives the Alameida Thouless line.

General Symmetric AMP

$$x^{(t+1)} = A m^{(t)} - b_t m^{(t-1)}$$

$$m^{(t)} = f_t(x^{(t)})$$

$$b_t = E\left[\operatorname{div} f_t(X^t)\right]$$

where $X^{(t)} \sim \mathcal{N}(0, \kappa_{t,t} I_N)$

and $\kappa_{s,t} = \frac{1}{N}\langle m^{(s-1)}, m^{(t-1)}\rangle$

Now let us talk about a signal processing / Bayesian inference problem very close to the TAP / spin glass problem.

The Wigner Spike model

$Y, W \in \mathbb{R}^{N \times N}, \quad x_* \in \mathbb{R}^N$

$$Y = \frac{\sqrt{\lambda}}{N} x_* x_*^T + W$$

observations

$$i \le j, \quad W_{ji} = W_{ij} \overset{iid}{\sim} \mathcal{N}(0, \tfrac{1}{N})$$

$$x_{*i} \overset{iid}{\sim} P_X$$

Want to recover $x_*$ from $Y$.

$$P(\underline{x} | Y) \propto e^{-N \| \underline{Y} - \underline{x}\underline{x}^T \|_F^2 / 2} \prod_i P_X(x_i)$$

$$\propto e^{N \underline{x}^T \underline{Y} \underline{x} - N \|\underline{x}\|^4 / 2} \prod_i P_X(x_i)$$

See the analogy with Ising! $\underline{x} \longleftrightarrow \underline{s}$

$\underline{Y} \longleftrightarrow \underline{J}$

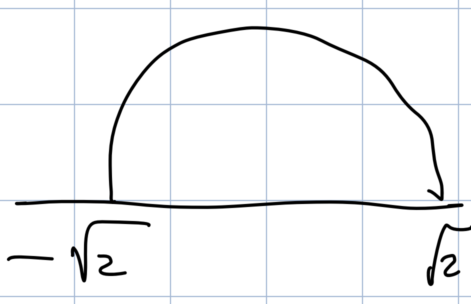
If we assume $P_X$ is Rademacher

$$P_X(x_i) = \frac{1}{2}\,\delta(x_i+1) + \frac{1}{2}\,\delta(x_i-1)$$

then it is exactly a disordered Ising model, with

$$\underset{\substack{\uparrow \\ \text{Single Hopfield pattern}}}{Y_{ij} = \sqrt{\tfrac{\lambda}{N}}\, x_i x_j} + \underset{\substack{\uparrow \\ \text{spin glass}}}{W_{ij}}$$

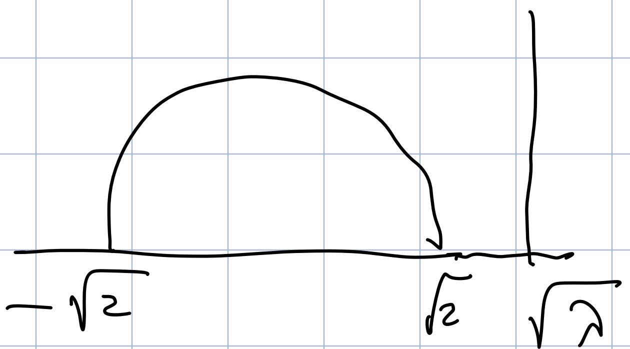
Eigenvalues of $\underline{\Xi}$



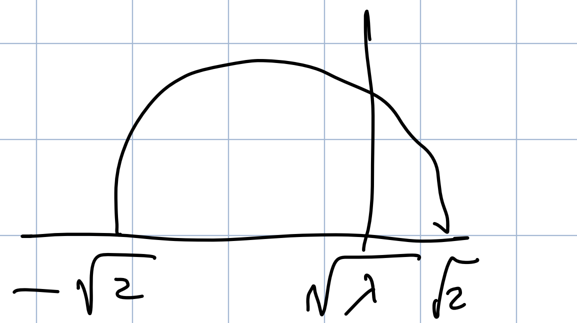
Eigenvalue of $\sqrt{\frac{\lambda}{N}}\, x_* x_*^T$ : $\frac{\sqrt{\lambda}}{N} \|x_*\|^2$

$= \sqrt{\lambda}$

(Rademacher)




easily detectable




Hard to detect

Analysis

$$\underline{x}^{(t+1)} = \left(\underline{W} + \frac{\sqrt{\lambda}}{N} x_* x_*^T\right)\underline{m}^{(t)} - b_t \underline{m}^{(t-1)}$$

$$\underline{m}^t = f_t(\underline{x}^t)$$

$$\underline{x}^{(t+1)} = \underline{\underline{W}}\,\underline{m}^{(t)} - b_t\,\underline{m}^{(t-1)} + \sqrt{\lambda}\,\underline{x}^{*} q_{0t}$$

$$q_{0t} = \frac{1}{N}\,\underline{x}_{*}^{T}\,\underline{m}^{t}$$

overlap of $E[\underline{x}]$ with $\underline{x}_{*}$

Define $\underline{\tilde{x}}^{(t+1)} = \underline{W}\,\underline{m}^{t} - b_t\,\underline{m}^{(t-1)}$

$$\tilde{f}_t(\tilde{x}) = f_t(\underline{x} + x_0 q_{0t})$$

$$\underline{\tilde{x}}^{(t+1)} = \underline{W}\,\underline{m}^{(t)} - b_t\,\underline{m}^{(t-1)}$$

$$\underline{m}^{(t)} = \tilde{f}_t(\tilde{x}^{t})$$

$$q_t = \mathbb{E}\left[\left(\tanh\left(\beta\sqrt{q_{t-1}}\,z + \beta\sqrt{\lambda}\,q_{0t-1}X_*\right)\right)^2\right]$$

$$q_{0t} = \mathbb{E}\left[\tanh\left(\beta\sqrt{q_{t-1}}\,z + \beta\sqrt{\lambda}\,q_{0t-1}X_*\right)X_*\right]$$

$\lambda > \lambda_c$ $\quad q_{0t} \to$ nonzero bound

$\lambda < \lambda_c$ $\quad q_{0t} \to 0$

# Extra Notes:

## 1) Iterated BP → Onsager/AMP

In the style of belief propagation, we could have written the eqn connecting $E_{N+1}[S_0]$ and $E_N[S_i]$ as

$$m_i^{(t+1)} = \text{th}\left(\beta h_i + \beta \sum_{j \in \partial i} J_{ij}\, m_{j\to i}^{(t)}\right)$$

Now, what about $m_{j\to}^{(t)}$ ?

Perhaps replace it by $m_j^{(t)}$ + correction

$$m_j^{(t)} = \mathrm{th}\left(\beta h_j + \beta \sum_{k \in \partial j \setminus i} J_{jk}\, m_{k\to j}^{(t-1)} + \beta J_{ji}\, m_{i\to j}^{(t-1)}\right)$$

$$= \mathrm{th}\left(\beta h_j + \beta \sum_{k \in \partial j \setminus i} J_{jk}\, m_{k\to j}^{(t-1)}\right)$$

$$+ \beta J_{ji}\left(1 - \mathrm{th}^2\left(\beta h_j + \beta \sum_{k \in \partial j \setminus i} J_{jk}\, m_{k\to j}^{(t-1)}\right)\right) m_{i\to j}^{(t-1)}$$

$$\simeq m_{j\to i}^{(t)} + \beta J_{ji}\left(1 - m_{j\to i}^{(t)\,2}\right) m_{i\to j}^{(t-1)}$$

Further approx

$$m_{k\to j}^{(t)} \simeq m_j^{(t)} - \beta J_{ji}\left(1 - m_j^{(t)\,2}\right) m_i^{(t-1)}$$

$$m_i^{(t+1)} = \mathrm{th}\left(\beta h_i + \beta \sum_j J_{ij}\left(m_j^{(t)} - \beta J_{ji}\left(1 - m_j^{(t)\,2}\right) m_i^{(t-1)}\right)\right)$$

$$= \mathrm{th}\left(\beta h_i + \beta \sum_j J_{ij}\, m_j^{(t)} - \beta^2 \sum_j J_{ij}^2 \left(1 - m_j^{(t)\,2}\right) m_i^{(t-1)}\right)$$

$$= \mathrm{th}\left(\beta h_i + \beta \sum_j J_{ij}\, m_j^{(t)} - \beta^2 J^2 (1 - q_t)\, m_i^{(t-1)}\right)$$

TAP eqn!

2) High temperature (small $\beta$) estimate of $\mathrm{Cov}[S_i; S_j]$ for RFIM

Let us start with the partition func

$$Z = \sum_{\{S\}} e^{\frac{\beta J}{N}\left(\sum S_i\right)^2 + \beta \sum_i h_i S_i}$$

$$\frac{1}{\beta} \frac{\partial \ln Z}{\partial h_i} = E[S_i]$$

$$\frac{1}{\beta^2} \frac{\partial^2 \ln Z}{\partial h_i \partial h_j} = \mathrm{Cov}[S_i, S_j]$$

We plan to expand $Z$ in the small $\beta$ limit, compute $\ln Z$, and take derivatives.

$$Z = \sum_{\{S\}} e^{\frac{\beta J}{N}\left(\sum S_i\right)^2 + \beta \sum_i h_i S_i}$$

$$= \sum_{\{S\}} \Big[ 1 + \frac{\beta J}{N} \sum (S_i)^2 + \beta \sum h_i S_i$$

$$+ \frac{1}{2} \frac{\beta^2 J^2}{N^2} \left(\sum S_i\right)^4 + \frac{\beta^2 J}{N} \left(\sum_i S_i\right)^2 \sum h_i S_i$$

$$+ \frac{1}{2} \beta^2 \left(\sum_i h_i S_i\right)^2 + \frac{1}{6} \frac{\beta^3 J^3}{N^3} \left(\sum_i S_i\right)^6$$

$$+ \frac{1}{2} \frac{\beta^3 J^2}{N^2} \left(\sum S_i\right)^4 \sum h_i S_i + \frac{1}{2} \frac{\beta^3 J}{N} \left(\sum_i S_i\right)^2 \left(\sum_i h_i S_i\right)^2$$

$$+ \frac{1}{6} \beta^3 \left(\sum h_i S_i\right)^3 + O(\beta^4) \Big]$$

$$= 1 + \frac{\beta J}{N} N + 0 + \frac{1}{2} \frac{\beta^2 J^2}{N^2} \left(N + 3N(N-1)\right)$$

$$+ 0 + \frac{1}{2} \beta^2 \sum_i h_i^2 + \frac{1}{6} \frac{\beta^3 J^3}{N^3} \left(N + 15N(N-1) + 15N(N-1)(N-2)\right)$$

$$+ 0 + \frac{1}{2} \frac{\beta^3 J}{N} N \sum_i h_i^2 + \frac{1}{2} \frac{\beta^3 J}{N} \sum_{i \neq j} h_i h_j + O(\beta^4) \Big]$$

$$= 2^N \Big[ 1 + \beta J + \frac{1}{2} \beta^2 \left( \frac{3N-2}{N} J^2 + \sum h_i^2 \right)$$

$$+ \frac{1}{6}\beta^3\left(\frac{15N^2-30N+16}{N^2}J^3 + 3J\sum h_i^2 + \frac{3J}{N}\sum_{i\neq j} h_i h_j\right)$$

$$+ O(\beta^4)\Big]$$

$$\ln Z = N\ln 2 + \beta J + \frac{1}{2}\beta^2\left(\frac{3N-2}{N}J^2 + \sum h_i^2\right)$$

$$+ \frac{1}{6}\beta^3\left(\frac{15N^2-30N+16}{N^2}J^3 + 3J\sum h_i^2 + \frac{3J}{N}\sum_{i\neq j} h_i h_j\right)$$

$$- \frac{1}{2}\left[\beta^2 J^2 + \frac{1}{2}\beta^3 J\left(\frac{3N-2}{N}J^2 + \sum h_i^2\right)\right]$$

$$+ \frac{1}{3}\beta^3 J^3 + O(\beta^4)$$

So, for $i \neq j$,

$$\mathrm{Cov}[S_i, S_j] = \frac{1}{\beta^2}\frac{\partial^2}{\partial h_i \partial h_j}\ln Z = \frac{\beta J}{N} + O(\beta^2).$$