# Application of message passing algorithm in network problems

Hou TianQi
Supervisor:WONG Kwok Yee Michael

Department of Physics
The Hong Kong University of Science and Technology

PhD Qualifying Examination, August 2018

# outline

1. Belief propagation in 2D Ising model
2. Message passing in network flow problems
3. Message passing in RBM

# Outline

1. Belief propagation in 2D Ising model
2. Message passing in network flow problems
3. Message passing in RBM

# Belief propagation

2D Ising model has spin variables $\{s_i\}$ in lattice networks ,where $s_i = +1$ represent this spin is 'up',and $s_i = -1$ represent 'down'. $J_{ij}$ is the strength of the interaction of two spins.

## Definition

The Hamilton is:

$$H = -\sum_{ij} J_{ij} s_i s_j$$

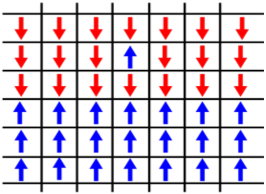


Figure: 2D Ising model toy model

# Bethe approximation

In 2D Ising model ,if we neglect long range interaction.

## Bethe approximation

In the Bethe approximation ,only $P(s_i, s_j)$ and $P(s_i)$ are considered.

$$P(\{s_i\}) = \frac{\prod_{ij} P(s_i, s_j)}{\prod_i P(s_i)}$$

# Belief propagation

The free energy of 2D Ising model is

$$\mathcal{F} = \mathcal{U} - T\mathcal{S}$$

By minimizing the free energy $\frac{\partial \mathcal{F}}{\partial P(s_i)} = 0$ and $\frac{\partial \mathcal{F}}{\partial P(s_i, s_j)} = 0$ we get belief propagation:

## belief propagation

$$p_{a\to i}(\sigma_i) = \frac{1}{Z_{a\to i}} \sum_{j\in\partial a\setminus i} \varphi(\partial a) \prod_{j\in\partial a\setminus i} q_{j\to a}(\sigma_j)$$

$$q_{i\to a}(\sigma_i) = \frac{1}{Z_{i\to a}} \prod_{b\in\partial i\setminus a} p_{b\to i}(\sigma_i)$$

HKUST

## cavity probability

- $q_{a\to i}$ : probability of $\sigma_i$ only interact with $a$
- $p_{i\to a}$ :$\sigma_i$ intercact with all factor node expect $a$

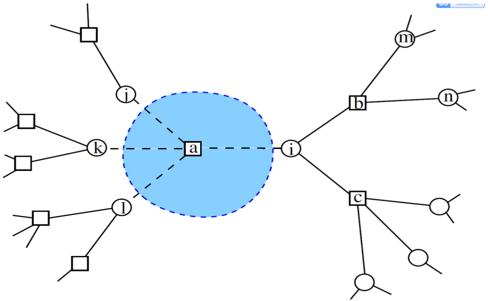



Figure: cavity probability $q_{a\to i}(\sigma_i)$ and$p_{i\to a}(\sigma_i)$

# Belief propagation

We parameterize the cavity probability $q_{a\to i}(\sigma_i)$ and $p_{i\to a}(\sigma_i)$ as:

$$p_{i\to a}(\sigma_i) = \frac{e^{\beta h_{i\to a}\sigma_i}}{2\cosh\beta h_{i\to a}}$$

$$q_{a\to i}(\sigma_i) = \frac{e^{\beta u_{a\to i}\sigma_i}}{2\cosh\beta u_{a\to i}}$$

we can propagate message in 2D Ising model:

## Belief propagation in 2D Ising model

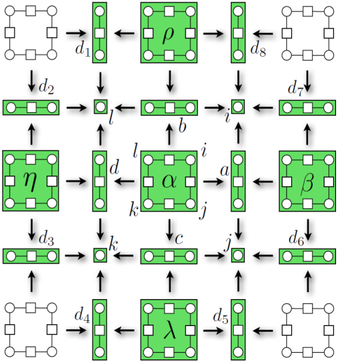
$$h_{i\to a} = \sum_{b\in\partial a\setminus i} u_{b\to i}$$

$$u_{a\to i} = \frac{1}{\beta}\arctan[\tanh\beta J_{ij}\tanh\beta h_{j\to a}]$$

# Generalize belief propagation(GBP)

cluster variation method is a advanced approximation method which consider the region cluster probability $P(s_i, s_j, s_k, s_l)$.In this approximation.

- node $i$ probability:$P_i(s_i)$
- link $a$ probability : $P_a(s_i, s_j)$
- region $\alpha$ probability: $P_\alpha(s_i, s_k, s_j, s_l)$

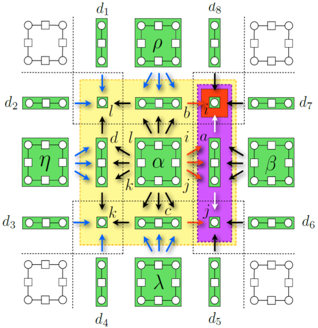


# Generalize belief propagation (GBP)

by minimizing the free energy ,we can get the propagation of two kinds probability $P_{a\to i}(\sigma_i)$ and $P_{\alpha\to a}(\sigma_i,\sigma_j)$

For link to node message:

$$P_{a\to i}(\sigma_i) = \frac{1}{Z_{a\to i}}\varphi_a(\sigma_i,\sigma_j)P_{\alpha\to a}(\sigma_i,\sigma_j)P_{\beta\to a}(\sigma_i,\sigma_j)$$
$$P_{c\to j}(\sigma_j)P_{d_5\to j}P_{d_6\to j}$$

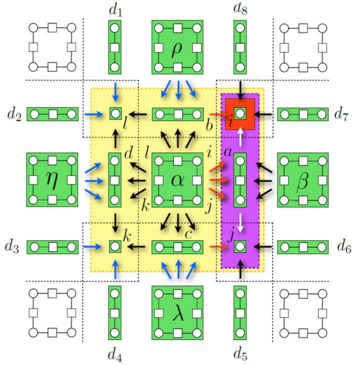


# Generalize belief propagation (GBP)

For region to link message :

$$P_{\alpha\to a}(\sigma_i,\sigma_j)P_{b\to i}(\sigma_i)P_{c\to j}(\sigma_j) = \frac{1}{Z_{\alpha\to a}}\sum_{\sigma_l,\sigma_k}\varphi_b(\sigma_i,\sigma_l)\varphi_c(\sigma_k,\sigma_j)\varphi_d(\sigma_k,\sigma_l)$$

$$P_{\rho\to b}(\sigma_i,\sigma_l)P_{\eta\to d}(\sigma_l,\sigma_k)P_{\lambda\to c}(\sigma_k,\sigma_j)P_{d_1\to l}(\sigma_l)P_{d_3\to k}(\sigma_k)P_{d_2\to l}(\sigma_l)$$



# Generalize belief propagation (GBP)

parameterize the cavity probability

$$P_{a\to i}(\sigma_i) = \frac{e^{\beta u_{a\to i}\sigma_i}}{2\cosh \beta u_{a\to i}}$$

$$P_{\alpha\to a}(\sigma_i, \sigma_j) \propto e^{u^i_{\alpha\to a}\sigma_i + u^j_{\alpha\to a}\sigma_j + U_{\alpha\to a}\sigma_i\sigma_j}$$

parameterize the marginal probability

$$P(\sigma_i, \sigma_j) = \frac{1 + m_i\sigma_i + m_j\sigma_j + c_{ij}\sigma_i\sigma_j}{4}$$

Compare with the simulation results,Generalize belief propagation will get more accurate critical temperature(in 2D Ising model, $T_c = 2.2$).

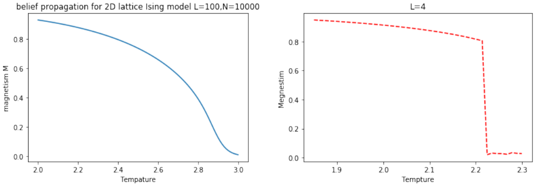



Figure: Magnetism of 2D Ising model by BP method and GBP method at different temperature

# Outline

# Network flow and Potential

In many network flow problems,as facility location or power grids, each node have some resource $P_i$, our aim is to minimize total cost by network flow $y_{ij}$.

$$\min E = \sum_{ij} \frac{R_{ij}}{2} y_{ij}^2$$

$$s.t. \sum_{j \in \partial i} y_{ij} + P_i = 0$$

The network flow can be seen as difference of potential of nodes$\mu_i, \mu_j$,

$$y_{ij} = \frac{1}{R_{ij}} (\mu_j - \mu_i)$$

The networks nodes potential satisfy laplacian equation:

$$\mathcal{L} \boldsymbol{\mu} = P$$

$$\mathcal{L}_{ij} = (\sum_{k \in \partial i} \frac{1}{R_{ik}}) \delta_{ij} - \frac{1}{R_{ij}}$$

# Message Passing for network flow

The network flow problem can be solved by solve network flow $y_{ij}$ directly, or by solving potential $\mu_i$.

## cavity method

As in the Ising model case, $E_{j\to i}(y_{ij})$ is the cavity energy in the absence of node $j$:

$$E_{j\to i}(y_{ij}) = min_{y_{ij}}[\sum_{k\in\partial j\backslash i} E_{k\to j}(y_{jk}) + \frac{1}{2}R_{ij}y_{ij}^2]$$

parameter the cavity energy $E_{j\to i}(y_{ij})$ by $E_{j\to i}(y_{ij}) = \frac{1}{2}r_{i\to j}(y_{ij} - y_{i\to j})^2$ we get flow messages propagate in the network by:

$$r_{j\to i} = (\sum_{k\in\partial j\backslash i} r_{k\to j}^{-1})^{-1} + R_{ij}$$

$$y_{j\to i} = \frac{P_j + \sum_{k\in\partial j\backslash i} y_{k\to j}}{1 + R_{ij}\sum_{k\in\partial j\backslash i} r_{k\to j}}$$

In the perspective of potential $\mu_i$, the cost function we optimize is :

$$E = \sum_{ij} \frac{1}{2} g_{ij} (\mu_i - \mu_j)^2 - \sum_i P_i \mu_i$$

we define potential cavity method $E_{j\to i}(\mu_i)$

$$E_{j\to i}(\mu_i) = \min_{\mu_j} [\sum_{k\in\partial j\backslash i} E_{k\to j}(\mu_j) + \frac{1}{2} g_{ij} (\mu_j - \mu_i)^2 - P_j \mu_j]$$

By parameter the potential cavity energy as

$$E_{j\to i}(\mu_i) = \frac{1}{2} g_{j\to i} (\mu_i - \mu_{j\to i})^2$$

# cavity method for potential

The cavity message for potentials propagate in the network is:

$$g_{j\to i} = \frac{(\sum_{k\in\partial j\backslash i} g_{k\to j})g_{ij}}{\sum_{k\in\partial j\backslash i} g_{k\to j} + g_{ij}}$$

$$g_{j\to i}\mu_{j\to i} = \frac{g_{ij}(\sum_{k\in\partial j\backslash i} g_{k\to j}\mu_{k\to j} + P_j)}{\sum_{k\in\partial j\backslash i} g_{k\to j} + g_{ij}}$$

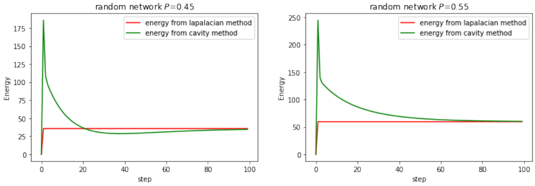



Figure: Results of cavity method in ER network

# Generalize belief propagation for network flow problem

we try to include high order correlation in network flow problem by consider region part, same as in 2D Ising model , we define two cavity energy:

- link to node cavity energy $E_{a\to i}(\mu_i)$ :

$$E_{a\to i}(\mu_i) = \frac{1}{2} g_{a\to i}(\mu_i - \mu_{a\to i})^2$$

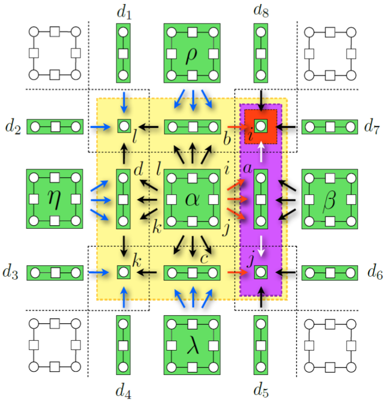
- region to link cavity energy $E_{\alpha\to a}(\mu_i, \mu_j)$ we define as

$$E_{\alpha\to a} = \frac{1}{2} g^i_{\alpha\to a}(\mu_i - \mu^i_{\alpha\to a})^2 + \frac{1}{2} g^j_{\alpha\to a}(\mu_j - \mu^j_{\alpha\to a})^2 + \frac{1}{2} h_{\alpha\to a}(\mu_i - \mu_j)^2$$

# Generalize belief propagation for network flow problem

link to node cavity energy is :

$$E_{a\to i}(\mu_i) = \min[\frac{1}{2}g_{ij}(\mu_i - \mu_j)^2 - P_j\mu_j + \sum_{b\in\partial j\backslash a} E_{b\to j}(\mu_j) + \sum_{\mathcal{R}\in\alpha,\beta} E_{\mathcal{R}\to a}(\mu_i,\mu_j)]$$



# Generalize belief propagation for network flow problem

region to link cavity energy is:

$$\begin{aligned} E_{\alpha\to a}(\mu_i,\mu_j) = \min[&\frac{1}{2}g_{ij}(\mu_i-\mu_j)^2+\frac{1}{2}g_{il}(\mu_i-\mu_l)^2+\frac{1}{2}g_{jk}(\mu_j-\mu_k)^2 \\ &-P_k\mu_k-P_l\mu_l+\sum_{i=1,2}E_{d_i\to l}(\mu_l)+\sum_{i=3,4}E_{d_i\to k}(\mu_k)+ \\ &E_{\rho\to b}(\mu_l,\mu_i)+E_{\eta\to d}(\mu_k,\mu_l)+E_{\lambda\to c}(\mu_k,\mu_j)] \\ &-E_{b\to i}(\mu_i)-E_{c\to j}(\mu_j) \end{aligned}$$

# Generalize belief propagation for network flow problem

- Avoid diverge,we can random choose base station (ground station in electric) $b$, we can set $\mu_b = 0$
- when $j$ is a dangling node(degree $k$=1), we set $g_{j\to i} = 0$ and $g_{j\to i}\mu_{j\to i} = P_j$
- we can use damping to accelerate algorithm and avoid diverge
- we use parallel update,update $E_{\alpha\to a}(\mu_i, \mu_j) + E_{b\to i}(\mu_i) + E_{c\to j}(\mu_j)$ together

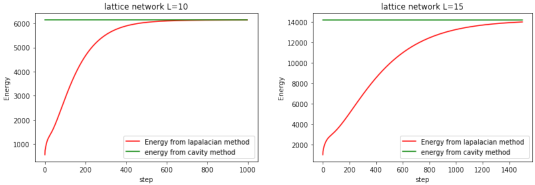



Figure: Results of GBP method in lattice network

# Message passing for network fluctuation

The robust of power system is import in modern society. let $P_i^0, y_{ij}^0, \mu_i^0$ be the initial state, when some fluctuation on the resource add to the system.

$$P_i = P_i^0 + \delta P_i$$
$$y_{ij} = y_{ij}^0 + \delta y_{ij}$$
$$\mu_i = \mu_i^0 + \delta \mu_i$$

# Message passing for network fluctuation

The fluctuation of potential by message passing algorithm is :

$$g_{j\to i}^2 \langle \delta\mu_{j\to i}^2 \rangle = \frac{g_{ij}^2 \sum_{k\in\partial j\backslash i} g_{k\to j}^2 \langle \delta\mu_{k\to j}^2 \rangle}{(\sum_{k\in\partial j\backslash i} g_{k\to j} + g_{ij})^2}$$

$$\langle \delta\mu_j^2 \rangle = \frac{\langle \delta P_j^2 \rangle + \sum_{k\in\partial j\backslash i} g_{k\to j}^2 \langle \delta\mu_{k\to j}^2 \rangle}{(\sum_{k\in\partial j} g_{k\to j})^2}$$

# Message passing for network fluctuation

The fluctuation of network flow by message passing algorithm is

$$\langle \delta y_{j\to i}^2 \rangle = \frac{\langle \delta P_j^2 \rangle + \sum_{k\in\partial j\setminus i} \langle \delta y_{k\to j}^2 \rangle}{(1 + R_{ij} \sum_{k\in\partial j\setminus i} r_{k\to j}^2)^2}$$

$$\langle \delta y_{ij}^2 \rangle = \frac{r_{j\to i}^2 \langle \delta y_{j\to i}^2 \rangle + r_{i\to j}^2 \langle \delta y_{i\to j}^2 \rangle}{(r_{i\to j} + r_{j\to i} - R_{ij})^2}$$

# Message passing for network fluctuation

Testing the message passing algorithm for network flow fluctuation and potential fluctuation,Because neglect correlation between cavity network or cavity potential,it shows bad performance.

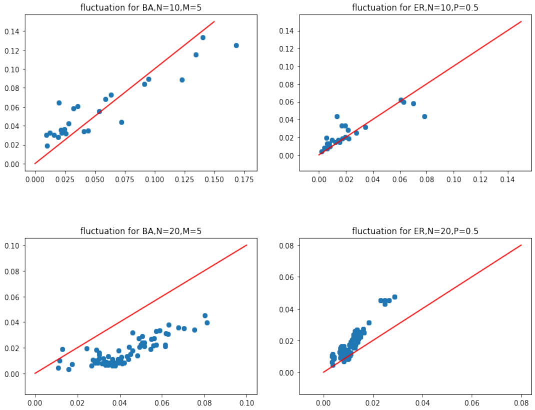



Figure: performance of message passing for network fluctuation in random network

# Message passing for network fluctuation

We construct a simple model include only one region to test how to improve the performance of network fluctuation by Generalize belief propagation method.

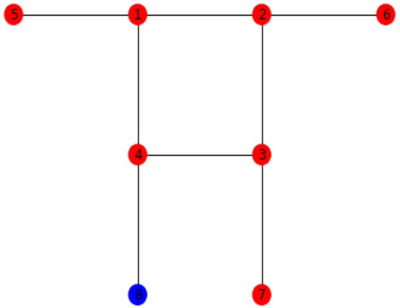



Figure: In our toy model,there is one region node,and node 8 is a base station, $\mu_8 = 0$

# Message passing for network fluctuation

Because we can not neglect the interaction of $\delta y_{j\to i}$, we use general Belief propagation to represent fluctuation of each node.In our toy model, set $S_i = \langle \delta P_i^2 \rangle$,

$$\langle \delta \mu_1^2 \rangle = k_1 S_1 + k_2 S_2 + k_s S_3 + k_4 S_4 + k_5 S_5 + k_6 S_6 + k_7 S_7 + k_8 S_8$$

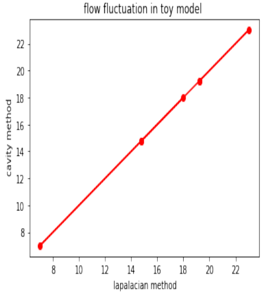



Figure: In our toy model, we get accurate network fluctuation by generalize belief propagation

# Open Problem

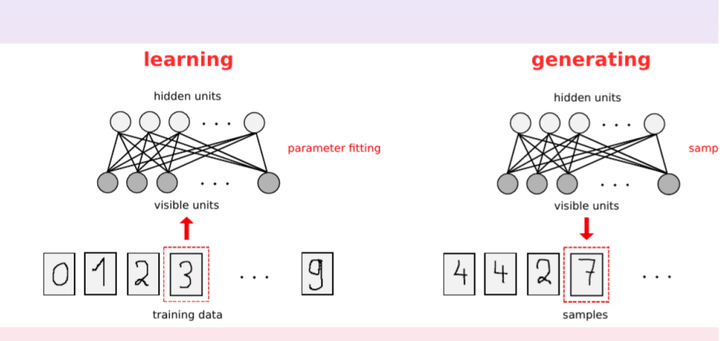
- Our Generalize belief propagation define in lattice network , we should explore more region interaction in random network ,ER network, BA network .etc.
- In our toy model, it gets accurate fluctuation prediction,how to apply this to large lattice network to get more accurate fluctuation will be explore.

# Outline

1. Belief propagation in 2D Ising model
2. Message passing in network flow problems
3. Message passing in RBM

# Background of RBM

Restricted Boltzmann machine (RBM) is a popular generative model in unsupervised learning. we can learn the feature ( internal representation ) $\xi$ from datasets $\{\sigma\}$,and then generative new pattern use the inferences $\hat{\xi}$.



# Background of RBM

In the RBM model,Given binary $\boldsymbol{h} \in \{0,1\}^M$,$\boldsymbol{\sigma} \in \{0,1\}^N$,$\boldsymbol{\xi} \in \{-1,1\}^{N\times M}$

$$E(\boldsymbol{h},\boldsymbol{v}) = -\boldsymbol{h}^T \xi \boldsymbol{\sigma} - \boldsymbol{b}^T \boldsymbol{h} - \boldsymbol{c}^T \boldsymbol{\sigma}$$

$$P(\boldsymbol{h},\boldsymbol{v}|W) = \frac{1}{Z} e^{-\beta E(\boldsymbol{h},\boldsymbol{v})}$$

## Remark

- $\beta$ is the inverse temperature ,represent the noise level of input
- $\boldsymbol{b}, \boldsymbol{c}$is the bias of visible $\boldsymbol{v}$,and hidden $\boldsymbol{h}$ as external filed in Ising model
- When $\boldsymbol{h}, \boldsymbol{\sigma}$is high dimension,partition function$Z$ is intractable!

# Training the RBM by message passing

For simplify,we use one feature $\boldsymbol{\xi}$, the feature we want to learn can be express as bayesian posterior:

$$P(\boldsymbol{\xi}|\{\boldsymbol{\sigma}^a\}) = \frac{\prod_{a=1}^{M} P(\boldsymbol{\sigma}^a|\boldsymbol{\xi})P(\boldsymbol{\xi})}{\prod_{a=1}^{M} P(\boldsymbol{\sigma}^a|\boldsymbol{\xi})P(\boldsymbol{\xi})} = \frac{1}{Z}\prod_{a=1}^{M} \cosh(\frac{\beta}{\sqrt{N}}\boldsymbol{\xi}^T\boldsymbol{\sigma}^a)$$

# Training the RBM by message passing

By consider each data $\boldsymbol{\sigma}^a$is a constraint to feature vector,cavity probability absence of data $\boldsymbol{\sigma}^a$,

$$P_{i\to a}(\xi_i) = \frac{1}{Z_{i\to a}} \prod_{b\in\partial i\backslash a} \mu_{b\to i}(\xi_i)$$

cavity probability only consider data$\boldsymbol{\sigma}^a$:

$$\mu_{b\to i}(\xi_i) = \sum_{\xi_j|j\in\partial b\backslash i} \cosh(\frac{\beta}{\sqrt{N}}\boldsymbol{\xi}\boldsymbol{\sigma}^b) \prod_{j\in\partial b\backslash i} p_{j\to b}(\xi_j)$$

# Training the RBM by message passing

By parameter cavity probability $P_{i\to a}(\xi_i)$ and $\mu_{b\to i}(\xi_i)$:

$$P_{j\to b}(\xi_j) = \frac{e^{\beta h_{j\to b}\xi}}{2\cosh \beta h_{j\to b}}$$

$$m_{i\to a} = \sum_{\xi_j} \xi_j P_{j\to b}(\xi_j)$$

$$\mu_{b\to i}(\xi_i) = \frac{e^{\beta u_{b\to i}\xi}}{2\cosh \beta u_{b\to i}}$$

We use the relaxed BP by only consider the first two moments of cavity probability by:

mean

$$G_{b\to i} = \frac{1}{\sqrt{N}} \sum_{j\in\partial b\backslash i} \sigma_j^b m_{j\to b}$$

variance

$$\mathcal{E}^2_{b\to i} = \frac{1}{N} \sum_{j\in\partial b\backslash i} (1 - m^2_{j\to b})$$

# Training the RBM by message passing

The message propagate in the RBM is :

$$m_{i\to a} = \tanh\Big(\sum_{b\in\partial i\backslash a} u_{b\to i}\Big)$$

$$u_{b\to i} = \arctan[\tanh(\beta G_{b\to i})\tanh(\frac{\beta}{\sqrt{N}}\sigma_i^b)]$$

When message passing converge in the RBM , we will learn the weight $\xi_i$

$$\hat{\xi}_i = \langle \xi_i \rangle = \tanh\Big(\sum_{a\in\partial i} u_{a\to i}\Big)$$

The overlap with original feature vector is defined as $q = \frac{1}{N}\sum_i \hat{\xi}_i \xi_i^{true}$ can measure the performance of the message passing algorithm.

# Training the RBM by message passing

Message passing algorithm propagate message in the M data $\boldsymbol{\sigma}^a$ and N feature node $\xi_i$, the time complexity is $\mathcal{O}(M \times N)$,By Thouless-Anderson-Palmer (TAP) equation(approximate message passing) , we can simplify the total messages only in feature nodes and data nodes themself.

$$G_a = \frac{1}{\sqrt{N}} \sum_{i \in \partial a} \sigma_i^a m_i - \beta(1 - \frac{1}{N} \sum_i m_i^2) \tanh \beta G_a$$

$$m_i = \tanh[\sum_{b \in \partial i} \frac{\beta \sigma_i^b}{\sqrt{N}} \tanh \beta G_b - \frac{\beta^2 m_i}{N} \sum_{b \in \partial i} (1 - \tanh(\beta G_b)^2)]$$

The time complexity of AMP(approximate message passing) In the RBM reduced to $\mathcal{O}(M + N)$.

# Replica symmetry method

In the large system $N \to \infty$, and $M \to \infty$ ,but $\alpha = \frac{M}{N} \propto \mathcal{O}(1)$, we can get the average free energy at the given feature distribution by replica method.

$$-\beta N\langle f\rangle = \lim_{n\to 0} \frac{\ln\langle Z^n\rangle}{n}$$

Use free entropy , we can get entropy of this system as

$$s = (1 - \beta\frac{\partial}{\partial\beta})f$$

# Replica symmetry method

In the replica framework, we can define another overlap $r = \frac{1}{N}\sum_i \langle \xi_i^a \xi_i^b \rangle$.The average free energy is determined by overlap $q, r$.

$$q = \int \mathcal{D}z \tanh(\hat{q} + \sqrt{\hat{r}}z)$$

$$r = \int \mathcal{D}z \tanh(\hat{q} + \sqrt{\hat{r}}z)^2$$

$$\hat{q} = \alpha\beta^2 e^{-\frac{1}{2}\beta^2} \int \mathcal{D}z \int \mathcal{D}y \sinh(\beta t) \tanh \beta(qt + \sqrt{r - q^2}y)$$

$$\hat{r} = \alpha\beta^2 e^{-\frac{1}{2}\beta^2} \int \mathcal{D}z \int \mathcal{D}y \cosh(\beta t) \tanh \beta(qt + \sqrt{r - q^2}y)^2$$

# Replica symmetry method

If $\xi^{true}$ and one inference feature $\xi^{a}$ is generated from some temperature $\beta$, the overlap $q, r$is same.

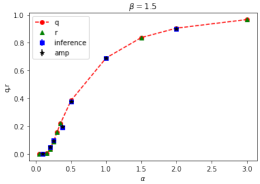



Figure: MP and AMP algorithm results $q, r$ vs prediction of replica method

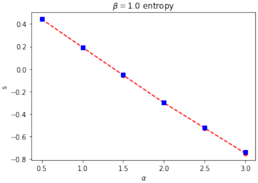



Figure: MP results of entropy $\mathcal{S}$ vs replica method result

# Training the RBM by message passing

## Open problem

- The entropy $S < 0$ when $\alpha$ is large means the solution has more complicated Structure,replica symmetry(RS) framework doesn't match ,it should be explored by replica symmetry breaking (RSB).
- if datasets are generated by different model,how to design heuristic algorithm to get more accurate feature $\boldsymbol{\xi}$.
- If datasets has two feature $\boldsymbol{\xi}^1$ and $\boldsymbol{\xi}^2$, how the relation of overlap of these feature ($q^{12} = \frac{1}{N}\sum_i \xi_i^1 \xi_i^2$) influence the learning process, we expect disentangle features will improve learning process.

Thanks for your attention!