Mean Field Theory

Solmaz Torabi

Dept. of Electrical and Computer Engineering
Drexel University
st669@drexel.edu

Advisor: Dr. John M. Walsh

Refrences

G. Parisi, "On the replica approach to glasses," *arXiv preprint*, 1997.

V. Dotsenko, "Introduction to the replica theory of the disordered statistical systems." *Cambridge University Press*, 2005.

L. Csat, M. Opper, and O. Winther, "Tap gibbs free energy, belief propagation and sparsity," *Advances in Neural Information Processing Systems*, 2001.

M. Opper and D. Saad, "Advanced mean field methods: Theory and practice," *MIT press*, 2001.

[1] [2] [3] [4]

Outline

- Some thermodynamic quantities
- Mean field approximation
- TAP (Thouless, Anderson and Palmer)
  - Cavity approach
  - Plefka expansion
- Self averaging property
- Replica method

Thermodynamic quantities

In statistical physics:

- we consider N discrete variables σ_i
- The cost function $H(\sigma)$ is called Hamiltonian.
- T is the temperature and $\beta = \frac{1}{T}$ is the inverse temperature.
- Each configuration or assignments of the variable has a weight $e^{-\beta H(\sigma)}$ (Boltzmann weight)
- The partition function is defined as

$$Z(\beta) = \sum_{\{\sigma\}} e^{-\beta H(\sigma)}$$

Thermodynamic quantities

- Average of any quantities $A(\sigma)$ that depends on $(\{\sigma\})$ with respect to the Boltzmann measure is

$$\langle A(\beta)\rangle = \sum_{\{\sigma\}} A(\sigma)\frac{e^{-\beta H(\sigma)}}{Z}$$

- The average of Hamiltonian itself is called the energy(internal energy):

$$E(\beta) = \langle H\rangle = \sum_{\{\sigma\}} H(\sigma)\frac{e^{-\beta H(\sigma)}}{Z}$$

- An unusable energy is given by the entropy.

$$S(\beta) = -\sum P(\sigma)\ln P(\sigma)$$

- Free energy: determines the amount of energy 'free' for work

$$F(\beta) = E(\beta) - S(\beta) = -\frac{1}{\beta}\log Z(\beta)$$

Third law of thermodynamics

- Notice: The zero temperature limit (when $\beta \to \infty$) of the energy give the value of the minimum of the Hamiltonian (which is called the ground state energy in physics)

$$\lim_{\beta \to \infty} E(\beta) = \min_{\{\sigma\}} H(\sigma)$$

- The entropy at zero temperature give us the (logarithm of the) number of configuration Ω of $\{\sigma\}$:

$$\lim_{\beta \to \infty} S(\beta) = \log \Omega$$

Connection with combinatorial optimization

- finding the minimal value of a function of discrete variable, and the corresponding number of assignment

Statistical Physics interpretation:

-

- computing the ground-state energy (the zero temperature energy) and entropy of the associated statistical physics model!

If we know how to solve the statistical physics model, we know the value of the best assignment

Example: Coloring problem

- We need to color a graph with q colors such that two neighboring nodes have a different color.

Let's make connection to statistical physics

- 1. The problem is equivalent to finding the ground state of anti-ferromagnet on a graph, a problem with Hamiltonian :

$$H = \sum_{i,j \in \mathcal{G}} J \sigma_i \sigma_j$$

- 2. Solve statistical problem
- 3. Configuration with zero energy is a proper coloring
- 4. If the zero temperature energy is zero, the graph is colorable.
- 5. The entropy allows to find the number of colors.

Ising model

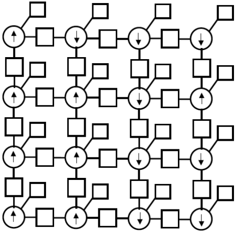
- An example of pairwise Markov random field with $\psi_{ij}(x_i, x_j) = \exp\{J_{ij}x_i x_j\}$, $\psi_i(x_i) = \exp\{\theta_i x_i\}$
- is a mathematical model of ferromagnetism in statistical mechanics.
- x_i represents magnetic dipole moments of atomic spins, $x_i \in \{+1, -1\}$, any two adjacent sites i, j has an interaction J_{ij}
- each site i has an external magnetic field θ_i

- The energy for each configuration is

$$H(\mathbf{X}) = -\sum_{i,j} J_{ij}x_i x_j - \sum_i \theta_i x_i$$

- The configuration probability is

$$P(\mathbf{X}) = \frac{e^{-\beta H(\mathbf{X})}}{Z} = \frac{e^{-\beta \sum_{i,j} J_{ij}x_i x_j - \beta \sum_i \theta_i x_i}}{Z}$$



What is spin glass?

- In spin-glasses, disorder is explicitly present in the Hamiltonian, typically under the form of random couplings J among the degrees of freedom X
- Disorder J is specified by it's probability distribution

example Sherrington and Kirkpatrick model (SK)

$$H(X) = -\sum_{i,j} J_{ij} x_i x_j$$

- spins $x_i = +1, -1$ are the degrees of freedom
- Couplings between neighboring spins follow a Gaussian distribution

Objective:

$$F = \frac{1}{\beta} \log Z$$

where

$$Z = \sum_x e^{-\beta H(x,J)}$$

- 1. Replica theory: typical macroscopic properties averaged over the quenched randomness

$$f = \lim_{n \to 0} -\frac{1}{\beta n} \langle \log Z^n \rangle_J$$

- 2. Cavity method: Correction of Naive mean field, by subtracting the self induced filed
- 3. TAP method: enables us to compute the thermal average of the dynamic variables for a given realization.

Outline

- Some thermodynamic quantities
- Mean field approximation
- TAP (Thouless, Anderson and Palmer)
  - Cavity approach
  - Plefka expansion
- Self averaging property
- Replica method
- Applications of statistical physics in information theory
  - Error correcting code
  - Lossy compression

Variational free energy

- Variational method approximates an intractable distribution $P(\mathbf{X})$ of random variables $\mathbf{X} = (S_1, ..., S_N)$ by a tractable distribution $Q(\mathbf{X})$
- Q is chosen to minimize certain distance measure.

$$KL(Q||P) = \sum_{\mathbf{X}} Q(\mathbf{X}) \ln \frac{Q(\mathbf{X})}{P(\mathbf{X})} = \left\langle \ln \frac{Q}{P} \right\rangle_Q$$

where $\langle . \rangle_Q$ denotes the expectation with respect to Q

Variational free energy

To find the best approximate to $P = \frac{e^{-H(\mathbf{X})}}{Z}$

$$KL(Q||P) = \ln Z + E[Q] - S[Q]$$

where

- $S[Q] = -\sum_{\mathbf{X}} Q(\mathbf{X}) \ln Q(\mathbf{X})$ is the entropy of Q
- $E[Q] = \sum_{\mathbf{X}} Q(\mathbf{X}) H[\mathbf{X}]$ is called average energy

$$\Longrightarrow \min_Q KL(Q||P) = \ln Z + \min_Q \underbrace{(E[Q] - S[Q])}_{\text{Variational free energy}}$$

Variational free energy for Ising model

- The model under consideration is a Boltzmann machine.

$$P(\mathbf{X}) = \frac{e^{-H(\mathbf{X})}}{Z} = \frac{e^{-\sum\limits_{i,j} J_{ij}x_ix_j - \sum\limits_i \theta_i x_i}}{Z}$$

- For binary variable it is convenient to reparametrize these marginals as follows,

$$m_j = \langle x_j \rangle_Q \ \Rightarrow Q_j(x_j) = \frac{1 + x_j m_j}{2}$$

Mean Field approximation

Find a factorized distribution that best describes the true distribution.

- For binary variable the most general factorized distribution has the form.

$$Q^{MF}(x) = \prod_i Q_i(x_i) = \prod_i \frac{(1 + x_i m_i)}{2}$$

- $KL(Q^{MF}||P) = E(Q^{MF}) - S(Q^{MF}) + \log(Z)$
- $E(Q^{MF}) = \sum Q^{MF} H(x) = -\sum_{ij} J_{ij} m_i m_j$
- $S(Q^{MF}) = -\sum_i Q^{MF} \ln Q^{MF} = -\sum_i \left(\frac{1+m_i}{2} \ln \frac{1+m_i}{2} + \frac{1-m_i}{2} \ln \frac{1-m_i}{2} \right)$

Mean Field approximation

$$\min_{m_i} KL(Q^{MF}||P)$$

- take derivative with respect to m_i
- $\frac{\partial}{\partial m_i}\Big\{ -\sum_{ij} J_{ij} m_i m_j + \sum_i \frac{1+m_i}{2} \ln \frac{1+m_i}{2} + \frac{1-m_i}{2} \ln \frac{1-m_i}{2} + \log(Z) \Big\}$

Mean Field fixed points

$$\frac{\partial KL}{\partial m_i} = -\sum_{j \in N(i)} J_{ij} m_j + \log\left(\frac{m_i}{1-m_i}\right) = 0$$

- Fixed points of MF approximation:

$$m_i = \frac{\exp(\sum_j J_{ij} m_j) - \exp(-\sum_j J_{ij} m_j)}{\exp(\sum_j J_{ij} m_j) + \exp(-\sum_j J_{ij} m_j)}$$

$$\Rightarrow \boxed{m_i = \tanh(\sum_j J_{ij} m_j), \qquad i = 1, \dots, N}$$

Mean Field

Naive approach:

$$\langle X_i \rangle = \Big\langle \tanh\big(\sum_j J_{ij} X_j\big) \Big\rangle_P, \qquad i = 1, ..., N$$

- In Naive approach the expectation is with respect to the difficult P.
- The expectation is outside the nonlinear function

Mean field approach :

$$\langle X_i \rangle = \tanh\big(\sum_j J_{ij} \langle X_j \rangle_Q\big), \qquad i = 1, ..., N$$

- MF replaces the field $h_i = \sum_j J_{ij} X_j$ by it's mean field

Mean Field

$$m_i = \tanh(\sum_j J_{ij} m_j), \qquad i = 1, ..., N \tag{1}$$

Note

- The intractable task of computing marginals has been replaced by the problem of solving a set of nonlinear equations.

Mean Field

$$m_i = \tanh(\sum_j J_{ij} m_j), \quad i = 1, ..., N \tag{1}$$

Note

- The intractable task of computing marginals has been replaced by the problem of solving a set of nonlinear equations.
- These MF equations are run sequentially, i.e. we fix all m_j except m_i.

Mean Field

$$m_i = \tanh(\sum_j J_{ij} m_j), \qquad i = 1, ..., N \tag{1}$$

Note

- The intractable task of computing marginals has been replaced by the problem of solving a set of nonlinear equations.
- These MF equations are run sequentially, i.e. we fix all m_j except m_i.
- In each step MF free energy is convex. Equation (1) finds minimum in one step.

Mean Field

$$m_i = \tanh(\sum_j J_{ij} m_j), \qquad i = 1, ..., N \tag{1}$$

Note

- The intractable task of computing marginals has been replaced by the problem of solving a set of nonlinear equations.
- These MF equations are run sequentially, i.e. we fix all m_j except m_i.
- In each step MF free energy is convex. Equation (1) finds minimum in one step.
- This procedure can be interpreted as coordinate descent in the m_i

Mean Field

$$m_i = \tanh(\sum_j J_{ij} m_j), \qquad i = 1, ..., N \tag{1}$$

Note

- The intractable task of computing marginals has been replaced by the problem of solving a set of nonlinear equations.
- These MF equations are run sequentially, i.e. we fix all m_j except m_i.
- In each step MF free energy is convex. Equation (1) finds minimum in one step.
- This procedure can be interpreted as coordinate descent in the m_i
- Alternatively, all parameters m_i can be updated in parallel.

Mean Field

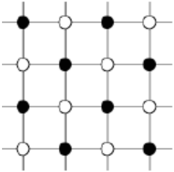
$$m_i = \tanh(\sum_j J_{ij} m_j), \qquad i = 1, ..., N \tag{1}$$

Note

- The intractable task of computing marginals has been replaced by the problem of solving a set of nonlinear equations.
- These MF equations are run sequentially, i.e. we fix all m_j except m_i.
- In each step MF free energy is convex. Equation (1) finds minimum in one step.
- This procedure can be interpreted as coordinate descent in the m_i
- Alternatively, all parameters m_i can be updated in parallel.
- Doesn't guarantee of decreasing the cost function at each iteration.
  - There might be many solutions to (1).
  - Some of the solutions may not be local minima

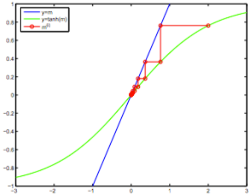
Mean Field

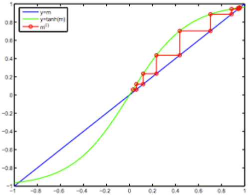
- In d-dimensional Ising model without the external magnetic field ($\theta = 0$) and having the same interaction $J_{ij} = \alpha$



$$m^{(t+1)} = \tanh(2d\alpha m^{(t)})$$

- For $\alpha < \frac{1}{2d}$, the iteration converges to $\lim_{t \to \infty} m^{(t)} = 0$ (left figure)
- For $\alpha > \frac{1}{2d}$, if $m^{(0)} \lessgtr 0 \Rightarrow \lim_{t \to \infty} m^{(t)} = \mp m^*$





[4]A. Montanari, Lecture notes for inference in graphical models,2011

Mean Field

- MF neglects the dependency between the random variables.

However,

- We get an upper bound on the exact free energy.

$$KL(Q^{MF}||P) = \underbrace{E(Q^{MF}) - S(Q^{MF})}_{=F[Q^{MF}]\text{ Variational MF energy}} - \underbrace{(-\log(Z))}_{\text{Exact free energy}}$$

Since $KL(Q^{MF}||P) \geq 0$

$$F(Q^{MF}) \geq -\log(Z)$$

Mean Field Method in general

- $P(x) = \frac{1}{Z} \prod_{a \in F} f_a(x_a)$ is True distribution
- $Q(x) = \prod_i q_i(x_i)$ is Approximate distribution

$$F^{MF}(Q) = \sum_i S(q_i) + \sum_{a \in F} \sum_{x_a} \prod_{x_i \in N(a)} q_i(x_i) \log f_a(x_a)$$

- We passed from $(|\mathcal{X}|^n - 1)$ to $n(|\mathcal{X}| - 1)$
- F^{MF} is no longer convex.

$$\min_Q F^{MF}(Q) \text{ subject to } \sum_{x_i} q_i(x_i) = 1$$

Mean Field Method in general

- Add Lagrange multiplier λ_i
- Find the stationary condition by $\frac{\partial L(Q,\lambda)}{\partial q_i(x_i)} = 0$

$$q_i(x_i) \propto \prod_{a \in N(i)} m_{a \to i}(x_i)$$

where

$$m_{a \to i}(x_i) = \exp\left(\sum_{x_j : j \in N(a) \setminus i} \log f_a(x_a) \prod_{j \in N(a) \setminus i} q_j(x_j) \right)$$

- A simple greedy algorithm for finding a stationary point consists in updating the q by iterating the above equations until convergence.

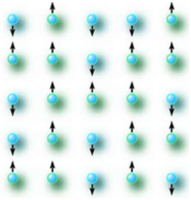
Outline

- Some thermodynamic quantities
- Mean field approximation
- TAP (Thouless, Anderson and Palmer)
  - Cavity approach
  - Plefka expansion
- Self averaging property
- Replica method
- Applications of statistical physics in information theory
  - Error correcting code
  - Lossy compression

TAP approximation

- Is derived after D.J. Thouless, P.W. Anderson and R. G. Palmer.
- Is derived for SK model,
  - Coupling J_{ij} are independent
  - Gaussian random variables for $i < j$ with variance J_0/N

- Cavity method
  (introduced by Parisi and Mezard ,1987)
- Plefka's expansion
  (introduced by Plefka, 1982)



Plefka Expansion

magnetization dependent Gibbs free energy by adding a set of external auxiliary field

$$G(\mathbf{m}) = \min_Q \{F[Q] \mid \langle \mathbf{X} \rangle_Q = \mathbf{m}\}$$

Define one parameter family of models

$$p_t(X) \propto \exp\left[t \sum_{ij} X_i X_j J_{ij} + \sum_i X_i \theta_i\right]$$

perturbative approach (Plefka): expand $G_t(m)$ to $\mathcal{O}(t^2)$ yields TAP equations.

Plefka Expansion

- Don't restrict the approximate distribution Q to be product distributions
- Minimize free energy in two steps:
  - Constrained minimization in the family of distributions satisfying $\langle \mathbf{X} \rangle_Q = \mathbf{m}$ for fixed $\mathbf{m}$

$$G(\mathbf{m}) = \min_Q \{ F[Q] = E[Q] - S[Q] \mid \langle \mathbf{X} \rangle_Q = \mathbf{m} \}$$

  - Minimize $G(\mathbf{m})$ with respect to $\mathbf{m}$

Plefka Expansion

$$G(\mathbf{m}) = \min_{Q}\{F[Q] \mid \langle \mathbf{X} \rangle_Q = \mathbf{m}\}$$

By adding Lagrange multiplier λ
Then Lagrangian

$$G(\mathbf{m}, \lambda) = E[Q] - S[Q] - \sum_i \lambda_i(\langle x_i \rangle_Q - m_i)$$

$$G(\mathbf{m}, \lambda) = \sum_{\mathbf{X}} Q(\mathbf{X})H[\mathbf{X}] - S[Q] - \sum_{\mathbf{x}} \sum_i \lambda_i x_i Q(\mathbf{X}) + \sum_i \lambda_i m_i$$

is the form of variational free energy, where $H[\mathbf{X}]$ is replaced by $H[\mathbf{X}] - \sum_i \lambda_i x_i$. We can construct such a gibbs free energy by adding a set of external auxiliary field.
$\Rightarrow Q_\lambda(\mathbf{X}) = \frac{1}{Z} e^{-H[\mathbf{X}] + \sum_i \lambda_i x_i}$

Plefka Expansion

The dual function is,

$$G(m_i) = \max_{\lambda_i} \left\{ \sum_i \lambda_i m_i - \log(Z(\lambda_i)) \right\}$$

- This equation known as Legendre transform between $\{\lambda_i\}$ and $\{m_i\}$.
- $Z(\lambda_i)$ is the normalizing constant for the Gibbs distribution

$$Q_\lambda(\mathbf{X}) = \frac{1}{Z_{\lambda_i}} e^{-H[\mathbf{X}] + \sum_i \lambda_i x_i} = \frac{1}{Z_{\lambda_i}} e^{-\sum_{i,j} J_{ij} x_i x_j - \sum_i \theta_i x_i + \sum_i \lambda_i x_i}$$

- Set $\theta \to 0$ by shifting the Lagrange multiplier $\lambda_i \to \lambda_i - \theta_i$
- $Z(\lambda_i) = \sum_{x_i} \exp(-\sum_{i,j} J_{ij} x_i x_j + \sum_i \lambda_i x_i)$

Plefka Expansion

$$G(m_i) = \max_{\lambda_i} \Big\{ \sum_i \lambda_i m_i - \log(\sum_{x_i} \exp(-\sum_{i,j} \beta J_{ij} x_i x_j + \sum_i \lambda_i x_i)) \Big\}$$

- Plefka expansion is derived by $J_{ij} \to \beta J_{ij}$, by Taylor expanding the Gibbs free energy around $\beta = 0$, where β is an inverse temperature in physics,

Notice

- For each term in Taylor expansion, one has to expand the Lagrange multiplier λ_i which maximize the Gibbs distribution as well as $\log(Z)$
- The auxiliary field is temperature dependent.

Plefka Expansion

- with $G_n = \frac{\partial^n}{\partial \beta^n} G(\mathbf{m})|_{\beta=0}$

$$G(\mathbf{m}) = G_0(\mathbf{m}) + \beta G_1(\mathbf{m}) + \frac{\beta^2}{2!} G_2(\mathbf{m}) + \ldots$$

Plefka Expansion

- with $G_n = \frac{\partial^n}{\partial \beta^n} G(\mathbf{m})|_{\beta=0}$

$$G(\mathbf{m}) = G_0(\mathbf{m}) + \beta G_1(\mathbf{m}) + \frac{\beta^2}{2!} G_2(\mathbf{m}) + \ldots$$

- $G_0(\mathbf{m}) = \sum_i \left\{ \frac{1+m_i}{2} \ln \frac{1+m_i}{2} + \frac{1-m_i}{2} \ln \frac{1-m_i}{2} \right\}$ Spins are entirely controlled by the auxiliary field.

Plefka Expansion

- with $G_n = \frac{\partial^n}{\partial \beta^n} G(\mathbf{m})|_{\beta=0}$

$$G(\mathbf{m}) = G_0(\mathbf{m}) + \beta G_1(\mathbf{m}) + \frac{\beta^2}{2!} G_2(\mathbf{m}) + ...$$

- $G_0(\mathbf{m}) = \sum_i \left\{ \frac{1+m_i}{2} \ln \frac{1+m_i}{2} + \frac{1-m_i}{2} \ln \frac{1-m_i}{2} \right\}$ Spins are entirely controlled by the auxiliary field.
- $G_1(\mathbf{m}) = -\sum_{i<j} J_{ij} m_i m_j$

Plefka Expansion

- with $G_n = \frac{\partial^n}{\partial \beta^n} G(\mathbf{m})|_{\beta=0}$

$$G(\mathbf{m}) = G_0(\mathbf{m}) + \beta G_1(\mathbf{m}) + \frac{\beta^2}{2!} G_2(\mathbf{m}) + \dots$$

- $G_0(\mathbf{m}) = \sum_i \left\{ \frac{1+m_i}{2} \ln \frac{1+m_i}{2} + \frac{1-m_i}{2} \ln \frac{1-m_i}{2} \right\}$ Spins are entirely controlled by the auxiliary field.
- $G_1(\mathbf{m}) = -\sum_{i<j} J_{ij} m_i m_j$
- $G_2(\mathbf{m}) = -\frac{1}{2} \sum_{ij} J_{ij}^2 (1 - m_i^2)(1 - m_j^2)$
- ...

Plefka Expansion

with $G_n = \frac{\partial^n}{\partial \beta^n} G(\mathbf{m})|_{\beta=0}$

$$G(\mathbf{m}) = G_0(\mathbf{m}) + \beta G_1(\mathbf{m}) + \frac{\beta^2}{2!} G_2(\mathbf{m}) + \ldots$$

- $G_0 = \sum_i \left\{ \frac{1+m_i}{2} \ln \frac{1+m_i}{2} + \frac{1-m_i}{2} \ln \frac{1-m_i}{2} \right\}$ $\Rightarrow$ MF variational entropy
- $G_1(\mathbf{m}) = -\sum_{i<j} J_{ij} m_i m_j$ $\Rightarrow$ MF variational energy
- $G_2(\mathbf{m}) = -\frac{1}{2} \sum_{ij} J_{ij}^2 (1 - m_i^2)(1 - m_j^2)$
- ...$\Rightarrow$ Takes into account the higher order dependencies

TAP approximation

TAP approximation= Minimizing $G(m)$ for $\beta = 1$ and keeping only terms up to second order

$$G^{TAP}(m_i) = -\sum_{(ij)} J_{ij} m_i m_j + \sum_i \left\{ \frac{1+m_i}{2} \ln \frac{1+m_i}{2} + \frac{1-m_i}{2} \ln \frac{1-m_i}{2} \right\}$$
$$\underbrace{- 1/2 \sum_{(ij)} J_{ij}^2 (1-m_i^2)(1-m_j^2)}_{\text{dependencies between rvs}}$$

- TAP takes in to account the dependencies between random variables.
- It's exact in the high temperature for certain classes of models (SK models).

TAP approximation

Fixed points of TAP approximation:

$$\langle x_i \rangle = \tanh\Big(\sum_{j \in N(i)} J_{ij} \langle x_j \rangle - \langle x_i \rangle \sum_{j \in N(i)} J_{ij}^2 (1 - \langle x_j \rangle^2) \Big)$$

- Running these equations doesn't guarantee that TAP-Gibbs free energy decreases. (m_i appears on both sides)
- There is danger that radius of convergence (of taylor expansion) will be too small to obtain result for values of β we are interested in.

TAP approximation

TAP equations:

- Sherrington-Kirkpatrick model for N Ising spins $X_i = +1, -1$ with random coupling $J_{i,j} = \mathcal{N}(0, 1/N)$

$$P(X) = \exp(\sum_{ij} X_i J_{ij} X_j + \sum_i X_i \theta_i)$$

- Mean field correction equations (TAP equations)

$$\langle X_i \rangle = \tanh\left(\sum_j J_{ij} \langle X_j \rangle - \langle X_i \rangle \sum_j J_{ij}^2 (1 - \langle X_j \rangle^2) + \theta_i\right)$$

TAP approximation

Cavity approach

TAP approximation -Cavity approach

$$P(X) \propto \exp\left(x_i \sum_{j \in N(i)} J_{ij} x_j\right) P_{\setminus i}(x \setminus i)$$

The only dependance between x_i and other variables is through the field $h_i = \sum_j J_{ij} x_j$

$$p_i(x_i, h_i) \propto \exp\left(x_i h_i\right) P_{\setminus i}(h_i)$$

TAP approximation -Cavity approach

'Cavity' distribution: (in physical context: h_i is the magnetic filed at the cavity which is left when spin i is removed from the system

$$P_{\setminus i}(h_i) = \sum_{X_{\mathcal{N}(i)}} \delta\Big(h_i - \sum_j J_{ij} X_j\Big) P_{\setminus i}(X_{\mathcal{N}}(i))$$

Approximate $p_{\setminus i}(h_i)$ by Gaussian (central limit theorem)

$$P_{\setminus i}(h_i) \propto \exp\Big(-\frac{(h_i - a_i)^2}{2V_i}\Big)$$

$\Rightarrow$

$$p_i(x_i) = \int p_i(x_i, h_i)\, dh_i \ \propto \int e^{x_i h_i} p_{\setminus i}(h_i)\, dh_i \ \propto \exp\Big(a_i x_i + \frac{V_i}{2} x_i^2\Big)$$

TAP approximation -Cavity approach

$$p_i(x_i) = \exp\left(a_i x_i + \frac{V_i}{2} x_i^2\right)$$

Task is to find the mean and variance of the distribution

$$\langle h_i \rangle = \int dx \int p_i(x_i, h_i) h_i \, dh_i$$

$$\langle h_i \rangle = \frac{1}{Z_i} \int dx \int e^{x_i h_i} p_{\setminus i}(h_i) h_i \, dh_i = a_i + V_i \langle x_i \rangle$$

$$\boxed{\langle h_i \rangle = \langle h_i \rangle_{\setminus i} + V_i \langle x_i \rangle}$$

TAP

Weak dependancies of the variables x_j, lead us to neglecting the non diagonal term.

$$V_i = \sum_{jk} J_{ij} J_{ik} \Big(\langle x_j x_k \rangle_{\setminus i} - \langle x_j \rangle_{\setminus i} \langle x_k \rangle_{\setminus i} \Big)$$
$$\approx \sum_{j} J_{ij}^2 (\langle x_j^2 \rangle_{\setminus i} - \langle x_j \rangle_{\setminus i}^2)$$
$$\approx \sum_{j} J_{ij}^2 (\langle x_j^2 \rangle - \langle x_j \rangle^2)$$

since J_{ij} is modeled as an independent rvs, we neglect the non-diagonal term $j \neq k$

Application

$$\boxed{m_i = \tanh\left(\sum_j J_{ij}\langle x_j\rangle - V_i\langle x_i\rangle\right)} \Rightarrow \text{TAP fixed points}$$

$$\boxed{m_i = \tanh\left(\sum_j J_{ij}\langle x_j\rangle\right)} \Rightarrow \text{Mean Field fixed points}$$

- TAP makes weak correlation assumption when computing the distribution of the cavity field h_i , for the case when X_i is disconnected
- Takes in to account the reaction of the neighbors X_j due to the correlations created by the presence of X_i.
- nontrivial dependencies by estimating the reaction of all other rv when a single variable is deleted from the system.

Adaptive TAP Equations by Opper/Winther

- TAP was developed to treat SK model of disordered magnetic materials. Couplings are assumed to be random with certain classes of distributions.
- ADATAP does not assume a specific randomness of the coupling, but rather adapts to the concrete data.

ADAPTIVE TAP

ADAPT Considers slightly more general class of models

$$P(X) = \frac{1}{Z} \prod_j f_j(x_j) \exp\Big(\sum_{ij} X_i J_{ij} X_j \Big)$$

allows for latent Gaussian models but also discrete variables (spins) by taking

$$f_j(x_j) = e^{\theta_j x_j} (\delta(x_j - 1) + \delta(x_j + 1))$$

x_j become Ising spin variables and the model is of the Boltzamann machine type.

ADATAP

The distribution $P(X) = \frac{1}{Z} \prod_j f_j(x_j) e^{\sum_{ij} X_i J_{ij} X_j}$ is approximated by $Q(X)$

- Mean field:

$$Q_i(X_i) = \frac{f_i(X_i)}{Z_0^{(i)}} e^{X_i(\sum_j J_{ij} \langle X_j \rangle_Q)}$$

- TAP approximating the variance by neglecting the non-dioagonal terms

$$Q_i(X_i) = \frac{f_i(X_i)}{Z_0^{(i)}} e^{X_i(\langle h_i \rangle_{\setminus i}) + \frac{V_i}{2} X_i^2}$$

- ADATAP computes the covariance V_i without explicit knowledge of the distribution of J_{ij}

Replica Method

The replica method

$$Z_J = \sum_{\sigma} \exp(-\beta H[J,\sigma])$$

- Consider the *integer* power n of the partition function.

$$Z_J^n = \Big(\prod_a \sum_{\sigma^a}\Big) \exp\Big(-\beta \sum_{a=1} H[J,\sigma_a]\Big)$$

- Z_J^n is the partition function of n non-interacting identical replicas of the original system
- The free energy of this system is:

$$F_n = -\frac{1}{\beta n} \ln \overline{(Z_J^n)}$$

- By taking the limit $n \to 0$ the original free energy can be recovered

$$\lim_{n\to 0} F_n = -\lim_{n\to 0} \frac{1}{\beta n} \ln \overline{(Z_J^n)} = -\lim_{n\to 0} \frac{1}{\beta n} \ln \overline{[\exp(n \ln Z_J)]} = -\frac{1}{\beta} \overline{\ln Z_J} = F$$

Replica method

$$x^n = e^{n \ln x}$$

for $n \to 0$

$$e^{n \ln x} = 1 + n \ln x \Rightarrow \boxed{\ln x = \lim_{n \to 0} \frac{x^n - 1}{n}}$$

$$\overline{\ln(Z_J)} = \lim_{n \to 0} \frac{\overline{Z_J^n - 1}}{n} = \lim_{n \to 0} \frac{\overline{Z_J^n} - 1}{n}$$

Using a well known approximation

$$nx = \ln(1 + nx) \ \text{ for } \ nx << 1$$

$$\boxed{\overline{\ln(Z_J)} = \lim_{n \to 0} \frac{1}{n} \ln \overline{Z_J^n}}$$

Thank you

Questions?