# FEDERICO RICCI-TERSENGHI SPARSE GRAPHS, BETHE APPROXIMATION, RS/RSB -

18/07/2017

Main reference: Mézard, Montanari: "Information, Physics, and Computation"

CHAPTERS 14, (22), 18, 19 → concise commented list of references at the end of each chapter -

## STATEMENT OF THE PROBLEM:

Let's consider the probability distribution (graphical model)

$$p(\underline{x}) = \frac{1}{Z} \prod_{i=1}^{N} \psi_i(x_i) \prod_{a=1}^{M} \psi_a(\underline{x}_{\partial a}) \quad \text{over} \quad \underline{x} = \{x_i\}_{i=1\dots N}$$

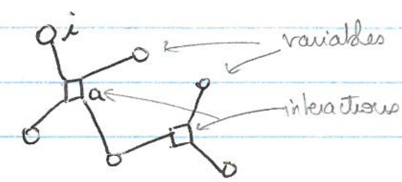
$\psi_a$: compatibility function

$\partial i$ = neighbors of $i$.

notation $\underline{x}_A = \{x_i \, ; \, i \in A\}$; $\partial a$ = neighbors of $a$

e.g. $\begin{cases} \psi_i(x_i) = e^{\beta H_i x_i} \\ \psi_a(x_a) = e^{\beta J_a x_{i_1} \dots x_{i_k}} \end{cases}$ (SPIN MODEL).

↳ graphical representation




We'll focus on sparse models: $\begin{cases} |\partial a| = O(1) = K & \to \text{regular} \\ M = \alpha N \quad \text{with } \alpha = O(1) & \Rightarrow \text{relation } \langle|\partial i|\rangle N = \langle|\partial a|\rangle M \end{cases}$

↳ This formalism contains CSPs:

e. $\psi_a(\underline{x}_{\partial a}) = \begin{cases} 0 & \text{allowed} \\ 1 & \text{broken} \end{cases}$ → zero temperature limit: $\lim_{\beta \to \infty} \psi_a \sim e^{-\beta E_a}$.

This models are hard to sample from for $N \to \infty$. Nevertheless we would like to be able to compute efficiently:

1) marginals / correlations $p_i(x_i) = \sum_{\underline{x} \setminus x_i} p(\underline{x})$ (exponential time in $N$ if exactly)

2) free entropy $\Phi = \overline{\ln Z}$

## THE BETHE APPROACH:

Mean field approximation that allows to factorize locally at least the complicated probability distribution

→ NAIVE MF: $p(\underline{x}) = \prod_{i=1}^{N} b_i(x_i)$ (belief)

→ belief parametrized, e.g. $b_i(x_i) = \frac{1 + m_i x_i}{2}$

↘ best self consistent guess for parameters

≡ minimize the Gibbs free energy

$$F_{Gibbs}[p] = -\beta \langle H(\underline{x}) \rangle_p - \langle \ln p \rangle_p$$

$p$: probability dist over the variables

$$\arg\min_p F_{Gibbs}(p) = \frac{e^{-\beta H(x)}}{Z} = p_{GB}$$

↳ $F_{Gibbs}(p_{GB}) = F = \ln Z$ -

Bethe $\xrightarrow{\text{dense limit}}$ TAP



→ THE BETHE APPROXIMATION: first correlations to take into account, is among the variables entering same interactions

$$p(\underline{x}) \approx \prod_{i=1}^{N} b_i(x_i) \prod_{a=1}^{M} \frac{b_a(\underline{x}_{\partial a})}{\prod_{i\in\partial a} b_i(x_i)} = \prod_{a=1}^{M} b_a(\underline{x}_{\partial a}) \prod_i b_i(x_i)^{1-|\partial_i|}$$

normalized (on trees)

N+M meanfield parameters > N for MF ← e.g. Ising pairwise $K=2$: $\begin{cases} b_i(x_i) = \frac{1+m_i x_i}{2} & \Rightarrow N \text{ param} \\ b_{ij}(x_i x_j) = \frac{1}{4}\left(1 + m_i x_i + m_j x_j + (C_{ij} + m_i m_j) x_i x_j\right) & \Rightarrow M \text{ param} \end{cases}$

<u>minimal requirements:</u>

- $0 \leq b(x_i) \leq 1$ ; $0 \leq b_a(\underline{x}_{\partial a}) \leq 1$.
- $\sum_{\underline{x}_{\partial a}\setminus i} b_a(\underline{x}_{\partial a}) = b_i(x_i)$.
- $\sum_{x_i} b_i(x_i) = 1$.

We define the <u>Bethe free energy</u> (~ free entropy) functional of $\{b_i ; b_a\}$: (

$$\boxed{F_{Bethe}(\underline{b}) = \sum_a \sum_{\underline{x}_{\partial a}} b_a(\underline{x}_{\partial a}) \ln \Psi_a(\underline{x}_{\partial a}) + \sum_i \sum_{x_i} b_i(x_i) \ln \Psi_i(x_i) - \sum_a \sum_{\underline{x}_{\partial a}} b_a(\underline{x}_{\partial a}) \ln b(\underline{x}_{\partial a}) - \sum_i (1-|\partial_i|) \ln b_i(x_i)}$$

Find $\underline{b}$ by minimizing, using the Lagrangian:

$$\mathcal{L}(\underline{b},\underline{\lambda}) = F_{Bethe}(\underline{b}) + \sum_i \lambda_i \left(\sum_{x_i} b_i(x_i) - 1\right) + \sum_{\substack{(i,a)\in E \\ x_i}} \lambda_{ia}(x_i) \left[\sum_{\underline{x}_{\partial a}\setminus i} b_a(\underline{x}_{\partial a}) - b_i(x_i)\right]$$

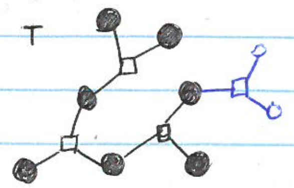
↳ after extremization:

$$\begin{cases} b_i(x_i) \propto \Psi_i(x_i) e^{-\frac{1}{|\partial_i|-1}\sum_{a\in\partial i}\lambda_{ia}(x_i)} \\ b_a(\underline{x}_{\partial a}) \propto \Psi_a(\underline{x}_{\partial a}) e^{-\sum_{i\in\partial a}\lambda_{ia}(x_i)} \\ \Psi_i(x_i) \exp\left(-\frac{1}{|\partial_i|-1}\sum_{a\in\partial i}\lambda_{ia}(x_i)\right) \propto \sum_{\underline{x}_{\partial a}\setminus i} b_a(\underline{x}_{\partial a}) \\ \lambda_i \quad \text{fix normalization} \end{cases}$$

⇒ highly non trivial set of equations that need to be solved.
(there are $\#\lambda_{ia}(x_i) = |E|(q-1)$ different lagrange multipliers ...).

↳ Bethe approximations on a tree:

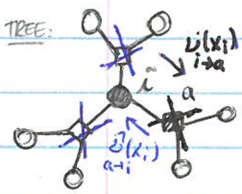
$p(\underline{x}) = \prod_a p_a(\underline{x}_{\partial a}) \prod_i p_i(x_i)^{1-|\partial i|}$ with $\begin{cases} p_i(x_i) \\ p_a(\underline{x}_{\partial a}) \end{cases}$ actual marginals. (proof by recursion) proof by recursion :(




$$P(\underline{x}) = P_T(\underline{x})\, P_a(\underline{x}_{\partial a \setminus i} | x_i) = P_T(\underline{x}) \frac{P_a(x_{\partial a})}{P_i(x_i)} \qquad \blacksquare$$

So that the Bethe free energy is exact on a tree!

## CAVITY FORMALISM



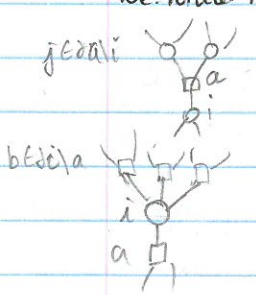

we define cavity marginals:

$$\begin{cases} \hat{\nu}_{a \to i}(x_i) = \text{marginal probability in a cavity graph without } \{\Psi_b\}_{b \in \partial i \setminus a} \\ \nu_{i \to a}(x_i) = \text{marginal prob in a cavity graph without } \Psi_a. \end{cases}$$

The marginals derive accordingly:


- $p_i(x_i) \propto \Psi_i(x_i) \prod_{a \in \partial i} \hat{\nu}_{a \to i}(x_i)$
- $p_a(\underline{x}_{\partial a}) \propto \Psi_a(\underline{x}_{\partial a}) \prod_{i \in \partial a} \hat{\nu}_{i \to a}(x_i)$


We write the recursive relations: BELIEF PROPAGATION




$$\hat{\nu}_{a \to i}(x_i) \propto \sum_{\underline{x}_{\partial a \setminus i}} \Psi_a(\underline{x}_{\partial a}) \prod_{j \in \partial a \setminus i} \nu_{j \to a}(x_j)$$

$$\nu_{i \to a}(x_i) \propto \Psi_i(x_i) \prod_{b \in \partial i \setminus a} \hat{\nu}_{b \to i}(x_i)$$

Rk: by comparing with the exact marginals by Bethe min see that message $\leftrightarrow$ lagrange multipliers:
$e^{-\lambda_{ia}(x_i)} = \nu_{ia}(x_i)$

Can we use these equations on a generic graph? (not tree)
↳ yes, the BP equations can be proven to be equivalent to the Bethe minimization.
So that the BP algorithm can be used to compute Bethe lagrange multipliers!

In practice BP are solved iteratively:
$$\begin{cases} \hat{\nu}_{a \to i}^{(t+1)}(x_i) = \hat{f}_a\left[\{\nu_{j \to a}^{(t)}\}_{j \in \partial a \setminus i}\right] \\ \nu_{i \to a}^{(t)}(x_i) = f_i\left[\{\hat{\nu}_{b \to i}^{(t)}\}_{b \in \partial i / a}\right] \end{cases}$$

hopefully, they converge to a fixed point: $\nu^{(t)}, \hat{\nu}^{(t)} \xrightarrow[t \to \infty]{} \underline{\nu}^*, \underline{\hat{\nu}}^*$
tricks: damping! random scheme of updates!

Can we rewrite the Bethe free entropy; as a function of messages?

$$b_i(x_i) = \frac{1}{Z_i} \psi_i(x_i) \prod_{a \in \partial i} \hat{\nu}_{a \to i}(x_i)$$
$$b_a(\underline{x}_{\partial a}) = \frac{1}{Z_a} \psi_a(x_{\partial a}) \prod_{i \in \partial a} \nu_{i \to a}(x_i)$$

↳ trick for rewriting: $\begin{cases} Z_{ia} = \sum_{x_i} \nu_{i \to a}(x_i)\, \hat{\nu}_{a \to i}(x_i) \\ Z_a = Z_i \quad Z_{a \to i} \\ Z_i = Z_{ia}\, Z_{i \to a} \\ b_i = \dfrac{\nu_{i \to a}\, \hat{\nu}_{a \to i}}{Z_{ia}} \end{cases}$

$$\Rightarrow \quad F[\nu, \hat{\nu}] = \sum_{a \in F} \ln Z_a + \sum_{i \in V} \ln Z_i - \sum_{ia \in E} \ln Z_{ia}$$

$$Z_a = \sum_{x_{\partial a}} \psi_a(\underline{x}_{\partial a}) \prod_{i \in \partial a} \nu_{i \to a}(x_i)$$
$$Z_i = \sum_{x_i} \psi_i(x_i) \prod_{a \in \partial i} \hat{\nu}_{a \to i}(x_i)$$

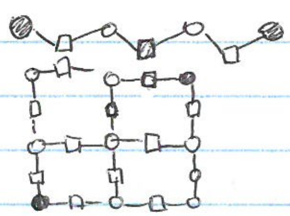
EXERCISE → explicitely
↳ show that same value for any normaliz° of the marginals.

↳ well defined for any graph!

COMPUTATION OF OBSERVABLES:

local: $\begin{cases} \langle x_i \rangle = \sum_{x_i} x_i\, b_i(x_i) \\ \langle \ln \psi_a(\underline{x}_{\partial a}) \rangle = \sum_{\underline{x}_{\partial a}} b_a(\underline{x}_{\partial a}) \ln \psi_a(\underline{x}_{\partial a}) \end{cases}$ → local energy term

non local:



→ tree or locally tree like ⇒ integrate out on
→ regular lattice: integrating over all the intermediate degrees of free is very costly!

instead: <u>linear response theorem</u>:

suppose $\psi_i(x_i) = e^{H_i x_i}$

$$\Rightarrow \langle x_i x_j \rangle - \langle x_i \rangle \langle x_j \rangle = \frac{\partial \langle x_i \rangle}{\partial H_j}$$

$\langle \; \rangle \equiv$ gibbs boltzmann

<u>Rk</u>: in a MF model, computing response or correlation are not consistent. Nevertheless, the response is less affected by the factorization → better MF approximation than just $\langle x_i x_j \rangle - \langle x_i \rangle \langle x_j \rangle$!