Lecture 11. Application of the Bethe free energy to Coding. The Connection between MAP and BP decoding.

We apply the formalism developed last time to coding. This leads us to a conjecture concerning the link between MAP & BP decoding. The ideas for the proof of this conjecture on the BEC are outlined.

1. Ensemble average of Bethe free energy and Replica Symmetric or Trial entropy.

Gallager ensemble LDPC (l, r).

BMS channel $p(Y | X)$

The associated graphical model has "Gibbs" or "MAP" distribution (see lecture no 4).

$$\mu(s_1 \dots s_N \mid y_1 \dots y_N) = \frac{1}{Z(\vec{h})} \prod_a \frac{1}{2}\left(1 + \prod_{i \in a} s_i\right) \prod_i e^{h_i s_i}$$

view these as fct nodes of degree 1.

$$h_i = \frac{1}{2} \ln \frac{p(y_i \mid s_i = +1)}{p(y_i \mid s_i = -1)}$$

(input word is "all zero codeword" : $s_i^{\text{input}} = +1$ all i).

Recall (lecture 4):

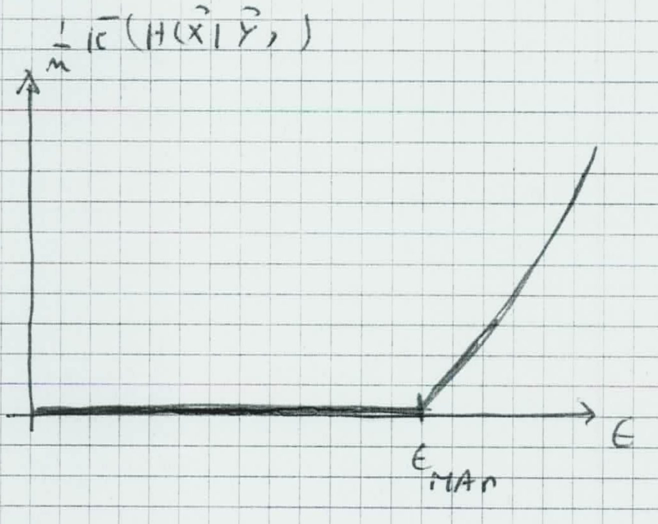
$$\frac{1}{m}\mathbb{E}_{LDPC}\left(H(\vec{X}\,|\,\vec{Y})\right) = \frac{1}{m}\mathbb{E}_{LDPC}\,\mathbb{E}_{\vec{h}}\left(\ln Z(\vec{h})\right) - \int dh\, h\, c(h)$$

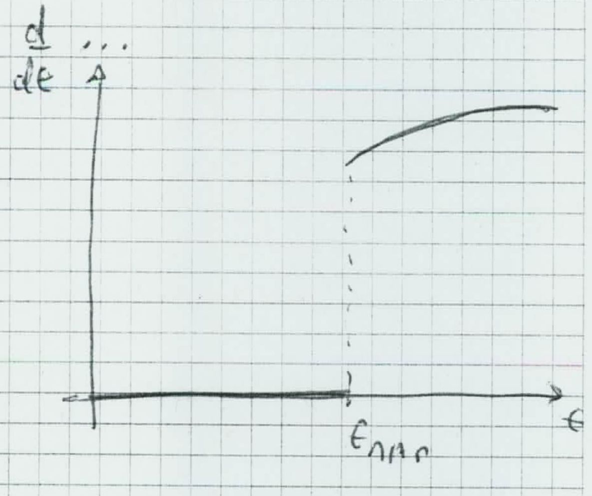
where $c(h)$ = channel density (in L-domain).

Performance curve (MAP - GEXIT function)

$$\frac{d}{d\epsilon}\,\frac{1}{m}\,\mathbb{E}_{LDPC}\left(H(\vec{X}\,|\,\vec{Y})\right) \propto \text{average bit prob of error for BEC}$$

Typical behavior that we expect (say (3,6) ensemble)





(In exercises no 4 formulas for GEXIT are discussed).

Goal: show that Averaged form of Bethe free energy allows to compute MAP performance curve. Since stat pts of Bethe are fixed pts of BP equations we will see that BP & MAP performance curves are intimately related.

Bethe expression for Coding; fixed code, fixed channel output.

Take Bethe expression of last time and adapt it to the graphical model of coding. You get the functional:

$$F_{Bethe}[h_{j\to a}, h_{a\to j}] = F_a + F_i - F_{ai}$$

$$F_a = \ln\{1 + \prod_{j\in\partial a} \tanh h_{j\to a}\}$$

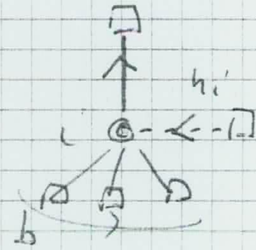
$$F_i = \ln\left\{ e^{h_i} \prod_{a\in\partial i} (1+\tanh h_{a\to i}) + e^{-h} \prod_{a\in\partial i} (1-\tanh h_{a\to i})\right\}$$

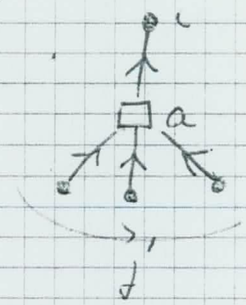
$$F_{ai} = \ln\left\{ 1 + (\tanh h_{i\to a})(\tanh \hat{h}_{a\to i})\right\}.$$

(up to a constant proportional to $\ln 2$).

Note that stat points of Bethe functional satisfy BP fixed point equations!

$$\begin{cases} h_{i\to a} = h_i + \sum_{b\in\partial i\setminus a} \hat{h}_{b\to i} \\[2ex] \hat{h}_{a\to i} = \operatorname{atanh}\left\{ \prod_{j\in\partial a\setminus i} \tanh h_{j\to a} \right\} \end{cases}$$





Replica Symmetric or Trial free energy/entropy:

- Take code (graph from LDPC (l, r)) at random.
- Take r.v $H \sim c(h)$ (channel output r.v).
- Take i.i.d $H_1, \dots H_r \sim Q(h)$ where $Q(h)$ is a "trial" distribution to be fixed later.
- Define r.v $\hat{H} = \operatorname{atanh}\left(\prod_{j=1}^{r-1} \tanh H_j \right)$.

  and form $\hat{H}_1, \dots \hat{H}_l$ i.i.d copies.

  These are distr $\sim \hat{Q}(\hat{h})$ which depends on $Q(\cdot)$.

Define

counts correctly nb of checks.

$$f_{Bethe}[H, \hat{H}] = \frac{l}{r} \ln \left\{ 1 + \prod_{j=1}^{r} \tanh H_j \right\}$$

$$+ \frac{1}{\cdot} \ln \left\{ e^{h} \prod_{a=1}^{l} (1 + \tanh \hat{H}_a) + e^{-h} \prod_{a=1}^{l} (1 - \tanh \hat{H}_a) \right\}$$

normalized by $\frac{1}{n}$.

$$- l \ln \left(1 + (\tanh H)(\tanh \hat{H}) \right)$$

(counts correctly nb of edges)

Definition. The RS or Trial free energy (entropy) is the functional of $Q(\cdot)$:

$$f_{RS}(Q(\cdot)) = \mathbb{E}_{LDPC} \, \mathbb{E}_{h, H, \hat{H}} \left\{ f_{Bethe}[H, \hat{H}] \right\}.$$

Note that ~~time~~ if LDPC is regular (say 3, 6) the $\mathbb{E}_{LDPC}$ is trivial. Otherwise it is an average over local degrees of nodes (l, r are random).

For the entropy we have

$$h_{RS}(Q(\cdot)) = - f_{RS}(Q(\cdot)) - \int dh \, h \, c(h).$$

2. Conjecture : relation between BP & MAP

definition :
$$f_{exact} \equiv \lim_{n \to +\infty} -\frac{1}{n} \mathbb{E}_{LDPC} \mathbb{E}_h (\ln Z(h))$$

conjecture !
$$f_{exact} = \min_{Q(\cdot)} f_{RS}(Q(\cdot))$$

or equivalently :

$$\lim_{n \to +\infty} \frac{1}{n} \mathbb{E}_{LDPC} \left(H(\vec{X} | \vec{Y}) \right) = \max_{Q(\cdot)} \left(h_{RS}(Q(\cdot)) \right)$$
$$= -\min_{Q(\cdot)} \left(f_{RS}(Q(\cdot)) - \int c(h) h \, dh \right)$$

formal stationary points are solutions of density evolution fixed point equations.

$$\begin{cases} \hat{Q}(\hat{H}) = \int \prod_j Q(H_j) \, \delta\left(\hat{H} - \operatorname{atanh} \prod_{j=1}^{k-1} \tanh H_j\right) \\ Q(H) = \int c(h) dh \int \prod_a \hat{Q}(\hat{H}_a) \, \delta\left(H - (h + \hat{H}_1 + \dots + \hat{H}_{l-1})\right) \end{cases}$$

To get the correct entropy we have to select the correct stat pt (the one that yields a global ~~minimum~~ maximum).

Main rigorous results:

* upper / lower bound : $f_{exact} \leq \min_{Q(\cdot)} f_{RS}(Q(\cdot))$

or $\lim_{n \to \infty} \frac{1}{n} \mathbb{E}_{LDPC}(H(\vec{X} \mid \vec{Y})) \geq \max_{Q(\cdot)} h_{RS}(Q(\cdot))$

(Proof by "interpolation method").

* Equality for LDGM codes for high noise.

* Equality for LDPC over the BEC : ideas will be developped by Ruediger

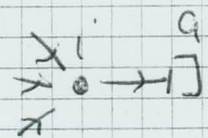
3. Explicit case of BEC.

From analysis of BP equations for the BEC we know that:

$$c(h) = \epsilon\,\delta(h) + (1-\epsilon)\,\Delta_\infty(h) \qquad \text{(channel law)}$$

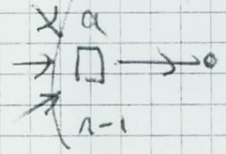
$$Q(H) = x\,\delta(H) + (1-x)\,\Delta_\infty(H)$$

prob of erasure message from



$$\hat{Q}(\hat{H}) = \hat{x}\,\delta(\hat{H}) + (1-\hat{x})\,\Delta_\infty(\hat{H})$$

prob of erasure message from



We must have $\hat{x} = 1 - (1-x)^{r-1}$

prob at least one incoming message is "erasure"

and x is the "trial" variable parametrising $f_{RS}(x)$ or $h_{RS}(x)$.

We have

$$h_{RS}(x) = \mathbb{E}_{h,H,\hat{H}}\left\{ \frac{l}{r} \ln\left(1 + \prod_{j=1}^{r} \tanh H_j\right)\right\} \qquad (a)$$

$$+\ \mathbb{E}_{h,H,\hat{H}}\left\{ \ln\left(\prod_{a=1}^{l}(1+\tanh\hat{H}_a) + e^{-2h}\prod_{a=1}^{l}(1-\tanh\hat{H}_a)\right)\right\} \qquad (b)$$

$$-\ \mathbb{E}_{h,H,\hat{H}}\left\{ l \ln\left(1 + \tanh H \tanh\hat{H}\right)\right\} \qquad (c)$$

<u>Term (a)</u> :

$$\frac{\ell}{r}(1-x)^{r} \cdot \ln 2 .$$

<u>Term (c)</u> :

$$-\ell(1-x)(1-\hat{x}) \ln 2 .$$

<u>Term (b)</u> :

$$\underbrace{(1-\epsilon)}_{(h=\pm\infty)} \sum_{E=0}^{\ell} \binom{\ell}{E} \hat{x}^{E} (1-\hat{x})^{\ell-E} \ln\left(2^{\ell-E}\right)$$

$$+ \underbrace{\epsilon}_{(h=0)} \sum_{E=0}^{\ell-1} \binom{\ell}{E} \hat{x}^{E} (1-\hat{x})^{\ell-E} \ln\left(2^{\ell-E}\right) \quad \text{"at least one non-erasure"}$$

$$+ \epsilon \binom{\ell}{\ell} \hat{x}^{\ell} (1-\hat{x})^{\ell-\ell} \underline{\ln 2} . \quad \text{"all erasures"}$$

$$= \underbrace{\left\{ \sum_{E=0}^{\ell} \binom{\ell}{E} \hat{x}^{E} (1-\hat{x})^{\ell-E} (\ell-E) \right\}}_{} \ln 2 + \epsilon \hat{x}^{\ell} \ln 2 .$$

$$(1-\hat{x}) \sum_{E=0}^{\ell} \binom{\ell}{E} \hat{x}^{E} \frac{d}{dy} y^{\ell-E} \Big|_{y=1-\hat{x}}$$

$$= (1-\hat{x}) \frac{d}{dy} (\hat{x}+y)^{\ell} \Big|_{y=1-\hat{x}}$$

$$= (1-\hat{x}) \cdot \ell .$$

So one finds (in units of ln2) :

$$h_{RS}(x) = \left\{ \frac{l}{r}(1-x)^{r} \right\} + \left\{ \epsilon\, \hat{x}^{l} + l(1-\hat{x}) \right\} - \left\{ l(1-x)(1-\hat{x}) \right\} - \underbrace{\frac{l}{r}}_{\text{comes from all extra } \ln 2\text{'s}}$$

$$\Rightarrow \boxed{ \begin{aligned} h_{RS}(x) &= + l\, x(1-\hat{x}) + \frac{l}{r}(1-x)^{r} + \epsilon\, \hat{x}^{l} - \frac{l}{r} \\ \text{with} \quad \hat{x} &= 1 - (1-x)^{r-1} \end{aligned} }$$

Stationary pts :

$$\frac{d}{dx} h_{RS}(x) = 0 \Rightarrow \frac{\partial}{\partial x} h_{RS} + \frac{\partial h_{RS}}{\partial \hat{x}} \cdot \frac{d\hat{x}}{dx} = 0$$

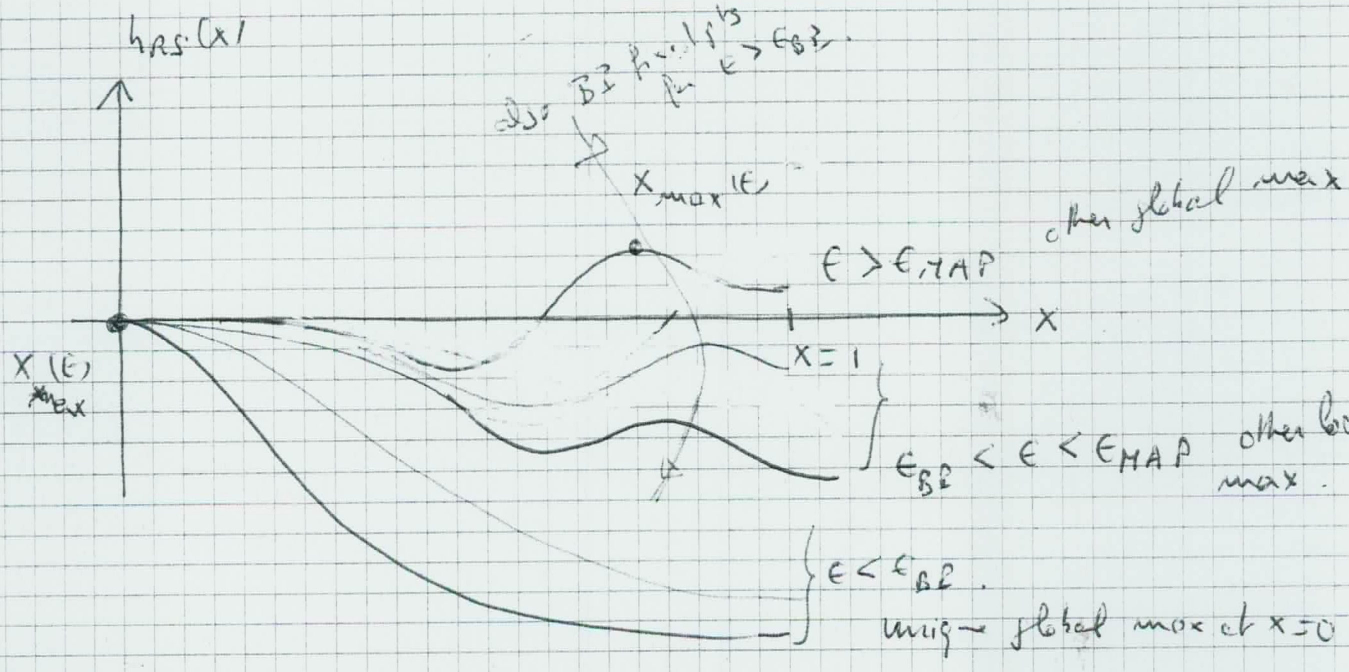
$$\Rightarrow \underbrace{\left(l(1-\hat{x}) - l(1-x)^{r-1} \right)}_{0} + \left(-lx + \epsilon\, l\, \hat{x}^{l-1} \right) \cdot (r-1) \cdot (1-x)^{r-2} = 0$$

$$\Rightarrow \left(-lx + \epsilon\, l \left(1 - (1-x)^{r-1} \right)^{l-1} \right) \cdot (r-1)(1-x)$$

$$\Rightarrow \boxed{ x = \epsilon \left(1 - (1-x)^{r-1} \right)^{l-1} }$$

usual DE equation for erasure probability.

Plot of $h_{RS}(x)$ as a fct of x for fixed ϵ.

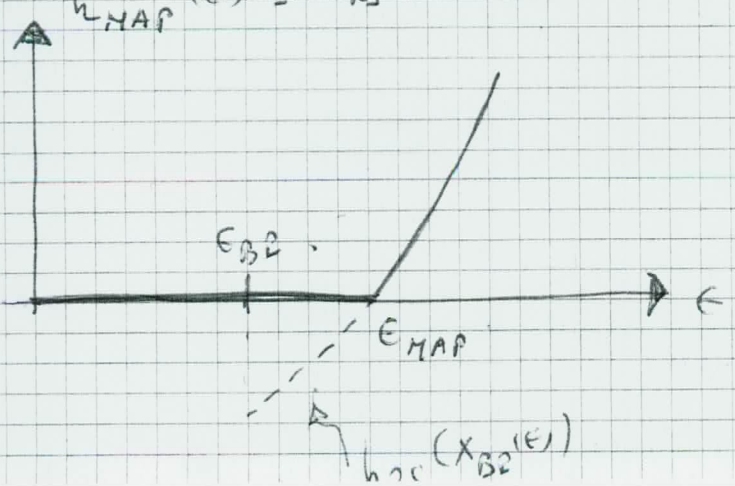


$\epsilon < \epsilon_{BP}$ $\quad x_{max}(\epsilon) = 0$. unique max

$\epsilon_{BP} < \epsilon < \epsilon_{MAP}$ $\quad \begin{cases} x_{max}(\epsilon) = 0 & \text{global max} \\ x_{BP}(\epsilon) \neq 0 & \text{local max} \end{cases}$.

$\epsilon_{MAP} < \epsilon$ $\quad \begin{cases} x_{max}(\epsilon) \neq 0 & \text{global max} \\ x_{BP}(\epsilon) = 0 & \text{local max} \end{cases}$.

$$h_{MAP}(\epsilon) = h_{RS}(x_{max}(\epsilon)).$$



MAP conditional entropy is:

$$h_{MAP}(\epsilon) = \max_{0 \le x \le 1} h_{RS}(x) = h_{RS}(x_{max}(\epsilon)),$$

where $x_{max}(\epsilon)$ satisfies $x = \epsilon\left(1-(1-x)^{r-1}\right)^{l-1}$.

There is a close analogy with CW model:

$$\begin{cases} f_{CW} = \min_m \phi(m) = \phi(m_{Min}) \quad (\text{lecture } 3) \\ \phi(m) = \left(-\frac{K}{2}m^2 - hm\right) - \underbrace{\mathcal{H}_2(m)}_{\text{binary entropy fct.}} \\ m_{Min} \text{ satisfies } m = \tanh(Km+h), \end{cases}$$

Probability of error.

MAP: Av Prob of bit error $= \epsilon \frac{d}{d\epsilon} h_{MAP}(\epsilon)$.

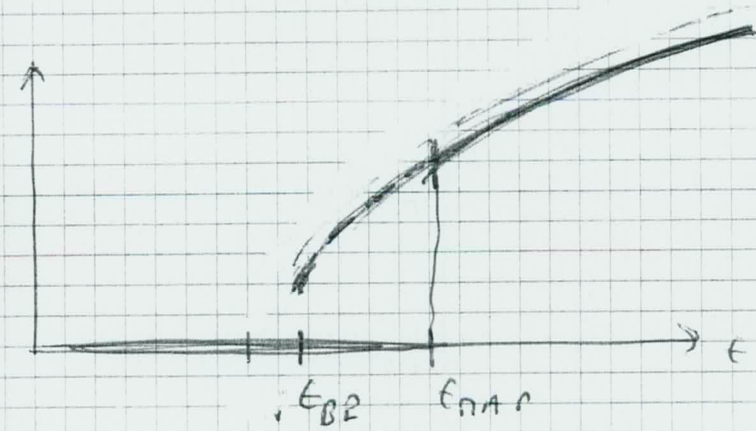
(see exercises no 4 & 5).

$$\frac{d}{d\epsilon} h_{MAP} = \underbrace{\frac{\partial}{\partial \epsilon} h_{RS}(x_{max}(\epsilon))}_{\hat{x}^l} + \underbrace{\frac{\partial}{\partial x_{max}} h_{RS}}_{0} \cdot \frac{d x_{max}}{d\epsilon}. \quad (\text{away from threshold})$$

$$= \left(1-(1-x_{max}(\epsilon))^{r-1}\right)^{l}.$$

So we get for MAP decoding (*)

$$\text{Av Prob bit error} = \epsilon \left(1 - \left(1 - x_{max}(\epsilon)\right)^{r-1}\right)^{l}.$$



Since for $\epsilon < \epsilon_{BP}$ & $\epsilon > \epsilon_{MAP}$ we have

$x_{max}(\epsilon)$ is the same than iterative soln $x_{BP}(\epsilon)$.

the BP & MAP curves are equal.

for $\epsilon_{BP} < \epsilon < \epsilon_{MAP}$; $x_{max}(\epsilon) \neq x_{BP}(\epsilon)$

and BP curve is an extension of MAP curve.

(*) Compare with

$$\text{Av Prob bit error for BP decoding} = \epsilon \left(1 - \left(1 - x_{BP}(\epsilon)\right)^{r-1}\right)^{l}.$$