Statistical Physics for Communication and Computer Science 17/03/2011

Lecture Notes 4: Statistical Physics Reformulation of Coding

Lecturer: Nicolas Macris *Scribe: Marc Desgroseilliers*

We reformulate the coding problem of lecture 1 in the language of statistical physics. As we will see the MAP decoder can be interpreted as a general spin system. Later in the course, we will come to the formulation of the Belief Propagation decoder and its relation to the MAP.

1 Generalization of the Ising Model

The Ising model can be generalized in various directions. Here we point out the structure of the most general models, keeping the same alphabet $\{-1,+1\}$. All the problems that will interest us in this course can be cast in this unified framework.

The most general function of N spin variables $s_i \in \{-1,+1\}$ is of the form

$$\mathcal{H}(\underline{s}) = -\sum_{A\subset\{1,\cdots,N\}} J_A \prod_{i\in A} s_i$$

The constants J_A are "Fourier coefficients" of the expansion on the basis $\prod_{i\in A} s_i$, $A \subset \{1,\cdots,N\}$ for functions of $\underline{s}$. There are 2^N basis functions. In order for $\mathcal{H}$ to be a "reasonable" Hamiltonian (that leads to nice thermodynamic behavior) one asks that

- The system has a large number of degrees of freedom: N tends to infinity.
- Interactions are *local*. Each term in the sum couples a finite number of spins, so that $|A| = O(1)$. Each spin enters in a finite number of terms. In particular since there are N spins the total number of non zero Fourier coefficients is $O(N)$.

The Gibbs distribution for finite N is defined as usual

$$\mu_N(\underline{s}) = \frac{1}{Z_N} e^{-\beta\mathcal{H}(\underline{s})} = \frac{1}{Z_N} \prod_{A\subset\{1,\cdots,N\}} e^{\beta J_A \prod_{i\in A} s_i}$$

where Z_N is the sum of the numerator over $\underline{s} \in \{-1,+1\}^N$. Locality implies that this measure factors into a product of terms with a small number of variables in each of them. Moreover each variable enters in a small number of factors.

For calculationnal purposes it is often useful to write the measure in the form

$$\mu_N(\underline{s}) = \frac{\prod_{A\subset\{1,\cdots,N\}}(1+(\tanh\beta J_A)\prod_{i\in A} s_i)}{\sum_{\underline{s}}\prod_{A\subset\{1,\cdots,N\}}(1+(\tanh\beta J_A)\prod_{i\in A} s_i)}$$

Bipartite factor graphs (Tanner graphs) are a convenient way to summarize the factored structure of the measure. Attach N spin variables to *variable nodes* $s_1,\cdots,s_N$. Consider

the collection of subsets $A \subset \{1, \cdots, N\}$ for which $J_A \neq 0$. The number of such subsets is $O(N)$. For each subset A we have a *function node* with the weight

$$\psi(s_i, i \in A) = 1 + (\tanh \beta J_A) \prod_{i \in A} s_i$$

Edges of the bipartite graph connect A to the variables $i \in A$. Locality translates in the finiteness of the degrees of variable and function nodes.

Coding and SAT problems can be formulated in this language with the extra feature that the collection of subsets A and J_A are random chosen in an ensemble. As we will see hard constraints imply that some of the J_A's are "infinite". This simply means that $\tanh \beta J_A = 1$ in the expression of the measure above.

2 Coding: Notations and Preliminaries

We take a code $\mathcal{C}$ from the Gallager Ensemble $LDPC(l, r, n)$ with l = degree of variable nodes, r =degree of check nodes and n = length of codewords. Consider a transmitted codeword $\underline{x}^{input} = (x_1^0, \ldots, x_n^0)$ through a binary, memoryless channel without feedback, with received word $\underline{y}^{input} = (y_1, \ldots, y_n)$. We will always assume that the codeword is selected uniformly at random. We will use the spin variable notation for the codebits $s_i = (-1)^{x_i}$. The channel is described by the transition probabilities

$$P(\underline{y}|\underline{s}) = \prod_{i=1}^{n} p(y_i|s_i) \tag{1}$$

Two examples will be mainly kept in mind, the Binary Symmetric Channel (BSC) and the Binary Input Additive White Noise Channel (BIAWNC) $y_i = x_i + \sigma n_i$ with n_i standard Gaussian, σ^2 the noise variance ($SNR = \sigma^{-2}$).

Bit-MAP decoding. Let $\mathbb{P}(s|y)$ be the posterior probability distribution given the received word $\underline{y}$. We take the hard estimate

$$\begin{aligned} \hat{s}_i(\underline{y}) &= argmax\ \mathbb{P}(s_i|\underline{y}) \\ &= sign(\mathbb{P}(s_i = 1|\underline{y}) - \mathbb{P}(s_i = -1|\underline{y})) \\ &= sign(\sum_{s_i} s_i \mathbb{P}(s_i|y)) \end{aligned} \tag{2}$$

where $\mathbb{P}(s_i|y)$ is the marginal of $\mathbb{P}(s|y)$. The argument of the sgn function is sometimes called the soft estimate. The average bit-probability of error is

$$\begin{aligned} P_{error} &= \frac{1}{n} \sum_{i=1}^{n} \mathbb{P}(\hat{s}_i(\underline{y}) \neq s_i^0) \\ &= \frac{1}{n} \sum_{i=1}^{n} \frac{1}{|\mathcal{C}|} \sum_{\underline{s}^0 \in \mathcal{C}} \mathbb{E}_{\underline{Y}|s^0}[\mathbf{1}(\hat{s}_i(\underline{y}) \neq s_i^0)] \\ &= \frac{1}{n} \sum_{i=1}^{n} \frac{1}{|\mathcal{C}|} \sum_{\underline{s}^0 \in \mathcal{C}} \mathbb{E}_{\underline{Y}|s^0}[\frac{1}{2}(1 - s_i^0 sign\langle s_i \rangle)] \end{aligned} \tag{3}$$

where $\mathbb{E}_{\underline{Y}|s^0}$ is the expectation over the channel realization when $\underline{s}^0$ is sent.

Block-MAP decoding. The estimate of the block is

$$\hat{\underline{s}}(y) = argmax\ \mathbb{P}(\underline{s}|\underline{y}) \tag{4}$$

and the block error

$$P^B_{error} = \mathbb{P}(\hat{\underline{s}}(\underline{y}) \neq \underline{s}^0) \tag{5}$$

As we will see below block decoding is equivalent to finding the ground states (minimum energy states) of a Hamiltonian.

3 MAP Decoding as a Spin Glass Problem

We show that the posterior distribution $\mathbb{P}(s|y)$ is a random Gibbs measure. Two preliminary observations are useful:

Observation 3.1. A code word $\underline{x}$ has to satisfy all parity check constraints $\sum_{i\in c} x_i = 0$. This is equivalent to $\prod_{i\in c} s_i = 1$. Thus the prior distribution over codewords can be written as

$$\mathbb{P}(\underline{s}) = \frac{1}{|\mathcal{C}|}\mathbb{I}(\underline{s}) = \frac{1}{|\mathcal{C}|}\prod_{c\in\mathcal{C}} \mathbb{I}(\underline{s} \text{ satisfies } c) = \frac{1}{|\mathcal{C}|}\prod_{c\in\mathcal{C}} \frac{1}{2}(1+\prod_{i\in c} s_i) \tag{6}$$

Observation 3.2. Channel outputs can be expressed in terms of their half-loglikelihoods

$$h_i = \frac{1}{2}\ln\frac{\mathbb{P}(y_i|+1)}{\mathbb{P}(y_i|-1)}. \tag{7}$$

One checks that

$$\mathbb{P}(y_i|s_i) = \mathbb{P}(y_i|s_i=+1)e^{-h_i}e^{h_i s_i} \tag{8}$$

Notation 3.3. We will abusively interchange the arguments $\underline{y}$ and $\underline{h}$ when no confusion is possible.

Using first Bayes law, second the channel law, and lastly the two observations, we obtain:

$$\begin{aligned}\mathbb{P}(\underline{s}|\underline{y}) &= \frac{P(\underline{y}|\underline{s})\mathbb{P}(\underline{s})}{\sum_{\underline{s}} P(\underline{y}|\underline{s})\mathbb{P}(\underline{s})} = \frac{\prod_{i=1}^n \mathbb{P}(y_i|s_i)\mathbb{P}(\underline{s})}{\sum_{\underline{s}}\prod \mathbb{P}(y_i|s_i)\mathbb{P}(\underline{s})} \\ &= \frac{1}{Z(\underline{h})}\prod_{c\in\mathcal{C}}\frac{1}{2}(1+\prod_{i\in c} s_i)\prod_{i=1}^n e^{h_i s_i}\end{aligned} \tag{9}$$

To get the last equality we pulled out the $\mathbb{P}(y_i|+1)e^{-h_i}$ terms from both numerator and denominator (since they do not depend on s_i), and cancelled them. The denominator (partition function) is simply

$$Z(\underline{h}) = \sum_{\underline{s}}\prod_{c\in\mathcal{C}}\frac{1}{2}(1+\prod_{i\in c} s_i)\prod_{i=1}^n e^{h_i s_i} \tag{10}$$

This is a random Gibbs distribution. By this we mean that for each channel realization $\underline{h}$ and each code $\mathcal{C}$ from the Gallager ensemble we have a measure over the spins $\underline{s} \in \{-1,+1\}^n$. An important feature (of all Gibbs distributions) is that it is factored into a product of "local" terms, i.e terms which depend on a finite number of spins.

Let us point out a few physics terms that are commonly used. Another name for random Gibbs measures is "spin-glass". This refers to the fact that glass is an amorphous material (amorphous = random ordering of atoms). There is no good theory for real glass, but random spin systems were initially invented as toy models to investigate properties of glass, hence the name spin-glass. The channel outputs and the code (or Tanner graph) are called "quenched" or "frozen" variables because they are fixed in the measure. The spins are called "annealed" variables because they adjust themselves in the environment of the quenched variables. These terms also refer to the vastly different dynamical time scales of the two type of degrees of freedom in amorphous materials like glass.

What is the Hamiltonian associated to the Gibbs measure of the MAP decoder ? One way to identify it is to write

$$\mathbb{P}(\underline{s}|\underline{y}) = \lim_{K_c \to +\infty} \frac{\prod_{c\in\mathcal{C}}(1+\tanh K_c \prod_{i\in c} s_i)\prod_{i=1}^n e^{h_i s_i}}{\sum_{\underline{s}} \prod_{c\in\mathcal{C}}(1+\tanh K_c \prod_{i\in c} s_i)\prod_{i=1}^n e^{h_i s_i}} \tag{11}$$

Using the identity (which is true because $\prod_{i\in c} s_i = \pm 1$)

$$e^{K_c \prod_{i\in c} s_i} = \cosh K_c + \sinh K_c \prod_{i\in c} s_i = \cosh K_c (1+\tanh K_c \prod_{i\in c} s_i) \tag{12}$$

we get

$$\mathbb{P}(\underline{s}|\underline{y}) = \lim_{K_c \to +\infty} \frac{e^{-\mathcal{H}_{\mathcal{C}}(\underline{s}|\underline{h})}}{\sum_{\underline{s}} e^{-\mathcal{H}_{\mathcal{C}}(\underline{s}|\underline{h})}} \tag{13}$$

for the Hamiltonian

$$\mathcal{H}_{\mathcal{C}}(\underline{s}|\underline{y}) = -\sum_{c\in\mathcal{C}} K_c \prod_{i\in c} s_i - \sum_{i=1}^n h_i s_i \tag{14}$$

This is the hamiltonian of the spin system associated to the Tanner graph of $\mathcal{C}$. The parity check constraints appear as "infinite" coupling constants K_c and the channel outputs as magnetic fields h_i. The "temperature" in this spin system is $k_B T = 1$ (more on this later). It is interesting to remark that it is *not a good analogy to think of channel noise as temperature*: channel noise measure the variance of the random magnetic field h_i.

Bit-MAP decoding. Recall that the average over $\underline{s}$ with respect to $\mathbb{P}(\underline{s}|\underline{y})$ are denoted by $\langle - \rangle$. The bit-MAP estimate (2) is now

$$\hat{s}_i(\underline{y}) = sign\langle s_i \rangle \tag{15}$$

The soft bit estimate is nothing else than a magnetization. The average bit probability of error (16) becomes,

$$P_{error} = \frac{1}{n}\sum_{i=1}^n \frac{1}{|C|} \sum_{\underline{s}^0 \in C} \mathbb{E}_{\underline{h}|s^0}\left(\frac{1}{2}(1 - s_i^0 sign\langle s_i \rangle)\right) \tag{16}$$

If the *sign* function was not here this would be an expected magnetization. The presence of the *sign* does not make this quantity very convenient to deal with, because it cannot be obtained as a derivative of a free energy[1]. We will see that derivatives of the free energy, which are more directly related to the magnetization, and are technically easier to deal with, are convenient performance measures. They have the same threshold as the probability of error.

Block-MAP decoding. The block estimate maximizes $\mathbb{P}(\underline{s}|\underline{y})$. Because of (**??**) this is equivalent to minimizing $\mathcal{H}_{\mathcal{C}}(\underline{s}|\underline{y})$.

$$\hat{\underline{s}}(\underline{y}) = argmin\, \mathcal{H}_{\mathcal{C}}(\underline{s}|\underline{h}) \tag{17}$$

Therefore block decoding is equivalent to finding the lowest energy state of a Hamiltonian. This also means that block decoding is a "zero temperature" problem.

Finite temperature decoder. There is a small literature on the so-called finite temperature decoder based on the Gibbs measure ($k_B T = \beta^{-1}$)

$$\mathbb{P}_{\beta}(\underline{s}|\underline{y}) = \lim_{K_c \to +\infty} \frac{e^{-\beta \mathcal{H}_{\mathcal{C}}(\underline{s}|\underline{h})}}{\sum_{\underline{s}} e^{-\beta \mathcal{H}_{\mathcal{C}}(\underline{s}|\underline{h})}} \tag{18}$$

This decoder interpolates between block and bit decoding. For $\beta = 1$ one gets the bit decoder and for $\beta = +\infty$ one gets the block decoder.

4 Channel Symmetry and Gauge Transformations

For symmetric channels all formulas simplify and there are some remarkable identities that hold. Channel symmetry is a special case of "gauge symmetry" also called Nshimori symmetry in the spin glass context.

Definition 4.1. A channel is said to be *output symmetric* if $P(y_i|s_1) = P(-y_i| - s_i)$.

From now on we limit ourselves to this class of channels. The BSC, BEC, BIAWNGC belong to it.

Let $\underline{\tau} = (\tau_1, \ldots, \tau_n)$ be a codeword in $\mathcal{C}$. It is immediate to see that under the transformation (change of variable)

$$s_i \mapsto \tau_i s_i$$
$$h_i \mapsto \tau_i h_i$$

$\mathbb{P}(\underline{s}|\underline{h})$ is invariant. Note that this works because τ is a codeword so that $\prod_{i \in c} \tau_i = 1$ for all c. The set of such transformations form a group (because the codewords from a group) and the transformations are local (because each variable gets multiplied by a different sign). In physics transformations with these two properties are called "gauge transformations" and when they leave invariant a Hamiltonian one speaks of "gauge symmetry"[2].

Let us now explore some consequences of this symmetry. We start with the simplest one.

[1] As we will see shortly in this context the free energy is essentialy equivalent to the cnditional entropy $H(\underline{X}|\underline{Y})$.

[2] The prototype of gauge symmetry is the invariance of Maxwell equations under choice of potentials. Gauge symmetry seems to be an underlying principle for all four fundamental forces.

4.1 Independence of input codeword.

Under this change of variables

$$\langle s_i \rangle \to \tau_i \langle s_i \rangle \tag{19}$$

where $\langle - \rangle$ is the same expectation on both sides (invariance of Gibbs measure). Therefore

$$\begin{aligned} \mathbb{E}_{\underline{h}|\underline{s}^0}(sgn(\langle s_i \rangle)) &= \mathbb{E}_{\dots\tau_i y_i \dots|\underline{s}^0}(\tau_i sgn(\langle s_i \rangle)) \\ &= s_i^0 \tau_i \mathbb{E}_{\underline{Y}|\dots\tau_i s_i^0}(sgn(\langle s_i \rangle) \end{aligned} \tag{20}$$

Using $\underline{\tau} = \underline{s}^0$ for the gauge, we get

$$s_i^0 \mathbb{E}_{\underline{h}|\underline{s}^0}(sgn(\langle s_i \rangle) = \mathbb{E}_{\underline{h}|+1}(sgn(\langle s_i \rangle)) \tag{21}$$

and so Equation (16) simplifies to

$$P_{error} = \frac{1}{n} \sum_{i=1}^{m} \mathbb{E}_{\underline{h}|+1}(\frac{1}{2}(1 - sgn(\langle s_i \rangle)) \tag{22}$$

Thus, for symmetric channels we can assume without loss of generality that the input word is the all +1 word (note that in the 0/1 language this is the all 0 codeword). From now on we denote $\mathbb{E}_{\underline{h}|+1}$ as $\mathbb{E}_{\underline{h}}$. The distribution of h_i is given by

$$c(h_i)dh_i = p(y_i| + 1)dy_i \tag{23}$$

We leave it as an exercise for the reader to check that,

Observation 4.2. For output symmetric channels, we have

$$c(-h) = c(h)e^{-2h} \tag{24}$$

Example 4.3. For the BSC(flip prob p) we have $c(h) = (1-p)\delta(h - \ln\frac{1-p}{p}) + p\delta(h - \ln\frac{p}{1-p})$

Example 4.4. For the BIAWGNC we have $c(h) = \frac{1}{\sqrt{2\pi\sigma^{-2}}} e^{-(h-\frac{1}{\sigma^2})^2/\frac{2}{\sigma^2}}$

4.2 Nishimori identities

Let us state a general proposition:

Proposition 4.5 (Nishimori identities for general spin systems)**.** *Consider the general spin system with Hamiltonian,*

$$\mathcal{H}(\underline{s}) = -\sum_A J_A \prod_{i \in A} s_i$$

and random i.i.d coupling constants satisfying

$$\mathbb{P}(-J_A) = \mathbb{P}(J_A)e^{-2J_A}$$

We have $\forall m_1 \dots m_l$ integers, the general identity

$$\mathbb{E}_J[\langle S_{X_1} \rangle^{m_1} \langle S_{X_2} \rangle^{m_2} \dots \langle S_{X_l} \rangle^{m_l}] = \mathbb{E}_J[\langle S_{X_1}^{m_1} S_{X_2}^{m_2} \dots S_{X_l}^{m_l} \rangle \langle S_{X_1} \rangle^{m_1} \langle S_{X_2} \rangle^{m_2} \dots \langle S_{X_l} \rangle^{m_l}]$$

where $S_X = \prod_{i \in X} s_i$.

Remark 4.6. These identities are valid for $\langle - \rangle$ associated to the Gibbs measure $\mu(s) = \frac{e^{\mathrm{H}(s)}}{Z}$ with $\beta = 1$. In the space of all possible parameters of the Hamiltonian the line $\beta = 1$ plays a special role because of these identities and is called the "Nishimori line".

We do not give the proof of this general proposition, since we only need special cases of it. Here we only prove the simplest possible such identity valid for the Gibbs distribution associated to a symmetric channel and any linear code.

Proposition 4.7 (Simplest Nishimori Identity for Coding)**.** *For a binary input, output symmetric, memoryless without feedback channel,*

$$\mathbb{E}_{\underline{h}}(\langle s_i \rangle) = \mathbb{E}_{\underline{h}}(\langle s_i \rangle^2)$$

Proof. For convenience we use the slightly abusive notation: $\underline{\tau h}$ *for the vector* $(\tau_j h_j)_{j=1\cdots n}$. Using a gauge transformation, $s_j \to \tau_j s_j$, $h_j \to \tau_j h_j$ for $\tau \in \mathcal{C}$, together with channel symmetry, we have,

$$\begin{aligned} \mathbb{E}_{\underline{h}}(\langle s_i \rangle) &= \mathbb{E}_{\underline{\tau h}}(\tau_i \langle s_i \rangle) \\ &= \mathbb{E}_{\underline{h}}[\prod_{i=1}^{n} e^{h_j \tau_j - h_j} \tau_i \langle s_i \rangle] \end{aligned} \tag{25}$$

Now, summing over all the codewords $\underline{\tau} \in \mathcal{C}$

$$\begin{aligned} \mathbb{E}_{\underline{h}}(\langle s_i \rangle) &= \frac{1}{|\mathcal{C}|} \mathbb{E}_{\underline{h}}[\sum_{\underline{\tau} \in \mathcal{C}} [\prod_{j=1}^{n} e^{h_j \tau_j} e^{h_j}] \langle s_i \rangle] \\ &= \frac{1}{|\mathcal{C}|} \mathbb{E}_{\underline{h}}[Z \prod_{j=1}^{n} e^{-h_j} \langle \tau_i \rangle \langle s_i \rangle] \\ &= \frac{1}{|\mathcal{C}|} \sum_{\underline{\eta} \in \mathcal{C}} \mathbb{E}_{\underline{h}}[\prod_{j=1}^{n} e^{h_j \eta_j} \prod_{j=1}^{n} e^{-h_j} \langle \tau_i \rangle \langle s_i \rangle] \end{aligned} \tag{26}$$

For each term in the sum over $\underline{\eta}$, we do a gauge transformation $s_j \to \eta_j s_j$, $\tau_j \to \eta_j \tau_j$, $h_j \to \eta_j h_j$. The sum last becomes:

$$\begin{aligned} &= \frac{1}{|\mathcal{C}|} \sum_{\underline{\eta} \in \mathcal{C}} \mathbb{E}_{\underline{\eta h}}[\prod_{j=1}^{n} e^{h_j \eta_j^2} \prod_{j=1}^{n} e^{-h_j \eta_j} \langle \tau_j \rangle \langle s_i \rangle \eta_j^2 \prod_{j=1}^{n} e^{h_j \eta_j - h_j}] \\ &= \frac{1}{|\mathcal{C}|} \sum_{\underline{\eta} \in \mathcal{C}} \mathbb{E}_{\underline{h}}[\langle \tau_i \rangle \langle s_i \rangle] \\ &= \mathbb{E}_{\underline{h}}[\langle s_i \rangle^2] \end{aligned} \tag{27}$$

To get the second equality we use channel symmetry. The last one is trivial because τ and s are dummy variables (so $\langle \tau \rangle = \langle s \rangle$). □

5 Relation Between Conditional Entropy and Free Energy

Often, in the large size limit $n \to \infty$ the free energy $\frac{1}{n} \ln Z(\underline{h})$ concentrates. It is therefore relevant to consider the *average free energy* over channel realizations:

$$\mathbb{E}_{\underline{\underline{h}}}[\frac{1}{n} \ln Z(\underline{\underline{h}})] \tag{28}$$

When the performance of the ensemble of codes has to be assessed one also consider a further averaging over the Gallager ensemble. The following result makes the connection with information theory:

Proposition 5.1. *We have the following relation*

$$H(\underline{X}|\underline{Y}) = \mathbb{E}_{\underline{\underline{h}}}[\ln Z(\underline{\underline{h}}] - n \int dh\, hc(h) \tag{29}$$

The last term depends only on the underlying channel: thus one may say that the average over channel outputs of the free energy and Shannon's conditional entropy are the same thing.

Proof in the case of a Gaussian channel. There are various ways to prove this relation. Here we show one that is valid for the BIAWGNC not because it is the simplest, but because it illustrates a nice use of Nishimori identities. The proof for general channels can be found in the literature.

First note that for the BIAWGNC the last term is equal to σ^{-2}.

$$\begin{aligned} H(\underline{X}|\underline{Y}) &= -\mathbb{E}_{\underline{Y}}[\sum_{\underline{s}} \mathbb{P}(\underline{s}|\underline{y}) \ln \mathbb{P}(\underline{s}|\underline{y})] \\ &= \mathbb{E}_{\underline{Y}}[lnZ(\underline{y})] - \mathbb{E}_{\underline{Y}}[\sum_{\underline{s}} \mathbb{P}(\underline{s}|\underline{y}) \ln \prod_{c\in\mathcal{C}} \frac{1}{2}(1 + \prod_{i\in c} s_i)] - \mathbb{E}_{\underline{Y}}[\sum_{\underline{s}} \mathbb{P}(\underline{s}|\underline{y})(\sum_{i=1}^{n} h_i s_i)] \\ &= \mathbb{E}_{\underline{Y}}[\ln Z(\underline{y})] - 0 - \sum_{i=1}^{n} \mathbb{E}_{\underline{Y}}[h_i \langle s_i \rangle] \end{aligned} \tag{30}$$

So it remains to show $\mathbb{E}_{\underline{Y}}[h_i\langle s_i\rangle] = \sigma^{-2}$, an identity that does not seems trivial at first sight. The trick is to use integration by parts and a Nishimori identity. Let us first consider the expectation over h_i.

$$\begin{aligned} \mathbb{E}_{\underline{h_i}}[h_i\langle s_i\rangle] &= \int c(h_i) h_i \langle s_i\rangle dh_i = \frac{1}{\sigma^2} \int \left(-\frac{\partial}{\partial h_i} c(h_i) + c(h_i)\right) \langle s_i\rangle dh_i \\ &= \frac{1}{\sigma^2} \int c(h_i) \frac{\partial}{\partial h_i} \langle s_i\rangle dh_i + \frac{1}{\sigma^2} \int c(h_i) \langle s_i\rangle dh_i \\ &= \frac{1}{\sigma^2} \mathbb{E}_{h_i}[\frac{\partial}{\partial h_i} \langle s_i\rangle + \langle s_i\rangle] \\ &= \frac{1}{\sigma^2} \mathbb{E}_{h_i}[\langle s_i^2\rangle - \langle s_i\rangle^2 + \langle s_i\rangle] \\ &= \frac{1}{\sigma^2} \mathbb{E}_{h_i}[1 - \langle s_i\rangle^2 + \langle s_i\rangle] = \sigma^{-2} \end{aligned} \tag{31}$$

Averaging over all other h_j's we find

$$\mathbb{E}_{\underline{h}}[h_i\langle s_i\rangle] = \sigma^{-2} \tag{32}$$

□

6 Performance Curves - Generalized MAP Exit function)

As seen in previous lectures, for usual spin systems, derivatives of the free energy with respect to magnetic fields or other sources yield the magnetization or other averages. First phase transitions are then identified as jump discontinuities (first order) of these averages. Second order phase transitions are identified as jump discontinuities in their derivatives (but these seem to be less relevant in coding (?)).

Here the control parameter is the noise intensity, call it ϵ, and it is a fruitful idea to look at derivatives of the conditional entropy/average free energy with respect to this parameter.

Proposition 6.1 (Generalized EXIT curve of a symetric channel)**.**

$$\begin{aligned}\frac{\partial}{\partial\epsilon}\frac{1}{n}H(\underline{X}|\underline{Y}) &= -\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}_{\underline{h}\setminus h_i}\int dh_i\frac{\partial}{\partial\epsilon}\{c(h_i)\}\ln\left(\frac{1-\langle s_i\rangle\tanh h_i}{1-\tanh h_i}\right)\\ &= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}_{\underline{h}\setminus h_i}\int dh_i\frac{\partial}{\partial\epsilon}\{c(h_i)\}\ln\left(\frac{1+\langle s_i\rangle_{h_i=0}\tanh h_i}{1+\tanh h_i}\right)dh_i\end{aligned}$$

When a phase transition between a decodable phase (small noise ϵ) and undecodable phase (large noise ϵ) is present it turns out that this function has a jump at the same threshold than the error probability. The proof will be done only for the BIAWGNC were everything becomes simpler. We refer to the literature fior the general case.

Remark 6.2. The first formula involves the soft bit estimate. In the coding theory literature it is often expressed in terms of the likelihood

$$H_i = \frac{1}{2}ln\frac{\mathbb{P}(s_i=+1|\underline{y})}{\mathbb{P}(s_i=-1|\underline{y})}, \qquad \tanh H_i = \langle s_i\rangle$$

The second formula involves the so-called extrinsic soft bit estimate (because $h_i = 0$). This is often expressed in terms of the extrinsic likelihood

$$H_i^0 = \frac{1}{2}ln\frac{\mathbb{P}(s_i=+1|\underline{y}\setminus y_i)}{\mathbb{P}(s_i=-1|\underline{y}\setminus y_i)}, \qquad \tanh H_i^0 = \langle s_i\rangle_{h_i=0}$$

For the BIAWGNC the formulas become much simpler and the Generalized EXIT curve is really an average magnetization. This is also essentialy the case for the BEC (see exercises).

Proposition 6.3. *For a BIAWGNC the above formula can be simplified to*

$$\frac{\partial}{\partial(1/\sigma^2)}H(X^n|Y^n) = \frac{1}{2n}\sum_{i=1}^{n}(\mathbb{E}_h[\langle s_i\rangle]-1) \tag{33}$$

Proof. The proof is again a nice application of integration by parts and Nishimori identities. For a BIAWGNC check the identity,

$$\frac{\partial}{\partial \sigma^{-2}} c(h) = (-\frac{\partial}{\partial h} + \frac{1}{2}\frac{\partial^2}{\partial h^2}) c(h) \tag{34}$$

Then proceeding similarly to beforehand intergartion by parts yields,

$$\frac{\partial}{\partial \sigma^{-2}} \mathbb{E}_h(lnZ) = \sum_{i=1}^{n} \mathbb{E}_{\underline{h}}[(+\frac{\partial}{\partial h_i} + \frac{1}{2}\frac{\partial^2}{\partial h_i^2}) lnZ] \tag{35}$$

Consequently,

$$\begin{aligned} \frac{\partial}{\partial \sigma^{-2}} \mathbb{E}_h(lnZ) &= \sum_{i=1}^{n} \mathbb{E}_{\underline{h}}[\langle s_i \rangle + \frac{1}{2}(\langle s_i^2 \rangle - \langle s_i \rangle^2] \\ &= \frac{1}{2} \sum_{i=1}^{n} \mathbb{E}_h(1 + \langle s_i \rangle) \end{aligned} \tag{36}$$

because of the Nishimori identity. Finaly use the relation between conditional entropy and average free energy to conclude. □

Second derivatives of $H(\underline{X}|\underline{Y})$ are related to correlation functions (covariances) $\langle s_i s_j \rangle - \langle s_i \rangle \langle s_j \rangle$. We do not mention them at this point but they may turn out to be useful later in the course.