Lecture 18
Density Evolution II

18.1. Introduction

We got familiar with the notion of *message passing algorithms* in the previous sessions, and saw the Gallager's algorithm A, (*Majority*), as a very basic example of this class of algorithms. We have also seen the *density evolution*, which is a useful method in order to investigate the performance of an iterative algorithm for a given ensemble of codes. The main task in this method is to find the probabilities of sending various messages through a random edge in any given round of the algorithm. Now, the question is how exact the probabilities computed by density evolution are. Namely, for a randomly chosen graph from the ensemble, G, and a random channel output, R, and a given iteration, ℓ, how far is the number of edges who carry the wrong messages, $U(G, R)$, from the value obtained using the density evolution method?

In this session which is a more technical one, we will find some bounds on the probability that a particular realization of $U(G, R)$ differ so much from what expected from the density evolution. We give the proof in two steps. First of all, we will find some bounds on the expectation of the number of incorrect messages sending along the edges. We will use the *depth ℓ tree-like* assumption in this step, which will be proven. In the second step, we will use a martingle argument and Azuma's inequality in order to show that the difference between a $U(G, R)$ for a particular choice of G and R does not differ so much from its expectation. We will finish the report by combining these two facts and proving the main theorem.

18.2. Preliminaries

Let G be a bipartite graph with n nodes on the left *(variable nodes)* and m nodes on the right hand side *(check nodes)*. Let also $\lambda(x)$ and $\rho(x)$ be the left and the right node degree distributions, respectively. We denote by E the number of edges in G, and by Δ_v and Δ_c the maximum node degrees of the variable and the check nodes, respectively.

Suppose that the all zero codeword of length n is transmitted over binary input memoryless channel and sequence R of the same length is received. We denote by $U(G, R)$, the number of incorrect messages among all E variable-to-check node messages passed at iteration ℓ of the algorithm. Let p be the expected probability of sending an incorrect message through a random edge which has been obtained by density evolution using the neighborhood tree-like assumption. In order to find the reliability of such assumption, we will show that for sufficiently large n, with high probability this neighborhood is tree-like.

Lecture by Amin Shokrollahi
Scribes by Soheil Mohajer

18.3. Tree Assumption

Let v be any particular variable node in the graph. We denote by $\mathcal{N}_v^\ell$ the neighborhood of depth ℓ of node v, where by *one* level depth, we mean the variable nodes which share an adjacent check node with v.

Lemma 18.1. *Let G be a random bipartite graph with n variable nodes. For constant maximum node degrees of the variable and the check nodes (independent of n) and a constant ℓ, the probability of a depth ℓ neighborhood of a variable node being tree-like is at least $1 - \frac{\gamma}{n}$ (for some constant γ), which goes to 1 as n increases.*

Proof. We will prove the lemma using a recursive argument. Indeed, we will find the probability of tree-like neighborhood in depth $t \leq \ell$, assuming that the depth $(t-1)$ neighborhood is tree-like.

Let Δ_v and Δ_c be the maximum node degree of the variable and the check nodes, respectively. For simplicity, we just consider regular graphs; for the irregular graphs, we can just add some edges in order to make all the variable (check) nodes of degree Δ_v (Δ_c), and just "increase" the probability of not being tree-like. So if we can upper bound this probability for regular graphs, it is also upper bounded for the irregular ones.

Note that there are

$$N_v(t) = \sum_{i=0}^{t} (\Delta_r - 1)^i (\Delta_l - 1)^i$$

variable nodes and

$$N_c(t) = \sum_{i=0}^{t-1} (\Delta_r - 1)^{i+1} (\Delta_l - 1)^i$$

check nodes in $\mathcal{N}_v^t$. Assume the depth $t-1$ neighborhood is tree-like and we have already revealed k additional edges in t-th level without creating a loop. Then, the probability of having no loop after revealing one more edge is

$$\begin{aligned} \frac{\#\text{remaining edges which do not create a loop}}{\#\text{remaining edges}} &= \frac{E - (N_c(t-1) + k)\Delta_c}{E - (N_c(t-1) + k)} \\ &= 1 - \frac{(N_c(t-1) + k)(\Delta_c - 1)}{E - (N_c(t-1) + k)} \\ &\geq 1 - \frac{N_c(\ell)}{m}. \end{aligned} \tag{18.1}$$

Thus, since the t-th level of neighborhood has $N_c(t) - N_c(t-1)$ edges, the conditional probability of tree-like neighborhood will be lower bounded by $(1 - \frac{N_c(\ell)}{m})^{(N_c(t) - N_c(t-1))}$. Using the chain rule, the whole probability that the chosen edges outgoing from the variable nodes do not create a loop, can be lower bounded by

$$(1 - \frac{N_c(\ell)}{m})^{N_c(\ell)}.$$

We can use a similar argument to bound the probability that the edges outgoing from the check nodes do not create a loop by

$$(1 - \frac{N_v(\ell)}{n})^{N_v(\ell)},$$

and so the total probability that the depth ℓ neighborhood being tree-like is lower bounded by

$$(1 - \frac{N_c(\ell)}{m})^{N_c(\ell)} (1 - \frac{N_v(\ell)}{n})^{N_v(\ell)},$$

which is greater than

$$1 - \frac{N_v^2(\ell) + \frac{\Delta_c}{\Delta_v} N_c^2(\ell)}{n}$$

for sufficiently large n. □

Remark 18.2. *A similar claim and proof can be also done for the neighborhood of an edge. We don't go through the details here, but may alternatively use both of the tree-like neighborhood assumptions.*

18.4. Concentration Around Tree-Like Case

We have already used density evolution to obtain the fraction of edges who carry an incorrect message in a given round of the algorithm, p. In this section we will see that for large enough n, the expected number of such edges is really close to the value obtained by density evolution.

Lemma 18.3. *Let G be a randomly chosen graph from the ensemble $\mathcal{G}(\rho, \lambda)$, and R be a random channel observation. Then, for sufficiently large n, the expected number of edges carrying incorrect messages is concentrated around the total number of edges in the graph times the probability of an edge with a tree-like neighborhood carrying an incorrect message, i.e.,*

$$| \mathbb{E}[U(G,R)] - Ep | \leq \frac{E\epsilon}{2}$$

for n greater than $2\gamma/\epsilon$.

Proof. We set an arbitrary ordering on the edges of the graph and denote by e_i, the i-th edge in this ordering. Define

$$U^{(i)}(G,R) := \mathbb{1}[\text{an incorrect message is passed along } e_i], \qquad i = 1, \ldots, E.$$

So,

$$\begin{aligned} \mathbb{E}[U(G,R)] &= \mathbb{E}[\sum_{i=1}^{E} U^{(i)}(G,R)] \\ &= E\,\mathbb{E}[U^{(1)}(G,R)], \end{aligned}$$

where the second equality follows from the fact that $U^{(i)}(G,R)$'s are identically distributed. Using the Bayes' rule, we have

$$\begin{aligned} \mathbb{E}[U^{(1)}(G,R)] &= \Pr[\text{an incorrect message passed along } e_1] \\ &= \Pr[\text{inc. mes. along } e_1 \mid \mathcal{N}^{\ell}_{e_1} \text{is tree-like}] \cdot \Pr[\mathcal{N}^{\ell}_{e_1} \text{is tree-like}] \\ &\quad + \Pr[\text{inc. mes. along } e_1 \mid \mathcal{N}^{\ell}_{e_1} \text{is not tree-like}] \cdot \Pr[\mathcal{N}^{\ell}_{e_1} \text{is not tree-like}]. \end{aligned}$$

The first term in RHS is exactly what we have already computed in density evolution and equals p. Using Lemma (18.1), we have $\Pr[\mathcal{N}^{\ell}_{e_1} \text{is not tree-like}] \leq \gamma/n$. Thus,

$$\mathbb{E}[U^{(1)}(G,R)] \geq p \cdot (1 - \frac{\gamma}{n}).$$

On the other hand, we have

$$\mathbb{E}[U^{(1)}(G,R)] \leq p \cdot 1 + 1 \cdot \frac{\gamma}{n}.$$

Therefore,

$$Ep(1 - \frac{\gamma}{n}) \leq \mathbb{E}[U(G,R)] \leq E(p + \frac{\gamma}{n})$$

and clearly,

$$| \mathbb{E}[U(G,R)] - Ep | < \frac{E\epsilon}{2} \tag{18.2}$$

for $n > \frac{2\gamma}{\epsilon}$. □

18.5. Concentration Around the Expectation

The next part of our task is to show that with high probability the number of incorrect messages passed along the edges concentrate around its expectation. In order to show that, we will use a martingale argument.

Lemma 18.4. *With high probability, the number of incorrect messages passed along the edges is concentrated around its expectation, i.e.,*

$$\Pr\left[U(G,R) - \mathbb{E}[U(G,R)]\right] \leq e^{-\beta n},$$

for some constant β.

Proof. We start with a completely random graph chosen from the ensemble $\mathcal{G}(\rho,\lambda)$ and a random channel output. Then in each step we reveal one more edge of the graph and find the expectation of number edges carrying an incorrect message. After E steps where the graph is completely determined, we will continue with the process and reveal the variable nodes value (channel observation) one-by-one and find the same expectation. Thus, after n more steps the channel output R is also determined and there is no more randomness.

More precisely, we define the sequence of random variables $\{U_i(G,R)\}_{i=0}^{i=E+n}$ as

$$\begin{aligned} U_0(G,R) &= \mathbb{E}_{G,R}[U(G,R)], \\ U_i(G,R) &= \mathbb{E}_{G,R}[U(G,R) \mid e_1,\ldots,e_i] \quad i=1,\ldots,E, \\ U_{E+j}(G,R) &= \mathbb{E}_{R,G}[U(G,R) \mid e_1,\ldots,e_E \,\&\, R_1,\ldots,R_j] \\ &= \mathbb{E}_{R}[U(G,R) \mid R_1,\ldots,R_j] \quad j=1,\ldots,n. \end{aligned} \tag{18.3}$$

Then $U_{E+n}(G,R) = U(G,R)$. It is easy to see that $\{U_i(G,R)\}_{i=0}^{i=E+n}$ form a *Doob Maringale* and the goal is to bound the difference between the first and the $(E+n)$-th random variable. We can use *Azuma*'s inequality to bound this difference, but before that, we need to show that the difference between any two consecutive random variables, $U_{i+1}(G,R)$ and $U_i(G,R)$, is bounded by some constant.

Let $\mathcal{G}(G,i) \subset \mathcal{G}(\rho,\lambda)$ be the set of graphs such that the first i edges are equal to the edges in G. Recall the process we have used for producing random graphs from the ensemble $\mathcal{G}(\rho,\lambda)$, wherein we had E sockets in both the left and right hand side, found a random matching for them, and merged the sockets in each side to come up with the node degree distributions. We denote by $\mathcal{G}_j(G,i)$ the subset of $\mathcal{G}(G,i)$ consisting of those graphs in which the $(i+1)$-th edge (connected to the $i+1$-th socket on the LHS) is connected to j-th check socket. Thus,

$$\mathcal{G}(G,i) = \bigcup_j \mathcal{G}_j(G,i)$$

and we have

$$\begin{aligned} U_i(G,R) &= \mathbb{E}[U(G,R) \mid G \in \mathcal{G}(G,i)] \\ &= \sum_j \mathbb{E}[U(G,R) \mid G \in \mathcal{G}_j(G,i)] \cdot \Pr[G \in \mathcal{G}_j(G,i) \mid G \in \mathcal{G}(G,i)] \end{aligned} \tag{18.4}$$

Now, let $G \in \mathcal{G}_j(G,i)$ and $G' \in \mathcal{G}_{j'}(G,i)$. We can easily convert any instantiation of $\mathcal{G}_j(G,i)$ to an instantiation of $\mathcal{G}_{j'}(G,i)$ using an exchanging in endpoints of two edges (if $(i+1,j)$ and (i',j') are two edges in $G \in \mathcal{G}_j(G,i)$, we just need to replace them with $(i+1,j')$ and (i',j)), and by any of these exchanges, we make change in a constant (depend on Δ_l, Δ_l, and ℓ, but not n) number of wrong messages passed along the edges. We know that an edge can contribute in at most some constant, let say $C_1 = C(\Delta_v,\Delta_c,\ell)$, of depth ℓ neighborhoods. We deal with four edges and so the change in $U(G,R)$ is upper bounded by $4C_1$, and so the difference of the expectations is also bounded. More precisely,

$$\left|\mathbb{E}[U(G,R) \mid G \in \mathcal{G}_j(G,i)] - \mathbb{E}[U(G',R) \mid G' \in \mathcal{G}_{j'}(G,i)]\right| \leq 4C_1.$$

Note that by definition of $U_{i+1}(G,R)$, it equals to $\mathbb{E}[U(G,R) \mid G \in \mathcal{G}_j(G,i)]$, for some j. Therefore,

$$\begin{aligned} |U_{i+1}(G,R) - U_i(G,R)| &\leq \max_j |\mathbb{E}[[U(G,R) \mid G \in \mathcal{G}_j(G,i)] - U_i(G,R)| \\ &\overset{(a)}{\leq} \max_{j,k} |\mathbb{E}[[U(G,R) \mid G \in \mathcal{G}_j(G,i)] - \mathbb{E}[[U(G',R) \mid G' \in \mathcal{G}_k(G,i)]| \\ &\leq 4C_1 \end{aligned} \tag{18.5}$$

where in (a) we have used the average expression for $U_i(G,R)$, given in (18.4). This proves the bound for the first E steps. For the next n steps, we can observe that revealing one more channel output, only the messages passed through the edges within the depth ℓ neighborhood can be changed. Thus, the change in the expected number of passed incorrect messages is at most $C_2 = \sum_{i=1}^{\ell} (\Delta_v \Delta_c)^i$. We can summarize these two facts by

$$|U_{i+1}(G,R) - U_i(G,R)| \leq C$$

where $C = \max\{4C_1, C_2\}$. Now, applying Azuma's inequality to the obtained martingale, we have

$$\begin{aligned} \Pr[|U(G,R) - \mathbb{E}[U(G,R)]| \geq \alpha] &= \Pr[|U_{E+n}(G,R) - U_0(G,R)| \geq \alpha] \\ &\leq 2e^{-\frac{\alpha^2}{2(E+n)C^2}} \end{aligned} \tag{18.6}$$

Set $\alpha = \frac{E\epsilon}{2}$, we have

$$\begin{aligned} \Pr[|U(G,R) - \mathbb{E}[U(G,R)]| \geq \frac{E\epsilon}{2}] &\geq 2e^{-\frac{E^2\epsilon^2/2}{2(E+n)C}} \\ &= 2e^{-\beta n} \end{aligned} \tag{18.7}$$

where $\beta = \frac{a^2\epsilon^2}{4(1+a)C}$ and a is the average node degree of the variable nodes. □

18.6. The Main Result

Now we are ready to prove the main theorem which guarantees that the density evolution is good enough for analyzing the convergence of the message passing algorithm over a sparse bipartite graph.

Theorem 18.5. *Let G be a randomly chosen graph from the ensemble $\mathcal{G}(\rho,\lambda)$ and R be the channel output which is also random. Further, let p be the probability of sending an incorrect message along an edge in the ℓ-th round of the algorithm, obtained by density evolution, and $U(G,R)$ be the exact number of such edges for a particular realization of G and R. Then, with high probability, $U(G,R)$ is close to Ep, i.e,*

$$\Pr[\mid U(G,R) - Ep \mid \geq E\epsilon] \leq 2e^{\beta n}.$$

Proof. We use the Lemmas (18.3) and (18.4) we have already proven.

$$\begin{aligned} \Pr[\mid U(G,R) - Ep \mid \geq E\epsilon] &\leq \Pr[\mid U(G,R) - \mathbb{E}[U(G,R)] \mid + \mid \mathbb{E}[U(G,R)] - Ep \mid \geq E\epsilon] \\ &\overset{(b)}{\leq} \Pr[\mid U(G,R) - \mathbb{E}[U(G,R)] \mid \geq \frac{E\epsilon}{2}] \\ &\overset{(c)}{\leq} 2e^{\beta n} \end{aligned} \tag{18.8}$$

where (b) follows from the Lemma (18.3) for $n > 2\gamma/\epsilon$, and (c) is exactly the result of Lemma (18.4). □