

Introduction to LDPC Codes

Lecture notes June 15-19 2003

Jeremy Thorpe

May 27, 2003

1 Summary

LDPC codes are linear codes over a finite field (henceforth F_2) can be described by a parity check matrix H . A code is a LDPC code if the average number of non-zero elements in each row of H is a small constant independent of the code's dimension n . The information contained in H can be conveniently expressed in a bipartite graph g , which contains variable nodes and check nodes. LDPC codes can be decoded by a fast, iterative algorithm called the Belief Propagation (BP) algorithm which is defined on g .

2 LDPC Codes

All linear codes over F_2 can be defined by their generator matrix G , or their parity check matrix H . LDPC codes are defined by their H matrix. A sequence of codes of fixed rate $R = k/n$ and increasing length n is a sequence of LDPC codes if the limit of the average weight in any row or column is a constant.

2.1 Tanner-Graph representation

A different representation of the parity-check matrix H is its so-called *Tanner graph*. A tanner graph is a bipartite graph in which nodes can be grouped into two types where every edge has one endpoint of one type and one endpoint of the other. The Tanner graph g corresponding to H has a node for each column of H (variable in the code), and each row of H (check equation in the code), and an edge connecting variable node i with check node j if and only if the $H_{ij} \neq 0$.

The code can be then defined in terms of g by: $C = \left\{ \vec{x} : \sum_{i:i\bar{j}} x_i = 0 \forall j \right\}$.

3 Probability

Probability distributions, likelihood functions, and Bayes' Rule are central to the understanding of the BP algorithm. Probability distributions of $X \in F_2$

can be characterized by a *probability difference* or alternatively by *probability ratio*, each of which is useful in performing different computations. Likelihood functions on $X \in F_2$ can be essentially characterized by *likelihood ratio* (down to a multiplicative constant that contains no useful information about X). The Bilinear transform takes a probability difference into a probability ratio and back.

3.1 Distributions

A *distribution* on a random variable X is a function $f : \mathcal{X} \rightarrow \mathbb{R}$ s.t. $p(X = x) = f(x)$, $\sum_x p(X = x) = 1$, that assigns a probability to each value x in the alphabet \mathcal{X} that the random variable X can take. As a matter of shorthand notation, when the random variable is clear, $p(x)$ is substituted for $p(X = x)$.

3.2 Likelihood Functions

A likelihood function on X is the probability $P(y|x)$ of some other observed variable Y for each possible value of X . Unlike a distribution, a likelihood function need not sum to 1. Graphically, the difference between probability function and a likelihood function can be understood by looking at a hypothetical table of $p(y|x)$ as follows:

$p(y x)$		
$y \backslash x$	0	1
a	.1	.6
b	.6	.2
c	.3	.2

it can be seen that while any column of the table gives a *conditional distribution* on Y (which sums to 1), any row of the table gives a *likelihood function* on X (which need not sum to 1).

3.3 Bayes' Rule

Bayes' Rule gives an explicit relationship between an a priori distribution $p(x)$, a likelihood function $p(y|x)$, and an a posteriori distribution $p(x|y)$. Bayes' Rule says:

$$p(x|y) = \frac{p(x) \cdot p(y|x)}{p(y)} \tag{1}$$

for fixed y :

$$p(x|y) \propto p(x) \cdot p(y|x) \tag{2}$$

A more explicit formulation of this is:

$$\frac{p(X = 1|y)}{p(X = 0|y)} = \frac{p(X = 1)}{p(X = 0)} \cdot \frac{p(y|X = 1)}{p(y|X = 0)} \tag{3}$$

3.3.1 Example

Suppose that the conditional distribution of Y given X is given by table 1. Suppose that the a priori distribution of X is $p(X = 0) = \frac{2}{3}, p(X = 1) = \frac{1}{3}$, and the observed value of Y is the a . The a priori probability ratio in (3) is $\frac{p(X=1)}{p(X=0)} = \frac{1}{2}$. The likelihood ratio $\frac{p(y|X=0)}{p(y|X=1)} = \frac{.6}{.1} = 6$. We compute the a posteriori probability ratio to be $\frac{1}{2} \times 6 = 3$. The a posteriori probability ratio is 3, which implies that $p(X = 1|y) = .75$.

3.4 Bayes Rule for independent evidence

replacing Y in equation (3) with Y_1, Y_2, \dots, Y_n :

$$\frac{p(X = 1|y_1, y_2, \dots, y_n)}{p(X = 0|y_1, y_2, \dots, y_n)} = \frac{p(X = 1)}{p(X = 0)} \cdot \frac{p(y_1, y_2, \dots, y_n|X = 1)}{p(y_1, y_2, \dots, y_n|X = 0)} \quad (4)$$

If the Y 's are independent conditioned on X :

$$p(y_1, y_2, \dots, y_n|x) = \prod_{i=1}^n p(y_i|x) \quad (5)$$

Which implies that

$$\frac{p(X = 1|y_1, y_2, \dots, y_n)}{p(X = 0|y_1, y_2, \dots, y_n)} = \frac{p(X = 1)}{p(X = 0)} \cdot \prod_{i=1}^n \frac{p(y_i|X = 1)}{p(y_i|X = 0)} \quad (6)$$

3.4.1 Example

Suppose the a priori probability and channel is defined as before, and we observe: $Y_1 = a, Y_2 = c, Y_3 = b$. We compute:

$$\begin{aligned} \frac{p(X = 1|y_1, y_2, \dots, y_3)}{p(X = 0|y_1, y_2, \dots, y_3)} &= \frac{p(X = 1)}{p(X = 0)} \cdot \prod_{i=1}^3 \frac{p(y_i|X = 1)}{p(y_i|X = 0)} \\ &= \frac{1}{2} \cdot \frac{.6}{.1} \cdot \frac{.2}{.3} \cdot \frac{.2}{.6} \end{aligned} \quad (7)$$

$$= \frac{2}{3} \quad (8)$$

Thus the a posteriori probability $p(X = 1|y_1, y_2, y_3) = .4$

3.5 Distribution of Sum

Suppose the distributions of several variables $X_i \in F_2$ are known. One required operation in BP is computing the distribution of $\sum_i X_i$. A brute force way to calculate this probability is to write:

$$P\left(\sum_i X_i = 1\right) = \sum_{x_i: \sum_i x_i = 1} \prod_{i'} P(X_{i'} = x_{i'})$$

This is inelegant, and worse, computationally expensive.

A better way begins with the observation that:

$$\begin{aligned} E\left((-1)^X\right) &= P(X=0) \cdot (-1)^0 + P(X=1) \cdot (-1)^1 \\ &= P(X=0) - P(X=1) \end{aligned} \tag{9}$$

characterizes the distribution of X .

Since $(-1)^{\sum_i X_i} = \prod_i (-1)^{X_i}$, it follows that for independent X_i 's,

$$\begin{aligned} E\left((-1)^{\sum_i X_i}\right) &= E\left(\prod_i (-1)^{X_i}\right) \\ &= \prod_i E\left((-1)^{X_i}\right) \end{aligned} \tag{10}$$

3.5.1 Example

Suppose we have $p(X_1=1) = 1$, $p(X_2=1) = .2$, $p(X_3=1) = .1$. Compute the distribution on the mod-2 sum $\sum_{i=1}^3 x_i$

$$\begin{aligned} E\left((-1)^{X_1}\right) &= P(X_1=0) - P(X_1=1) = -1 \\ E\left((-1)^{X_2}\right) &= .6 \\ E\left((-1)^{X_3}\right) &= .8 \end{aligned}$$

So

$$\begin{aligned} E\left((-1)^{\sum_i X_i}\right) &= -1 \cdot .6 \cdot .8 \\ &= -.48 \end{aligned}$$

We can invert to find $p\left(\sum_{i=1}^3 x_i = 1\right) = .74$

3.6 The Bilinear Transform

The bilinear transform relates quantities in (3) with those in (9). The bilinear transform is:

$$B(x) = \frac{1-x}{1+x}$$

It has the property that:

$$\begin{aligned} B\left(\frac{N}{D}\right) &= \frac{1 - \frac{N}{D}}{1 + \frac{N}{D}} \\ &= \frac{D - N}{D + N} \end{aligned}$$

And that:

$$\begin{aligned}
 B(B(x)) &= B\left(\frac{1-x}{1+x}\right) & (11) \\
 &= B\frac{(1+x) - (1-x)}{(1+x) + (1-x)} \\
 &= \frac{2x}{x} \\
 &= x
 \end{aligned}$$

It also has the property that it takes a probability ratio into a probability difference:

$$\begin{aligned}
 B\left(\frac{p(x=1)}{p(x=0)}\right) &= \frac{(px=0) - p(x=1)}{(px=0) + p(x=1)} & (12) \\
 &= (px=0) - p(x=1)
 \end{aligned}$$

and back:

$$B((px=0) - p(x=1)) = \frac{p(x=1)}{p(x=0)} \quad (13)$$

4 The Belief Propagation Algorithm

The Belief Propagation (BP) algorithm is an iterative algorithm which is defined on g . In this algorithm, the i th variable node computes probability distributions on code symbols X_i , and the j th node computes likelihood functions on code symbols X_i , where i is adjacent to j in g . BP is carried out for a number of iterations T (which may be fixed beforehand, or be decided at runtime by a *stopping rule*). For each time $t = 0, 1, \dots, T$, for each edge \overline{ij} , a message $m_{ij}^{(t)}$ is computed and "sent" from variable node i and "received" at check node j , and then for each edge \overline{ij} a message $m_{ji}^{(t)}$ is computed and "sent" from check node j and "received" at variable node i .

4.1 System Idealization

Ordinarily, an ECC decoder assumes that X has been picked uniformly at random from the code C . The BP decoder makes a formally different, but equivalent assumption. The BP decoder assumes that X is distributed iid uniform over F_2 , and that the sums $S_j = \sum_{i:\overline{ij}} x_i$ have all been observed to be 0. This

assumption is equivalent to the usual assumption, because once all the sums S_j are known to be 0, the distribution on $X|S$ is uniform over C ($x \notin C$ has a posteriori probability 0 because it has likelihood 0, and for $\{x : x \in C\}$, a posteriori probability is , because a priori distribution is uniform, and likelihood distribution is 1).

In addition, the decoder knows y which is assumed to have been produced by the channel with conditional probability $p(y_i|x_i)$, independent conditioned on X .

4.2 Ideal Message Definitions

The ideal message m_{ij} from the i th variable node to be passed to the j th check node is message about the distribution of X_i . The form of this message is:

$$m_{ij} \approx p\left(X_i = 0|y_i, S_{j'::j' \neq j, \bar{i}\bar{j}'} = 0\right) - p\left(X_i = 1|y_i, S_{j'::j' \neq j, \bar{i}\bar{j}'} = 0\right)$$

This message tells the probability distribution on X_i given the channel information y_i and the messages supplied from each adjacent node j' except for j itself. The message from node j itself is excluded because it is not independent of S_j , which would violate an assumption to be stated.

At time t , the j th check node computes message $m_{ji}^{(t)}$ to be passed to the i th variable node:

$$m_{ji} \approx \frac{p(S_j = 0|X_i = 1)}{p(S_j = 0|X_i = 0)}$$

This message gives the likelihood ratio of X_i , given the observation that $S_j = 0$.

4.3 Actual Message Definitions

Starting with the ideal definition of m_{ij} :

$$m_{ij} \approx p\left(X_i = 0|y_i, S_{j'::j' \neq j, \bar{i}\bar{j}'} = 0\right) - p\left(X_i = 1|y_i, S_{j'::j' \neq j, \bar{i}\bar{j}'} = 0\right)$$

By (13):

$$= B \left(\frac{p\left(X_i = 1|y_i, S_{j'::j' \neq j, \bar{i}\bar{j}'} = 1\right)}{p\left(X_i = 0|y_i, S_{j'::j' \neq j, \bar{i}\bar{j}'} = 0\right)} \right)$$

By (6), assuming appropriate conditional independence of all S_j :

$$= B \left(\frac{p(X_i = 1)}{p(X_i = 0)} \cdot \frac{p(y|x = 1)}{p(y|x = 0)} \cdot \prod_{j'::j' \neq j, \bar{i}\bar{j}'} \frac{p(S_{j'} = 0|X_i = 1)}{p(S_{j'} = 0|X_i = 0)} \right)$$

Since the a priori distribution on X_i is uniform:

$$= B \left(\frac{p(y|x = 1)}{p(y|x = 0)} \cdot \prod_{j'::j' \neq j, \bar{i}\bar{j}'} \frac{p(S_{j'} = 0|X_i = 1)}{p(S_{j'} = 0|X_i = 0)} \right)$$

Substituting $m_{j'i}$ for its definition:

$$= B \left(\frac{p(y|x = 1)}{p(y|x = 0)} \cdot \prod_{j'::j' \neq j, \bar{i}\bar{j}'} m_{j'i} \right)$$

To put this in the context of the *iterative* message passing algorithm, we let $m_{ij}^{(t)}$ be expressed in terms of $m_{j'i}^{(t-1)}$.

$$m_{ij}^{(t)} = B \left(\frac{p(y|x=1)}{p(y|x=0)} \cdot \prod_{j':j' \neq j, \bar{i}\bar{j}'} m_{j'i}^{(t-1)} \right) \quad (14)$$

Starting with the ideal definition of m_{ji} :

$$m_{ji} = \frac{p(S_j = 0|X_i = 1)}{p(S_j = 0|X_i = 0)}$$

By(12):

$$= B(p(S_j = 0|X_i = 0) - p(S_j = 0|X_i = 1))$$

Since $S_j = 0$ if and only if $\sum_{i':i' \neq i, \bar{i}\bar{j}} X_{i'} = X_i$:

$$= B \left(p \left(\sum_{i':i' \neq i, \bar{i}\bar{j}} x_{i'} = 0 \right) - p \left(\sum_{i':i' \neq i, \bar{i}\bar{j}} x_{i'} = 1 \right) \right)$$

By (10), and assuming of independence of all X_i .

$$= B \left(\prod_{i':i' \neq i, \bar{i}\bar{j}} p(x_{i'} = 0) - p(x_{i'} = 1) \right)$$

Substituting the ideal definition of $m_{i'j}$:

$$= B \left(\prod_{i':i' \neq i, \bar{i}\bar{j}} m_{i'j} \right)$$

Again, putting this in the context of iterative message passing, $m_{ji}^{(t)}$ is expressed in terms of $m_{i'j}^{(t)}$:

$$m_{ji}^{(t)} = B \left(\prod_{i':i' \neq i, \bar{i}\bar{j}} m_{i'j}^{(t)} \right) \quad (15)$$

4.4 Initialization and Final Decision

Since the messages from variable to check $m_{ij}^{(t)}$ depend on $m_{j'i}^{(t-1)}$, define $m_{ji}^{(-1)}$ is defined as the message that does not bias the computation at the variable node, namely the likelihood ratio 1.

In making the final decision, the following quantity is computed:

$$\begin{aligned} & \frac{p\left(X_i = 1|y_i, S_{j:\bar{i}j} = 0\right)}{p\left(X_i = 0|y_i, S_{j:\bar{i}j} = 0\right)} \\ & \approx \frac{p(y|x=1)}{p(y|x=0)} \cdot \prod_{j:\bar{i}j} m_{ji}^{(T)} \end{aligned}$$

If the quantity is bigger than 1, then the algorithm declares that $X_i = 1$ is more probable (a posteriori) and hence $\hat{x}_i = 1$, and otherwise $\hat{x}_i = 0$.

5 Statement of the BP Algorithm

The BP algorithm can be briefly stated as follows:

For each $\bar{i}j$ in g , set:

$$m_{ji}^{(-1)} = 1$$

For $t = 0, 1, \dots, T$
for each $\bar{i}j$ in g , set:

$$m_{ij}^{(t)} = B \left(\frac{p(y|x=1)}{p(y|x=0)} \cdot \prod_{j':j' \neq j, \bar{i}j'} m_{j'i}^{(t-1)} \right)$$

and then for each $\bar{i}j$ in g , set:

$$m_{ji}^{(t)} = B \left(\prod_{i':i' \neq i, \bar{i}'j} m_{i'j}^{(t)} \right)$$

Finally, for each $i \in \{1, 2, \dots, n\}$ set:

$$\hat{x}_i = \begin{cases} 0 & : \frac{p(y|x=1)}{p(y|x=0)} \cdot \prod_{j:\bar{i}j} m_{ji}^{(T)} < 1 \\ 1 & : \text{otherwise} \end{cases}$$

6 Discussion

There is a theorem which says that if BP is run for T iterations, and the neighborhood including all nodes less than or equal to $2T$ around the i th variable node has no loops, then the BP algorithm computes the exact distribution on X_i given the evidence in that neighborhood. The key points are that all of the assumptions about conditional evidence hold, and all of the evidence in that neighborhood is included in the computation (exactly once).

However, the BP algorithm is rarely run for such a short number of iterations that the above theorem applies. There is considerable variance in professional opinion, but it is quite typical to run BP for 20-50 or so iterations, even when the maximum number permitted by the theorem for most nodes is 3. In almost all cases, number of iterations is determined by engineering considerations with the belief that performance only gets (marginally) better with increasing number of iterations.