Faster convergence rates for error probabilities of message passing decoders for low-density parity-check codes

Peter L. Bartlett
Department of Statistics and Division of Computer Science
University of California, Berkeley
367 Evans Hall, CA 94720-3860, USA
bartlett@stat.berkeley.edu

Evan Greensmith
Research School of Information Sciences
Australian National University
ACT 0200, Australia
evan@discus.anu.edu.au

July 9, 2002

Abstract

We consider the convergence of the error probability of low-density parity-check codes under message passing decoding, as a function of the length of the code. Suppose that the number of decoding iterations is fixed. We show that, for almost all bipartite graphs defining a LDPC code and almost all received messages, the error probability decreases to a constant times its asymptotic value at the rate $\log n/n$, where n is the length of the code.

1 Background

A (d_v, d_c)-regular LDPC code of length n is a binary linear code for which, when the code is expressed as the set of solutions x to the parity-check equation $Hx = 0$, the $m \times n$ matrix H (where $m = nd_v/d_c$) has exactly d_v ones in each column and d_c ones in each row. For each parity-check matrix H, we can construct a bipartite graph, with n variable nodes (one for each bit of the codeword) and m check nodes (one for each parity-check equation). An edge between a variable node and a check node corresponds to a non-zero entry in the parity-check matrix H. Every variable node has degree d_v, and every check node has degree d_c. Thus, the set of (d_v, d_c)-regular LDPC codes of length n corresponds to a set $\mathcal{G}(d_v, d_c)$ of bipartite graphs of this type.

Define $N = d_v n = d_c m$. We can represent each graph as a permutation on the set $[N] = \{1, \ldots, N\}$, as follows: Partition the set $[N]$ into n distinct subsets of size d_v, and assign each of these subsets to a distinct variable node. (Think of the subset as a set of d_v 'variable sockets' connected to the variable node.) For $i \in [nd_v]$, let $v(i)$ be the variable node corresponding to the subset containing i. Similarly, partition $[nd_v] = [md_c]$ into m distinct subsets of size d_c, and assign each subset to a distinct check node. (Again, think of the subset as a set of d_c 'check sockets' connected to the check node.) For $j \in [md_c]$, let $c(i)$ be the check node corresponding to the subset containing j. Then for any permutation π on $[nd_v]$, construct a bipartite graph in $\mathcal{G}(d_v, d_c)$ by placing an edge between variable node $v(i)$ and check node $c(j)$ if and only if $\pi(i) = j$. (That is, the edge connects the variable node and the check node via variable socket i and check socket j.)

Luby, Mitzenmacher, Shokrollahi and Speilman [2] consider LDPC codes for which this permutation is chosen uniformly at random. This corresponds to a particular probability distribution over the set $\mathcal{G}(d_v, d_c)$. They analyzed the binary symmetric channel and binary erasure channel with a one-bit message passing algorithm for a random code, and showed that the performance of the decoder is concentrated around its expected value. That is, if the number of decoding rounds is fixed, for almost all randomly chosen codes and almost all channel realizations, the error probability for a code of length n is close to the asymptotic error probability as n grows large.

Richardson and Urbanke [3] extended this result to broad classes of memoryless channels with a binary input alphabet and of message-passing decoders. They considered channels and decoders satisfying certain symmetry assumptions (which ensure, for example, that the channel output is i.i.d.), and developed a general approach to proving conditions under which the error probability converges to zero as the number of decoding iterations increases. The key condition on the code and its underlying graph was that the neighborhood of a particular edge e connecting a variable node to a check node is 'tree-like'. For a directed edge $e = (v, c)$ in a graph, the directed neighborhood of depth d of e, denoted $\mathcal{N}_e^d$, is defined as the subgraph containing all directed paths $e_1, \ldots, e_d$ of length d starting from v but with $e_1 \neq e$. We say $\mathcal{N}_e^d$ is tree-like if it contains no cycles.

Richardson and Urbanke [3] showed for a variety of channels and decoders that, provided the channel parameters are suitably favourable, for a fixed graph G (and hence code) and a fixed variable-to-check edge in G whose directed neighborhood of depth $2l$ is tree-like, the probability that the message passed over e at decoding iteration l is incorrect converges to zero as l increases. They did not consider rates of convergence, but it is easy to see that, for many channels and decoding algorithms, there is an exponential bound on the probability. For example, for discrete message alphabets, if we denote the probability that the message passed over e at iteration l is k by p_k^l, then the reasoning in [3] shows that there is a bound of the form

$$p_k^l \le f_k(p_1^{l-1}, \ldots, p_K^{l-1}),$$

for some function f_k. For incorrect messages and suitably favourable channels, this probability converges to zero. In that case, we can find an upper bound for f_k that is a polynomial with all monomials containing every p_i^{l-1} with degree at least 1. This implies that, for some constants $c_1, c_2 > 1$ (which depend on d_v, d_c and properties of the channel),

$$\Pr\left(\text{message passed over } e \text{ at iteration } l \text{ is incorrect}\right) \le c_1 c_2^{-l},$$

where the probability is over the message received at the decoder. They also proved the following two results. (The constants are slightly modified to simplify the presentation, but they follow from their proof.)

Lemma 1.1 *For a fixed variable node socket and randomly chosen graph G, the edge e_1 from that socket has*

$$\Pr\left(\mathcal{N}_e^{2l} \text{ not tree-like}\right) \le \left(\frac{d_v d_c}{d_v d_c - d_v - d_c}\right)^2 \frac{(d_v^2 d_c^2)^l}{n}.$$

Lemma 1.2 *Let Z_l be the number of messages passed over all variable-to-check node edges at decoding iteration l that are incorrect. If the channel output is i.i.d. (which follows from the symmetry conditions if the channel is memoryless), then with probability at least $1 - \delta$ (over the random choice of the graph and the received message),*

$$\frac{Z_l}{n d_v} \le \frac{\mathbb{E} Z_l}{n d_v} + \sqrt{\frac{16(d_v + 1)(d_v d_c)^l \ln(2/\delta)}{n d_v^2}}.$$

This result (and the more specific Theorem 1 from [2] shows that the error probability converges to its asymptotic value at least as quickly as $1/\sqrt{n}$, where n is the code length. This result is a consequence of the Hoeffding-Azuma inequality, which shows that certain martingale processes are concentrated about their expectation. This inequality, which uses only boundedness of the martingale differences, cannot give a faster rate of convergence than $1/\sqrt{n}$. However, if the expectation of the error probability is small, we should expect a faster rate of convergence. In this note, we use recent concentration results due to Boucheron, Lugosi and Massart [1] to prove that the convergence occurs at the faster rate of $\log n/n$.

2 Faster Concentration

The following lemma is the main result of this note. It improves on the concentration result from [3] (Lemma 1.2).

Lemma 2.1 *If the channel output is i.i.d., with probability at least $1-\delta$ (over the random graph and the received message), the mistake rate on the lth decoding round satisfies*

$$\frac{Z_l}{nd_v} \leq 2\frac{\mathbb{E}Z_l}{nd_v} + \frac{c(d_v d_c)^{3l}\ln(2ed_v n/\delta)\ln(2/\delta)}{nd_v}.$$

To illustrate the consequences of this improvement, consider how the two results differ in the rate at which the mistake rate converges to zero when the number of decoding iterations is chosen to optimize the bound. For the upper bounds from [3], we obtain the following estimate for this rate.

Theorem 2.2 *Suppose the channel and decoding algorithm are such that*

- *The channel output is i.i.d., and*
- *There are constants $c_1, c_2 > 1$ such that, for any graph G and any fixed variable-to-check edge e in G that has neighborhood $\mathcal{N}_e^{2l}$ tree-like, the probability that the message passed over e at decoding iteration l is incorrect is no more than $c_1 c_2^{-l}$.*

Then there is a constant c_3 such that, with probability at least $1-\delta$,

$$\frac{Z_l}{nd_v} \leq c_1 c_2^{-l} + c_3 (d_v d_c)^l \sqrt{\frac{\ln(2/\delta)}{n}}.$$

Setting l to

$$l_n = \frac{\beta}{\ln(d_v d_c)} \ln\left(\sqrt{\frac{n}{\ln(2/\delta)}}\frac{1-\beta}{\beta}\right)$$

where

$$\beta = \frac{\ln(d_v d_c)}{\ln(c_2 d_v d_c)},$$

shows that, for some constant c_4, with probability at least $1-\delta$,

$$\frac{Z_{l_n}}{nd_v} \leq c_4 \left(\frac{\ln(2/\delta)}{n}\right)^{(1-\beta)/2}.$$

Proof: The first inequality follows from the combination of Lemmas 1.1 and 1.2, and from an estimate of the form

$$\frac{(d_v^2 d_c^2)^l}{n} + \sqrt{\frac{(d_v d_c)^l \ln(2/\delta)}{n}} \leq 2\sqrt{\frac{(d_v^2 d_c^2)^l \ln(2/\delta)}{n}},$$

whenever the first expression is smaller than 1.

■

Using Lemma 2.1 gives the following improvement.

Theorem 2.3 *Suppose the channel and decoding algorithm satisfy the same conditions as in Theorem 2.2. Then there is a constant c_3 such that, with probability at least $1 - \delta$,*

$$\frac{Z_l}{nd_v} \leq c_1 c_2^{-l} + c_3 (d_v^3 d_c^3)^l \frac{\ln(n/\delta) \ln(2/\delta)}{n}.$$

Setting l to

$$l_n = \frac{\beta}{\ln(d_v d_c)} \ln \left(\frac{n}{\ln(2/\delta)} \frac{1-\beta}{\beta} \right)$$

where

$$\beta = \frac{\ln(d_v^3 d_c^3)}{\ln(c_2 d_v^3 d_c^3)},$$

shows that, for some constant c_4, with probability at least $1 - \delta$,

$$\frac{Z_{l_n}}{nd_v} \leq c_4 \left(\frac{\ln(n/\delta) \ln(2/\delta)}{n} \right)^{1-\beta}.$$

3 Proof

The following concentration inequality can give tighter bounds than the Hoeffding-Azuma inequality, provided that we have additional variance information (beyond boundedness) about the relevant random variables. It is due to Boucheron, Lugosi, and Massart [1]. Notice that the result is applicable to functions of independent random variables, rather than martingales, which is a more restrictive setting.

Theorem 3.1 *Suppose $f : \mathbb{R}^n \to \mathbb{R}$, and $X_1, \ldots, X_n, X_1', \ldots, X_n'$ are independent random variables, with X_i and X_i' identically distributed. Define*

$$\begin{aligned} Z &= f(X_1, \ldots, X_n) \\ Z^{(i)} &= f(X_1, \ldots, X_{i-1}, X_i', X_{i+1}, \ldots, X_n). \end{aligned}$$

Notice that Z and $Z^{(i)}$ have the same distribution. Define

$$V_+ = \mathbb{E} \left[\sum_{i=1}^n (Z - Z^{(i)})^2 \left[\!\left[Z > Z^{(i)} \right]\!\right] | X_1^n \right],$$

where $X_1^n = (X_1, \ldots, X_n)$ and $[\![\phi]\!]$ takes value 1 when the predicate ϕ is true.

If there are positive constants a, b such that, almost surely, $V_+ \leq aZ + b$, then for all $t > 0$,

$$\Pr\left(Z > \mathbb{E}Z + t\right) \leq \exp\left(\frac{-t^2}{4a\mathbb{E}Z + 4b + 2at} \right).$$

For LDPC codes, we need to choose a random bipartite graph with N edges, where $N = d_v n$ (d_v is the out-degree of the variable nodes, and n is the number of variable nodes). We can construct this graph using a permutation of $\{1, \ldots, N\}$ chosen uniformly at random, and to construct this, we can use a sequence of N independent random variables. In particular, for $1 \le i \le N$, let X_i be chosen uniformly on $\{i, \ldots, N\}$ and define $p_0(i) = i$. Then for $j = 1, \ldots, N$, define

$$p_j(i) = \begin{cases} p_{j-1}(X_j) & \text{if } i = j \\ p_{j-1}(j) & \text{if } i = X_j \\ p_{j-1}(i) & \text{otherwise} \end{cases}$$

That is, we start out with the identity permutation p_0, and in going from p_{j-1} to p_j we use X_j to introduce a random swap (of the jth and X_jth elements). At the end, p_N is uniform on the set of permutations. To see this, notice that $p_i(j) = p_N(j)$ for $j \le i$, and that $p_N(i) = p_i(i)$ is chosen uniformly from among those $1 \le j \le N$ not in the set

$$\{p_N(1), \ldots, p_N(i-1)\}.$$

The function f we'll consider is the number of mistakes (incorrect messages) that occur on the variable-to-check edges in the lth decoding round. For now, we fix the input to the decoder and consider only the effect of the random variables $X_1, \ldots, X_N$ that determine the graph.

Consider the random graph $G = g(X_1, \ldots, X_N)$ that corresponds to a particular choice of this random permutation. That is, we fix an order on the $d_v n$ outgoing edge sockets from the variable nodes, and fix an order on the $d_c m$ incoming edge sockets to the check nodes, and construct the random graph by connecting variable socket i to check socket $p_N(i)$. Consider the change as we go from this graph to $G^{(i)} = g(X_1, \ldots, X_i', \ldots, X_N)$. In particular, to apply the concentration theorem, we need to consider the expectation (under the choice of X_i') of

$$\left(Z - Z^{(i)}\right)^2 \left[\!\left[Z > Z^{(i)}\right]\!\right].$$

This is an expectation over a set of $N-i+1$ graphs. Each graph is generated by starting with G, picking a pair of edges that join the variable sockets j_i, j_i' to the check sockets $p_N(j_i), p_N(j_i')$ (call these edges e_i, e_i'), swapping the check sockets between these edges, and then swapping the check socket $p_N(j_i')$ (from the edge connecting j_i to $p_N(j_i')$) with the check socket of another edge. Obviously, this set includes the graph G. (It is easy to see from the definitions how these quantities depend on the random variables $X_1, \ldots, X_N$ and X_i'. In particular, $j_i = p_N^{-1}(p_{i-1}(i))$, $j_i' = p_N(p_{i-1}(X_i))$, $p_N(j_i) = p_{i-1}(i)$, $p_N(j_i') = p_{i-1}(X_i)$, and the new edge has check socket chosen uniformly from $\{p_{i-1}(i), \ldots, p_{i-1}(N)\}$. In general, e_i and e_i' might be the same edge, but this doesn't affect the argument.)

Now, let B be a bound on the number of edges in the l-neighborhood of a particular edge. We have $B \le c(d_v d_c)^l$ for some constant c. Then, for the random graph G, some Z edges have mistakes. We say that an edge e of G is *bad* if it is in an l-neighborhood of an edge with a mistake. So the number k of bad edges of G is no more than ZB. Clearly, to have the number of mistakes decrease when we move from graph G to graph $G^{(i)}$ (that is, to have $Z > Z^{(i)}$), we must change at least one of the k bad edges.

For each i, we say that the random graph $G^{(i)}$ is bad given G if either of the edges e_i or e'_i are bad. (Notice that this is a property of the random variables $X_1, \ldots, X_N$ and not of X'_i.) If the graph $G^{(i)}$ is not bad given G, only k of the $N - i + 1$ possible graph values that $G^{(i)}$ could take could cause a decrease of the mistakes in G, and so in that case

$$\mathbb{E}\left((Z - Z^{(i)})^2 \left[\left[Z > Z^{(i)}\right]\right] |X_1^N\right) \leq c\frac{k}{N - i + 1}B^2.$$

(This is true because even when we change an edge near a mistake, we cannot change $Z - Z^{(i)}$ by more than a constant times B when we switch the edges.) If $G^{(i)}$ is bad given G, we cannot rule out $Z > Z^{(i)}$ for any value of X'_i, and so the strongest statement we can make for such a graph is

$$\mathbb{E}\left((Z - Z^{(i)})^2 \left[\left[Z > Z^{(i)}\right]\right] |X_1^N\right) \leq cB^2,$$

for some c.

Suppose that M of the N graphs $G^{(i)}$ are bad given G. Then we have shown that

$$\begin{aligned}
\mathbb{E}\left[\sum_{i=1}^{N}(Z - Z^{(i)})^2 \left[\left[Z > Z^{(i)}\right]\right] |X_1^N\right] &\leq c\left(M + k\sum_{i=1}^{N}\frac{1}{N - i + 1}\right)B^2 \\
&= c\left(M + k\sum_{i=1}^{N}\frac{1}{i}\right)B^2 \\
&\leq c\left(M + k\ln(e(N-1))\right)B^2.
\end{aligned}$$

Lemma 3.2 *With probability at least* $1 - \delta$*,*

$$M \leq 7\ln(eN/\delta)k.$$

Proof: Say that check socket j is swapped by X_i if $j \in \{p_{i-1}(i), p_{i-1}(X_i)\}$, that is, if the random variable X_i is such that the permutations p_{i-1}^{-1} and p_i^{-1} have different values at j. Let n_j be the number of X_is that swap j. We show that, with probability at least $1 - Ne^{-t/6}$, $\max_j n_j$ is no more than $2\ln(e^2(N-1)) + t$. When this occurs, we have

$$M < (2\ln(e^2(N-1)) + t)k.$$

Rearranging and making some approximations gives the result.

Notice that permuting the check socket labels does not change the distribution of $\max_j n_j$, so it has the same distribution as $\max_j n'_j$, with n'_j defined as follows. Start with a multiset S_0 containing N copies of 0, and proceed as follows. At step j, pick two elements $a_j, b_j \in S_{j-1}$ uniformly at random. Set

$n'_j = a_j + 1$, and if a_j and b_j are the same element, set S_j to S_{j-1} with the element a_j removed, otherwise set S_j to S_{j-1} with a_j removed and with b_j replaced by the element $b_j + 1$.

Now, let V_j, W_j be the elements a_j, b_j of S_{j-1} chosen at the jth step. Fix $V_1, \ldots, V_N$ and j, and for the random variable $V = n'_j$, let $V^{(i)}$ be defined using an independent copy W'_i of W_i. Then set

$$V_+ = \mathbb{E}\left(\sum_{i=1}^{j} (V^{(i)} - V)^2 \left[\!\left[V^{(i)} < V \right]\!\right] |W_1^j \right).$$

Then it is easy to see that there are no more than V values of i for which changing W_i to W'_i can lead to a decrease, and for these values, the change can be no more than 1. Thus, $V_+ \leq V$, and the concentration theorem implies that

$$\Pr\left(n'_i > 2\mathbb{E}n'_i + t\right) \leq e^{-t/6}.$$

Since this is true for any $V_1, \ldots, V_N$, the average over these random variables of this probability is similarly bounded. Furthermore, for any i,

$$\begin{aligned} \mathbb{E}n'_i &= 1 + \sum_{j=1}^{i-1} \frac{1}{N-j} \\ &\leq 1 + \ln(e(N-2)). \end{aligned}$$

It follows that

$$\Pr\left(\max_i n'_i > 2\ln(e^2(N-2)) + t\right) \leq Ne^{-t/6}.$$

■

Combining shows that, with probability at least $1 - \delta$,

$$\mathbb{E}\left[\sum_{i=1}^{N} (Z - Z^{(i)})^2 \left[\!\left[Z > Z^{(i)} \right]\!\right] |X_1^N \right] \leq cB^3 \ln\left(\frac{eN}{\delta}\right) Z.$$

Now consider the effect of the received messages. Since the n inputs to the decoder, $Y_1, \ldots, Y_n$, are independent, we can use a similar analysis. Let $\tilde{Z}^{(i)}$ denote the number of mistakes made on any variable-to-check edge at decoding round l when the ith input is changed, for $1 \leq i \leq n$. Clearly, $\tilde{Z}^{(i)} < Z$ only if the ith input arrives at a variable node in an l-neighborhood of a mistake (that is, if the variable node is bad). There can be no more than k such nodes, and for each, changing the input can decrease Z by no more than B. Together with the previous inequality, this implies

$$\begin{aligned}
V^+ &= \mathbb{E}\left[\sum_{i=1}^{N}(Z - Z^{(i)})^2 \left[\left[Z > Z^{(i)}\right]\right] |X_1^N\right] \\
&\quad + \mathbb{E}\left[\sum_{i=1}^{n}(Z - \tilde{Z}^{(i)})^2 \left[\left[Z > \tilde{Z}^{(i)}\right]\right] |Y_1^N\right] \\
&\leq cB^3 \ln\left(\frac{eN}{\delta}\right) Z + B^2 k \\
&\leq cB^3 \ln\left(\frac{eN}{\delta}\right) Z.
\end{aligned}$$

Applying the concentration theorem, with $t = \mathbb{E}Z + \epsilon N$, gives

$$\begin{aligned}
&\Pr\left(\frac{Z}{N} > 2\mathbb{E}\frac{Z}{N} + \epsilon\right) \\
&\leq \exp\left(\frac{-c(\mathbb{E}Z + \epsilon N)^2}{B^3 \ln(eN/\delta)(\mathbb{E}Z + \epsilon N)}\right) + \delta \\
&= \exp\left(\frac{-c(\mathbb{E}Z + \epsilon N)}{B^3 \ln(eN/\delta)}\right) + \delta \\
&\leq \exp\left(\frac{-c\epsilon N}{B^3 \ln(eN/\delta)}\right) + \delta,
\end{aligned}$$

and this is no more than 2δ for

$$\epsilon \geq \frac{cB^3 \ln(eN/\delta) \ln(1/\delta)}{N}.$$

References

[1] S. Boucheron, G. Lugosi, and P. Massart. Concentration inequalities using the entropy method. Technical report, CNRS-Université Paris-Sud, 2002.

[2] M. G. Luby, M. Mitzenmacher, M. A. Shokrollahi, and D. A. Spielman. Analysis of low density codes and improved designs using irregular graphs. *IEEE Transactions on Information Theory*, 47(2):585–598, February 2001.

[3] Tom Richardson and Rüdiger Urbanke. The capacity of low-density parity-check codes under message-passing decoding. *IEEE Transactions on Information Theory*, 47(2):599–681, February 2001.