Relations between Belief Propagation on Erasure and Symmetric Channels (Extended Abstract)

Omid Etesami *
Sharif University of Technology
Department of Computer Engineering
etesami@ce.sharif.edu

November 30, 2003

Abstract

We will prove an upper bound on the performance of belief propagation in decoding an ensemble of codes on a symmetric channel, given the threshold of the ensemble on the erasure channel. We will show a duality between this bound and a lower bound known for the thresholds of belief propagation decoding on symmetric channels [1]. We will also show that if a raptor code is optimized for the erasure channel, then a fixed reception overhead is enough for its decoding on arbitrary symmetric channels.

1 Introduction

The *belief propagation* (BP) algorithm has succeeded in decoding graphical codes at rates very close to Shannon capacity [2]. The asymptotic performance of BP for many classes of codes can be determined by *density evolution* [3]. However, except for the erasure channel [4], theoretical results about density evolution are not satisfactory.

The accessibility of density evolution on erasure channels motivates exploring how the performance of BP over different channels relates to its performance over the erasure channel. In this paper we will show that for a code ensemble the BP decoding threshold on the binary erasure channel gives an upper bound on the success of BP decoding on binary-input symmetric channels. To derive this result we will assume that BP runs almost cycle-free, which allows us to use density evolution.

It is too much to expect that tight estimates of the performance of BP on a general channel be obtained from decoding thresholds on the erasure channel. However, we will show that the bound reveals the exact reception overhead needed for successful decoding of a sequence of raptor codes on an erasure channel, given that the sequence achieves the capacity of a symmetric channel.

*Part of this work was done while the author was a summer student at École Polytechnique Fédérale de Lausanne.

The upper bound is similar to a lower bound proved by Khandekar in [1] on the performance of BP on symmetric channels.

Here is an outline of this paper: In Section 2 we will review the basics of belief propagation and density evolution for symmetric channels. In Section 3 we will derive our main result that upper-bounds the decoding performance of BP over symmetric channels. We will also state its dual, the lower bound. In Section 4 we will apply the results of Section 3 to raptor codes that are used as fountain codes on symmetric channels.

2 Belief Propagation and Density Evolution

In this section we describe the BP algorithm using the terminology of *low-density parity-check* (LDPC) codes [5]. An LDPC code can be visualized as a bipartite graph with two sets of nodes called the variable nodes and the check nodes. Variable nodes correspond to codeword bits, and the sum of variable nodes adjacent to each check node has an even parity in LDPC codewords.

At each iteration of BP, nodes communicate with their neighbors through edges. The message on an edge conveys the belief of the sender about the value of the variable node incident to that edge, given the incoming messages of the sender at the previous iteration. This belief is expressed in the form of a *log-likelihood ratio* (LLR), where the LLR of a binary random variable x is defined as $\ln(\Pr[x=0]/\Pr[x=1])$.

Let v be a variable node and let c be a check node adjacent to v. Suppose that m_v denotes the LLR that has been obtained for v from the channel. Further suppose that $\mathrm{m}_{vc}^{(\ell)}$ ($\mathrm{m}_{cv}^{(\ell)}$) denotes the message passed from v to c (from c to v) at the ℓth iteration of the algorithm. It is clear that for the zeroth iteration we have $\mathrm{m}_{vc}^{(0)} = \mathrm{m}_v$. For the next iterations the outgoing messages are given by

$$\begin{aligned} \mathrm{m}_{vc}^{(\ell)} &= \mathrm{m}_v + \sum_{c' \neq c} \mathrm{m}_{c'v}^{(\ell-1)} \\ \tanh(\frac{\mathrm{m}_{cv}^{(\ell)}}{2}) &= \prod_{v' \neq v} \tanh(\frac{\mathrm{m}_{v'c}^{(\ell)}}{2}). \end{aligned}$$

To guarantee that extrinsic information is only passed along, the above sum is over all neighbors of v other than c, and the above product is over all neigbors of c other than v.

At iteration ℓ we can decode a variable node v by looking at the sign of the sum of m_v and all LLRs communicated to v at iteration ℓ. We say that BP succeeds in decoding an ensemble of codes if the decoding error probability tends to zero as the number of iterations increases.

BP can be used for decoding many classes of codes. In this paper we study those classes for which the asymptotic behavior of BP for large block lengths can be determined by density evolution. Density evolution assumes the *independence assumption* that all messages communicated to a node at a certain iteration are independent. This mainly means that for any constant number of iterations, BP should run cycle-free except for a negligible fraction of nodes. And this requirement is usually satisfied by codes that are constructed from sparse graphs (i.e. graphs in which the number of edges is linear in the number of nodes) with enough randomness in their structure [3]. Two examples of such codes are *LT codes* [6] that have bounded degree (used as the main code of *raptor codes* [7]) and *irregular repeat-accumulate* (IRA) codes [8].

Throughout this paper we study memoryless channels with binary input and symmetric output, simply called *symmetric channels*. Examples of such channels are the binary erasure channel (BEC), the binary symmetric channel (BSC), and the binary-input additive gaussian noise (BIAGN) channel. For symmetric channels we assume that the transmitted codeword is the all-zero codeword without affecting the decoding error probability.

Let Z denote the LLR of a random bit recieved from the channel, and let f denote the probability density function (pdf) of Z on $\overline{\mathbb{R}} = \mathbb{R} \cup \{\pm\infty\}$. It is easy to see that f is *symmetric* in the following sense: $f(x) = \mathrm{e}^x f(-x)$ for all $x \in \overline{\mathbb{R}}$. Using the independence assumption, the pdf of messages passed along random edges remains symmetric during BP [2]. This fact will play an important role in the proof of our main result.

We finish this section by mentioning BP over the BEC. In this case, messages either contain no information (erasure messages) or contain the value of the variable bit with complete certainty. This means that in the case of all-zero codewords, LLR messages are in $\{0, +\infty\}$, and density evolution reduces to a one-dimensional recursion for the probability of erasure messages at different iterations.

3 Relationships between Symmetric and Erasure Channels

For a symmetric channel C, define $\mathrm{E}(C) = \mathrm{E}[\tanh(Z/2)]$, where Z is the LLR obtained for a zero bit from channel C. The following properties can be proved about $\mathrm{E}(C)$ [9].

Proposition 1. *For a symmetric channel C, we have*

1. $\mathrm{Cap}(C) \leq \mathrm{E}(C) \leq \ln(4)\mathrm{Cap}(C)$.
2. $\mathrm{E}(\mathrm{BEC}(\epsilon)) = 1 - \epsilon$, *where ϵ is the erasure probability.*
3. $\mathrm{E}(\mathrm{BSC}(\epsilon)) = (1-2\epsilon)^2$, *where ϵ is the cross-over error probability.*
4. $\mathrm{E}(BIAGN(\sigma)) = \frac{1}{2}m - \frac{1}{4}m^2 + \frac{5}{24}m^3 + O(m^4)$, *where $m = 2/\sigma^2$.*
5. $\mathrm{E}(\mathcal{L}(\mu)) = \frac{5}{24}\mu^2 + O(\mu^3)$, *where $\mathcal{L}(\mu)$ is the Laplace channel with parameter μ.*

The following theorem is an upper bound on the performance of BP on symmetric channels.

Theorem 2. *If BP succeeds in decoding an ensemble of codes on a symmetric channel C, then it succeeds in decoding the ensemble on a BEC with erasure probability* $1 - \mathrm{E}(C)$.

The result can also be interpreted as a lower bound on the performance of BP on the BEC. To prove the theorem we need the following result [9].

Proposition 3. *Suppose that X and Y are independent random variables with symmetric pdf's on $\overline{\mathbb{R}}$. Then $X + Y$ has a symmetric pdf, and we have*

$$\mathrm{E}[\tanh(\frac{X+Y}{2})] \leq \mathrm{E}[\tanh(\frac{X}{2})] + \mathrm{E}[\tanh(\frac{Y}{2})] - \mathrm{E}[\tanh(\frac{X}{2})]\,\mathrm{E}[\tanh(\frac{Y}{2})].$$

Proof. (Of Theorem 2, Sketch) We will track $\mathrm{E}[\tanh(X/2)]$ during the course of density evolution, where X is a message communicated along an edge. In the case of the BEC, the quantity $\mathrm{E}[\tanh(X/2)]$ represents the probability that a message is not an erasure message.

We will use terminology of LDPC codes, i.e., variable and check nodes. Let $X_1, \ldots, X_{j-1}$ denote the incoming messages along $j-1$ edges adjacent to a check node of degree j. The outgoing message X along the remaining edge is given by $\tanh(X/2) = \prod_{i=1}^{j-1} \tanh(X_i/2)$. Since density evolution assumes that $X_1, \ldots, X_{j-1}$ are independent, we have

$$\mathrm{E}[\tanh(\frac{X}{2})] = \prod_{i=1}^{j-1} \mathrm{E}[\tanh(\frac{X_i}{2})].$$

Now let $X_1, \ldots, X_{j-1}$ denote the incoming messages along $j-1$ edges adjacent to a variable node of degree j, and let Z denote the LLR obtained for that variable from the channel. The outgoing message X along the remaining edge is given by $X = Z + \sum_{i=1}^{j-1} X_i$. We claim that

$$1 - \mathrm{E}[\tanh(\frac{X}{2})] \geq (1 - \mathrm{E}[\tanh(\frac{Z}{2})]) \prod_{i=1}^{j-1} (1 - \mathrm{E}[\tanh(\frac{X_i}{2})]).$$

Random variables $X_1, \ldots, X_{j-1}$, and Z are independent and have symmetric pdf's. Thus, for $j = 2$, the claim is a restatement of the inequality of Proposition 3. The claim follows by induction on j.

In the case of the BEC the above inequality is tight, because the outgoing erasure probability of variable nodes is the product of incoming erasure probabilities. Since the erasure probability of the BEC is $1 - \mathrm{E}(C)$, the quantity $\mathrm{E}[\tanh(X/2)]$ is never more than what it would be in the case of the BEC. Clearly, if the probability of error tends to 0 on the symmetric channel, then the distribution of X tends to Δ_∞, and $\mathrm{E}[\tanh(X/2)]$ tends to 1. It follows that if BP succeeds on the symmetric channel, then the probability of erasure messages converges to 0 on the BEC as the number of iterations increases. That is, the decoding is successful over the BEC. □

For a symmetric channel C, let $\gamma(C) = \mathrm{E}[\mathrm{e}^{-Z/2}]$ be the Bhattacharya parameter of C. Theorem 2 is the analogue of the following lower bound on the performance of BP on symmetric channels [1].

Theorem 4. *If BP does not succeed in decoding an ensemble of codes on a symmetric channel, then it does not succeed in decoding the ensemble on a BEC with erasure probability $\gamma(C)$.*

Theorems 2 and 4 are similar in that they both relate the performance of BP on symmetric channels to its performance on the BEC. While Theorem 2 was derived by an analysis that is tight at check nodes (but not tight at variable nodes), Theorem 4 is based on an analysis that is tight at variable nodes (and not tight at check nodes). Moreover, the analysis of Theorem 2 is quite accurate at iterations where messages have a high uncertainty, whereas the analysis of Theorem 4 is more accurate when messages have a low uncertainty.

This last point manifests itself in the *stability condition* of LDPC codes and its analogue for LT codes, the dual of LDPC codes. The stability condition gives an upper bound on the fraction of check nodes of degree 2 in terms of $\gamma(C)$, if the error probability of density evolution converges to zero when it is initialized by a density with a small enough probability of error [2]. On the other

hand, the analogue of the stability condition for LT codes gives a lower bound on the fraction of output nodes of degree 2 in terms of $\mathrm{E}(C)$, if an LT code successfully starts decoding when it is initialized by a density with a probability of error close to 1/2 (assuming that the fraction of output nodes of degree 1 is very small).

4 Applications to Fountain Codes

Fountain codes are a class of codes designed for reliable transmission over a channel with unknown quality. A fountain code produces, for a given vector of k input bits, a potentially limitless stream of output bits. Each output bit is produced independently and randomly from the k input bits. The receiver collects output bits of the encoder from the channel, and with each bit, it records the reliability of the bit. The receiver collects bits until the sum of information of individual bits is $k(1+\epsilon)$, where ϵ is an appropriate constant, called the *overhead*, that allows the receiver to recover the correct input bits with high probability.

LT and raptor codes are two efficient classes of fountain codes that achieve the capacity of any BEC. To generate an output bit in an LT code, a "degree distribution" is sampled to obtain a degree d, and then d randomly chosen input bits are XOR-ed to obtain the value of the output bit. A simple balls-and-bins argument shows that the average output degree of an LT code should be at least logarithmic in k to recover all input bits. Raptor codes circumvent this inefficiency by first encoding input bits using a block code (e.g. an LDPC code) with rate close to 1, and then transmitting the encoded message using an LT code. Unlike LT codes, raptor codes are constructed from sparse graphs, which allows us to use density evolution for their BP analysis.

Since fountain codes may be used at situations where some receivers have noiseless channels (a BEC with no erasure) and some receivers have symmetric channels with arbitrary quality, the results of Section 3 are suitable for the analysis of raptor codes on arbitrary symmetric channels.

Theorem 5. *Assume that a raptor code has an ϵ overhead on the BEC. Then collecting a total of $(1+\epsilon)k/(1-\gamma(C))$ output raptor bits from a symmetric channel C is enough for decoding the k input bits.*

Proof. Follows from Theorem 4. □

For a sequence of degree distributions of raptor codes that achieves the capacity of the BEC, Theorem 5 shows that the sequence simultaneously beats the so-called *computational cut-off rate* on all symmetric channel. The rate was conjectured to be a limit for "practical communication" before the advent of turbo codes and iterative decoding [1]. The interesting point about this result is that this performance is achieved while the encoder is totally oblivious about the underlying channel.

Furthermore if a raptor code is optimized for the erasure channel, then a fixed reception overhead is enough for its decoding on arbitrary symmetric channels.

Corollary 6. *For arbitrary small $\epsilon > 0$, the reception overhead of raptor codes optimized for the BEC is at most $\log_2(\mathrm{e}) - 1 + \epsilon = 0.442\ldots$ on any binary-input symmetric channel.*

The following result shows that Theorem 2 provides the exact overhead that is necessary for raptor codes on the BEC, assuming that they are capacity achieving on a binary-input symmetric channel.

Theorem 7. *If a sequence of raptor codes achieves the capacity of a binary-input symmetric channel C, then the overhead of the sequence on the BEC is exactly equal to* $\mathrm{E}(C)/\mathrm{Cap}(C) - 1$.

Proof. (Sketch) The sufficiency of the stated overhead follows from Theorem 2. The necessity of the stated overhead is proved by looking at the limit Ω_2 of the fraction of output bits of degree 2 in a capacity achieving sequence (which has been derived in [9]), and using the analogue of the stability condition for raptor codes on the BEC. □

In fact, the above result shows that for raptor codes that achieve the capacity of a symmetric channel, the decoding bottleneck on the BEC is at the beginning of BP.

Acknowledgement

I would like to thank Amin Shokrollahi for introducing me to fountain codes and his support regarding this paper.

References

[1] A. Khandekar, *Graph-based Codes and Iterative Decoding*, Ph.D. thesis, California Institute of Technology, 2002.

[2] T. J. Richardson, A. Shokrollahi, R. Urbanke, "Design of Capacity-Approaching Irregular Low-Density Parity-Check Codes," *IEEE Trans. Info. Theory*, vol. 47, pp. 619-637, 2001.

[3] T. J. Richardson, R. Urbanke, "Capacity of Low-Density Parity-Check Codes under Message-Passing Decoding" *IEEE Trans. Info. Theory*, vol. 47, pp. 599-618, 2001.

[4] M. Luby, M. Mitzenmacher, A. Shokrollahi, D. Spielman, and V. Stemann, "Practical Loss-Resilient Codes," *Proc. 29th ACM Symposium on Thoery of Computing*, pp. 150-159, 1997.

[5] R. G. Gallager, *Low-Density Parity-Check Codes*, Cambridge, Massachussetts: M.I.T. Press, 1963.

[6] M. Luby, "LT Codes," *Proc. 43rd Annual IEEE Symposium on Foundations of Computer Science*, pp. 271-280, 2002.

[7] A. Shokrollahi, "Raptor Codes," Preprint, 2002.

[8] H. Jin, A. Khandekar, and R. McEliece, "Repeat Accumulate Codes," *Proc. 2nd International Symposium on Turbo Codes*, pp. 1-18, 2000.

[9] O. Etesami, M. Molkaraie, A. Shokrollahi, "Raptor Codes on Symmetric Channels," Preprint, 2003.