6.440 Essential Coding Theory March 12, 2008

Lecture 11

Lecturer: Madhu Sudan *Scribe: Jelani Nelson*

1 Overview

In these notes we will discuss Forney's general minimum distance (GMD) decoding of concatenated codes [1]. We will also discuss a method of Forney for algorithmically achieving capacity on the binary symmetric channel (BSC) using concatenated codes.

2 Naïve Decoding of Concatenated Codes

Suppose we have a concatenated code where the outer code is an $[N, K, D]_Q$ code, and the inner code is an $[n, k, d]_q$ code, with $Q = q^k$. In section 4 we will see that the concatenated code is actually an $[Nn, Kk, Dd]_q$ code, but for now we will show the weaker property that it is an $[Nn, Kk, Dd/2]_q$. Consider the naïve decoding procedure which uses the inner decoding scheme on each received symbol, then applies the outer decoding scheme.

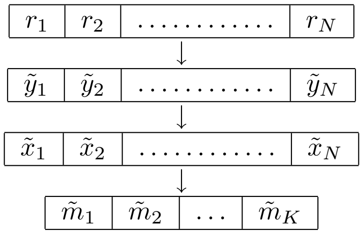
r_1 r_2 .... r_N
$\tilde{y}_1$ $\tilde{y}_2$ .... $\tilde{y}_N$
$\tilde{x}_1$ $\tilde{x}_2$ .... $\tilde{x}_N$
$\tilde{m}_1$ $\tilde{m}_2$ ... $\tilde{m}_K$

Figure 1: A diagram of the naïve decoding scheme.

In Figure 1, the received encoded message is r, which may have errors. Each block r_i is decoded using the inner code, to arrive at a new message $\tilde{y}$. If r had no errors, this stage of decoding would yield a message y. We now map each codeword $\tilde{y}_i$ of the inner code to the appropriate symbol of the outer code, giving $\tilde{x}$. We then decode $\tilde{x}$ to the nearest codeword of the outer code, at which point we can recover a message $\tilde{m}$ corresponding to that codeword.

Let e_i denote the number of errors in r_i. We state the following claims without proof; their proofs follow easily from definitions.

Claim 1 *If $e_i < d/2$, then $\tilde{y}_i = y_i$.* ■

Claim 2 *If there are fewer than $D/2$ blocks i with $\tilde{y}_i \neq y_i$, then the overall decoding procedure is successful.* ■

Corollary 3 *Fewer than $Dd/4$ errors in the encoded message implies that fewer than $D/2$ blocks $i \in [N]$ have $e_i \geq d/2$.* ■

By Claim 2 and Corollary 3, fewer than $Dd/4$ errors implies successful decoding. Thus, the concatenated code is a $[Nn, Kk, Dd/2]_q$ code.

3 Achieving Capacity on the BSC

Recall the binary symmetric channel from Lecture 2 depicted in Figure 2. A binary message is sent, and each bit is flipped independently at random with some probability p.

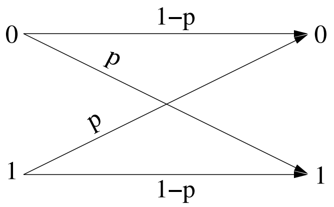

Figure 2: A binary symmetric channel where each bit flips with probability p.

We showed in Lecture 2 a theorem of Shannon which states that we can transmit at a rate R as long as $R = 1 - H(p) - \varepsilon$ for some $\varepsilon > 0$. Furthermore, we showed that such a rate is possible with decoding error probability at most $\exp(-\varepsilon N)$, where N is the block length. We will now investigate a method of Forney to achieve this result algorithmically. First though, we show how to achieve a failure probability of at most $1/\text{poly}(N)$.

First, by brute force we find a code with message length $n = c' \log N$ for some constant c' to be determined later with rate $R = 1 - H(p) - \varepsilon$ and decoding failure probability at most $\exp(-\varepsilon c' \log N)$. We know such a code exists by Shannon's theorem. We then break our message into $N/(c' \log N)$ blocks each of size $c' \log N$ then encode each block separately. The probability of a decoding error on block i is at most $\exp(-\varepsilon c' \log N) \leq N^{-(c+1)}$ for $c' = O((c+1)/\varepsilon)$. Thus, by a union bound, every block is successfully decoded with probability at least $1 - 1/N^c$.

To declare that this encoding procedure can be done in polynomial time, we need to discuss how to find a code with message length $n = c' \log N$ in polynomial time with the properties Shannon's theorem guarantees. We cannot try all subsets of codewords as there are too many. We omit the details, but one can prove that a linear code exists achieving Shannon capacity. Linear codes are however still not a small enough space of codes to search over as the generator matrix will have $\Omega(\log^2 N)$ entries, and thus there would be $N^{\Omega(\log N)}$ over which to search. It turns out that not only does a linear code exist achieving Shannon capacity, but a linear code exists achieving capacity whose generator matrix is a Toeplitz matrix. Toeplitz matrices are $m \times n$ matrices A satisfying $A_{i,j} = A_{i-1,j-1}$ for all $i, j > 1$. Thus, such a matrix is fully specified by $O(c' \log N)$ bits, and we can thus brute force try every such generator matrix in polynomial time. We omit the details, but our proof from Lecture 2 can be modified to show that capacity-achieving codes with Toeplitz generator matrices exist (essentially the reason is that each codeword is random and codewords will be pairwise independent, and our proof from Lecture 2 only relied on pairwise independence of codewords), and we can also calculate the code's decoding error probability in polynomial time.

The above procedure only achieved decoding error probability at most $1/\text{poly}(N)$. We now show Forney's method of achieving capacity. The outer code will be a Reed-Solomon code with rate $1 - \varepsilon$ and block length and alphabet size N. Such a code has $K = (1 - \varepsilon)N$ and distance εN. The inner code will be an $[n, k, d]_2$ code with $2^k = N$ and rate $1 - H(p) - \varepsilon$; we can use the above discussion to find such a code by brute force. The concatenated code has rate $(1 - \varepsilon)(1 - H(p) - \varepsilon) \geq 1 - H(p) - 2\varepsilon$. Each block of the concatenated code has length $n = \log N/(1 - H(p) - \varepsilon)$ blocks, and there are N such blocks. The probability that we err in decoding the ith block is at most $1/\text{poly}(N)$, so by Chernoff bounds the probability that at least $\varepsilon N/2$ blocks are incorrectly decoded is at most exponentially small, giving the desired result.

4 Generalized Minimum Distance Decoding

GMD decoding is a decoding procedure for the same concatenated code from section 2 where the outer code is a Reed-Solomon code. Recall that an erasure is an error where a symbol is corrupted to a '?' so that we know that an error occurred at a particular symbol. The key idea in GMD decoding is that it is easier to decode a channel with erasures than one with errors since we can restrict decoding to the unerased symbols. More concretely, suppose we use a Reed-Solomon code with distance D and have at most s erasures and t errors.

Claim 4 *If $s + 2t < D$ then we can successfully decode.*

Proof Ignoring the erased symbols, we are left with output that is equivalent to a codeword of an $[N - s, K, D - s]$ Reed-Solomon code, so we can imagine that we have t errors on such a codeword. Thus, we can decode as long as $t < (D - s)/2$, which is the claim. ■

We use the same notation from Figure 1. We also define $\tilde{e}_i$ to be the "apparent" number of errors in block i, i.e. $\Delta(r_i, \tilde{y}_i)$.

The decoding algorithm is now defined as follows. After computing $\tilde{y}$, with probability $\min\{\frac{\tilde{e}_i}{d/2}, 1\}$ we declare r_i to be an erasure. Otherwise, we keep $\tilde{y}_i$. We will now show that in expectation, the number of erasures plus twice the number of errors is at most D as long as $\sum_i e_i < Dd/2$. By Claim 4, this will show that the concatenated code is an $[Nn, Kk, Dd]$ code.

Forney's analysis is as follows. Define u_i to be the event that an erasure is declared on block i. Let v_i be the event that we keep $\tilde{y}_i$ and $\tilde{y}_i \neq y_i$ (thus, we have an error). We thus want to show that $E[\sum_i u_i + 2\sum_i v_i] < D$. First we observe that if $y_i = \tilde{y}_i$, then $e_i = \tilde{e}_i$. Also, if $y_i \neq \tilde{y}_i$, then $e_i \geq d - \tilde{e}_i$ by the triangle inequality on $r_i, y_i, \tilde{y}_i$, using the fact that $\Delta(y_i, \tilde{y}_i) \geq d$.

Claim 5

$$E[u_i + 2v_i] \leq \frac{e_i}{d/2}$$

Proof We split the proof into two cases.

Case 1 ($\tilde{y}_i = y_i$)**:** In this case we have $E[u_i] = \tilde{e}_i/(d/2) = e_i/(d/2)$, and $E[v_i] = 0$.

Case 2 ($\tilde{y}_i \neq y_i$)**:** Here we have $E[u_i] = \tilde{e}_i/(d/2)$ and $E[v_i] = 1 - \tilde{e}_i/(d/2)$, so $E[u_i + 2v_i] = 2 - \tilde{e}_i/(d/2)$. We now use that $-\tilde{e}_i \leq e_i - d$. Thus, $2 - \tilde{e}_i/(d/2) \leq 2 + e_i/(d/2) - d/(d/2) = e_i/(d/2)$. ■

Overall, we thus have $\sum_i E[u_i + 2v_i] \leq \sum_i e_i/(d/2) = (\sum_i e_i)2/d$. Thus, as long as $\sum e_i < dD/2$, the expected sum of erasures and twice the number of errors is at most D. We can thus say that there exists a configuration of declaring erasures such that the number of erasures plus twice the number of errors is at most D so that by Claim 4 the concatenated code is an $[Nn, Kk, Dd]$ code. We now show how to derandomize the decoding algorithm. We note that nowhere did we even use pairwise independence between the blocks. Thus we pick a random number $x \in [0, 1]$ and use the same x as a threshold in comparison to $\tilde{e}_i/(d/2)$ to decide for each block whether to declare it an erasure or not. The number of such thresholds that actually affect our decision is simply the number of blocks, so we can afford to try all such thresholds x rather than picking one at random. We then decode using each such x, and given our overall decoding we can encode again to determine whether the sum of erasures and twice the number of errors was in fact at most D. At least one x must satisfy this constraint, and we use the decoding from that x.

References

[1] G. David Forney, Jr. Generalized minimum distance decoding. *IEEE Trans. Inform. Theory*, vol. IT-12, pp. 125–131, 1966.