18.409 Notes (10/29/98): Concatenated codes, the Golay code

Patrick Maurer

December 14, 1998

1 Review of last lecture

In the last class, concatenated codes were introduced and used to show a method for constructing good codes.

Reed Solomon (RS) Codes:

$w \to (P_w(x_1), P_w(x_2), \ldots P_w(x_n))$

$P_w(x_i) \in \mathrm{GF}(2^k)$ $x_i \in \mathrm{GF}(2^k)$

We really want these over bits; $C : \{0,1\}^k \to \{0,1\}^j$.

If the RS code has rate R, and the inner code C has rate r, then the concatenated code has rate $R_0 = rR$.

If the RS code has relative minimum distance $\frac{1-R}{2}$, and the inner code C has relative minimum distance δ, then the concatenated code has relative minimum distance $\delta\frac{1-R}{2}$.

We also observed Utison's (sp??) scheme for making this explicit:

$C_{x_i} : w \to (w, x_i w)$

where almost all of the random linear codes C_{x_i} are good (all but an exponentially small number of them).

$\Rightarrow$ codes of this type are good.

$r < 1 - H(\delta)$ $\frac{1}{2} < 1 - H(\delta)$

2 Concatenated codes

Concatenated codes were found by Forney in 1966. Concatenated codes represent an attempt to achieve Shannon's promise with reasonable complexity.

2.1 Random Linear Codes

If $r < 1 - H(p)$, then a random linear code of rate r, with maximum likelihood (ML) decoding, can have a probability of error approaching zero as n grows large. To see this, look at the earlier proof of random linear codes and note that $P(\text{error}) \to 0$ exponentially quickly. I.e., $P(\text{error}) \sim e^{-cn}$ for some constant c.

Unfortunately, the decoding complexity of this method is exponential! With concatenated codes, we can get similar results, but the decoding complexity will be polynomial in n ($\text{poly}(n)$).

To see this, we can view the inner code C as producing the correct symbol with "pretty good" reliability. Some inner codewords will be wrong, but these can then be "cleaned up" by the RS code. We can show that it is highly unlikely for many of the inner codewords to be wrong, so the RS code can have fairly high rate.

The inner code C has a rate $r < 1 - H(p)$. The probability of a decoding error using code C is $P_e = 2^{-n(E_p(r)+o(1))}$, where $o(1)$ approaches zero as n grows large. P_e goes to zero exponentially quickly.

If we let the rate of the outer code $R = 1 - 10P_e$, then the RS code can correct up to a fraction $5P_e$ errors. I.e., an RS code of this rate will correct all the errors in a block of $N = q - 1$ symbols provided fewer than $5P_eN$ symbol errors occurred in that block.

Theorem 1 (Chernoff bound) *For any $\alpha > 0$ and for any real s,*
$P(e^{sx} \geq \alpha) \leq \frac{E\{e^{sx}\}}{\alpha}$

Proof: We note that $e^{sx} \geq 0, \forall x$ and then apply the Markoff inequality.

By the Chernoff bound, the probability of more than a $5P_e$ fraction of inner codes returning the wrong symbol is exponentially small in n. So, the probability that the RS code makes errors is *really* small.

The rate of the concatenated code is $r \cdot R$. For our choice of R, the rate $r \cdot R = r \cdot (1 - 10P_e) \xrightarrow{n\to\infty} r$.

The complexity of decoding concatenated codes is pretty low. The decoding effort can be broken down into the time to decode the RS code ($\leq n^2$) plus the time to decode the code C ($\leq n \log n$) n times (once for each inner codeword). The time to decode is then $\mathcal{O}(n^2 \log n)$ for $n \log n$ bits. Thus we have polynomial complexity, and we still have exponential dropoff in error probability.

Concatenated codes are also very useful in practice. Good examples of this are the cross-interleaved redundancy code (two shortened RS codes with an interleaver between them) and perhaps more notably NASA's deep space error control code (an RS outer code, interleaver, and an inner convolutional code).

Without sounding too greedy, what are some of the "flaws" with concatenated codes?

When trying to communicate at a rate very close to capacity, the error probability, while still dropping exponentially, is a very slow exponential. This means that to achieve a given probability of error, the choice of n must be very large; hence, the complexity must be very large.

Let us define the following: Let R and N be the rate and length of the RS code respectively, and r be the rate of the inner code.

[need the computational cutoff issues]

3 The Hadamard matrix

A few lectures ago, we proved the Plotkin bound, which states that one can generate a collection of at most $2n$ codewords with the restriction they be of length n and have minimum distance $\frac{n}{2}$. But how do we *construct* such a collection? One way is with a Hadamard matrix.

Definition 1 *A Hadamard matrix* $\mathbf{H}$ *of order* n *is an* $n \times n$ *matrix whose elements* $h_{ij} \in \{-1, +1\}$ *and such that* $\mathbf{H}\mathbf{H}^T = n\mathbf{I}$ *(i.e., all rows are orthogonal).*

Two ways to generate Hadamard matrices are the Sylvester construction and with quadratic residues. The Sylvester construction starts with a Hadamard matrix of order 2, and continues by way of the following:

Theorem 2 (Sylvester Construction) *If* $\mathbf{H}_n$ *is a Hadamard matrix, then so is* $\mathbf{H}_{2n} = \begin{pmatrix} \mathbf{H}_n & \mathbf{H}_n \\ \mathbf{H}_n & -\mathbf{H}_n \end{pmatrix}$.

Proof: We can show that $\mathbf{H}_{2n}\mathbf{H}_{2n}^T = 2n\mathbf{I}$*:*

$$\begin{aligned}
\mathbf{H}_{2n}\mathbf{H}_{2n}^T &= \begin{pmatrix} \mathbf{H}_n & \mathbf{H}_n \\ \mathbf{H}_n & -\mathbf{H}_n \end{pmatrix} \begin{pmatrix} \mathbf{H}_n^T & \mathbf{H}_n^T \\ \mathbf{H}_n^T & -\mathbf{H}_n^T \end{pmatrix} \\
&= \begin{pmatrix} \mathbf{H}_n\mathbf{H}_n^T + \mathbf{H}_n\mathbf{H}_n^T & \mathbf{H}_n\mathbf{H}_n^T - \mathbf{H}_n\mathbf{H}_n^T \\ \mathbf{H}_n\mathbf{H}_n^T - \mathbf{H}_n\mathbf{H}_n^T & \mathbf{H}_n\mathbf{H}_n^T + \mathbf{H}_n\mathbf{H}_n^T \end{pmatrix} \\
&= \begin{pmatrix} 2\mathbf{H}_n\mathbf{H}_n^T & 0 \\ 0 & 2\mathbf{H}_n\mathbf{H}_n^T \end{pmatrix} = \begin{pmatrix} 2n\mathbf{I}_n & 0 \\ 0 & 2n\mathbf{I}_n \end{pmatrix} = 2n\mathbf{I}_{2n}
\end{aligned}$$

So, we start with $\mathbf{H}_2 = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$, and can recursively produce Hadamard matrices of order 2^k:

$$\mathbf{H}_{2^k} = \begin{pmatrix} \mathbf{H}_{2^{k-1}} & \mathbf{H}_{2^{k-1}} \\ \mathbf{H}_{2^{k-1}} & \mathbf{H}_{2^{k-1}} \end{pmatrix}$$

If we take the Hadamard matrix $\mathbf{H}_n$ and map each element by $\{(1 \to 1), (-1 \to 0)\}$, then the rows of our matrix become our codewords. It is clear that we have n codewords. To show $d_{\text{min}} = \frac{n}{2}$, we first note that the inner product of any two (distinct) rows of $\mathbf{H}_n$ is zero. Observe also that multiplying any two elements of $\mathbf{H}_n$ results in a $+1$ if the elements are the same, and a -1 if the elements differ. These two facts show that the number of positions in which two codewords agree (match) must equal the number in which they disagree. Thus, the distance between any two codewords derived from $\mathbf{H}_n$ is precisely $\frac{n}{2}$.

To achieve the Plotkin bound, we must double the number of codewords. This is easily done, simply be taking the complement of each codeword. Each codeword in the resulting collection is precicely $\frac{n}{2}$ away from all other codewords except its complement, from which it is n away. As an example, the codebook derived from $\mathbf{H}_4$ would be $\{1111, 1010, 1100, 1001, 0000, 0101, 0011, 0110\}$. This is a collection of 8 codewords of length 4, all at least 2 apart in Hamming distance.

Another way to construct Hadamard matrices is by the Paley construction (quadratic residues).

Definition 2 *Let p be prime. The quadratic residues* $\bmod p$ *are the nonzero squares* $\bmod p$.

Definition 3 (Quadratic Residue Character) *define the function*

$$\chi(x) = \begin{cases} 0 & \textit{if } x \equiv 0 \bmod p \\ 1 & \textit{if } \exists a \textit{ such that } a^2 \equiv x \bmod p \\ -1 & \textit{otherwise} \end{cases}$$

This is known as the quadratic residue character, or the Legendre symbol.

We assume $p \equiv -1 \bmod 4$. That is, $n = p + 1$, p prime, and n is a multiple of 4.

To construct a $n \times n$ Hadamard matrix, we first construct the matrix $\mathbf{M}_p = \mathbf{Q}_p - \mathbf{I}_p$, where the elements of $\mathbf{Q}_p \ni q_{ij} = \chi(i - j)$, and we index the elements of $\mathbf{Q}_p$ from 0 to $p - 1$. Note that the zero elements of $\mathbf{Q}_p$ (along the diagonal) become -1 in $\mathbf{M}_p$.

Theorem 3 *If* $p \equiv -1 \bmod 4$, *then*
$\mathbf{H}_n = \begin{pmatrix} 1 & \mathbf{1}_{1\times p} \\ \mathbf{1}_{p\times 1} & \mathbf{M}_p \end{pmatrix}$ *is a Hadamard matrix.*

Proof: It is clear that the elements of $\mathbf{H}_n$ *only take on values in the set* $\{-1. +1\}$. *We must show that every row of* $\mathbf{H}_n$ *is orthogonal to every other row. We first observe that the first column of* $\mathbf{H}_n$ *is all ones. Knowing that orthogonality amounts to an equal number of agreements and disagreements (i.e., same sign versus opposite sign), we would like to show that any two rows of* $\mathbf{M}_p$ *agree in precisely* $\frac{(p-1)}{2}$ *places.*

Consider any two rows x *and* y *of* $\mathbf{M}_p$. *These rows agree at column* i *when* $\chi(x-i) = \chi(y-i)$, *or equivalently when* $\chi(\frac{x-i}{y-i}) = 1$.

???

4 References

Papoulis, Athanasios, "Probability, Random Variables, and Stochastic Processes". 3rd ed. McGraw-Hill. copyright 1991.

Wicker, Stephen R., "Error Control Systems for Digital Communication and Storage". Prentice-Hall. copyright 1995.