6.895 Essential Coding Theory October 18, 2004

Lecture 11

Lecturer: Madhu Sudan *Scribe: Elena Grigorescu*

1 Overview

Today we will be relating Shannon's capacity to coding and decoding algorithms that could achieve this capacity. We will be mainly concerned with correcting random and adversarial errors in binary codes. We deal with the following two cases

1. decoding from fraction p random error
2. decoding from fraction $\frac{1}{2} - \epsilon$ adversarial error.

2 Goals

- Recall the random error channel with bits flipped independently at random w.p. p. Shannon's result stated that information transmission is feasible at a rate of $R = 1 - H(p) - \epsilon$. We will show that this is achievable with efficient encoding and decoding.
- In the adversarial error model, since we cannot obtain an optimal bound as above, we will focus on a $(\frac{1}{2} - \epsilon)$ fraction of flipped bits and would like to produce efficient encoding and *list*-decoding algorithms, for rate > 0 ($R(\epsilon) > 0$ for $\epsilon > 0$).

At this point we have already seen a coding scheme that will be helpful in achieving our goals. As we will show here, the Forney concatenation method will be enough for our purposes.

3 The random error case

Recall Forney's concatenation coding scheme. A message is first encoded using an outer RS code, and then block-wise encoded using an inner encoding. From an outer RS code $[N, K, N-K]_N$ and an inner code $[n, k, d]_2$, with $k = log_2 N$, results a $[Nn, Kk, Dd]_2$ code. The decoding is a block-by-block decoding followed by a RS decoding.

Using the above as our framework, we would like to produce poly time encoding and decoding algorithms, which would allow us to correct p errors.

Shannon's theorem stated:

Theorem 1 *$\exists E : \{0,1\}^k \to \{0,1\}^n$ and $D : \{0,1\}^n \to \{0,1\}^k$ with $k = 1 - H(p) - \epsilon n$ s.t.*

$$Pr[\textit{decoding error given } BSC_p] < exp(-n).$$

It can be assumed that in this case the encoding and decoding algorithms are in exponential time, by picking a linear encoding function.

We obtain the following corollary to Shannon's theorem.

Corollary 2 *$\exists E' : \{0,1\}^K \to \{0,1\}^N$ and $D' : \{0,1\}^N \to \{0,1\}^K$ with similar parameters s.t.*

$$Pr[\textit{decoding error}] < \frac{1}{N},$$

and E', D' are poly time algorithms.

Proof To encode a message $m \in \{0,1\}^K$, divide it into $\frac{K}{k}$ blocks of size k, where $k = \log K$. Then use the Shannon encoding function E to encode each of these blocks into words of size n, and concatenate the results. We have thus obtained $N = \frac{K}{k}n$. Therefore $\frac{K}{N} = \frac{k}{n}$ so the rate is preserved by this new encoding.

We have then

$$Pr[\text{decoding failure of } (E', D')_K] \leq \frac{K}{k} Pr[\text{decoding failure of } (E, D)_k].$$

Then by picking large enough constant c s.t. $k = c \log K$, the above probability can be made $\leq \frac{1}{KN}$

Therefore, the new E' and D' are $poly(N)$ which is what we hoped for.

■

Although this result is nice, we would like to be able to decode correctly with higher probability, and obtain exponentially small error in the number of blocks, without changing much the rate.

3.1 Implementation using Forney's code

Consider a Forney code given by:

- outer RS-code: $[N, (1-\epsilon)N, \epsilon N]_N$
- inner code (assuming BSC_p): $[n, k = (1 - H(p) - \epsilon)n, d]_2$. We should be able to correct p fraction of random error with $exp(-n)$ decoding failure. These requirements impose that $k \geq log_2 N$.

We have obtained a code with the following parameters:

- $R = (1 - H(p) - \epsilon)(1 - \epsilon) \geq 1 - H(p) - 2\epsilon$. So we had to compromise a little in the rate.
- $Pr[\text{decoding failure}] \leq Pr[\frac{\epsilon N}{2}$ inner blocks leading to decoding failure]

  $\leq \binom{N}{\frac{\epsilon N}{2}}(exp(-n))^{\frac{\epsilon N}{2}} \leq exp(-nN)$ (which is what we wanted!)
- Running time $= exp(k)poly(N)$. In later lectures we will see that we can get $exp(k)linear(N)$ running time.

Since k is a parameter more intrinsic to our solution, it would be nice to substitute it with something that is more intrinsic to the channel. This is one of the challenges of current research in this area. The formal problem can be stated as follows:

Theorem 3 *Given BSC_p, ϵ and n, design encoding E and decoding D schemes of rate $R = 1 - H(p) - \epsilon$, block length n, such that the encoding and decoding algorithms run in $poly(\frac{1}{\epsilon}, n)$ and the probability of decoding failure is exponentially small in n.*

This completes our discussion of the random error model.

4 Adversarial error model

We will be presenting partial progress towards solving the question of correcting $\frac{1}{2} - \epsilon$ fraction of errors.

4.1 List decoding from $(\frac{1}{2}-\epsilon)$ fraction of errors

The first question to ask is for what p we will need list decoding. We know that for a $\frac{d}{2n}$ fraction we can correctly uniquely decode. In a previous lecture we alluded to the following fact, known as the Plotkin bound, that will give us insight into a bound for p. We will later provide two proofs of the Plotkin bound.

Theorem 4 *If $C=[n,k,d]_2$ such that $\frac{d}{n}\geq\frac{1}{2}$, then $k\leq\log_2(2n)$.*

The above bound is tight, since the Hadamard codes achieve exactly these parameters. However, the rate is very poor.

Corollary 5 *If $E:\{0,1\}^k\rightarrow\{0,1\}^n$, and $D:\{0,1\}^n\rightarrow\{0,1\}^k$ correcting $p\geq\frac{1}{4}$ fraction of adversarial error with unique decoding, then $k\leq\log_2(2n)$.*

Since we aim for arbitrarily close to $\frac{1}{2}$ errors, unique decoding is not possible and therefore need list decoding.

We can now state the same goal as we did for random errors, namely:

Theorem 6 *Given $\epsilon>0$, design $E:\{0,1\}^k\rightarrow\{0,1\}^n$ and $D:\{0,1\}^n\rightarrow\{\{0,1\}^k\}^l$, (with $l<poly(n)$) s.t. $\forall m\in\{0,1\}^k$ and error $e\in\{0,1\}^n$, s.t. $wt(e)\leq(\frac{1}{2}-\epsilon)n$, we have $m\in D(E(m)+e)$, and*

- *$\frac{k}{n}=f(\epsilon)>0$, and*
- *$poly(n)$ running time.*

Before we can construct such E and D, we need to obtain some helpful combinatorial results.

4.2 Combinatorial results

Suppose we are not interested in the running time of the encoding and decoding functions. The question is whether we can achieve the above requirements. The answer is yes, and there are 2 ways to show this:

1. non-constructive: $\forall p\in[0,\frac{1}{2}]$, $\exists$ code of rate $1-H(p)$ that corrects p-fraction adversarial error with $poly(n)$ size list. This is tight in some sense, since we cannot have $R=1-H(p)+\epsilon'$ to correct p fraction of errors from $poly(n)$ size lists. It would contradict Shannon's theorem, since we would be getting $\frac{1}{poly(n)}$ probability of decoding failure.

2. distance vs. list decoding connection: Take a code of $\frac{d}{n}=\frac{1}{2}-\epsilon^2$. Then it can list decode a $\frac{1}{2}-\epsilon$ fraction of errors with poly size lists.

In order to make these results algorithmic, we start with some impossibility results.

Theorem 7 (Plotkin) *Let $c_1,\ldots,c_m\in\{0,1\}^n$ be codewords s.t. $\forall i\neq j$ we have $\Delta(c_i,c_j)\geq\frac{n}{2}$. Then $m\leq 2n$.*

One can restate the above as follows:

Theorem 8 *Let $c'_1,\ldots,c'_m\in\{-1,1\}^n$ be vectors s.t. $\forall i\neq j$ we have $<c'_i,c'_j>\leq 0$. Then $m\leq 2n$.*

Note that this conversion can be done using the encoding $0\rightarrow 1$ and $1\rightarrow -1$, of the bits of a codeword c_i into the entries of the vector c'_i. Also, $<c'_i,c'_j>=\sum_{k=1}^{n}c'_{i,k}c'_{i,j}=n-2\Delta(c_i,c_j)\leq 0$, if $\Delta(c_i,c_j)\geq\frac{n}{2}$.

Proof Geometric approach:

The proof is inductive. Start with vector c_m and consider its normal hyperplane H. Project all the other $m-1$ vectors to H. Say $v_1,\ldots,v_{m-1}$ are the projections. Note that at most one of them can be 0 since the c_i's were different. Also, note that the v_is satisfy the same property as the c_i's, namely the inner product of any two is ≤ 0. Therefore we obtained at least $m-2$ vectors of dimension $n-1$ satisfying the same conditions as we started with. Inductively, the theorem follows.

Linear Algebraic approach (to be completed in the next lecture):

If $m \geq n+1$ then $\exists \lambda_1, \ldots, \lambda_{n+1} \in \mathbb{R}$ s.t. $\sum_{i=1}^{i=n+1} \lambda_i c_i' = 0$. Suppose $\lambda_j > 0$ for $j \leq i$ and $\lambda_j < 0$ for $j > i$. Then, let $v = \sum_{j \leq i} \lambda_j c_j' = -\sum_{j>i} \lambda_j c_j'$. Therefore $0 << v, v >= -\sum_{j \leq i, j' > i} \lambda_j \lambda_{j'} < c_j, c_{j'} > \leq 0$.

It remains to deal with the $v = 0$ case.

$\blacksquare$