6.897 Algorithmic Introduction to Coding Theory September 9, 2002

Lecture 2

Lecturer: Madhu Sudan *Scribe: Deniss Čebikins*

1 Introduction

In this lecture, we consider classical scenarios in the theory of communication, in which a sender (which we will refer to as the *source*) wants to transmit a piece of information, which we will represent as a binary string, to a receiver. The transmission is to be conducted through a *channel*, which has certain properties depending on the scenario we are considering. The natural goal is to set up a scheme that requires transmission of as little data as possible and that minimizes the chance of error.

The first situation we deal with concerns the *noiseless channel*. In this case there is no need to account for transmission errors, so we can concentrate on minimizing the number of bits that have to be sent to the receiver. In the second scenario we have a *noisy channel*, which alters every bit that is transmitted through it with a small probability p. We set a condition that the probability of the receiver getting an incorrect message is at most ϵ, and we show the existence of a certain transmission scheme that satisfies the condition, and we also show that this scheme has the best possible efficiency in terms of the amount of data being transmitted.

2 Noiseless Channel

Suppose that the source produces a sequence of n independent bits such that each bit has value 0 with probability $1 - p$ and value 1 with probability p. We use the notation $x \leftarrow \mathrm{BSC}_p$ to indicate that a sequence (string) x is generated according to this principle. Since we would like to minimize the number of bits that we send to the receiver, it is reasonable for the sender to encode the message using an encoding function E, then send the encoded message, so that the receiver can decode the message using a decoding function D. Formally, we need to construct a pair of functions

$$E \ : \ \{0,1\}^k \to \{0,1\}^*$$

$$D \ : \ \{0,1\}^* \to \{0,1\}^n$$

such that $D(E(x)) = x$ for all $x \in \{0,1\}^n$. Our goal is to minimize the value of the expected size $\mathrm{Exp}_{x \leftarrow \mathrm{BSC}_p}[|E(x)|]$ of the encoded message.

We define a *distribution* on a finite set S to be a function $\mathcal{D} \ : \ S \to [0,1]$ such that $\sum_{x \in S} \mathcal{D}(x) = 1$. The function

$$H(\mathcal{D}) = -\sum_{x \in S} \mathcal{D}(x) \log \mathcal{D}(x)$$

is the *entropy function*. The *binary entropy function* is defined as follows:

$$H_2(p) = -p \log p - (1-p) \log(1-p).$$

Notice that $H_2(p) = H(\mathcal{D})$ for the distribution $\mathcal{D}(0) = 1-p$, $\mathcal{D}(1) = p$ on the set $\{0,1\}$.

The following result solves the above problem of minimizing data transmission in a noiseless channel.

Theorem 1 (Noiseless Coding Theorem) *There exist functions E and D satisfying the condition $D(E(x)) = x$ such that*

$$Exp_{x \leftarrow BSC_p}[|E(x)|] = (H_2(p) + o(1)) \cdot n,$$

where $n = |x|$ is the size of the original message.

3 Noisy Channel

Suppose that the source generates purely random strings (i.e. bits have values 0 or 1 with equal probability). Consider a noisy channel which alters every bit transmitted through it with probability $p < 1/2$. We can view the transmission process as follows: on input z, the channel generates a string $\eta \leftarrow \mathrm{BSC}_p$ of length $|z|$ and outputs the modified string $z + \eta$.

As in the previous scenario, we would like to obtain functions

$$E \ : \ \{0,1\}^k \rightarrow \{0,1\}^n$$

$$D \ : \ \{0,1\}^n \rightarrow \{0,1\}^k$$

for encoding and decoding the message, respectively. One of our goals is to obtain a pair of such functions such that the probability that the receiver gets a wrong message after decoding the transmission is bounded above by a prescribed number $\epsilon > 0$. Formally, this condition can be written as follows:

$$\Pr_{\substack{\eta \leftarrow BSC_p \\ x \leftarrow \{0,1\}^k}} [D(E(x) + \eta) \neq x] \leq \epsilon.$$

Having stated the condition, our problem is to maximize the value of k/n for which such E and D exist.

Theorem 2 (Noisy Coding Theorem) *Given $\delta > 0$ and $\epsilon > 0$, there exist functions E and D satisfying the above conditions such that*

$$\frac{k}{n} = 1 - H_2(p) - \delta$$

if k and n are sufficiently large.

We use the well-known lemma to prove the theorem.

Lemma 3 (Chernoff Bound) *Let $\mathcal{D}$ be a distribution on $\{0,1\}$, and let $x_1, \ldots, x_N$ be independent bits chosen from $\mathcal{D}$. If $\mu = Exp_{x \leftarrow \mathcal{D}}[x]$, then for any λ the following inequality holds:*

$$\Pr_{x_i \leftarrow \mathcal{D}}\left[\left|\frac{\sum_{i=1}^{N} x_i}{N} - \mu\right| \geq \lambda\right] \leq e^{-\lambda^2 \cdot \frac{N}{2}}.$$

Proof of Theorem 2. Let k and n satisfy the inequality

$$\frac{k}{n} = 1 - H_2(p) - \delta,$$

and suppose that $E : \{0,1\}^k \to \{0,1\}^n$ is an arbitrary function. Define $D(y)$ to be a string x for which the Hamming distance $\Delta(E(x), y)$ is the smallest possible.

Choose a small constant $\gamma > 0$ such that

$$H_2((p+\gamma)) = H_2(p) + \delta' < H_2(p) + \delta.$$

Now suppose that a string $x \in \{0,1\}^k$ and the value $E(x) \in \{0,1\}^n$ are fixed, and the rest of values of E are chosen at random. For a string $\eta \leftarrow \mathrm{BSC}_p$ of length n, we say that η is *bad* if $wt(\eta) > (p+\gamma) \cdot n$. It follows from Lemma 3 that there exists a constant $c > 1$ such that

$$\Pr_{\eta \leftarrow BSC_p}[\eta \text{ is bad}] < c^{-n}. \tag{1}$$

(Indeed, we have $\mathrm{Exp}_{\eta \leftarrow BSC_p}[wt(\eta)/n] = p$, so setting $\lambda = \gamma$ implies that the probability of η being bad is less than $e^{-\gamma^2 \cdot \frac{n}{2}} = c^{-n}$, where $c > 1$.)

Define $\mathrm{Ball}(z, r)$ to be the set of all strings z' such that $\Delta(z, z') \leq r$. Let $y = E(x) + \eta$. Note that

$$|\mathrm{Ball}(y, (p+\gamma)\cdot n)| = \sum_{i=0}^{(p+\gamma)\cdot n} \binom{n}{i} \leq 1 + (p+\gamma)\cdot n \cdot \binom{n}{(p+\gamma)\cdot n} =$$

$$= 1 + (p+\gamma)\cdot n \cdot 2^{(H_2(p+\gamma)+o(1))\cdot n} = 2^{(H_2(p)+\delta'+o(1))\cdot n}$$

since $(p+\gamma)\cdot n = 2^{o(1)\cdot n}$.

Fix $x' \neq x$. Then

$$\Pr_E\left[E(x') \in \mathrm{Ball}\,(y, (p+\gamma)\cdot n)\right] = \frac{|\mathrm{Ball}\,(y, (p+\gamma)\cdot n)|}{2^n} \approx \frac{2^{(H_2(p)+\delta'+o(1))\cdot n}}{2^n}.$$

We conclude that

$$\Pr_E[\exists\ x' \neq x \text{ s.t. } E(x') \in \mathrm{Ball}(y, (p+\gamma)\cdot n)] \leq \frac{2^k \cdot 2^{(H_2(p)+\delta'+o(1))\cdot n}}{2^n} = 2^{(\delta'-\delta+o(1))\cdot n} < a^{-n} \tag{2}$$

for some $a > 1$. If η is not bad, i.e. $wt(\eta) \leq (p+\gamma)\cdot n$, and there is no string $x' \neq x$ such that $\Delta(y, x') \leq (p+\delta)\cdot n$, then we have $D(y) = D(E(x)+\eta) = x$, so the receiver gets the correct message. We conclude from equations (1) and (2) that

$$\Pr_{E,\eta}[\text{ the receiver gets a wrong message }] < a^{-n} + c^{-n}$$

for any fixed x. It follows that

$$\Pr_{x,\eta}[\text{ the receiver gets a wrong message }] < a^{-n} + c^{-n}$$

for some function E. Finally, observe that $a^{-n} + c^{-n} < \epsilon$ for sufficiently large n. The theorem follows.

4 Converse Theorem

The following result shows that the bound in Theorem 2 is in fact tight.

Theorem 4 *Let k and n satisfy the relation*

$$\frac{k}{n} = 1 - H_2(p) + \delta(n)$$

for some $\delta(n) = \Omega(1)$. Then

$$\lim_{k,n\to\infty} \Pr_{\eta}[D(E(x) + \eta) = x] = 0$$

for any pair of functions (E, D).

We briefly describe the idea of the proof. If $\eta \leftarrow \mathrm{BSC}_p$, then with high probability we have $wt(\eta) \geq pn$. There are at least $\binom{n}{pn}$ strings z such that $wt(z) \geq pn$, and the string η equals each such string with probability of at most $1/\binom{n}{pn}$. Thus in order for the probability of successful decoding to be bounded below by a positive constant, we need to have for every x about $c \cdot \binom{n}{pn}$ strings y such that $D(y) = x$ (here $c > 0$). However,

$$2^k \cdot c \cdot \binom{n}{pn} \approx 2^k \cdot c \cdot 2^{H_2(p)n+o(1)} = c \cdot 2^{n+n\delta(n)+o(1)} > 2^n$$

for sufficiently large n. This is a contradiction, as there are only 2^n strings of length n.