CS378: Randomized Algorithms
David Zuckerman

Lecture Notes
April 4, 2018

Lecture Notes on Coding Theory

An error-correcting code encodes a message by a somewhat longer string, in such a way that even if some bits of the longer string are corrupted the original message can be recovered.

There are two types of noise models. The first is probabilistic, and is covered in the Mitzenmacher-Upfal text. The second is adversarial, which we cover here.

1 Codes and Minimum Distance

Definition 1. *A code is a subset $C \subseteq \mathbb{F}_q^n$. The most important choice of q is $q = 2$, i.e., $\mathbb{F}_2 = \{0, 1\}$, and $C \subseteq \mathbb{F}_2^n$ is called a binary code. Sometimes a code is given implicitly by an encoding function* $\text{Enc} : \mathbb{F}_q^k \to \mathbb{F}_q^n$*, in which case the code is the range of* Enc.

Definition 2. *The Hamming distance $\Delta(v, w) = |\{i : v_i \neq w_i\}|$.*

Definition 3. *The minimum distance $d(C) = \min_{v,w \in C, v \neq w} \Delta(v, w)$.*

Proposition 1. *If the distance of a code = d, then it can be used to:*

1. *correct $d - 1$ erasures (i.e. missing symbols),*
2. *detect $d - 1$ errors (i.e. changed symbols), and*
3. *correct $\frac{d-1}{2}$ errors.*

2 Example: Repetition Code

In the repetition code R, each symbol is repeated ℓ times for some parameter ℓ. This implies $n = \ell k$ and $d(R) = \ell$. (Taking $\ell = c \log k$ for a large enough constant c suffices for the binary symmetric channel.)

3 Example: Even Weight Code

In the even weight code E, a single parity bit is appended at the end. In other words,

$$E = \{x = x_1 x_2 \dots x_n \in \{0, 1\}^n | x_1 \oplus x_2 \oplus \dots \oplus x_n = 0\}.$$

Here we have $n = k + 1$ and $d(E) = 2$.

4 Asymptotically Good Codes

We can achieve much better binary codes, called asymptotically good codes. Such codes have:

$$\begin{aligned} k &= \Omega(n) \\ d &= \Omega(n) \end{aligned}$$

Proposition 2. *There exists asymptotically good binary codes.*

Proof. Choose $v_1, ..., v_K \in \{0,1\}^n$, where $K = 2^k$. By a Chernoff bound:

$$\Pr[\Delta(v_i, v_j) \leq (\frac{1}{2} - \alpha)n] \leq e^{-2\alpha^2 n}$$
$$\Pr[(\exists i \neq j : \Delta(v_i, v_j) \leq (\frac{1}{2} - \alpha)n] \leq \binom{K}{2} e^{-2\alpha^2 n}$$

The code will exist if the above probability is < 1, or in other words, if $\binom{K}{2} < e^{2\alpha^2 n}$. □

5 Example: Reed-Solomon Codes

While the existence of asymptotically good codes is important, it is much more important and useful to have explicit, efficient codes. Here we describe a code over the alphabet $\mathbb{Z}_p$, the integers modulo a prime p. In particular, the encoding functions maps $\mathbb{Z}_p^k$ to $\mathbb{Z}_p^n$, where $n \geq p$.

The message symbols $(a_0, a_1, \ldots, a_{k-1})$ of a Reed-Solomon code define a polynomial $f(x) = a_0 + a_1 x + \ldots + a_{k-1} x^{k-1}$ of degree at most $k-1$ over $\mathbb{Z}_p$. Its encoding is the vector

$$(f(0), f(1), \ldots, f(n-1)) \in \mathbb{Z}_p^n.$$

The important point is that since a degree $k-1$ polynomial has at most $k-1$ roots, any two distinct polynomials of degree $k-1$ agree on at most $k-1$ points. This implies that the minimum distance is at least $n - (k-1)$.

Thinking of $k = n/2$, we get $d \geq n/2$, and hence these are asymptotically good over the large alphabet. It's more challenging to construct an asymptotically good code over the binary alphabet $\{0,1\}$.