Lecture 26: title

instructor: Jin-Yi Cai Scribe: Bess Berg, Michael Kowalczyk

1 Codeword Spacing

Consider a set $S \subseteq \{0,1\}^n$ of codewords for some error-correcting code with relative distance d/n. We would like to bound the number of codewords in a ball of radius w. For $x, y \in \{0,1\}^n$, let $dist(x,y)$ denote the Hamming distance between x and y, and let $wt(x)$ denote the Hamming weight of x.

LEMMA 1.1 *Let $S \subseteq \{0,1\}^n$ such that for all $x, y \in S$, $x \neq y$, we have $dist(x,y) \geq d$. For some $w \in \{0,1,\ldots,n\}$, let N be the number of points in S of Hamming distance at most ω from the origin 0^n; that is, $N = |\{x \in S : wt(x) \leq w\}|$. Then we can give the following upper bound for N:*

$$N \leq \frac{1-2\beta}{4\alpha^2 - 2\beta},$$

where $\alpha = 1/2 - w/n$, and $\beta = 1/2 - d/n$.

Note that this lemma in fact applies to *any* ball of radius w: We can simply shift the set S and the ball so that the ball is centered at the origin, and then the upper bound on the intersection size will still hold.

Proof. Let

$$\Sigma = \sum_{i,j=1}^{N} \langle v_i, v_j \rangle,$$

where $v_1, \ldots, v_N$ are the points in $\{x \in S : wt(x) \leq w\}$. Note that we have

$$\langle v_i, v_j \rangle = \sum_{k=1}^{n} v_{ik} v_{jk}.$$

Also, let w_i denote the set of 1-bits in v_i so that $|w_i| = wt(v_i)$. By noticing that $\langle v_i, v_j \rangle = |w_i \cap w_j|$ and also that $dist(v_i, v_j) = |w_i| + |w_j| - 2|w_i \cap w_j|$ we can derive $\langle v_i, v_j \rangle = \frac{1}{2}(wt(v_i) + wt(v_j) - dist(v_i, v_j))$. Furthermore, because $dist(v_i, v_j) \geq d$ we have $\langle v_i, v_j \rangle \leq \frac{1}{2}(wt(v_i) + wt(v_j) - dist(v_i, v_j))$

for all $i \neq j$. Therefore we can rewrite Σ, letting $\bar{w}$ denote the average weight over all v_i:

$$\begin{aligned}
\Sigma &= \sum_{i=1}^{N}\langle v_i, v_i\rangle + \sum_{1\le i\neq j\le N}\langle v_i, v_j\rangle \\
&= \sum_{i=1}^{N} wt(v_i) + \sum_{1\le i\neq j\le N}\langle v_i, v_j\rangle \\
&\le \sum_{i=1}^{N} wt(v_i) + \sum_{1\le i\neq j\le N}\frac{wt(v_i)+wt(v_j)-d}{2} \\
&= N\bar{w} + N(N-1)(\bar{w}-\frac{d}{2}).
\end{aligned}$$

This gives an upper bound on Σ. On the other hand, we have:

$$\Sigma = \sum_{k=1}^{n}(\sum_{i=1}^{N} v_{ik})(\sum_{j=1}^{N} v_{jk}) \tag{1}$$

$$= \sum_{k=1}^{n}(\sum_{i=1}^{N} v_{ik})^2. \tag{2}$$

Let $X_k = \sum_{i=1}^{N} v_{ik}$, and note that X_k is the number of v_i that have a 1 in the k^{th} position. Then $\sum_k X_k$ is the total number of 1's over all v_i, and therefore $\sum_k X_k = N\bar{w}$. We obtain the lower bound:

$$\begin{aligned}
(2) &= \sum_{k=1}^{n} X_k^2 \\
&\ge \sum_{k=1}^{n}(\frac{N\bar{w}}{n})^2 \\
&= n(\frac{N\bar{w}}{n})^2.
\end{aligned}$$

The inequality follows because the convex sum $\sum_k X_k^2$ subject to $\sum_k X_k = N\bar{w}$ is minimized when all X_k are equal. Finally, combining the upper and lower bounds yields the following equivalent inequalities:

$$\begin{aligned}
n(\frac{N\bar{w}}{n})^2 &\le N\bar{w} + N(N-1)(\bar{w}-\frac{d}{2}) \\
N\frac{\bar{w}^2}{n^2} &\le \frac{\bar{w}}{n} + (N-1)(\frac{\bar{w}}{n}-\frac{d}{2n}) \\
N &\le \frac{\frac{1}{2}\frac{d}{n}}{(\frac{\bar{w}}{n})^2 - (\frac{\bar{w}}{n}) + \frac{1}{2}\frac{d}{n}} \\
N &\le \frac{1-2\beta}{4(\alpha')^2-2\beta},
\end{aligned}$$

where $\beta = \frac{1}{2} - \frac{d}{n}$ and $\alpha' = \frac{1}{2} - \frac{\bar{w}}{n}$. Note that the average weight of a set of weights where each is upperbounded by w must be at most w, and therefore $\bar{w} \leq w$. So for $\alpha = \frac{1}{2} - \frac{w}{n}$ we have $\alpha' > \alpha$, and it follows that

$$N \leq \frac{1-2\beta}{4(\alpha)^2 - 2\beta}.$$

LEMMA 1.2 *If every pair of points in $S \subseteq \{0,1\}^n$ has relative distance at least $\frac{1-\gamma}{2}$, then every ball of relative radius at most $1/2 - \sqrt{\gamma}$ contains at most $1/\gamma$ points in S.*

Proof. Set $d = (\frac{1-\gamma}{2})n$ and $w = (1/2 - \sqrt{\gamma})n$ and apply the previous lemma to obtain $\alpha = \sqrt{\gamma}$ and $\beta = \gamma/2$. We get:

$$N \leq \frac{1-2\beta}{4\alpha^2 - 2\beta} = \frac{1-\gamma}{3\gamma} < \frac{1}{3\gamma} < \frac{1}{\gamma}.$$

2 Reed-Soloman Code

Let $\mathbb{F} = GF[2^m]$, and let $\alpha_0, \alpha_1, \ldots, \alpha_{k-1} \in \mathbb{F}$. The Reed-Solomon (RS) encoding of $\alpha = \alpha_0 \ldots \alpha_{k-1}$ interprets the α_i as coefficients to a polynomial of degree $k-1$; that is, α defines the polynomial

$$f_\alpha(x) = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \ldots + \alpha_{k-1} x^{k-1}.$$

Then the RS encoding of α is the concatenation of the function f_α evaluated at all points a_i in the field $\mathbb{F}$:

$$RS : \alpha \mapsto (f_\alpha(a_1), f_\alpha(a_2), \ldots, f_\alpha(a_{|\mathbb{F}|})).$$

Equivalently, we may consider the $2^m \times k$ matrix

$$M = \begin{pmatrix} 1 & a_1 & a_1^2 & a_1^3 & \ldots & a_1^{k-1} \\ 1 & a_2 & a_2^2 & a_2^3 & \ldots & a_2^{k-1} \\ 1 & a_3 & a_3^2 & a_3^3 & \ldots & a_3^{k-1} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \end{pmatrix}$$

and let $\bar{\alpha} = (\alpha_0, \ldots, \alpha_{k-1})^T$, so that the RS encoding of α is $M\bar{\alpha}$. Note that if $\alpha \neq \beta$ then the corresponding polynomials $f_\alpha \neq f_\beta$ can agree on at most $k-1$ points. It follows that the RS encodings $M\bar{\alpha}$ and $M\bar{\beta}$ must disagree on at least $|\mathbb{F}| - (k-1) = 2^m - (k-1)$ entries. The distance of this RS encoding is therefore $2^m - (k-1)$.

3 Hadamard Code

Our goal is to concatenate the RS encoding with a Hadamard encoding. In the Hadamard encoding, we interpret the original input $x \in \{0,1\}^k$ as an index into the rows of a $2^k \times 2^k$ matrix H_k, and the encoding of x is the row indexed by x. H_k, the Hadamard matrix, can be defined recursively:

$$\begin{aligned} H_0 &= 1 \\ H_1 &= \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \\ &\vdots \\ H_{k+1} &= \begin{pmatrix} H_k & H_k \\ H_k & -H_k \end{pmatrix}. \end{aligned}$$

Alternately, we may give an entrywise definition of H_k. For $x, y \in \{0,1\}^k$, let h_{xy} be the entry contained in the row indexed by x and the column indexed by y. Then $h_{xy} = (-1)^{\langle x,y\rangle \pmod 2}$, where $\langle x,y\rangle = \sum_{i=1}^{n} x_i y_i$.

We will work with the $0-1$ version of the Hadamard matrix; that is, H_k will be exactly as defined above except that in place of every 1 there will be a 0 and in place of every -1 there will be a 1. Then the Hadamard encoding of $x \in \{0,1\}^k$ will be the row of the $0-1$ H_k indexed by x. It is not hard to show that codewords have a relative distance of $\frac{1}{2}$ (we leave this to the reader as an exercise).

4 Concatenation of codes

We concatenate the Reed-Soloman code with the Hadamard code to prove the following lemma.

LEMMA 4.1 *For all $n \in \mathbb{N}$ and $\delta > 0$, there exists an error-correcting code $EC_{n,\delta} : \{0,1\}^n \to \{0,1\}^{\bar{n}}$, where $\bar{n} = poly(n, \frac{1}{\delta})$, such that every Hamming ball of relative radius $\frac{1}{2} - \delta$ in $\{0,1\}^n$ contains at most $\frac{1}{\delta^2}$ codewords. Furthermore, $EC_{n,\delta}$ can be computed in time $poly(n, \frac{1}{\delta})$, and $\bar{n}$ can be assumed to be a power of 2.*

Proof. (Outline) Pick $m = \lceil \log(\frac{n}{\delta^2}) \rceil$, so that 2^m is $\frac{n}{\delta^2}$ or a little greater. The code length will be $\bar{n} = 2^{2m} \approx \frac{n^2}{\delta^4}$. For $\alpha \in \{0,1\}^n$, let $\alpha = \alpha_0\alpha_1 \ldots \alpha_{k-1}$, where $\alpha_i \in GF[2^m]$ and $k = \lfloor \frac{n}{m} \rfloor$. Then we define:

$$\begin{aligned} EC : \alpha = \alpha_0\alpha_1 \ldots \alpha_{k-1} &\overset{RS}{\mapsto} (f_\alpha(a_1), \ldots, f_\alpha(a_{2^m})) \\ &\overset{Had.}{\mapsto} (Had(f_\alpha(a_1)), \ldots, Had(f_\alpha(a_{2^m}))) \end{aligned}$$