Error-correcting codes: Basic notions

Yuval Filmus

December 7, 2017

A *linear code* of length n is a subspace of the vector space $\mathbb{Z}_2^n$, where $\mathbb{Z}_2 = \{0, 1\}$ is the additive group modulo 2. The vectors in the code are known as *codewords*. The *rate* of a code of dimension k and length n is k/n. The *minimal distance* of a code is the minimal d such that any two different codewords differ in at least d places. The *relative distance* of a code is d/n, where d is the minimal distance. An $[n, k, d]$ code is a linear code of length n, dimension k, and minimal distance d.

1. Show that the minimal distance of a code is the smallest Hamming weight of a non-zero codeword.

2. **Generating matrix**. Let C be an $[n, k, d]$ code.

   (a) Show that C is the row span of some full rank $k \times n$ matrix G over $\mathbb{Z}_2$. Such a matrix is known as a *generator matrix*.

   (b) Show that (possibly after rearranging columns), C has a generator matrix in *standard form* $G = [I_k|P]$, where I_k is the $k \times k$ identity matrix.

3. **Communication using linear codes.** Let C be an $[n, k, d]$ code with generating matrix G. For $x \in \mathbb{Z}_2^k$, define $C(x) = xG$.

   Suppose that A sends a k-bit message x to B by transmitting $C(x)$ via a noisy channel that is allowed to flip up to e bits.

   (a) **Error detection.** For which values of e can B determine (with certainty) whether the channel flipped any bits or not?

   (b) **Error correction.** For which values of e can B recover A's message (even in the presence of errors)?

4. **Code extension.** Let C be an $[n, k, d]$ code, and construct a set $C' \subseteq \{0, 1\}^{n+1}$ obtained by extending each codeword of C by a parity bit, equal to the sum of all bits in the original codeword.

   (a) Show that C' is a k-dimensional linear code of length $n + 1$.

   (b) Show that if d is even then C' is an $[n + 1, k, d]$ code, and if d is odd then C' is an $[n + 1, k, d + 1]$ code.

5. **Parity-check matrix.** Let C be an $[n, k, d]$ code.

   (a) Show that there exists a full rank $(n - k) \times n$ matrix H over $\mathbb{Z}_2$ such that $C = \{x : Hx = 0\}$. Such a matrix is known as a *parity-check matrix*.

   (b) Show that d is the minimum number of columns of H which are linearly dependent.

   (c) Deduce the *Singleton bound*: $d \leq n - k + 1$. Codes achieving this bound are known as *maximum distance separable* (MDS) codes.

6. **Dual code.** Let C be an $[n,k,d]$ code, and define the *dual code* $C^\perp = \{x \in \mathbb{Z}_2^n : x \perp y \text{ for all } y \in C\}$, where $x \perp y$ if $\langle x, y \rangle = 0$.

   (a) Show that $C^\perp$ is an $[n, n-k, D]$-code for some D (called the *dual distance*).

   (b) Show that $(C^\perp)^\perp = C$. (Hint: Show that $C \subseteq (C^\perp)^\perp$.)

   (c) Show that if C is generated by G, then a full rank matrix H is a parity-check matrix of C iff $HG^T = 0$.

   (d) Show that $C^\perp$ is generated by any parity-check matrix of C.

   (e) Suppose that C is generated by a matrix in standard form $[I_k|P]$. Show that $C^\perp$ is generated by $[-P^T|I_{n-k}]$.

7. **Hamming bound.** Let C be an $[n,k,d]$ code.

   (a) For every $w \in C$, let $B(w)$ consist of all points at Hamming distance at most $\frac{d-1}{2}$ from w. Show that the balls $\{B(w) : w \in C\}$ are disjoint.

   (b) Deduce that $2^k \leq 2^n / \binom{n}{\leq \lfloor \frac{d-1}{2} \rfloor}$, where $\binom{n}{\leq k} = \binom{n}{0} + \cdots + \binom{n}{k}$.

   (c) Use Stirling's approximation $m! \sim \sqrt{2\pi m}(m/e)^m$ to show that for constant $\theta \in (0,1)$,

   $$\binom{m}{\theta m} \sim \frac{2^{h(\theta)m}}{\sqrt{2\pi\theta(1-\theta)m}}.$$

   (d) Conclude that if C_n is a sequence of codes with rate approaching R and relative distance approaching δ then

   $$R \leq 1 - h(\delta/2).$$

8. **Gilbert–Varshamov bound.** Fix $\delta \in (0, 1/2)$ and $R \in (0,1)$.

   (a) Let G be a random $Rn \times n$ matrix over $\mathbb{Z}_2$. Show that for every $w \neq 0$, wG is a uniformly random vector in $\mathbb{Z}_2^n$.

   (b) Show that for any $w \neq 0$, the probability that wG has Hamming weight at most δn is at most $2^{(h(\delta)-1)n}$. (Hint: use the bound on $\binom{n}{\leq \delta n}$ we have seen on Week 6.)

   (c) Deduce that the probability that the code generated by G has minimum distance at most δn is at most $2^{(R+h(\delta)-1)n}$.

   (d) Conclude that for every $R < 1 - h(\delta)$ there exist arbitrarily long codes with rate R and relative distance at least δ.

   There is a gap between the Hamming and Gilbert–Varshamov bounds. While the Hamming bound can be improved (using the elementary Elias–Bassalygo bound and the spectral MRRW bound, also known as the linear programming bound), a gap still remains. This is probably the most important open question in coding theory.

Exercise: Perfect codes. An $[n,k,d]$ code is *perfect* if the inequality in 7b is tight.

1. Construct an $[n, 0, \infty]$ code, and show that it is perfect (with $\lfloor \frac{d-1}{2} \rfloor$ replaced by n).

2. Construct an $[n, n, 1]$ code, and show that it is perfect.

3. When n is odd, construct an $[n, 1, n]$ code, and show that it is perfect.

Other perfect codes include the Hamming codes (discussed next week) and the binary Golay [gɔˈlɛ] code. Every other perfect code has the same parameters as one of these codes.