Error-correcting codes: Some codes

Yuval Filmus

December 31, 2017

1. **Hamming codes.** Let $r \geq 2$. Define C to be the collection of all vectors $x \in \mathbb{Z}_2^{2^r-1}$ (indexed by non-zero r-bit vectors) such that for all $0 \leq t \leq r-1$ we have $\sum_{i \colon i_t=1} x_i = 0$.

   (a) Show that C is a linear code of dimension $2^r - r - 1$.

   (b) Show that C has minimum distance 3, and so is a $[2^r - 1, 2^r - r - 1, 3]$ code.

   (c) Show that C is a perfect code.

   (d) Construct from C a $[2^r, 2^r - r - 1, 4]$ code.

2. **Hadamard codes.** Let $k \geq 0$. For $x \in \mathbb{Z}_2^k$, let $f_x \colon \mathbb{Z}_2^k \to \mathbb{Z}_2$ be the function $f_x(y) = \sum_{i=1}^k x_i y_i$. We can identify f_x with a binary vector of length 2^k. Define H to be the collection of all vectors of the form f_x.

   (a) Show that H is a linear code of dimension k.

   (b) Show that H has minimum distance 2^{k-1}, and so is a $[2^k, k, 2^{k-1}]$ code.

3. **Polynomials over finite fields.** Let $\mathbb{F}_q$ be the finite field with q elements.[1] An n-variate polynomial over $\mathbb{F}_q$ is a sum of monomials with coefficients from $\mathbb{F}_q$ in which the degree of each variable is less than q. Each n-variate polynomial over $\mathbb{F}_q$ defines a function from $\mathbb{F}_q^n$ to $\mathbb{F}_q$, which we often identify with the polynomial.

   (a) Calculate the number of n-variate polynomials over $\mathbb{F}_q$.

   (b) Show that every function $f \colon \mathbb{F}_q^n \to \mathbb{F}_q$ can be represented as a polynomial. One way is to use the formula

   $$\sum_{y_1,\ldots,y_n \in \mathbb{F}_q} f(y_1,\ldots,y_n) \prod_{i=1}^n \prod_{z \neq y_i} \frac{x_i - z}{y_i - z}.$$

   (c) Show that every function $f \colon \mathbb{F}_q^n \to \mathbb{F}_q$ has a *unique* polynomial representation. One way is to show that the dimension of the space of functions coincides with the dimension of the space of polynomials.

   When $q = 2$, an n-variate polynomial over $\mathbb{F}_2$ is a sum of monomials of the form $x_I = \prod_{i \in I} x_i$, where $I \subseteq \{1,\ldots,n\}$. The *support* of a polynomial P, written $\operatorname{supp} P$, is the collection of subsets of $\{1,\ldots,n\}$ such that $P = \sum_{I \in \operatorname{supp} P} x_I$.

[1] A finite field with q elements exists if and only if q is a prime power, and all finite fields with q elements are isomorphic. It often suffices to consider only the finite fields $\mathbb{F}_p$ for primes p, which are just $\{0,\ldots,p-1\}$ with addition and multiplication modulo p.

4. **Reed–Muller codes.** Let $0 \le r \le m$ be parameters. Define $RM(r,m)$ to be the collection of functions $\mathbb{F}_2^m \to \mathbb{F}_2$ (encoded using their truth table) which can be described as polynomials over $\mathbb{F}_2$ of degree at most r.

   (a) Show that $RM(r,m)$ is a linear code, and determine its length and dimension.

   (b) Show that its minimum distance is at most 2^{m-r} by considering the polynomial $x_1 \cdots x_r$.

   (c) Suppose that $f \in RM(r,m)$ is non-zero, and let I be an inclusion-maximal set in $\operatorname{supp} f$ (that is, no set in $\operatorname{supp} f$ strictly contains I). Show that for each assignment for the variables $\{x_i : i \notin I\}$ there is at least one assignment to the variables $\{x_i : i \in I\}$ under which $f = 1$.

   (d) Conclude that the minimum distance of $R(r,m)$ is 2^{m-r}.

   So far we have considered codes over $\mathbb{F}_2$. However, we can consider codes over any finite field $\mathbb{F}_q$. An $[n,k,d]_q$-code is a subspace of $\mathbb{F}_q^n$ of dimension k in which the Hamming distance between any two codewords (equivalently, the Hamming weight of any non-zero codeword) is at least d.

5. **Reed–Solomon codes.** Let $q \ge n \ge k$ be parameters, where q is a prime power. Choose n distinct elements $\alpha_1, \ldots, \alpha_n \in \mathbb{F}_q$. Define $RS(q,n,k)$ to be the set of n-tuples $(f(\alpha_1), \ldots, f(\alpha_n))$, where f goes over all univariate polynomials of degree less than k.

   (a) Show that $RS(q,n,k)$ is a linear code, and determine its length and dimension.

   (b) It is known that a degree d polynomial over $\mathbb{F}_q$ has at most d roots. Use this to determine the minimum distance of $RS(q,n,k)$, and conclude that it is an MDS code.

6. **Schwartz–Zippel lemma.** The Schwartz–Zippel lemma states that if f is a non-zero multivariate polynomial over $\mathbb{F}_q$ then $\Pr[f = 0] \le \deg f/q$, where $\deg f$ is the total degree of f (the maximum of $\sum_i d_i$ over all monomials $\prod_i x_i^{d_i}$ appearing in f) and the probability is over a uniformly random input.

   (a) Prove the lemma in the univariate case.

   (b) Suppose that f depends on n variables $x_1, \ldots, x_n$, and write $f = \sum_{i=0}^{d} x_n^i f_i(x_1, \ldots, x_{n-1})$, where $f_d \not\equiv 0$. Suppose that $\alpha_1, \ldots, \alpha_{n-1}$ are such that $f_d(\alpha_1, \ldots, \alpha_{n-1}) \neq 0$. Show that $\Pr[f(\alpha_1, \ldots, \alpha_{n-1}, x_n) = 0] \le d/q$, where the probability is over x_n.

   (c) Prove the Schwartz–Zippel lemma by induction on n.

   The lemma can be improved somewhat, though in many applications the bound it gives is good enough.

7. **Non-binary Reed–Muller codes.** Reed–Muller codes can be generalized to arbitrary fields. Define $RM(q,r,m)$ to be the collection of functions $\mathbb{F}_q^m \to \mathbb{F}_q$ which can be described as polynomials of total degree at most m.

   (a) Show that $RM(q,r,m)$ is a linear code, and determine its length and dimension.

   (b) Use the Schwartz–Zippel lemma to lower bound the minimum distance of $RM(q,r,m)$.

(Exercise on the following page.)

Exercise: Communication protocol for the equality function.

(a) Let $n = 3k$. Show that there is a prime power q such that $n \leq q < 2n$. (In fact, Bertrand's postulate states that there is a *prime* q satisfying $n \leq q < 2n$.)

Alice and Bob each hold a k-bit string, and they wish to determine whether their strings are identical. They encode their inputs using the $RS(q, n, k)$ code. Alice sends a random $i \in \{1, \dots, n\}$, and Alice and Bob exchange the ith element of the encoding of their strings. They accept if the elements are identical.

(b) Show that if the inputs are equal then Alice and Bob always accept.

(c) Give an upper bound on the probability that Alice and Bob accept when their inputs are different.

(d) How many bits do Alice and Bob communicate?