Reed-Solomon Codes

(Handout February 20, 2013)

Let $\mathbb{F}_q$ be a finite field of order q. Here $q = p^r$ for some prime p and integer $r \geq 1$, and that $\mathbb{F}_q \supseteq \mathbb{F}_p$ is an extension of degree r. If your experience with finite fields is minimal, think of the special case $r = 1$ so that $\mathbb{F}_q = \mathbb{F}_p = \{0, 1, 2, \ldots, p-1\}$ is the field of integers mod p.

Lemma 1. For every nonzero $a \in \mathbb{F}_q$, we have $a^{q-1} = 1$.

Proof. Consider the product of all nonzero elements of $\mathbb{F}_q$, thus:

$$u = \prod_{0 \neq x \in \mathbb{F}_q} x \ \in \ \mathbb{F}_q \, .$$

Note that this product has $q - 1$ factors; and since the field elements commute, the value of the product u does not depend on the order in which we multiply together these factors. In particular if we substitute $y = ax$, we simply permute the factors in this product (the map $x \mapsto ax$ is bijective since its inverse is the map $x \mapsto a^{-1}x$). So

$$u = \prod_{0 \neq x \in \mathbb{F}_q} (ax) \ \in \ \mathbb{F}_q \ = \ a^{q-1} \prod_{0 \neq x \in \mathbb{F}_q} x \ = \ a^{q-1} u \ \in \ \mathbb{F}_q \, .$$

Moreover $u \neq 0$ since it is a product of nonzero field elements; so we may cancel u's to obtain $a^{q-1} = 1$. □

By convention, we agree that $0^0 = 1$.

Lemma 2. Let $k \in \{0, 1, 2, \ldots, q-1\}$. Then

$$\sum_{a \in \mathbb{F}_q} a^k = \begin{cases} 0, & \text{if } k \in \{0, 1, 2, \ldots, q-2\}; \\ -1, & \text{if } k = q-1. \end{cases}$$

Proof. First suppose $k \in \{0, 1, 2, \ldots, q-2\}$ and consider the sum of k-th powers of the field elements:

$$S_k = \sum_{x \in \mathbb{F}_q} x^k .$$

For every nonzero value $a \in \mathbb{F}_q$, the map $x \mapsto ax$ is bijective (just as in the proof of Lemma 1) so we may substitute $y = ax$ for x in the definition of S_k (which merely permutes the terms in the sum). Thus

$$S_k = \sum_{x \in \mathbb{F}_q} (ax)^k = a^k \sum_{x \in \mathbb{F}_q} x^k = a^k S_k$$

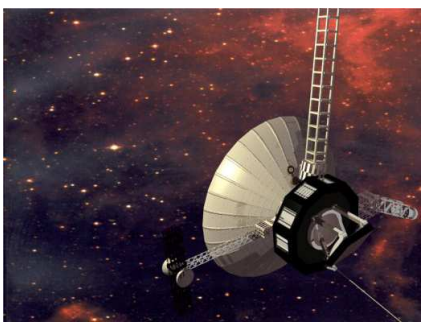
for all nonzero values of $a \in \mathbb{F}_q$. If $S_k \neq 0$ then we may cancel S_k's to obtain $a^k = 1$ for all nonzero $k \in \mathbb{F}_q$. This means that the polynomial $X^k - 1 \in \mathbb{F}_q[X]$ has $q-1$ distinct roots in $\mathbb{F}_q$. This is impossible since the polynomial $X^k - 1$ has degree only $k \leq q-2$. Our assumption that $S_k \neq 0$ has led to this contradiction; therefore in fact we must have $S_k = 0$ for all $k \in \{0, 1, 2, \ldots, q-2\}$.

Finally consider the case $k = q-1$. We have

$$S_{q-1} = \sum_{x \in \mathbb{F}_q} x^{q-1} = q - 1$$

since $0^{q-1} = 0$, whereas the remaining $q-1$ terms in the sum all have value 1 by Lemma 1. Here we must interpret the right hand side $q-1$ as an element of $\mathbb{F}_q$, and in fact an element of the subfield $\mathbb{F}_p$ where we reduce mod p to obtain $p-1$, or simply -1, as our final answer. □

Irving Reed (1923-2012) Gustave Solomon (1930-1996)



Voyager spacecraft (1977)

In 1960, Irving Reed and Gustave Solomon designed an infinite family of MDS codes using properties of finite fields. The first significant application of these codes was for the encoding of transmissions from the Voyager spacecraft launched during the 1970's. Beginning in the late 1980's, Reed Solomon codes found application in the encoding of digital information used in CD's, DVD's, video game devices, electronic data storage media, and a host of other devices. Most of these applications use fields of order $q = 2^r$ for convenience (the field of order $256 = 2^8$ being a popular choice; note that elements of $\mathbb{F}_{256}$ are conveniently representable as computer bytes, i.e. bitstrings of length 8). For ease of notation, we begin with the case of prime fields, and later point out that general finite fields are really no harder to work with than prime order fields.

Consider the field $\mathbb{F}_p = \{0, 1, 2, \ldots, p-1\}$ of prime order p and fix $k \in \{1, 2, \ldots, p-2\}$. Consider the code

$$\mathcal{C} = \big\{(f(0), f(1), f(2), \ldots, f(p-1)) \,:\, f(t) \in \mathbb{F}_p[t] \text{ of degree } < k\big\} \leq \mathbb{F}_p^p \,.$$

This is actually a linear code of length p over $\mathbb{F}_p$; for example if we add the codewords corresponding to two polynomials $f(t)$ and $g(t)$, we obtain the codeword corresponding to the polynomial $f(t)+g(t)$. Every polynomial $f(t) \in \mathbb{F}_p[t]$ of degree less than k has the form

$$f(t) = a_0 + a_1 t + a_2 t^2 + \cdots + a_{k-1} t^{k-1}$$

where $a_0, a_1, \ldots, a_{k-1} \in \mathbb{F}_p$. There are p^k such polynomials $f(t)$ (since there are p choices for each coefficient a_i). If $f(t)$ is not the zero polynomial, it has at most $\deg(f(t)) \leq k-1$ zeroes, so the corresponding codeword $(f(0), f(1), \ldots, f(p-1))$ has at least $p - (k-1)$ nonzero entries. Thus $\mathcal{C}$ has minimum weight at least $p-k+1$. The same argument shows that the linear map $f(t) \mapsto (f(0), f(1), f(2), \ldots, f(p-1))$ is nonsingular, so the code $\mathcal{C}$ has dimension k. Now $\mathcal{C}$ is a $[p, k, p-k+1]$-code, called a *Reed-Solomon code.* Since it meets the Singleton bound, it is an MDS code.

A generator matrix for the Reed-Solomon code of length p and dimension k is given by

$$G = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 0 & 1 & 2 & \cdots & p-1 \\ 0^2 & 1^2 & 2^2 & \cdots & (p-1)^2 \\ \vdots & \vdots & \vdots & & \vdots \\ 0^{k-1} & 1^{k-1} & 2^{k-1} & \cdots & (p-1)^{k-1} \end{bmatrix}.$$

A typical message word $x = (a_0, a_1, a_2, \ldots, a_{k-1}) \in \mathbb{F}_p^k$ is encoded as

$$xG = ((f(0), f(1), f(2), \ldots, f(p-1)) \in \mathbb{F}_p^p$$

where

$$f(t) = a_0 + a_1 t + a_2 t^2 + \cdots + a_{k-1} t^{k-1} \in \mathbb{F}_p[t].$$

By Lemma 2, the rows of the matrix

$$H = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 0 & 1 & 2 & \cdots & p-1 \\ 0^2 & 1^2 & 2^2 & \cdots & (p-1)^2 \\ \vdots & \vdots & \vdots & & \vdots \\ 0^{p-k-1} & 1^{p-k-1} & 2^{p-k-1} & \cdots & (p-1)^{p-k-1} \end{bmatrix}$$

are orthogonal to the rows of G. But from what we have already seen, H is a generator matrix for a Reed-Solomon $[p, p-k, k+1]$-code; in particular the rank of H is $p-k$ and H is a parity check matrix for $\mathcal{C}$.

We have just observed that the dual of a Reed-Solomon code is another Reed-Solomon code, of the complementary dimension. It is not hard to show that, more generally, the dueal of every MDS code is MDS.

Finally, the entire description above works just as well for an arbitrary finite field $\mathbb{F}_q = \{u_0, u_1, u_2, \ldots, u_{q-1}\}$. (Only the notation is a little more involved since the field elements are not simply $0, 1, 2, \ldots, q-1$ unless q is prime.) The Reed-Solomon code of dimension $k \in \{1, 2, \ldots, q-2\}$ is the $[q, k, q-k+1]$-code over $\mathbb{F}_q$ consisting of all vectors of the form

$$(f(u_0), f(u_1), f(u_2), \ldots, f(u_{q-1})) \in \mathbb{F}_q^q$$

where the polynomial $f(t) \in \mathbb{F}_q[t]$ has degree less than k.

Example. Take $q = 7$, $k = 4$. The matrix

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 & 4 & 5 & 6 \\ 0 & 1 & 4 & 2 & 2 & 4 & 1 \end{bmatrix}$$

is a generator matrix for a Reed Solomon $[7, 3, 5]$-code $\mathcal{C}$ over $\mathbb{F}_7$. We obtain a reduced row echelon form

$$G = \begin{bmatrix} 1 & 0 & 0 & 1 & 3 & 6 & 3 \\ 0 & 1 & 0 & 4 & 6 & 6 & 4 \\ 0 & 0 & 1 & 3 & 6 & 3 & 1 \end{bmatrix}$$

as an alternative generator matrix for $\mathcal{C}$. Note that $\mathcal{C}$ is a 2-error correcting code since $\left\lfloor \frac{5-1}{2} \right\rfloor = 2$.