Codes and Cryptography

Jorge L. Villar

MAMME, Fall 2015

PART VI

Outline

Linear Codes

A q-ary **linear code** $\mathcal{C}$ is a vector subspace of $\mathbb{F}_q^n$. (Now, q must be prime or a prime power.)

- The length of $\mathcal{C}$ is n and the size is q^k, where k is the dimension of the subspace.
- $\mathbf{0} = (0, \ldots, 0)$ is always a codeword.
- The minimum distance is now

$$d_{\mathcal{C}} = \min_{\mathbf{x},\mathbf{y}\in\mathcal{C},\, \mathbf{x}\neq\mathbf{y}} d(\mathbf{x},\mathbf{y}) = \min_{\mathbf{x}\in\mathcal{C},\, \mathbf{x}\neq\mathbf{0}} d(\mathbf{x},\mathbf{0})$$

  The distance $d(\mathbf{x},\mathbf{0})$ is the number of non-zero components, also called the **weight** of $\mathbf{x}$. Thus, $d_{\mathcal{C}}$ is the minimum positive weight.

We call the code a $[n,k,d]_q$-linear code.

Generating Matrix

A linear code $\mathcal{C}$ can be described by a basis of the subspace. Arranging the basis as rows of a matrix we obtain a **generating matrix** G for $\mathcal{C}$, which is a $k \times n$ full-rank matrix.

In a previous example in $\mathbb{F}_3^4$ the linear code $\mathcal{C} = \{0000, 0122, 0211, 1012, 1101, 1220, 2021, 2110, 2202\}$ can be generated from $\{0122, 1220\}$, defining the generating matrix

$$G = \begin{pmatrix} 0 & 1 & 2 & 2 \\ 1 & 2 & 2 & 0 \end{pmatrix}$$

Encoding is simply multiplying by G:
For any $\mathbf{s} \in \mathbb{F}_q^k$, $\mathrm{Enc}(\mathbf{s}) = \mathbf{s}G$

Systematic Codes

A linear code $\mathcal{C}$ is called **systematic** if the source word is a prefix of the corresponding codeword.

For linear codes, the generating matrix has the form

$$G = \left(I \middle| A\right)$$

In the example,

$$G = \left(\begin{array}{cc|cc} 1 & 0 & 1 & 2 \\ 0 & 1 & 2 & 2 \end{array}\right)$$

Then $\text{Enc}((s_1, s_2)) = (s_1, s_2, s_1 + 2s_2, 2s_1 + 2s_2)$

"Parity Check" Matrix

For a $[n, k, d]_q$-linear code $\mathcal{C}$ with generating matrix G, we call **"parity-check" matrix**, or better the error-checking matrix, to any $(n-k) \times n$ full-rank matrix H such that

$$GH^\top = 0^{k \times (n-k)}$$

that is, the rows of H span the orthogonal subspace of $\mathcal{C}$.

In the example,

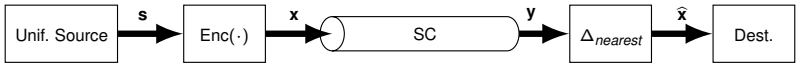
$$H = \begin{pmatrix} 1 & 0 & 1 & 2 \\ 1 & 1 & 0 & 1 \end{pmatrix}$$

If $G = (I \mid A)$ then we simply set $H = (-A^\top \mid I)$.

Curiously, in the example we can just take $H = G$ as $GG^\top = 0^{2\times 2}$

Syndromes and Error Correction

Decoding a linear code is simply projecting a vector (or corrupted codeword) onto the code.



For a $[n, k, d]_q$-linear code $\mathcal{C}$ with generating matrix G and error-checking matrix H, a corrupted codeword can be written as $\mathbf{y} = \mathbf{x} + \mathbf{e}$. The weight of $\mathbf{e}$ is at most $t = \lfloor \frac{d-1}{2} \rfloor$.

$$\mathbf{y}H^\top = \mathbf{x}H^\top + \mathbf{e}H^\top = \mathbf{e}H^\top$$

$\mathbf{e}H^\top$ is called the **syndrome** of $\mathbf{e}$ (which is independent of $\mathbf{x}$).

If we store the syndromes of all vectors $\mathbf{e}$ with weight up to t, we can easily recover $\mathbf{e}$ (and then, correct errors) by table lookup.

It requires storage $\left(\binom{n}{1} + \ldots + \binom{n}{t}\right) n \log_2 q$.

The Dual Code

For a $[n, k, d]_q$-linear code $\mathcal{C}$ we define its **dual code** $\mathcal{C}^\perp$ as the orthogonal subspace of $\mathcal{C}$.

- an error-checking matrix H for $\mathcal{C}$ is a generating matrix of $\mathcal{C}^\perp$, and vice versa
- $(\mathcal{C}^\perp)^\perp = \mathcal{C}$
- $\mathcal{C}^\perp$ is a $[n, n-k, d^\perp]_q$-linear code, for some $d^\perp$
- $d^\perp$ is the minimal number of linearly dependent columns in G

A code $\mathcal{C}$ is called **self-dual** if $\mathcal{C}^\perp = \mathcal{C}$ (as occurs with the previous example).

Notice that a linear code can only be self-dual if $n = 2k$.

Outline

Hamming Codes

Hamming codes are $[n, k, d]_q$-linear perfect codes with $d = 3$. The rows of the parity check matrix H are all nonzero vectors in $\mathbb{F}_q^m$ with the first nonzero component equal to 1, for a suitable dimension m.

Therefore, H is a $m \times n$ matrix in which any two columns are linearly independent, where $n = \dfrac{q^m - 1}{q - 1}$.

The dimension of the code is $k = n - m$.

$\mathcal{C}$ is perfect, since it meets the sphere packing bound:

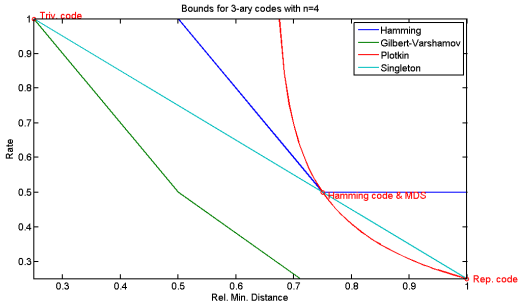
$$q^k = \frac{q^n}{(1 + n(q-1))}$$

They can correct only one error

Example of a Hamming $[4, 2, 3]_3$ Code

The previously used example: $q = 3$, $m = 2$, $n = 4$, $k = 2$

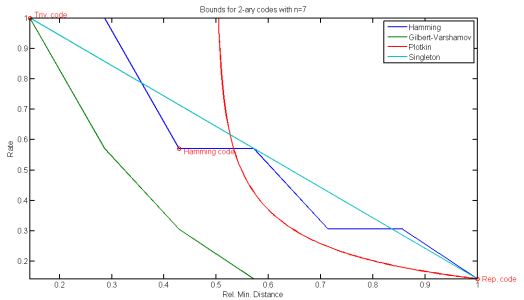
$$H = \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 2 \end{pmatrix} \qquad G = \begin{pmatrix} 1 & 0 & 1 & 2 \\ 0 & 1 & 2 & 2 \end{pmatrix}$$



Example of a Hamming $[7, 4, 3]_2$ Code

$q = 2$, $m = 3$, $n = 7$, $k = 4$

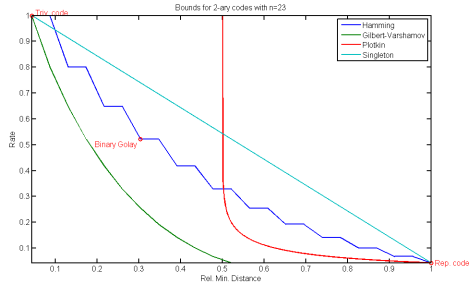
$$H = \begin{pmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{pmatrix} \quad G = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix}$$



Binary Golay Code

$[23, 12, 7]_2$-linear perfect code with generator matrix

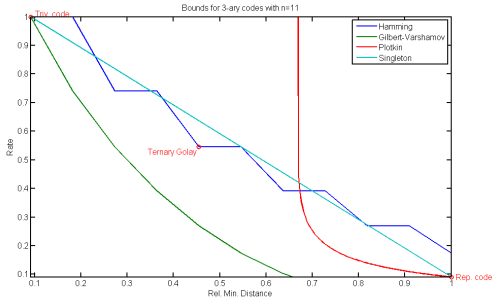
$$\begin{pmatrix}
1&0&0&0&0&0&0&0&0&0&0&0&1&1&1&1&1&1&1&1&1&1&1\\
0&1&0&0&0&0&0&0&0&0&0&0&1&1&0&1&1&1&0&0&0&1&0\\
0&0&1&0&0&0&0&0&0&0&0&0&1&0&1&1&1&0&0&0&1&0&1\\
0&0&0&1&0&0&0&0&0&0&0&0&1&1&1&0&0&0&1&0&1&1&0\\
0&0&0&0&1&0&0&0&0&0&0&0&1&1&1&0&0&0&1&0&1&1&0\\
0&0&0&0&0&1&0&0&0&0&0&0&1&1&0&0&0&1&0&1&1&0&1\\
0&0&0&0&0&0&1&0&0&0&0&0&1&0&0&0&1&0&1&1&0&1&1\\
0&0&0&0&0&0&0&1&0&0&0&0&0&0&1&0&1&1&0&1&1&1&1\\
0&0&0&0&0&0&0&0&1&0&0&0&0&0&1&0&1&1&0&1&1&1&0\\
0&0&0&0&0&0&0&0&0&1&0&0&0&1&0&1&1&0&1&1&1&0&0\\
0&0&0&0&0&0&0&0&0&0&1&0&1&0&1&1&0&1&1&1&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&1&0&1&1&0&1&1&1&0&0&0&1
\end{pmatrix}$$



Ternary Golay Code

$[11, 6, 5]_3$-linear perfect code

$$G = \begin{pmatrix}
1&0&0&0&0&0&1&1&1&1&1\\
0&1&0&0&0&0&2&2&1&1&0\\
0&0&1&0&0&0&2&1&2&0&1\\
0&0&0&1&0&0&1&2&0&2&1\\
0&0&0&0&1&0&1&0&2&1&2\\
0&0&0&0&0&1&0&1&1&2&2
\end{pmatrix}$$



Outline

Linear MDS Codes

Singleton bound for a $[n, k, d]_q$-linear code: $k \leq n - d + 1$.
$[n, k, n - k + 1]_q$-linear codes are the MDS linear codes.

- G is the generating matrix of a $[n, k, n - k + 1]_q$-linear MDS code if and only if every k columns of G are linearly independent
  Removing columns from G does not break the property
- A linear code $\mathcal{C}$ is MDS if and only if $\mathcal{C}^\perp$ is MDS
- For a linear MDS code, if $k \geq q$ then $n \leq k + 1$
  E.g., $\mathcal{C} = \mathbb{F}_q$ or $\mathcal{C} = \{(x_1, \ldots, x_n) \mid x_1 + \ldots + x_n = 0\}$
- For a linear MDS code, if $1 < k < q$ and q is prime, then $n \leq q + 1$
  E.g., the (extended) Reed-Solomon code

Reed-Solomon Codes (I)

Let $\mathbb{F}_q[X]_{\leq k-1}$ be the set of polynomials in $\mathbb{F}_q[X]$ with degree at most $k-1$, for $k < q$. It is a $\mathbb{F}_q$-vector space of dimension k, and there is a natural bijection from $\mathbb{F}_q^k$ to $\mathbb{F}_q[X]_{\leq k-1}$.

$$(a_1, \ldots, a_k) \mapsto a_1 X^{k-1} + \ldots + a_k X^0$$

Consider now the evaluation map $ev : \mathbb{F}_q[X] \to \mathbb{F}_q^q$ defined as

$$ev(h) = (h(0), h(\alpha^0), \ldots, h(\alpha^{q-2}))$$

where α is a primitive element of $\mathbb{F}_q$.

The image $\mathcal{C} = ev(\mathbb{F}_q[X]_{\leq k-1})$ is the Reed-Solomon code, which is a $[q, k, q-k+1]_q$-linear MDS code.

Reed-Solomon Codes (II)

- We can easily extend the code to a $[q+1, k, q-k+2]_q$-linear MDS code, adding the an extra component a_{k-1} to the codeword.
- We can also shorten the code to a $[n, k, n-k+1]_q$-linear MDS code, for any $n < q$, by evaluating the polynomial at a subset of the finite field.

Reed-Solomon codes are MDS:
Any nonzero codeword with weight d corresponds to a polynomial with at least $n-d$ roots. But then $n-d \leq k-1$, the maximum degree of the polynomial. Therefore, $d \geq n-k+1$, but by Singleton's bound, only the equality holds.

The extended Reed-Solomon code is proven to be MDS in a similar way.

Reed-Solomon Codes (III)

The generating matrix when the polynomials are evaluated at $\{0, 1, \alpha, \ldots, \alpha^{n-2}\}$ is the Vandermonde matrix:

$$G = \begin{pmatrix} 1 & 1 & 1 & 1 & \cdots & 1 \\ 0 & 1 & \alpha & \alpha^2 & \cdots & \alpha^{n-2} \\ 0 & 1 & \alpha^2 & \alpha^4 & \cdots & \alpha^{2(n-2)} \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ 0 & 1 & \alpha^{k-1} & \alpha^{2(k-1)} & \cdots & \alpha^{(k-1)(n-2)} \end{pmatrix}$$

In the extended code the extra column $(0, \ldots, 0, 1)^\top$ is added to the matrix.

The columns can be reordered without affecting the code properties

Reed-Solomon Codes (IV)

Reed-Solomon codes can be made systematic by using Lagrange interpolation instead of direct polynomial evaluation.

Indeed, we only need to determine the interpolation map $L : \mathbb{F}_q^k \to \mathbb{F}_q^n$

$$(h(\xi_1), \ldots, h(\xi_k)) \mapsto (h(\xi_1), \ldots, h(\xi_n))$$

for any set of n different evaluation points $\xi_1, \ldots, \xi_n \in \mathbb{F}_q$.

The map is well-defined, because $(h(\xi_1), \ldots, h(\xi_k))$ uniquely determines a polynomial $h \in \mathbb{F}_q[X]$ of degree $\leq k-1$.

An Example of Reed-Solomon Code

The $[8,4,5]_7$ Reed-Solomon code:
Finding a primitive element in $\mathbb{F}_7$:
Trying $\alpha = 2$: $\alpha^2 = 4$, $\alpha^3 = 1$ NO!
Trying $\alpha = 3$: $\alpha^2 = 2$, $\alpha^3 = 6$ YES! $\alpha^4 = 4$, $\alpha^5 = 5$
Building the generating matrix:

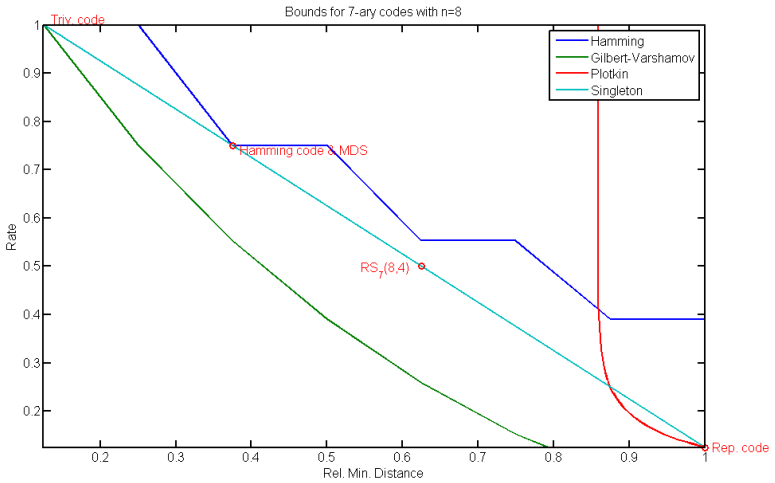
$$G = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 \\ 0 & 1 & 3 & 2 & 6 & 4 & 5 & 0 \\ 0 & 1 & 2 & 4 & 1 & 2 & 4 & 0 \\ 0 & 1 & 6 & 1 & 6 & 1 & 6 & 1 \end{pmatrix}$$

Turning it into systematic by Gauss elimination:

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

The code is selfdual! . . . Choosing $H = G$

The $[8,4,5]_7$ Reed-Solomon code:



Codes and Cryptography

Jorge L. Villar

MAMME, Fall 2015

END OF PART VI