236608 - Coding and Algorithms for Memories
Winter Semester 2018/2019

Linear Codes


Linear Codes

For a prime \( p \), \( GF(p) \) (Galois field of size \( p \)) denotes the integer ring modulo \( p \) which is well-known to be a field.

**Remark 1.** In this class we will only deal with fields of the form \( GF(p) \), where \( p \) is a prime number.

Let \( F \) be a field of size \( p \), where \( p \) is a prime number. An \((n, M, d)\) code \( \mathcal{C} \) over a field \( F \) is called linear if for all \( c_1, c_2 \in \mathcal{C} \) and \( a_1, a_2 \in F \), it holds that \( a_1 c_1 + a_2 c_2 \in \mathcal{C} \). That is \( \mathcal{C} \) is a linear sub-space of \( F^n \) over \( F \).

The *dimension* of a linear \((n, M, d)\) code \( \mathcal{C} \) over \( F \) is the dimension of \( \mathcal{C} \) as a linear sub-space of \( F^n \) over \( F \). Linear codes will be denoted by \([n, k, d]\), where \( k \) denotes the dimension. The difference \( n - k \) is called the *redundancy* of the code and will be denoted by \( r \).

Every basis of a linear \([n, k, d]\) code \( \mathcal{C} \) over \( F \) contains \( k \) codewords, and all the linear combinations of these \( k \) codewords generate the code \( \mathcal{C} \). Therefore \( M = |\mathcal{C}| = |F|^k \) and the code rate is \( R = \frac{\log_{|F|} M}{n} = k/n \).

**Generator Matrix**

A *generator matrix* of a linear \([n, k, d]\) code over \( F \) is a \( k \times n \) matrix whose rows form a basis of the code. The generator matrix is denoted by \( G \), and usually a code can have more than one generator matrix.

**Example 1.**

The \((5, 2, 5)\) code \( C_{5}^{rep} = \{00000, 11111\} \) is a binary linear \([5, 1, 5]\) code which is spanned by the vector \((1, 1, 1, 1, 1)\). The generator matrix of this code is \( G = (1 1 1 1 1) \).
The $(3,4,2)$ binary parity code $C_{3}^{\text{par}} = \{000,011,101,110\}$ is a binary $[3,2,3]$ code. It is spanned by the vectors $(0,1,1)$ and $(1,0,1)$ so the matrix

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

is a generator matrix of $C_{3}^{\text{par}}$. Note that also the matrix

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

is a generator matrix of $C_{3}^{\text{par}}$. \[\square\]

**Theorem 1.** Let $\mathcal{C}$ be a linear $[n,k,d]$ code over $F$. Then

$$d = \min_{c \in \mathcal{C}, c \neq 0} \{w_H(c)\}.$$

**Proof.** The proof appears in Proposition 2.1. Given in class.

An **encoder** to a binary linear $[n,k,d]$ code $\mathcal{C}$ with a generator matrix $G$ is a mapping $E_{\mathcal{C}} : \{0,1\}^k \to \mathcal{C}$ which is defined as follows: For every $u \in \{0,1\}^k$, $E_{\mathcal{C}}(u) = u \cdot G$.

Since the rank of $G$ is $k$, it is possible to apply elementary operations to the rows and obtain a $k \times n$ matrix that contains as a sub-matrix a $k \times k$ identity matrix. In case this generator matrix will have the form $(I_k|A)$, where $I_k$ is the $k \times k$ identity matrix and $A$ is a $k \times (n-k)$ matrix, then it will be called a **systematic** generator matrix.

If the code has a systematic generator matrix $G = (I_k|A)$ which is used for the encoder $E_{\mathcal{C}} : \{0,1\}^k \to \mathcal{C}$, then we get for all $u \in \{0,1\}^k$,

$$E_{\mathcal{C}}(u) = u \cdot G = u \cdot (I_k|A) = (u,u \cdot A).$$

**Parity-Check Matrix**

A **parity-check matrix** of a linear $[n,k,d]$ code $\mathcal{C}$ over $F$ is an $r \times n$ matrix $H$ over $F$ such that for every $c \in F^n$

$$c \in \mathcal{C} \iff H \cdot c^T = 0.$$

Hence, the code $\mathcal{C}$ is the right kernel of $H$, which is denoted by $\ker(H)$. Hence,

$$\text{rank}(H) = n - \dim \ker(H) = n - k.$$

If $H$ is of full rank, then $r = n - k$.

Let $G$ be a generator matrix of $\mathcal{C}$. The rows of $G$ span $\ker(H)$ and in particular,

$$H \cdot G^T = 0 \implies G \cdot H^T = 0.$$
Furthermore,
\[ \dim \ker(G) = n - \text{rank}(G) = n - k, \]
so the rows of \( H \) span \( \ker(G) \), and a parity-check matrix of a linear code can be computed by finding a basis of the kernel of a generator matrix of the code.

In the special case where \( G \) is a systematic matrix of the form \( G = (I_k|A) \), then the matrix \( H = (-A^T|I_{n-k}) \) is a parity-check matrix of the code.

**Theorem 2.** Let \( H \) be a parity-check matrix of a non-trivial linear code \( C \) (i.e. \( C \neq \{0\} \)). The minimum distance of the code \( C \) is the largest integer \( d \) such that every set of \( d - 1 \) columns in \( H \) is linearly independent.

**Proof.** The proof appears in theorem 2.2. Given in class.

**Example 2.**

1. A parity-check matrix of the \([5,1,5]\) code \( C_5^{rep} \) is
\[
H = \begin{pmatrix}
1 & 1 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 \\
1 & 0 & 0 & 1 & 0 \\
1 & 0 & 0 & 0 & 1 \\
\end{pmatrix}
\]

This matrix is constructed using the property that for a systematic matrix \( G = (I|A) \), the matrix \( H = (-A^T|I) \) is a parity-check matrix of the code.

2. A parity check matrix of the \([3,2,2]\) binary parity code \( C_3^{par} \) is
\[
H = \begin{pmatrix}
1 & 1 & 1 \\
\end{pmatrix}
\]
which is derived from the systematic generator matrix of the code.

\[
G = \begin{pmatrix}
1 & 0 & 1 \\
0 & 1 & 1 \\
\end{pmatrix}.
\]

**Example 3.** The linear \([7,4,3]\) binary Hamming code is defined by the parity check matrix
\[
H = \begin{pmatrix}
0 & 0 & 0 & 1 & 1 & 1 & 1 \\
0 & 1 & 1 & 0 & 0 & 1 & 1 \\
1 & 0 & 1 & 0 & 1 & 0 & 1 \\
\end{pmatrix}.
\]

A respective generator matrix is given by
\[
G = \begin{pmatrix}
1 & 1 & 1 & 1 & 1 & 1 & 1 \\
0 & 0 & 0 & 1 & 1 & 1 & 1 \\
0 & 1 & 1 & 0 & 0 & 1 & 1 \\
1 & 0 & 1 & 0 & 1 & 0 & 1 \\
\end{pmatrix}.
\]
It is possible to verify that $H \cdot G^T = 0$. The minimum distance of the code is 3, which can be verified using Theorem ??.

For any integer $m > 1$, the $[2^m - 1, 2^m - 1 - m, 3]$ binary Hamming code is defined by an $m \times (2^m - 1)$ parity check matrix whose columns range over all the nonzero vectors of $\{0, 1\}^m$. According to Theorem ??, it also results that the minimum distance of the code is 3. ■

Decoding of Linear Codes

The nearest-codeword decoder of a code $C$ returns for every word $y \in F^n$ the closest codeword $c \in C$ in Hamming distance. That is, for a word $y \in F^n$ the decoder returns a codeword $c \in C$ that minimizes the value $d_H(y, c)$. This is equivalent to finding a word $e \in F^n$ of minimum Hamming weight such that $y - e \in C$.

Assume that $H$ is an $(n - k) \times n$ parity-check matrix of $C$ (i.e., the rows of $H$ are linearly independent). For a word $y \in F^n$, we define its syndrome to be the vector $s \in F^n$ by

$$s = H \cdot y^T.$$  

For example, if $y$ is a codeword in $C$, then its syndrome is the zero vector $0$, since $H \cdot y^T = 0$ when $y \in C$. Furthermore, for every two words $y_1, y_2 \in F^n$,

$$y_1 - y_2 \in C \iff H \cdot y_1^T = H \cdot y_2^T.$$  

A set of words $y$ which all the have same syndrome is called a coset. For example, the set of words which corresponds to the coset $0$ is the code $C$. In general, there are $|F|^{n-k}$ cosets.

The nearest-codeword decoding for linear codes can be performed in the following three steps:

1. Compute the syndrome of the received word $y$ according to: $s = H \cdot y^T$.
2. Find a minimum-weight word $e \in F^n$ such that $s = H \cdot e^T$.
3. Decode $y$ to the codeword $c = y - e$.

Note that using this method all words in the same coset are decoded in the same way, and the vector $e$ which is used for decoding these words is called the coset leader.

Example 4. Consider the linear $[7, 4, 3]$ binary Hamming code from Example ???. Assume that the codeword $c = (0, 0, 0, 0, 0, 0)$ was transmitted and there was a single error in the fourth bit so the received word is $y = (0, 0, 0, 1, 0, 0, 0)$. The nearest-codeword decoder is operated as follows:
1. Compute \( s = H \cdot y^T \):

\[
\begin{pmatrix}
0 & 0 & 0 & 1 & 1 & 1 & 1 \\
0 & 1 & 1 & 0 & 0 & 1 & 1 \\
1 & 0 & 1 & 0 & 1 & 0 & 1
\end{pmatrix}
\begin{pmatrix}
0 \\
0 \\
1 \\
0 \\
0 \\
0
\end{pmatrix}
= \begin{pmatrix}
1 \\
0 \\
0
\end{pmatrix}
\]

2. The vector \( e \) with the same syndrome \((1, 0, 0)^T\) and minimum Hamming weight is \( e = (0, 0, 1, 0, 0, 0) \).

3. The decoded codeword is \( y + e = (0, 0, 0, 0, 0, 0) \).

Note that in this case we can find the vector \( e \) simply by the corresponding column of the parity check matrix. The full decoding procedure for each syndrome is listed in the following table.

<table>
<thead>
<tr>
<th>Syndrome</th>
<th>Coset Leader</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0, 0, 1)</td>
<td>(1, 0, 0, 0, 0, 0)</td>
</tr>
<tr>
<td>(0, 1, 0)</td>
<td>(0, 1, 0, 0, 0, 0)</td>
</tr>
<tr>
<td>(0, 1, 1)</td>
<td>(0, 0, 1, 0, 0, 0)</td>
</tr>
<tr>
<td>(1, 0, 0)</td>
<td>(0, 0, 0, 1, 0, 0)</td>
</tr>
<tr>
<td>(1, 0, 1)</td>
<td>(0, 0, 0, 0, 1, 0)</td>
</tr>
<tr>
<td>(1, 1, 0)</td>
<td>(0, 0, 0, 0, 0, 1)</td>
</tr>
<tr>
<td>(1, 1, 1)</td>
<td>(0, 0, 0, 0, 0, 1)</td>
</tr>
</tbody>
</table>

Problem 1. Show that the number of distinct generator matrices of a linear \([n, k, d]\) code over \( F = GF(p) \) is \( \prod_{i=0}^{k-1} (p^k - p^i) \).

Solution: We count the number of options to create a generator matrix for the code, while building the matrix row by row. For the first row, we can place every non-zero codeword so there are \( p^k - 1 \) options. For the second row, we can place every codeword which is not linearly dependent on the first row, so there are \( p^k - p \) options. Similarly, for the third row we avoid codewords which are in the span of the first two rows, and there are \( p^k - p^2 \) options, and so on.

Problem 2. For a positive integer \( m \) let \( n = 2^m - 1 \) and let \( C \) be the binary \([n, n - m, 3]\) Hamming code. Let \( H \) be a parity-check matrix of the code \( C \).

1. Show that for every two distinct columns \( h_1 \) and \( h_2 \) in \( H \), there is a unique third column in \( H \) that equals the sum \( h_1 + h_2 \).
**Solution:** Since the parity-check matrix $H$ consists of all non-zero vectors of $\{0,1\}^m$, the column $h_1 + h_2$ also appears in $H$. Every column appears exactly once in $H$, hence the uniqueness.

2. Show that the number of codewords of Hamming weight three in $C$ is $\binom{n}{2}/3 = n(n-1)/6$.

**Solution:** For every $0 \leq i_1, i_2 \leq n-1$ there exists exactly one codeword of weight three in $C$ with ones at position $i_1, i_2, i_3$, where the value of $i_3$ is calculated as follows: If $h_{i_1}, h_{i_2}$ is the $i_1, i_2$-th column in $H$, then $i_3$ is the location of the column $h_{i_1} + h_{i_2}$ in $H$. According to Section 1, such $i_3$ exists and is unique. Since we count every triple $i_1, i_2, i_3$ three times, the number of codewords of weight three is $\binom{n}{2}/3$.

3. Show that $C$ contains a codeword of Hamming weight $n$ (that is the all-one codeword $1$).

**Solution:** First we show that $\sum_{h \in \{0,1\}^m} h = 0$. Divide all $h \in \{0,1\}^m$ into pairs $\{h, \overline{h}\}$. Each pair sums to 1 and there are $2^{m-1}$ such pairs. $2^m - 1$ is even, hence the sum of all the pairs equals 0. If $\sum_{h \in \{0,1\}^m} h = 0$ then also $\sum_{h \in \{0,1\}^m \setminus \{0\}} h = 0$. But

$$\sum_{h \in \{0,1\}^m \setminus \{0\}} h = \sum_{h \in H} h = H \cdot 1^T = 0$$

4. How many codewords are there in $C$ of Hamming weight $n-1, n-2, n-3$?

**Solution:** According to Section 3, the all-one codeword is in $C$. Since $C$ is linear, if a word $c$ of Hamming weight $k$ is in $C$ then $\overline{c}$ of Hamming weight $n-k$ is also in $C$. The minimum distance of $C$ is 3. There are no codewords of Hamming weight 1,2 and as such no codewords of Hamming distance $n-1, n-2$. According to Section 2, there are $\binom{n}{2}/3$ codewords of Hamming weight 3, therefore the same amount of codewords have Hamming weight $n-3$.

**Final answer:** $0, 0, \binom{n}{2}/3$. 