Lecture 2

Linear Codes

2.1. Linear Codes

From now on we want to identify the alphabet Σ with a finite field $\mathbb{F}_q$. For general codes, introduced in the last section, the description is hard. For a code of rate R, we need to write down 2^{Rn} codewords, each of length n. Linear codes allow for an exponentially smaller description size.

A *linear code* C of *dimension* k is a set of vectors of $\mathbb{F}_q^n$ which forms a k-dimensional vector space over $\mathbb{F}_q$. The parameter n is called the block-length of C. A linear code of length n and dimension k over $\mathbb{F}_q$ is called an $[n,k]_q$-code.

Any matrix G whose rows form a basis of C is called a *generator matrix* for C. Any matrix H with independent rows and $\mathrm{rker}(H) = C$ is called a *check matrix* for C.

The inner product of vectors x and y in $\mathbb{F}_q^n$ is defined as $\langle x, y\rangle := \sum_{i=1}^n x_i y_i$. For a code C its dual $C^\perp$ is defined as the set of all y such that $yG = 0$.

How can we guarantee that decoding on $\mathrm{BSC}(\varepsilon)$ is possible without (or with very little) errors? On the BSC the codeword that has the smallest Hamming distance to the received word is the Maximum-likelihood estimate of the transmitted word. If any two elements of C are "far apart" and if ε is small enough, then with very high probability there is a unique codeword that is closest to the received word, and this one is the maximum likelihood estimate.

We call the minimum Hamming distance between any two different codewords the *minimum distance* of the code. If the code is linear, this is the minimum distance of a nonzero codeword from the zero word, or the minimum Hamming weight of a nonzero codeword. Denoting the Hamming distance of x and y by $d(x,y)$, and the Hamming weight of x by $\mathrm{wgt}(x)$, we see that

$$\min_{x,y\in C, x\neq y} d(x,y) = \min_{0\neq x\in C} \mathrm{wgt}(x),$$

if C is linear.

An $[n,k]_q$-code with minimum distance d is called an $[n,k,d]_q$-code.

A generator matrix for a code is called *systematic* if it is of the form $(I_k|G_1)$, where I_k is the $k\times k$-identity matrix.

We now state two insightful and simple lemmas, whose proofs are left as exercise.

Lemma 2.1. (1) *Let C be an $[n,k]_q$-code with check matrix H. If any set of $d-1$ columns of H are linearly independent over $\mathbb{F}_q$, then the minimum distance of C is at least d.*

(2) *Let C be an $[n,k,d]_q$-code, and let $e = \lfloor (d-1)/2 \rfloor$. Then, for any $z\in C$ and any $y\in\mathbb{F}_q^n$ such that $d(y,z)\leq e$ and any $z\neq x\in C$ we have $d(y,x) > e$. In other words, z is the unique closest codeword to y.*

The second part of the lemma is usually stated as follows: if z is transmitted and y is received, then ML-decoding will have zero error probability.

For a code C, the homogeneous polynomial $A_C(x,y) = \sum_{i=0}^n A_i x^i y^{n-i}$ is called the *weight enumerator* of C, where A_i is the number of codewords of weight i. The weight enumerator of C and the weight enumerator of $C^\perp$ are intimately related.

Lecture and scribes by Amin Shokrollahi

Theorem 2.2 (MacWilliams identities)**.** *Let C be an $[n,k]_q$-code. Then we have*

$$A_{C^\perp}(x,y) = \frac{1}{q^k} A_C(y-x, y+(q-1)x).$$

Proof. A function $\chi\colon \mathbb{F}_q \to \mathbb{C}$ is called an *additive character* if $\chi(a+b) = \chi(a)\chi(b)$ and $\chi(0) = 1$. It is called *trivial* if $\chi(x) = 1$ for all $x \in \mathbb{F}_q$. For a nontrivial additive character, we have

$$\sum_{\alpha \in \mathbb{F}_q} \chi(\alpha) = 0. \tag{2.1}$$

To see this, suppose that $\beta \in \mathbb{F}_q$ is such that $\chi(\beta) \neq 1$. Then the sum in question equals $\sum_{\alpha\in\mathbb{F}_q} \chi(\alpha+\beta)$ which in turn equals $\chi(\beta)$ times the sum in question. This implies $(1-\chi(\beta))\sum_{\alpha\in\mathbb{F}_q}\chi(\alpha) = 0$, from which we obtain $\sum_\alpha \chi(\alpha) = 0$.

For $x \in \mathbb{F}_q^n$ let

$$g(x) := \sum_{y\in\mathbb{F}_q^n} \chi(\langle x,y\rangle) z^{\mathrm{wgt}(y)}.$$

Then we have

$$\sum_{x\in C} g(x) = \sum_{y\in\mathbb{F}_q^n} z^{\mathrm{wgt}(y)} \sum_{x\in C} \chi(\langle x,y\rangle).$$

If $y \in C^\perp$, then $\chi(\langle x,y\rangle) = 1$. If $y \notin C^\perp$, then $\langle x,y\rangle$ takes every value in $\mathbb{F}_q$ the same number of times, say T, and hence $\sum_{x\in C}\chi(\langle x,y\rangle) = T\sum_{\alpha\in\mathbb{F}_q}\chi(\alpha) = 0$, by (2.1). Hence,

$$\sum_{x\in C} g(x) = |C| A_{C^\perp}(z,1). \tag{2.2}$$

For $\alpha \in \mathbb{F}_q$ let $\mathrm{wgt}(\alpha) = 1$ if $\alpha \neq 0$, and $\mathrm{wgt}(0) = 0$, so that $\mathrm{wgt}(x_1,\ldots,x_n) = \sum_i \mathrm{wgt}(x_i)$. Then

$$\begin{aligned} g(x) &= \sum_{y_1,\ldots,y_n\in\mathbb{F}_q} z^{\mathrm{wgt}(y_1)+\cdots+\mathrm{wgt}(y_n)} \chi(\sum_i x_i y_i) \\ &= \sum_{y_1,\ldots,y_n\in\mathbb{F}_q} z^{\mathrm{wgt}(y_1)+\cdots+\mathrm{wgt}(y_n)} \prod_{i=1}^n \chi(x_i y_i) \\ &= \prod_{i=1}^n \sum_{\alpha\in\mathbb{F}_q} z^{\mathrm{wgt}(\alpha)} \chi(x_i\alpha). \end{aligned}$$

If $x_i = 0$, then the last expression is equal to $1+(q-1)z$. If $x_i \neq 0$, then $\sum_\alpha z^{\mathrm{wgt}(\alpha)}\chi(x_i\alpha) = 1 + z\sum_{\alpha\neq 0}\chi(\alpha) = 1-z$. Therefore,

$$g(x) = (1-z)^{\mathrm{wgt}(x)}(1+(q-1)z)^{n-\mathrm{wgt}(x)},$$

and so $\sum_{x\in C} g(x) = A_C(1-z, 1+(q-1)z)$. Putting this and (2.2) together, we see that

$$A_C(1-z,1+(q-1)z) = |C| A_{C^\perp}(z,1).$$

Replacing z with x/y yields the result. □

Let us give a first example of this theorem: consider the $[n,k,2]_2$-parity code C, i.e., the set of all $x = (x_1,\ldots,x_n) \in \mathbb{F}_2^n$ such that $\sum_{i=1}^n x_i = 0$. This is the set of all even-weight words in $\mathbb{F}_2^n$, and hence we have

$$A_C(x,y) = \sum_{i=0}^{\lfloor n/2\rfloor} \binom{n}{2i} x^{2i} y^{n-2i}$$

The dual $C^\perp$ of C is one dimensional and equal to $\langle(1,1,\ldots,1)\rangle$, hence has the weight distribution $B(x,y) = y^n + x^n$. By the MacWilliams identities, we have

$$A_C(x,y) = \frac{1}{2}\left((y-x)^n + (y+x)^n\right),$$

which is true.

2.2. Relationship Between Distance Distribution and the Error Probability of the ML-decoder

The weight distribution of a linear code is a convenient means for upper bounding the error probability of the ML-decoder of a linear code on a binary symmetric channel.

For a discrete memoryless binary communication channel $\mathcal{C}$ with output alphabet I the quantity

$$\gamma(\mathcal{C}) = \sum_{y \in I} \sqrt{p(y,0)p(y,1)}$$

is called the *Bhattacharya parameter* of the channel. The goal of this section is to provide an upper bound on the word error probability of the ML-decoder using the weight distribution of a linear code.

Let the code C have weight distribution $\sum_i A_i x^i y^{n-i}$. If a vector $x \in C$ is transmitted over $\mathcal{C}$, let $P(x,z)$ denote the probability that the ML-decoder decodes the received word to a vector z of distance d from x. Because of the symmetry of the channel, this probability is the same for all pairs of distance d, and it will be denoted by P_d. By the union bound, the error probability P_e of the ML-decoder is upper bounded by

$$P_e \leq \sum_{d=1}^{n} A_d P_d.$$

Our task is thus to upper bound P_d.

Theorem 2.3. *We have $P_d \leq \gamma(\mathcal{C})^d$, hence $P_e \leq A(\gamma(\mathcal{C}), 1) - 1$.*

Proof. Let $D = \{x_1, \ldots, x_M\}$ be the set of codewords of distance d from z, and let $I_j := \{y \mid p(y,x_j) \geq p(y,z)\}$, i.e., the set of all received words that cause the ML-decoder to decode to x_j rather than z. Suppose that z was transmitted through the channel. Then

$$P_d = \sum_{y \in I_j} p(y,z) \leq \sum_{y \in I_j} \sqrt{p(y,z)p(y,x_j)},$$

where the inequality follows from $\sqrt{p(y,x_j)/p(y,z)} \geq 1$ by assumption. We thus have

$$\begin{aligned} P_d &\leq \sum_{y \in I_j} \sqrt{p(y,z)p(y,x_j)} \\ &\leq \sum_{y \in I^n} \sqrt{p(y,z)p(y,x_j)} \\ &= \sum_{y \in I^n} \prod_{i=1}^{n} \sqrt{p(y_i,z_i)p(y_i,x_{ji})} \\ &\leq \prod_{i=1}^{n} \sum_{y \in I} \sqrt{p(y,z_i)p(y,x_{ji})} \\ &= \gamma(\mathcal{C})^d \end{aligned}$$

For the last step, note that the inner sum is 1 if $z_i = x_{ji}$, and $\gamma(\mathcal{C})$ otherwise. □

This theorem means that the error probability of the ML-decoder is mostly influenced by the lower A_i's, i.e., by the number of pairs of words that have a small Hamming distance. This is one of the reasons why researchers often study codes for which any pair of vectors have a large Hamming distance (often also called codes with large minimum distance). The next lecture will make some of these concepts more precise.

2.3. First Examples of Codes

2.3.1. The Hamming Code

Consider the matrix H with r rows and $2^r - 1$ columns in which the columns are all the nonzero binary vectors of length r. We first show that

Lemma 2.4. *The matrix H described above has rank r.*

Proof. Note that the $r \times r$-identity matrix is a sub-matrix of H. □

The code which has H as its check matrix is called the *binary Hamming code*. It is a $[2^r - 1, 2^r - r - 1]_2$-code. Since all the columns of H are distinct, any two columns are independent. Hence, the minimum distance of C is at least 3. Moreover, there are 3 dependent columns (take any three binary vectors x_1, x_2, x_3 that sum up to zero), hence the minimum distance is exactly 3.

Example 2.5. The following is a systematic check matrix for the $[7, 4, 3]_2$-Hamming code:

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

Since Hamming codes have distance 3, they are one-error correcting. The following algorithm calculates the error position, provided that the columns are H are the binary expansions of the integers $1, 2, \ldots, 2^k - 1$:

1. If y denotes the received word, calculate $z = Hy^\top$ (This vector is often called the "syndrome" of the received word.)
2. Interpret z as the binary expansion of the integer i. Then the error is at position i.

It is easily seen that the algorithm is correct: if $y = c + e$ for a codeword c and an error vector e of weight one, then $H \cdot y^\top = H \cdot e^\top = h_j$, where j is the position of the nonzero entry of e, and h_j is the jth column of H. But according to the construction of H, the column h_j is the binary expansion of the integer j.

2.3.2. Hadamard Codes

Consider the vector space $\mathbb{F}_2^k$, and its dual space $(\mathbb{F}_2^k)^*$. This is the space of linear forms on $\mathbb{F}_2^k$, and is an k-dimensional vector space over $\mathbb{F}_2$. Define the following vector space morphism:

$$\begin{array}{rcl} \varphi\colon (\mathbb{F}_2^k)^* & \to & \mathbb{F}_2^{2^k-1} \\ \lambda & \mapsto & (\lambda(x) \mid x \in \mathbb{F}_2^k \setminus \{0\}). \end{array}$$

The image of φ is called a Hadamard code.

Theorem 2.6. (1) *φ is injective.*

(2) *The minimum distance of* $\mathrm{Im}\varphi$ *is* 2^{k-1}.

Proof. (1) Obviously, only the zero linear form vanishes on all the elements of $\mathbb{F}_2^k$.

(2) Let λ be a nonzero linear form. Then the dimension of $\ker \lambda$ is $n - 1$, and hence λ vanishes on exactly 2^{k-1} elements of $\mathbb{F}_2^k$, i.e., it does not vanish on exactly 2^{k-1} elements of this vector space. This shows that the weight of any nonzero element in the image of φ is 2^{k-1}. □

We will now show the following.

Theorem 2.7. (1) *The Hadamard code of length $2^k - 1$ and dimension k is the dual of the Hamming code of length $2^k - 1$ and co-dimension k.*

(2) *The weight distribution of the binary Hamming code of length $2^k - 1$ is*

$$\frac{1}{2^k}(y - x)^{2^k-1} + \frac{2^k - 1}{2^k}(y^2 - x^2)^{2^{k-1}-1}(y - x).$$

Proof. (1) First, note that for any nonzero point $x \in \mathbb{F}_2^n$ there is a linear form λ on $\mathbb{F}_2^n$ such that $\lambda(x) \neq 0$. Let $\lambda_1, \ldots, \lambda_n$ form a basis of the dual space $(\mathbb{F}_2^n)^*$. It follows that for any two $x, y \in \mathbb{F}_2^n \setminus \{0\}$, $x \neq y$, there is an i such that $\lambda_i(x) \neq \lambda_i(y)$. Therefore, in the matrix H in which the ith row is the vector $(\lambda_i(x) \mid x \in \mathbb{F}_2^n \setminus \{0\})$, all the columns are distinct. Moreover, H has $2^k - 1$ columns, so H is the check matrix of a binary Hamming code. On the other hand, H is clearly a generator matrix for the first order Reed-Muller code. Hence the duality result.

(2) The weight enumerator for the Hadamard code is $A_C(x, y) = y^{N-1} + (N - 1)x^{N/2}y^{N/2-1}$, where $N = 2^k$. It follows by the MacWilliams identities that the weight enumerator for the Hamming code is

$$\frac{1}{N}(y - x)^{N-1} + \frac{N - 1}{N}(y^2 - x^2)^{N/2-1}(y - x).$$

□

2.3.3. ML-Decoding of the Hadmard Code

The normal algorithm for ML-decoding of Hadamard codes on $\mathrm{BSC}(\varepsilon)$ would do the following: upon reception of a word y, go over all the 2^k codewords, and check which one is closest to y. This operation uses $O(2^{2k})$ time: for each check, we need to compare the $2^k - 1$ coordinate positions. Here we introduce a faster algorithm, which uses only $O(k2^k)$ operations.

In the received vector y replace each 0 by $+1$ and each 1 by -1. Call the resulting vector $\hat{y}$. Note that $\hat{y}_i = (-1)^{y_i}$. Let H be the Hadamard matrix of size 2^k. The rows and columns of this matrix are indexed by elements $(u, v) \in \mathbb{F}_2^k \times \mathbb{F}_2^k$, and the entry (u, v) of the matrix is $(-1)^{\langle u,v\rangle}$. Let z denote the vector $H \cdot \hat{y}^\top$ (we view $\hat{y}$ as a row vector). This vector is called the *Hadamard-Walsh* transform of $\hat{y}$. The entry of z corresponding to $u \in \mathbb{F}_2^k$ is

$$\sum_{v\in\mathbb{F}_2^k} (-1)^{\langle u,v\rangle}(-1)^{y_v} = |\{v \mid \langle u, v\rangle = y_v\}| - |\{v \mid \langle u, v\rangle \neq y_v\}.$$

The vector $(\langle u, v\rangle \mid v \in \mathbb{F}_2^k)$ is a codeword of the Hadamard code. Call it $E(u)$. Then, the above says that

$$(H \cdot \hat{y})_u = N - 2d(y, E(u))$$

Hence, if we can calculate $H \cdot \hat{y}$ efficiently, then we can find vector closest to y by finding the maximum entry of the Hadamard-Walsh transform of $\hat{y}$. It is well-known that the computation of the Hadamard-Walsh transform of y can be accomplished with $O(n \log(n)) = O(k2^k)$ operations. The pre- and postprocessing steps are clearly $O(2^k)$. Hence the results follows.

2.3.4. First Order Reed-Muller Code

Consider the space $V := \mathbb{F}_q[X_1, \ldots, X_m]_{\leq 1}$ of m-variate polynomials of degree at most 1 over $\mathbb{F}_q$. Let $N = q^m$, and map this space into $\mathbb{F}_q^N$ via

$$\begin{array}{rccl} \varphi\colon V & \to & \mathbb{F}_q^N \\ f & \mapsto & (f(z_1), \ldots, f(z_N)), \end{array}$$

where $z_1, \ldots, z_N$ constitute the elements of $\mathbb{F}_q^m$ in some ordering. This is clearly a linear mapping, and hence its image is a linear code of length N. As for the dimension, note that the kernel of the map is trivial, hence the dimension is the dimension of V, i.e., it is $m + 1$. This image is called the first order Reed-Muller code over $\mathbb{F}_q$. Can you see how, in the case of $q = 2$, this case is related to the Hadamard code?

2.4. Graph Representation of Binary Codes

Let H be a check matrix for a binary code. We can interpret H as a bipartite graph with n left nodes (n being the number of columns of H) and r right nodes (r being the number of rows of H), in the following way: if the element at position (i, j) of H is one, then we put an edge between the ith right and the jth left node.

In this representation the left nodes are called the "variable nodes" while the right nodes are called the "check nodes". Given such a graph, the code is defined as the set of all binary settings on the variable nodes such that for all check nodes the sum of the settings of the adjacent variable nodes is zero.

If this graph is sparse (i.e., if we have a sequence of such graphs for which the number of variable and check nodes goes to infinity, the ratio between the variable and check nodes is fixed, and and the average variable node degree is constant), then these codes are called low-density parity-check (LDPC) codes. We will deal with these codes later in the course.

Figure 2.1 shows a graph representation of the Hadamard code represented by the check matrix

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

Sometimes, certain properties of the code can be inferred from graph properties of the graph representation. Here is a very simple example.

Lemma 2.8. *Suppose that the variable nodes of a graph corresponding to a binary code C have weight at least 2, and that the graph does not have cycles of length 4. Then the minimum distance of the code is at least 3.*

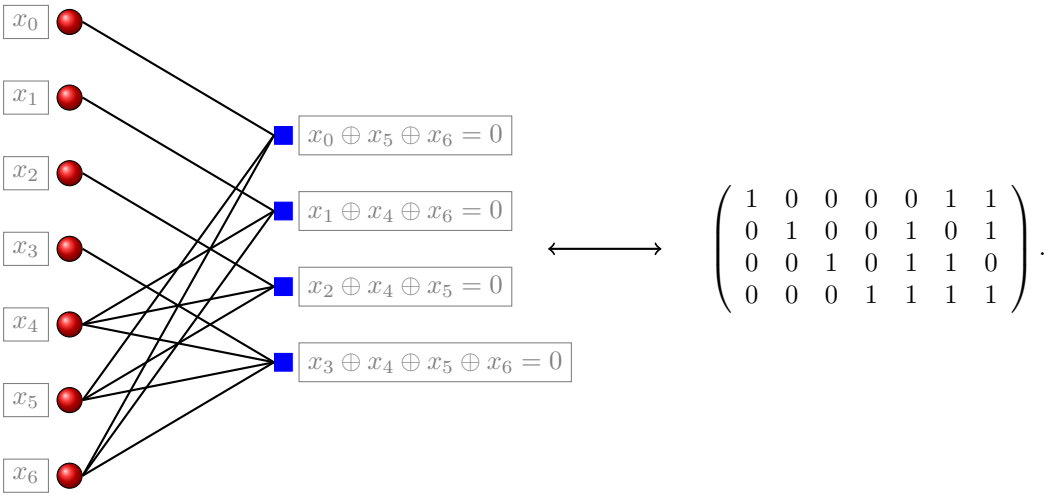

Figure 2.1: Graph representation of the $[7, 3, 4]_2$-Hadamard code

Proof. The minimum distance is not one: otherwise, there is a variable node that is not connected to any check node, which is a contradiction to the fact that the degree of the variable nodes is larger than one.

Suppose that the minimum distance is 2, and w.l.o.g. assume that minimum weight word is $(1, 1, 0, \ldots, 0)$. Consider the subgraph induced by the two first variable nodes. All check nodes in this graph must have even degree (or else they would not be satisfied). Moreover, there are at least two check nodes in this graph of degree greater than zero, since the degrees of the variable nodes is supposed to be ≥ 2. Then the graph formed by the two first variable nodes, and these two check nodes, is a cycle of length 4, contrary to the assumption. □

2.5. Creating New Codes from Old Ones

Let C be an $[n, k, d]_q$-code. In this section we describe methods to obtain $[n', k', d']_q$-codes from this code for values $n' \leq n$, $k' \leq k$, and $d' \leq d$.

Puncturing: The punctured code at position i, denoted C^i is the set of words $(x_1, x_2, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$. The length of this code is $n - 1$, its dimension is $k - 1$ or k, and its minimum distance is $d - 1$ or d. Note that if there is a codeword with nonzero entry at position i, then the dimension of this code is $k - 1$. Moreover, the minimum distance of the punctured code is $d - 1$ if and only if there is a codeword of weight d which has a nonzero entry at position i.

Shortening: In this construction the length and the dimension are decreased by 1, and the minimum distance will at least remain at its old value. Let C' be the intersection of C with the hyperplane of words that are zero at position i. The shortened version C_i of C at position i is the puncturing of C' at position i. Since C' is a sub-code of C, the minimum distance of C_i is at least d. Moreover, if C is such that there is a codeword with nonzero entry at position i, then the dimension of C_i is $k - 1$.

As a useful exercise, we propose that the reader prove the following:

Lemma 2.9. *The Hadamard code is obtained as a shortening of the binary first order RM-code.*

There is a simple connection between punctured and shortened versions of codes, due to P. Delsarte.

Proposition 2.10. *Let C be a linear code. Then*

$$\left(C^{\perp}\right)_i = \left(C^i\right)^{\perp}.$$

Proof. By inspection. □

2.6. Projecting Codes over Large Alphabets to Codes over Small Alphabets

In this section we assume that we have an $[n, k, d]_q$-code and that we want to produce from this code another coder over $\mathbb{F}_p$ where p is the characteristic of $\mathbb{F}_q$. For simplicity we will assume that $p = 2$.

The first method we discuss is the *binary image method*. Fix a vector space basis $\alpha_1, \ldots, \alpha_m$ of $\mathbb{F}_q/\mathbb{F}_2$ (so we assume that $q = 2^m$). The binary image of C, denoted $\mathrm{Im}_{\mathbb{F}_2}(C)$ is obtained by replacing each entry in a codeword by its representation in the above basis. The following is then obvious.

Proposition 2.11. *$\mathrm{Im}_{\mathbb{F}_2}(C)$ is an $[mn, mk, d']_q$-code where $d' \geq d$.*

The second method is the *subfield-subcode method*. The binary subfield-subcode of C is the code $C|_{\mathbb{F}_2} := C \cap \mathbb{F}_2^n$, i.e., we consider in C vectors that happen to have binary entries. In this method the length of the code is not altered, and obviously, the minimum distance cannot be smaller than that of the original code, simply because the subfield-subcode is in particular a subcode of C. As far as the dimension goes, we have the following result.

Proposition 2.12. *The dimension of $C|_{\mathbb{F}_2}$ is at least $mk - (m-1)n$.*

Proof. Suppose that $(x_1, \ldots, x_n)$ is an element of the subfield-subcode, and consider a check equation $\sum_i \mu_i x_i = 0$, wherein $\mu_i \in \mathbb{F}_q$. Then, for all automorphisms σ of $\mathbb{F}_q/\mathbb{F}_2$ we have $\sum_i \sigma(\mu_i) x_i = 0$. Moreover, $(x_1, \ldots, x_n)$ is in the subfield-subcode iff it satisfies all the automorphic images of all the checks. Since there are $r = n - k$ independent checks on C, there are at most mr independent checks on $C|_{\mathbb{F}_2}$, and hence the dimension of the subfield-subcode is at least $n - mr = mk - (m-1)n$. □

The bound given in the proposition is sharp. Consider the $[m, m-1]_{\mathbb{F}_q}$-code with check matrix $(\mu_1, \mu_2, \ldots, \mu_m)$, and suppose that the elements in this row vector are independent over $\mathbb{F}_2$. A check matrix for $C|_{\mathbb{F}_2}$ is then given by the $m \times m$-matrix

$$\begin{pmatrix} \mu_1 & \mu_2 & \cdots & \mu_{m-1} & \mu_m \\ \mu_1^2 & \mu_2^2 & \cdots & \mu_{m-1}^2 & \mu_m^2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mu_1^{2^{m-1}} & \mu_2^{2^{m-1}} & \cdots & \mu_{m-1}^{2^{m-1}} & \mu_m^{2^{m-1}} \end{pmatrix}.$$

The right kernel of this matrix does not contain any nontrivial binary vector, because of the independence assumption. Hence, the $\mathbb{F}_2$-rank of this matrix is full, and the dimension of $C|_{\mathbb{F}_2}$ is 0. On the other hand, $mk - (m-1)n = m(m-1) - (m-1)m = 0$, which shows the sharpness of the bound.

A much easier way for constructing the subfield-subcode is as follows. Fix a basis $\mu_1, \ldots, \mu_m$ of $\mathbb{F}_q/\mathbb{F}_2$. In the check matrix H of C, replace any entry by its representation in this basis as an m-dimensional column vector. Denote the resulting matrix by $\hat{H}$.

Proposition 2.13. *$\hat{H}$ is a check matrix for $C|_{\mathbb{F}_2}$.*

Proof. Let h_{ij} denote the (i,j)-entry of H, and $(x_1, \ldots, x_n) \in \mathbb{F}_2^n$. We need to show that if $\sum_j h_{ij} x_j = 0$ iff $\sum_j \tau_{ijk} x_j = 0$, where $h_{ij} = \sum_k \tau_{ijk} \mu_k$. This follows immediately from the independence of the μ_j. □

Example 2.14. Let $\mathbb{F}_4 = \{0, 1, \alpha, \alpha^2\}$, where $\alpha^2 + \alpha + 1 = 0$. Consider the $[8,3]_4$-code with check matrix

$$\begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & \alpha & \alpha & \alpha & \alpha^2 & \alpha^2 & \alpha^2 \\ 0 & 0 & 1 & \alpha & \alpha^2 & 1 & \alpha & \alpha^2 \end{pmatrix}.$$

It can be shown that the minimum distance of this code is 3. A basis of $\mathbb{F}_4/\mathbb{F}_2$ is given by (α, α^2). With respect to this basis a binary check matrix for $C|_{\mathbb{F}_2}$ is

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

The rank of this matrix is 5, so that $C|_{\mathbb{F}_2}$ is an $[8,3]_2$-code. The minimum distance of this code is at least 3. In fact, this code has minimum distance 4. Among all codes of length 8 and dimension 3, this code has the largest possible minimum distance.

The code on $\mathbb{F}_4$ is obtained from an algebraic geometric construction using a maximal elliptic curve over the field. Such codes will be the topic of one of the later classes.