Coding Theory

Alec Zabel-Mena

March 2, 2022

Contents

Chapter 1

Linear Codes.

1.1 Definitions, Generator, and Check Matrices.

Definition. Wed define an (n, k)**-linear code)** over a field F to be a k-dimensional subspace $\mathcal{C}$ of the n-dimensional vector space F^n over F.

Remark. We shall be focusing exclusively on the finit fields $\mathbb{F}_p$ where $p = 2, 3$. Then in this case, we can consider the vector spaces to be extension fields of $\mathbb{F}_p$. We shall prove theorems and lemmas however, for general fields, unless specified.

Definition. Let $\mathcal{C}$ be an (n, k)-linear codeover a field F. We we call a $k \times n$matrix G a **generator matrix** for $\mathcal{C}$ if its row space is $\mathcal{C}$.

Example 1.1. [1]

(1) A $(5, 1)$-linear code, $\mathcal{C}_1$, over $\mathbb{F}_2$ with generator matrix:

$$G_1 = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \end{pmatrix}$$

It contains the codewords 00000 and (11111); and has rate $\frac{1}{5}$. We call $\mathcal{C}_1$ the **binary repitition code**.

(2) The $(5, 3)$-code $\mathcal{C}_2$ with generator matrix:

$$G_2 = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix}$$

$\mathcal{C}_2$ has rate $\frac{3}{5}$.

(3) The $(7, 4)$**-Hamming Code**, $\mathcal{C}_3$ over $\mathbb{F}_2$ with generator matrix:

$$G_3 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix}$$

The $(7, 4)$-Hamming code has rate $\frac{4}{7}$.

Lemma 1.1.1. *If $\mathcal{C}$ is an (n,k)-code over a field F, and if G is a generator matrix for $\mathcal{C}$, then so is any matrix row-equivalent to G.*

Proof. Let A be an $k \times n$ matrix row-equivalent to G. Then, take $A \to G$ via the sequence of elementary matrices $\{E_i\}_{i=1}^m$. That is, $G = E_m \dots E_2 E_2 A$. Then for any $v \in F^n$. we can take $Av \to Gv$ via this same sequence; that is $Gv = E_m \dots E_2 E_1 Av$. Thus, A generates the same set of vectors as G, and hence has the same row space. ■

Remark. Thus, using this lemma, one would ideally like to find a generator matrix in Row-Reduced-Echelon form, for ease of computation.

Definition. If $\mathcal{C}$ is an (n,k)-code over a field F, we define a **check** for $\mathcal{C}$ to be the equation:

$$a_1x_1 + \dots + a_nx_n = 0 \tag{1.1}$$

satisfied for all $x \in \mathcal{C}$. We define the **dual code** of $\mathcal{C}$ to be the orthogonal complement

$$\mathcal{C}^\perp = \{a \in F^n : \langle a, x \rangle = 0\} \tag{1.2}$$

Where $\langle a, x \rangle$ is the inner product of a and x.

Proof. If $\mathcal{C}$ is an $(n,k) - code$, then $\mathcal{C}^\perp$ is an $(n, n-k)$-linear code. ■

Proof. We have by a result from [2] (theorem $4.I$), that $F^n = \mathcal{C} \oplus \mathcal{C}^\perp$, ($\oplus$ the direct sum). Then $\dim F^n = \dim \mathcal{C} + \dim \mathcal{C}^\perp$. Therefore, $\dim \mathcal{C}^\perp = n - k$. ■

Definition. Let $\mathcal{C}$ be an (n,k)-linear code over a field F. We define a **check** matrix for $\mathcal{C}$ the be an $n \times (n-k)$ matrix H such that $Hx^T = 0$.

Lemma 1.1.2. *If H is a check matrix for the (n,k)-code $\mathcal{C}$, then H is a generator matrix for the dual code $\mathcal{C}^\perp$.*

Proof. For any $x = (x_1, \dots, x_n) \in \mathcal{C}$, we have that $Hx^T = 0$, by definition. Thus, for any row $a = (a_1, \dots, a_n)$ of H. That is, $a_1x_1 + \dots + a_nx_n = \langle a, x \rangle = 0$, making $a \in \mathcal{C}^\perp$. Since a is an arbitrary row of H, this holds for every row of H. Thus the row space of H is equal to $\mathcal{C}^\perp$. ■

Lemma 1.1.3. *Let $\mathcal{C}$ be an (n,k)-code over a field F, and let G be a generator matrix for the code. If G has the form $G = (I_{k\times k}|A)$, then the check matrix for $\mathcal{C}$, corresponding to G has the form*

$$H = (-A^T | I_{(n-k)\times(n-k)}) \tag{1.3}$$

Example 1.2. [1] Consider the generator matrices for the codes in example 1.1, then:

(1) $H_1 = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \end{pmatrix}$

(2) $H_2 = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 1 \end{pmatrix}$

(3) $H_3 = \begin{pmatrix} 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 \end{pmatrix}$

Theorem 1.1.4. *Let $\mathcal{C}$ be an (n, k)-code over a field F. Then there is a unique $k \times n$ Row-Reduced-Echelon matrix G such that $x \in \mathcal{C}$ if, and only if x is in the row space of G. Likewise, there exists an $(n-k) \times n$ matrix H such that $x \in \mathcal{C}$ if, and only if $Hx^T = 0$.*

Corollary. *If $\mathcal{C}$ is used on a memoryless channel, then $G = (I_{k\times k}|A)$ and $H = (-A^T|I_{(n-k)\times(n-k)})$.*

1.2 Syndrome Decoding.

Definition. Let $\mathcal{C}$ be an (n, k)-code over a field F. For $x \in \mathcal{C}$, and $y \in F^n$ we call $z = y - x$ an **error pattern**. If $z_i \neq 0$, we say that the i-th component of x has **error**. If H is a check matrix, we call $s = Hy^T$ the **syndrome** of the vector $y = x + z$.

Lemma 1.2.1. *If H is a check matrix for an (n, k)-code $\mathcal{C}$ over a field F, and $x \in \mathcal{C}$, with error pattern $y = x + z$ for $z \in F^n$, then $Hy^T = Hz^T$.*

Proof. We have $Hy^T = H(x + z)^T = H(x^T + z^T) = Hx^T + Hz^T = Hz^T$, since $x \in \mathcal{C}$. ■

Corollary. *x has error if, and only if $s \neq 0$.*

Definition. Let $\mathcal{C}$ be an (n, k)-code over a field F, with check matrix H. Let $z \in F^n$ such that $y = x + z$. We define the set of all solutions to the syndrome $Hz^T = s$ to be a **coset** of the code $\mathcal{C}$.

Lemma 1.2.2. *Every coset of an (n, k)-code $\mathcal{C}$ has the form $\mathcal{C}_0 = \mathcal{C} + z_0$.*

Proof. Let $\mathcal{C}_0$ be a coset for the code $\mathcal{C}$, given by $z_0 \in F^n$. We have that for any $x \in \mathcal{C}$, $x + z_0 \in \mathcal{C}_0$, since $H(x + z_0)^T = Hz_0^T = s$; that is, $x + z_0$ is also a solution to the syndrome. Now, if y is a solution to the syndrome $Hz_0^T = s$, then $y = x + z_0$ for some $x \in \mathcal{C}$, so $y \in \mathcal{C} + z_0$. ■

Definition. We define a **q-ary symmetric channel** (qSC) to be a DMC with $A_X = A_Y = \mathbb{F}_q$, taking $X \to Y = X + Z$ with Z a random vector with independently distributed components and $P(Z = 0) = 1 - (q - 1)\epsilon$ and $P(Z = z) = \epsilon$ whenever $z \neq 0$.

Definition. Let $\mathcal{C}$ be an (n, k)-code. Then **Hamming weight** of a codeword $x \in \mathcal{C}$ is the number of nonzero components of x. That is:

$$w_H(x) = |\{x_i : x \not= 0\}| \tag{1.4}$$

Example 1.3. Let $X \to Y = X + Z$ over a qSC. Then for $z \in n\mathbb{F}_q^n$, $P(Z = z) = (1 - (q - 1)\epsilon)^{n-w_H(z)}\epsilon^{w_H(z)}$ If $\epsilon \leq \frac{1}{q}$, then $P(Z = z)$ is decreasing with respect to $w_H(z)$.

Definition. Let $\mathcal{C}$ be an (n, k)-code with check matrix H. We define the **standard matrix** for $\mathcal{C}$ to be the $(n - 1) \times (k + n)$ matrix whoes rows are the cosets of $\mathcal{C}$ The vector of least weight in each row is called the **coset leader**.

Example 1.4. Consider the $(5, 3)$-code of example 2.1 with check matrix

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

H has the syndromes 00, 01, 10, and 11; so we calculate the standard matrix for H to be:

$$\begin{pmatrix} 00000 & 00011 & 00101 & 00110 & 11001 & 11010 & 11100 & 11111 \\ 00100 & 00111 & 00001 & 00010 & 11101 & 11110 & 11000 & 11011 \\ 01000 & 01011 & 01101 & 01110 & 10001 & 10010 & 10100 & 10111 \\ 10000 & 10011 & 10101 & 10110 & 01001 & 01010 & 01100 & 01111 \end{pmatrix}$$

This brings us to an algorithm for syndrom decoding with the cosets of linear codes.

Algorithm 1.1 (Syndrome Decoding for a qSC.)**.** *Given an (n, k)-linear code over a field F and check matrix H. Assume that x is transmitted over a qSC, and is recieved as y.*

(1) Compute the syndrome $s = Hy^T$

(2) Find a minimum weight vector in the coset of s, label it z_0.

(3) Return $\hat{x} = y - z_0$.

Remark. If n and k are both small, we can go about step (2) using a standard matrix, implemented as a lookup table.

1.3 The Hamming Metric.

Definition. Let F be a field. The **Hamming distance** of two vectors $x, y \in F^n$ is the number of components for which $x_i \neq y_i$. That is, $d_H(x, y) = w_H(y - x)$.

Lemma 1.3.1. *Hamming distance makes F^n into a metric space.*

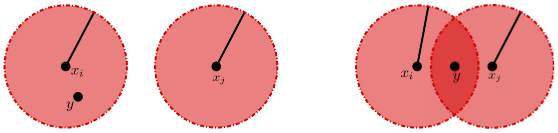

Figure 1.1: Hamming spheres around adjacent codewords [1].

Definition. Let $\mathcal{C}$ be an (n, k)-code. We define the **Hamming sphere** of radius r about a codeword $x \in \mathcal{C}$ to be the open ball $B_r(x) = \{y \in \mathcal{C} : d_H(x, y) < r\}$, where d_H is the Hamming distance of x and y.

Definition. Let $\mathcal{C}$ be a code (not necessarily linear). The **minumum distance** of $\mathcal{C}$ is the smalles Hamming distance across all codewords; i.e.

$$d = \min \{d_H(x, y) : x, y \in \mathcal{C} \text{ and } x \neq y\} \tag{1.5}$$

Similarly, we define the **minimum weight** of $\mathcal{C}$ to be:

$$w = \min \{w_H(x) : x \in \mathcal{C} \text{ and } x \neq 0\} \tag{1.6}$$

Theorem 1.3.2. *A code $\mathcal{C}$ with minumum distance d can correct e errors across all error patterns if, and only if $d \geq 2e+1$*

Proof. Let $\mathcal{C}$ be a code of length n (not necessarily linear), and suppose the codeword x_i is sent through a qSC as $y = x_i + z$, with $w_H(z) = e$. Then if each codeword is sent through the channel with probability $\frac{1}{M}$, pick the codeword y such that $d_H(x_i, y)$ is the smallest possible. Notice that $d_H(x, y) = w_H(z)$.

Now, for $x_j \in \mathcal{C}$ such that, $x_j \neq x_j$ if $d_H(x_i, y) \geq 2e+1$, then the hammng spheres $B_e(x_i)$ and $B_e(x_j)$ are disjoint; see figure 1.1. For if not, and there is a $y \in B_e(x_i) \cap B_e(x_j)$, then $d_H(x_i, x_j) \leq d_H(x_i, y) + d_H(y, x_j) < e + e = 2e$, which contradicts the assumption. Thus, if $d_H(x, y) \leq e$, then $y \in B_e(x_i)$. This means, if the difference between the components of x_i and y is less than e, y cannot be closer to x_j than it is to y, and so we can choose to decode it as x_i.

On the other hand, if $d_H(x_i, x_j) \leq 2e$, then we have that $B_e(x_i) \cap B_e(x_j) \neq \emptyset$. Then if $d_H(x_i, y) \leq e$, then $y \in B_(x_i) \cap B_e(x_j)$, and so y is as close to x_i as it is to x_j, so we cannot make a reasonable choice as how to decode y.

Therefore we have that the code $\mathcal{C}$ can correct e errors if, and only if the distance between any two codewords is atleast $2e+1$, that is, the minimum distance has to be atleast $2e+1$. ■

Example 1.5. If $\mathcal{C}$ is a code with minimum distance $d = 7$, then $\mathcal{C}$ can correct up to 3 errors, since $7 = 2 \cdot 3 + 1$. Likewise, if $d = 22$, $\mathcal{C}$ can correct up to 10 errors.

Remark. Due to this theorem, when analyzing codes, it is desirable to have codes with as large minumum distance as possible. Having large minimum distance will allow us to correct large ammounts of errors.

Lemma 1.3.3. *In an (n, k)-code $\mathcal{C}$ with minimum distance d and minimum weight w, we have $d = w$.*

Proof. Recall that for any two codewords $x, y \in \mathcal{C}$, $d_H(x, y) = w_H(y - x)$. ■

Definition. We call a code $\mathcal{C}$ **e-error-correcting** if it can correct at most e errrs. I.e. if $d \geq 2e+1$ for the largest possible e.

Theorem 1.3.4. *If $\mathcal{C}$ is an (n, k)-code over a field F with check matrix H, then the minimum distance of $\mathcal{C}$ is the smallest number of linearly independent columns of H.*

Proof. For all $x \in \mathcal{C}$, we have that $Hx^T = 0$. Now, notice that Hx^T is the linear combination of columns $\{a_1, \dots, a_n\}$ of H with the components of x. Thus, $Hx^T = x_1 a_1 +_n a_n = 0$. Thus, if $w_H(x) = w$, then Hx^T is linearly dependent in w columns and conversely. ■

Corollary. *If every subset of $2e$ columns of H are linearly independent, then $\mathcal{C}$ is e-error-correcting.*

Corollary. *If $\mathcal{C}$ is a code over $\mathbb{F}_2$, and all possible combinations less than e, of columns of H are distinct, then $d \geq 2e+1$.*

1.4 Hamming Codes.

Definition. Let H be an $m \times q^m - 1$ matrix with entries in a finite field $\mathbb{F}_q$ of q elements, such that the columns of H are the $q^m - 1$ vectors of $\mathbb{F}_q^m$. We define the **q-ary Hamming code** to be the $(q^m - 1, q^m - 1 - m)$-linear code over $\mathbb{F}_q$ whose check matrix is H.

Remark. When $q = 2$, we call H the parity check matrix of the binary $(2^m - 1, 2^m - 1 - m)$ Hamming code.

Definition. (1) Consider the $(2^m - 1, 2^m - 1 - m)$-binary Hamming code. If we have $y = x + z$, for a codeword x and error pattern z, and $z = 0$, then $s = 0$. On the otherhand, if $w_H(z) = 1$, then s is the i-th column of the parity check matrix H. This makes syndrome decoding for the Hamming code easy.

(2) The binary $(2^m - 1, 2^m - 1 - m)$ Hamming code can $\mathcal{C}$ correct 1 error, if and only if all Hamming spheres of radius 1 are disjoint. Notice a given Hamming sphere $B_1(x)$ has $n + 1$ codewords, so $\mathcal{C}$ can have atmost $\frac{2^n}{n+1}$ codewords. If $n = 2^m - 1$, we then get $2^{2^m - 1 - m}$ possiible codewords.

Definition. We call a code $\mathcal{C}$ over a field F, with minimum distance $d = 2e + 1$, **perfect** if every $x \in F^n$ has distance at most e to any other codeword of $\mathcal{C}$.

1.5 Weight Enumerators.

Bibliography

[1] R. McEliece, *The theory of information and coding.* Cambridge: Cambridge University Press, 2001.

[2] I. N. Herstein, *Topics in algebra.* New York: Wiley, 1975.

[3] D. J. Welsh, *Codes and cryptography.* Oxford Oxfordshire New York: Clarendon Press Oxford University Press, 1988.

[4] K. Hoffman and R. Kunze, *Linear algebra.* Englewood Cliffs, NJ: Prentice-Hall, 1971.

[5] J. Lint, *Introduction to coding theory.* Berlin New York: Springer, 1999.

[6] J. Justesen and T. Høholdt, *A course in error-correcting codes.* Zurich, Switzerland: European Mathematical Society, 2017.