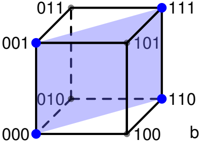
WEEKS 12&13: VECTOR SPACES AND LINEAR CODES (TEXTBOOK CHAPTER 13)

Prof. Michael Gastpar

Slides by Prof. M. Gastpar and Prof. em. B. Rimoldi

EPFL

Spring Semester 2025



binary code

Outline

Introduction and Organization

Entropy and Data Compression

Cryptography

Channel Coding

Error Detection and Error Correction

Finite Fields and Vector Spaces

Linear Codes

Reed Solomon Codes

Summary of Chapter 3

LAST WEEK

- LINEAR ALGEBRA, BUT IN MODULO.
  - → WORKS IF WE HAVE A "FINITE FIELD".
  - → PARTICULARLY SIMPLE IF WE DO "MODULO PRIME".

$\vec{x} \in \mathbb{F}^n$

$\vec{x} = (x_1, x_2, \dots, x_n)$

- SUBSPACES

$$S \subseteq \mathbb{F}^n$$

$$S = \left\{ \vec{s} : \vec{s} = a\vec{v}_1 + b\vec{v}_2 + c\vec{v}_3 \text{ for } a, b, c \in \mathbb{F} \right\}$$

IF $\vec{v}_1, \vec{v}_2, \vec{v}_3$ LIN. INDEP.,

THEN $\dim(S) = 3$

$$= \left\{ \vec{s} : \begin{array}{l} s_1 h_{11} + s_2 h_{12} + \dots + s_n h_{1n} = 0 \\ \vdots \\ s_1 h_{m1} + s_2 h_{m2} + \dots + s_n h_{mn} = 0 \end{array} \right\}$$

- LINEAR ALGEBRA LIKE IN $\mathbb{R}^n$

- <u>EXCEPT</u> THAT THE # OF POINTS IS <u>FINITE</u>:

$$\left(\text{card}(\mathbb{F})\right)^{\dim(\text{Space})}$$

Why do we care about linear codes?

Linear codes have more structure.

We use that structure to simplify our tasks, notably:

- To determine the code's performance (d_{min} in particular).
- To simplify the encoding.
- To simplify the decoding.

Definition (Textbook def. 13.1)

A block code is a **linear code** if the codewords form a subspace of $\mathbb{F}^n$ for some finite field $\mathbb{F}$.

code $\mathcal{C}$
$\vec{c}_0 = 0000000$
$\vec{c}_1 = 0011100$
$\vec{c}_2 = 0111011$
$\vec{c}_3 = 1110100$
$\vec{c}_4 = 0100111$
$\vec{c}_5 = 1101000$
$\vec{c}_6 = 1001111$
$\vec{c}_7 = 1010011$

IS THIS A <u>SUBSPACE</u> ?

= LIN. CODE ?

1) QUICK CHECKS

- IS THE ALL-ZERO VECTOR IN? ✓

- NUMBER OF CODEWORDS MUST BE 2^k $k=3$ ✓

code $\mathcal{C}$

$\vec{c}_0 = 0000000$

$\vec{c}_1 = 0011100$

$\vec{c}_2 = 0111011$

$\vec{c}_3 = 1110100$

$\vec{c}_4 = 0100111$

$\vec{c}_5 = 1101000$

$\vec{c}_6 = 1001111$

$\vec{c}_7 = 1010011$

IS THIS A SUBSPACE ?

2) FULL CHECK: FIND BASIS!

$\vec{v}_1 = 0011100$

$\vec{v}_2 = 0111011$

$\vec{v}_3 = 1110100$

code $\mathcal{C}$
$\vec{c}_0 = 0000000$
$\vec{c}_1 = 0011100$
$\vec{c}_2 = 0111011$
$\vec{c}_3 = 1110100$
$\vec{c}_4 = 0100111$
$\vec{c}_5 = 1101000$
$\vec{c}_6 = 1001111$
$\vec{c}_7 = 1010011$

IS THIS A <u>SUBSPACE</u> ?

2) FULL CHECK: FIND **BASIS**!

$k = 3$ ⇒ LOOKING FOR 3 BASIS VECTORS

$\vec{c}_1 = 0011100$

$\vec{c}_2 = 0111011$

$\vec{c}_3 = 1110100$

code $\mathcal{C}$
$\vec{c}_0 = 0000000$
$\vec{c}_1 = 0011100$
$\vec{c}_2 = 0111011$
$\vec{c}_3 = 1110100$
$\vec{c}_4 = 0100111$
$\vec{c}_5 = 1101000$
$\vec{c}_6 = 1001111$
$\vec{c}_7 = 1010011$

IS THIS A <u>SUBSPACE</u> ?

3) FIND A SET OF <u>EQUATIONS</u> !

$$m + k = n$$

$$k = 3 \quad , \quad n = 7$$

$\Rightarrow$ WE NEED

$$m = n - k = 7 - 3 = 4$$

EQUATIONS.

Example

Let $\mathcal{C} \subset \mathbb{F}_2^7$ be the block code that consists of the listed codewords. Is it linear?

Solution

We have

$$\begin{aligned}
\vec{c}_4 &= \vec{c}_1 + \vec{c}_2 \\
\vec{c}_5 &= \vec{c}_1 + \vec{c}_3 \\
\vec{c}_6 &= \vec{c}_2 + \vec{c}_3 \\
\vec{c}_7 &= \vec{c}_1 + \vec{c}_2 + \vec{c}_3
\end{aligned}$$

Therefore $\mathcal{C} = \mathrm{span}(\vec{c}_1, \vec{c}_2, \vec{c}_3) \subset \mathbb{F}_2^7$ is a linear code (over the finite field $\mathbb{F}_2$).

code $\mathcal{C}$
$\vec{c}_0 = 0000000$
$\vec{c}_1 = 0011100$
$\vec{c}_2 = 0111011$
$\vec{c}_3 = 1110100$
$\vec{c}_4 = 0100111$
$\vec{c}_5 = 1101000$
$\vec{c}_6 = 1001111$
$\vec{c}_7 = 1010011$

Example

What is the dimension of code $\mathcal{C}$ (i.e., the dimension of the subspace formed by the codewords)?

Solution

The set $(\vec{c}_1, \vec{c}_2, \vec{c}_3)$ is a basis of $\mathcal{C}$. Hence $\mathrm{dim}(\mathcal{C}) = 3$.

code $\mathcal{C}$
$\vec{c}_0 = 0000000$
$\vec{c}_1 = 0011100$
$\vec{c}_2 = 0111011$
$\vec{c}_3 = 1110100$
$\vec{c}_4 = 0100111$
$\vec{c}_5 = 1101000$
$\vec{c}_6 = 1001111$
$\vec{c}_7 = 1010011$

SIZE VS DIMENSION

We have seen that a k-dimensional subspace of $\mathbb{F}^n$ has cardinality $[\mathrm{card}(\mathbb{F})]^k$. (Count the number of linear combinations you can form with the vectors that form the basis, with coefficients in $\mathbb{F}$.)

EXAMPLE

If the size of a binary block code is not of the form 2^k, then the code is not linear.

Hamming Weight

Definition

Let $\vec{x} = (x_1, \ldots, x_n)$ be an n-tuple with components in a finite field.

The **(Hamming) weight** of $\vec{x}$, denoted $w(\vec{x})$, is the number of its non-zero components in $(x_1, \ldots, x_n)$, i.e.

$$w(\vec{x}) = d\big((0, \ldots, 0), (x_1, \ldots, x_n)\big).$$

Exercise ("academic" question)

In the definition of Hamming weight, we are requiring that the components of $\vec{x}$ take value in a (finite) field $\mathbb{F}$. Why?

Solution

Otherwise there is no guarantee that the alphabet contains the 0 element.

Recall that in a finite field $\mathbb{F}$, no matter how we label its elements, one is the 0 element (the identity element with respect to addition). Hence the Hamming weight is well-defined.

Example

- The weight of $(1, 0, 1, 1, 0)$ is 3.
- The weight of $(3, 0, 4, 1, 1, 2)$ is 5.

Theorem

The minimum distance of a linear code $\mathcal{C}$ is the smallest weight of a codeword in $\mathcal{C}$, zero-vector excluded, i.e.,

$$d_{min}(\mathcal{C}) = \min_{\vec{c}\in\mathcal{C};\vec{c}\neq\vec{0}} w(\vec{c})$$

FACT 1:

$$\boxed{d(\vec{u}, \vec{v}) = \sum_{i=1}^{n} d(u_i, v_i)} = \boxed{w(\vec{u} - \vec{v}) = \sum_{i=1}^{n} w(u_i - v_i)}$$

<u>OBS</u>: $d(u_i, v_i) = w(u_i - v_i)$

FACT 2:

IF $A \subseteq B$ THEN:

$$\min_{x \in A} f(x) \geq \min_{x \in B} f(x)$$

PROOF PART 1:

CLAIM: $$d_{\min}(\mathcal{C}) \geq \min_{\substack{\vec{c} \in \mathcal{C} \\ \vec{c} \neq \vec{0}}} w(\vec{c})$$

PROOF:

$$d_{\min}(\mathcal{C}) = \min_{\substack{\vec{u}, \vec{v} \in \mathcal{C} \\ \vec{u} \neq \vec{v}}} d(\vec{u}, \vec{v})$$

$$= \min_{\vec{u}, \vec{v} \in \mathcal{C}, \vec{u} \neq \vec{v}} w(\overbrace{\vec{u} - \vec{v}}^{\in \mathcal{C}}) \quad \boxed{\text{BY LINEARITY}}$$

$$\geq \min_{\vec{c} \in \mathcal{C}, \vec{c} \neq \vec{0}} w(\vec{c})$$

PROOF PART 2 :

CLAIM: $$d_{\min}(\mathcal{C}) \leq \min_{\substack{\vec{c} \in \mathcal{C} \\ \vec{c} \neq \vec{0}}} w(\vec{c})$$

PROOF :

$$\min_{\substack{\vec{c} \in \mathcal{C} \\ \vec{c} \neq \vec{0}}} w(\vec{c}) = \min_{\substack{\vec{c} \in \mathcal{C} \\ \vec{c} \neq 0}} d(\vec{c}, \vec{0})$$

$$\geq \min_{\vec{c} \in \mathcal{C}} \min_{\vec{v} \in \mathcal{C},\, \vec{v} \neq \vec{c}} d(\vec{c}, \vec{v})$$

$$= d_{\min}(\mathcal{C}) .$$

In the proof that follows, we use the following two facts:

- For all $\vec{u}, \vec{v} \in \mathbb{F}^n$,

  $$d(\vec{u}, \vec{v}) = w(\vec{u} - \vec{v})$$

  (Reason: $\vec{u}$ and $\vec{v}$ are different at position i iff $\vec{u} - \vec{v}$ is non-zero at position i.)

- Let $f : \mathcal{B} \to \mathbb{R}$ be an arbitrary function and $\mathcal{A} \subseteq \mathcal{B}$ be finite sets. Then

  $$\min_{x \in \mathcal{A}} f(x) \geq \min_{x \in \mathcal{B}} f(x)$$

  (We might find a smaller minimum if we enlarge the set.)

Proof:

$$\begin{aligned} d_{min}(\mathcal{C}) &= \min_{\vec{u},\vec{v}\in\mathcal{C};\vec{u}\neq\vec{v}} d(\vec{u},\vec{v}) \\ &= \min_{\vec{u},\vec{v}\in\mathcal{C};\vec{u}\neq\vec{v}} w(\vec{u}-\vec{v}) \\ &\geq \min_{\vec{c}\in\mathcal{C};\vec{c}\neq\vec{0}} w(\vec{c}) \qquad \text{(reason: } \mathcal{C} \text{ is a vector space, so } \vec{u}-\vec{v}\in\mathcal{C}\text{).} \end{aligned}$$

$$\begin{aligned} \min_{\vec{c}\in\mathcal{C};\vec{c}\neq\vec{0}} w(\vec{c}) &= \min_{\vec{c}\in\mathcal{C};\vec{c}\neq\vec{0}} d(\vec{c},0) \\ &\geq \min_{\vec{c},\vec{v}\in\mathcal{C};\vec{c}\neq\vec{v}} d(\vec{c},\vec{v}) \qquad \text{(reason: we have equality with } \vec{v}=0\text{)} \\ &= d_{min}(\mathcal{C}). \end{aligned}$$

□

code $\mathcal{C}$
$\vec{c}_0 = 0000000$
$\vec{c}_1 = 0011100$
$\vec{c}_2 = 0111011$
$\vec{c}_3 = 1110100$
$\vec{c}_4 = 0100111$
$\vec{c}_5 = 1101000$
$\vec{c}_6 = 1001111$
$\vec{c}_7 = 1010011$

EXERCISE

Find the minimum distance of the linear code $\mathcal{C}$.

$(w(\vec{c}_0)=0)$

$w(\vec{c}_1)=3$

$w(\vec{c}_2)=5$

Solution

- Compute the weight of each non-zero codeword:

$$\begin{aligned} w_1 &= w(0011100) = 3 \\ w_2 &= w(0111011) = 5 \\ w_3 &= w(1110100) = 4 \\ w_4 &= w(0100111) = 4 \\ w_5 &= w(1101000) = 3 \\ w_6 &= w(1001111) = 5 \\ w_7 &= w(1010011) = 4 \end{aligned}$$

- $d_{min}(\mathcal{C}) = 3$.

- Note: compare this to the work needed to compute $d(\vec{v}_i, \vec{v}_j)$ for all $i \neq j$.

code $\mathcal{C}$
$\vec{c}_0 = 0000000$
$\vec{c}_1 = 0011100$
$\vec{c}_2 = 0111011$
$\vec{c}_3 = 1110100$
$\vec{c}_4 = 0100111$
$\vec{c}_5 = 1101000$
$\vec{c}_6 = 1001111$
$\vec{c}_7 = 1010011$

EXAMPLE: $\mathcal{C} \leq \mathbb{F}_2^n$:

$\mathcal{C} = \{$ all sequences with an even number of 1's $\}$

$= \{ \vec{s} : s_1 + s_2 + \ldots + s_n = 0 \}$

$\dim(\mathcal{C}) = n-1. = k$

$d_{min}(\mathcal{C}) = 2$

MEETS SINGLETON'S BOUND.

Exercise ((Binary) Parity-Check Code)

The parity-check code $\mathcal{C} \subset \mathbb{F}_2^n$ consists of those elements of $\mathbb{F}_2^n$ that have an even number of 1s, i.e.,

$$\mathcal{C} = \Big\{(c_1, \ldots, c_n) \in \mathbb{F}_2^n : \sum_i c_i = 0\Big\}.$$

(Addition is in $\mathbb{F}_2$, i.e., mod 2.)

Determine k and d_{min}.

Solution

- The code is a subset of $\mathbb{F}_2^n$ that satisfies an homogeneous linear equation.
- Hence the code is linear, and $k = n - 1$.
- (We can also tell that $k = n - 1$, by observing that we are free to choose the first $n - 1$ bits and satisfy the constraint with the last symbol.)
- For a linear code, d_{min} is the minimum non-zero weight.
- It is achieved by any codeword that has exactly two 1s.
- Hence $d_{min} = 2$.

Note: In this example, linearity allows us to determine d_{min} via deductive reasoning rather than by inspection.

EXAMPLE: $\mathcal{C} \subseteq \mathbb{F}_2^n$

REPETITION CODE

$$\mathcal{C} = \{ 000\ldots0, 111\ldots1 \}$$

$$= \text{span}\{ (11\ldots1) \}$$

$$\left.\begin{array}{l} \Rightarrow k = 1 \\ d_{\min} = n \end{array}\right\} \begin{array}{l} d_{\min} \leq n-k+1 \\ \qquad\quad = n \\ \Rightarrow \text{MDS code } \textcircled{!} \end{array}$$

$\Rightarrow$ HOW MANY EQUATIONS?

$m = n-1.$

EXAMPLE: $\mathcal{C} \subseteq \mathbb{F}_p^n$

REPETITION CODE

$$\mathcal{C} = \{ 000\ldots0,\ 111\ldots1,\ 222\ldots2,\ \ldots (p-1, p-1, \ldots p-1) \}$$

$$= \text{span}(111\ldots1)$$ LINEAR ✓

$k = 1$, $d_{min} = n$

ALSO MDS CODE.

Exercise ((Binary) Repetition Code)

It is the subset of $\mathbb{F}_2^n$ that consists of two codewords, namely $(0, \ldots, 0)$ and $(1, \ldots, 1)$.

Determine k and d_{min}.

Solution

- The code is linear: it is the subspace of $\mathbb{F}_2^n$ spanned by $(1, \ldots, 1)$.
- $k = 1$.
- $d_{min} = n$ (the weight of the only non-zero codeword).

The above two codes are not terribly useful, but they have an interesting property shared by the next code which is even less useful.

EXERCISE (THE CODE $\mathbb{F}_2^n$)

$\mathbb{F}_2^n$ satisfies the definition of a linear code.

Determine k and d_{min}.

SOLUTION

- $k = n$.
- $d_{min} = 1$.

$$d_{min} \le n - k + 1$$
$$= 1$$

⇒ IS ALSO MDS!

BINARY LINEAR MDS CODES

The above three code families fulfill the Singleton bound with equality. Hence they are MDS codes. Moreover, they are also linear codes.

No other family of binary linear codes is MDS.

But there are non-binary codes that are MDS, e.g., the family of Reed-Solomon codes (studied next week).

$$\mathcal{C} = \{(00\ldots0),(11\ldots1)\}$$

$$\downarrow$$

$$\mathcal{C}' = \{(0101\ldots01),(1010\ldots10)\}$$

NOT LINEAR BUT IT'S MDS.

GENERATOR MATRIX

DEFINITION (TEXTBOOK DEFINITION 13.3)

Let $(\vec{c}_1, \ldots, \vec{c}_k)$ be a basis of a linear code $\mathcal{C} \subset \mathbb{F}^n$ over some finite field $\mathbb{F}$. The $k \times n$ matrix that has $\vec{c}_i$ as its ith row is called a **generator matrix** of $\mathcal{C}$.

Example

$(\vec{c}_1, \vec{c}_2, \vec{c}_3)$ form a basis for $\mathcal{C}$. Therefore

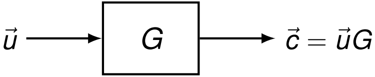
$$G = \begin{pmatrix} \vec{c}_1 \\ \vec{c}_2 \\ \vec{c}_3 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 & 1 & 0 & 0 \end{pmatrix}$$

is a generator matrix of $\mathcal{C}$.

code $\mathcal{C}$
$\vec{c}_0 = 0000000$
$\vec{c}_1 = 0011100$
$\vec{c}_2 = 0111011$
$\vec{c}_3 = 1110100$
$\vec{c}_4 = 0100111$
$\vec{c}_5 = 1101000$
$\vec{c}_6 = 1001111$
$\vec{c}_7 = 1010011$

ENCODING

A $k \times n$ generator matrix specifies an **encoding map** that sends an **information vector** $\vec{u} \in \mathbb{F}^k$ to the corresponding codeword $\vec{c} = \vec{u}G$.



EXAMPLE

$$\vec{u} = (1,0,1) \to \vec{c} = (1,0,1)\begin{pmatrix} 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 & 1 & 0 & 0 \end{pmatrix}$$

$$= (1,1,0,1,0,0,0).$$

code $\mathcal{C}$
$\vec{c}_0 = 0000000$
$\vec{c}_1 = 0011100$
$\vec{c}_2 = 0111011$
$\vec{c}_3 = 1110100$
$\vec{c}_4 = 0100111$
$\vec{c}_5 = 1101000$
$\vec{c}_6 = 1001111$
$\vec{c}_7 = 1010011$

We have seen the generator matrix

$$G_1 = \begin{pmatrix} \vec{c}_1 \\ \vec{c}_2 \\ \vec{c}_3 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 & 1 & 0 & 0 \end{pmatrix}.$$

Here is another one:

$$G_2 = \begin{pmatrix} \vec{c}_1 \\ \vec{c}_1 + \vec{c}_2 \\ \vec{c}_1 + \vec{c}_3 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 & 0 & 0 \end{pmatrix}.$$

A linear code $\mathcal{C}$ has as many generator matrices as the number of bases of the vector space $\mathcal{C}$.

Each generator matrix determines an encoding map:

$$
\begin{array}{rcl}
\vec{u} & \to & \vec{u}G_1 \\
000 & \to & 0000000 = \vec{c}_0 \\
001 & \to & 1110100 = \vec{c}_3 \\
010 & \to & 0111011 = \vec{c}_2 \\
011 & \to & 1001111 = \vec{c}_6 \\
100 & \to & 0011100 = \vec{c}_1 \\
101 & \to & 1101000 = \vec{c}_5 \\
110 & \to & 0100111 = \vec{c}_4 \\
111 & \to & 1010011 = \vec{c}_7
\end{array}
\qquad\qquad
\begin{array}{rcl}
\vec{u} & \to & \vec{u}G_2 \\
000 & \to & 0000000 = \vec{c}_0 \\
001 & \to & 1101000 = \vec{c}_5 \\
010 & \to & 0100111 = \vec{c}_4 \\
011 & \to & 1001111 = \vec{c}_6 \\
100 & \to & 0011100 = \vec{c}_1 \\
101 & \to & 1110100 = \vec{c}_3 \\
110 & \to & 0111011 = \vec{c}_2 \\
111 & \to & 1010011 = \vec{c}_7
\end{array}
$$

Exercise

How many generator matrices for a binary linear code of block-length $n = 7$ and dimension $k = 3$?

Solution

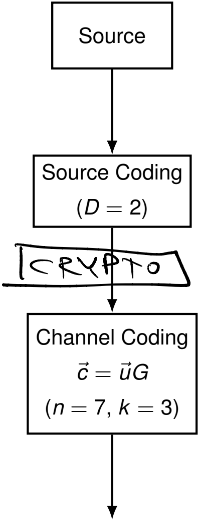
It is the number of lists that form a basis. A q-ary linear code of dimension k has q^k codewords and the number of bases is

$$(q^k - 1)(q^k - q) \cdots (q^k - q^{k-1}).$$

For a binary code ($q = 2$) we have

$$(2^3 - 1)(2^3 - 2)(2^3 - 2^2) = 7 \times 6 \times 4 = 168.$$

THE BIG PICTURE (TRANSMITTER)



dinner is served

$\underbrace{000}_{d}\,\underbrace{0111111}_{i}\,\underbrace{00101}_{n}\ldots$

$$\left\{\begin{array}{l}000\\011\\111\\100\\101\end{array}\right. \xrightarrow{G} \begin{array}{l}0000000\\1001111\\1010011\\0011100\\1101000\end{array}$$

00000001001111101001100111001101000

Exercise

Is $\{\vec{c}_2 + \vec{c}_3, \vec{c}_1 + \vec{c}_2, \vec{c}_1\}$ a basis of $\mathcal{C}$?

If yes,

- Specify the generator matrix.
- Explicitly specify the map $u_1 u_2 u_3 \rightarrow c_1 c_2 c_3 c_4 c_5 c_6 c_7$.

code $\mathcal{C}$
$\vec{c}_0 = 0000000$
$\vec{c}_1 = 0011100$
$\vec{c}_2 = 0111011$
$\vec{c}_3 = 1110100$
$\vec{c}_4 = 0100111$
$\vec{c}_5 = 1101000$
$\vec{c}_6 = 1001111$
$\vec{c}_7 = 1010011$

Solution (basis)

Let $$G' = \begin{pmatrix} \vec{c}_2 + \vec{c}_3 \\ \vec{c}_1 + \vec{c}_2 \\ \vec{c}_1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 0 & 0 \end{pmatrix}.$$

- From the first three columns we see that rank (G') = 3, so $\{\vec{c}_2 + \vec{c}_3, \vec{c}_1 + \vec{c}_2, \vec{c}_1\}$ are linearly independent.
- *n* linearly independent vectors of an *n*-dimensional space always form a basis of the space.
- G' is the generator matrix associated to this basis.

code $\mathcal{C}$
$\vec{c}_0 = 0000000$
$\vec{c}_1 = 0011100$
$\vec{c}_2 = 0111011$
$\vec{c}_3 = 1110100$
$\vec{c}_4 = 0100111$
$\vec{c}_5 = 1101000$
$\vec{c}_6 = 1001111$
$\vec{c}_7 = 1010011$

Solution (encoding map)

$$G' = \left(\begin{array}{ccc|cccc} 1 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 0 & 0 \end{array}\right)$$

$$(c_1, c_2, c_3, c_4, c_5, c_6, c_7) = (u_1, u_2, u_3)G'$$

Therefore:

$$\begin{aligned} c_1 &= u_1 & \qquad c_4 &= u_1 + u_3 \\ c_2 &= u_2 & c_5 &= u_1 + u_2 + u_3 \\ c_3 &= u_3 & c_6 &= u_1 + u_2 \\ & & c_7 &= u_1 + u_2 \end{aligned}$$

$$\vec{c} = (\underbrace{c_1, c_2, c_3}_{\text{message bits}}, \underbrace{c_4, c_5, c_6, c_7}_{\text{parity bits}}).$$

code $\mathcal{C}$
$\vec{c}_0 = 0000000$
$\vec{c}_1 = 0011100$
$\vec{c}_2 = 0111011$
$\vec{c}_3 = 1110100$
$\vec{c}_4 = 0100111$
$\vec{c}_5 = 1101000$
$\vec{c}_6 = 1001111$
$\vec{c}_7 = 1010011$

SYSTEMATIC FORM

The above matrix G' is in systematic form.

DEFINITION (SYSTEMATIC FORM)

A generator matrix G_s is in **systematic form** if

$$G_s = \left(I_k, P_{k\times(n-k)}\right).$$

Notice that a systematic generator matrix is a matrix in reduced echelon form.

When the generator matrix is in systematic form, each codeword is written as

$$\vec{c} = \vec{u}G_s = (\underbrace{u_1, \ldots, u_k}_{\vec{u}I=\vec{u}}, \underbrace{c_{k+1}, \ldots, c_n}_{\vec{u}P}).$$

EXAMPLE:

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

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

$$G'' = \left(\right.$$

EXAMPLE:

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

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

$$G^{(2)} = \begin{pmatrix} 1 & 0 & \boxed{1} & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 \end{pmatrix}$$

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

Given a linear code $\mathcal{C}$, how does one find the systematic form G_s?

1. Find a basis $\{\vec{c}_1, \dots, \vec{c}_k\}$ of $\mathcal{C}$.

2. Form the generator matrix: $G = \begin{pmatrix} \vec{c}_1 \\ \vdots \\ \vec{c}_k \end{pmatrix}$.

3. Row-reduce G (Gaussian elimination on rows) to obtain a matrix in reduced echelon form.

pivots (leading coeff.)

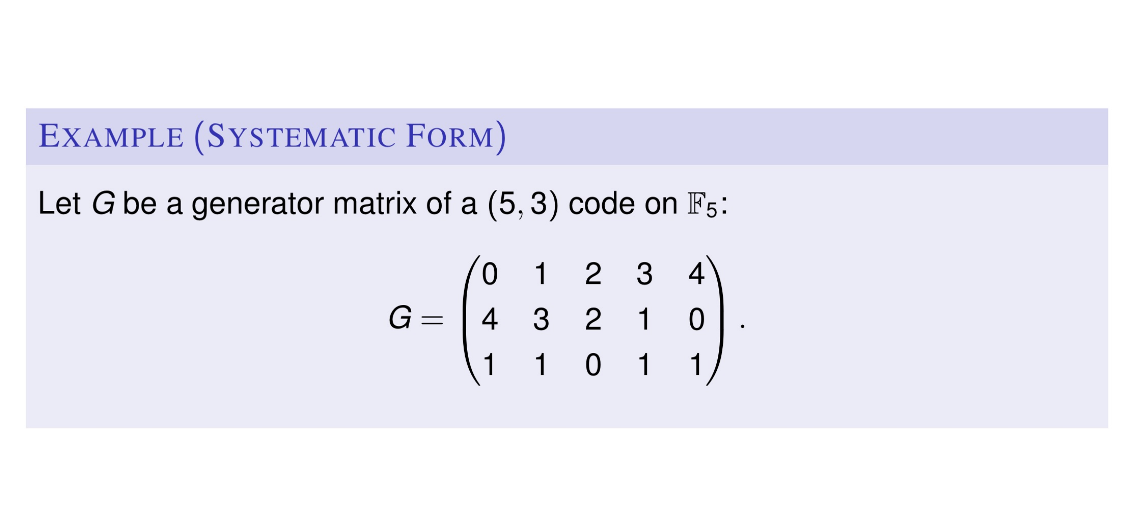
$$\underbrace{\begin{bmatrix} \star & \star & \star & \star & \star & \star & \star \\ 0 & \star & \star & \star & \star & \star & \star \\ 0 & 0 & 0 & \star & \star & \star & \star \\ 0 & 0 & 0 & 0 & \star & \star & \star \end{bmatrix}}_{\text{echelon form}} \qquad \underbrace{\begin{bmatrix} 1 & 0 & \star & 0 & 0 & \star & \star \\ 0 & 1 & \star & 0 & 0 & \star & \star \\ 0 & 0 & 0 & 1 & 0 & \star & \star \\ 0 & 0 & 0 & 0 & 1 & \star & \star \end{bmatrix}}_{\text{reduced echelon form}}$$

This procedure uses the three operations below (that do *not* modify the vector space spanned by the rows of G):

T_{ij}: transpose (swap) rows i and j;

αS_i: scale row i by α;

αA_{ij}: scale row i by α and add to row j.



Example (Systematic Form)

Let G be a generator matrix of a $(5,3)$ code on $\mathbb{F}_5$:

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

$$G^{(1)} = \begin{pmatrix} 1 & 1 & 0 & 1 & 1 \\ 4 & 3 & 2 & 1 & 0 \\ 0 & 1 & 2 & 3 & 4 \end{pmatrix}$$

$$G^{(2)} = \begin{pmatrix} 1 & 1 & 0 & 1 & 1 \\ 0 & 4 & 2 & 2 & 1 \\ 0 & 1 & 2 & 3 & 4 \end{pmatrix}$$

$$G^{(3)} = \begin{pmatrix} 1 & 1 & 0 & 1 & 1 \\ 0 & 4 & 2 & 2 & 1 \\ 0 & 0 & 4 & 0 & 0 \end{pmatrix}$$

$$G^{(4)} = \begin{pmatrix} 1 & 1 & 0 & 1 & 1 \\ 0 & 4 & 0 & 2 & 1 \\ 0 & 0 & 4 & 0 & 0 \end{pmatrix}$$

$$G^{(5)} = \begin{pmatrix} 1 & 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 3 & 4 \\ 0 & 0 & 4 & 0 & 0 \end{pmatrix}$$

The result is a generator matrix for the same code.

To make sure that we have the identity matrix on the left, we may have to swap columns.

If we swap columns, we obtain a different code (different set of codewords) that has the same parameters (n, k, d_{min}) as the original code.

Example (Systematic Form)

Let G be a generator matrix of a $(5, 3)$ code on $\mathbb{F}_5$:

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

ENCODING

EXAMPLE (SYSTEMATIC FORM (CONT.))

$$G = \begin{pmatrix} 0 & 1 & 2 & 3 & 4 \\ 4 & 3 & 2 & 1 & 0 \\ 1 & 1 & 0 & 1 & 1 \end{pmatrix} \xrightarrow{T_{13}} \begin{pmatrix} 1 & 1 & 0 & 1 & 1 \\ 4 & 3 & 2 & 1 & 0 \\ 0 & 1 & 2 & 3 & 4 \end{pmatrix}$$

$$\xrightarrow{T_{23}} \begin{pmatrix} 1 & 1 & 0 & 1 & 1 \\ 0 & 1 & 2 & 3 & 4 \\ 4 & 3 & 2 & 1 & 0 \end{pmatrix} \xrightarrow{4A_{21}} \begin{pmatrix} 1 & 0 & 3 & 3 & 2 \\ 0 & 1 & 2 & 3 & 4 \\ 4 & 3 & 2 & 1 & 0 \end{pmatrix}$$

$$\xrightarrow{A_{13}} \begin{pmatrix} 1 & 0 & 3 & 3 & 2 \\ 0 & 1 & 2 & 3 & 4 \\ 0 & 3 & 0 & 4 & 2 \end{pmatrix} \xrightarrow{2A_{23}} \begin{pmatrix} 1 & 0 & 3 & 3 & 2 \\ 0 & 1 & 2 & 3 & 4 \\ 0 & 0 & 4 & 0 & 0 \end{pmatrix}$$

$$\xrightarrow{4S_{3}} \begin{pmatrix} 1 & 0 & 3 & 3 & 2 \\ 0 & 1 & 2 & 3 & 4 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix} \xrightarrow{3A_{32}} \begin{pmatrix} 1 & 0 & 3 & 3 & 2 \\ 0 & 1 & 0 & 3 & 4 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix}$$

$$\xrightarrow{2A_{31}} \begin{pmatrix} 1 & 0 & 0 & 3 & 2 \\ 0 & 1 & 0 & 3 & 4 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix} = G_s.$$

Example (Systematic Form (cont.))

With the generator matrix in systematic form

$$G_s = \begin{pmatrix} 1 & 0 & 0 & 3 & 2 \\ 0 & 1 & 0 & 3 & 4 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix},$$

$\mathbb{F}_5$

the map $\underbrace{(u_1, u_2, u_3)}_{\text{information word}} \mapsto \underbrace{(c_1, c_2, c_3, c_4, c_5)}_{\text{codeword}}$ is

$u_i \in \mathbb{F}_5$

$$\begin{aligned} c_1 &= u_1 \\ c_2 &= u_2 \\ c_3 &= u_3 \\ c_4 &= 3u_1 + 3u_2 \\ c_5 &= 2u_1 + 4u_2 \end{aligned}$$

YESTERDAY

- DEF: (n, k) LINEAR BLOCK CODE $=$ k-DIM SUBSPACE OF $\mathbb{F}^n$
- HAS EXACTLY $(\mathrm{CARD}(\mathbb{F}))^k$ CODEWORDS
- $d_{min} =$ minimum weight of any non-zero codeword

- GENERATOR MATRIX G :

EVERY CODEWORD CAN BE WRITTEN AS

$$\vec{C} = \underset{\in \mathbb{F}^k}{\vec{u}}\, G$$

- ALSO GIVES ENCODING MAP
- PARTICULARY ATTRACTIVE :

$$G = \left(I_k \,\vdots\, P \right), \quad P \in \mathbb{F}^{k \times (n-k)}$$

"SYSTEMATIC" GENERATOR MATRIX.

- **DEF:** AN (n, k) LINEAR BLOCK CODE IS CALLED "SYSTEMATIC" IF IT HAS A SYSTEMATIC GENERATOR MATRIX.

- **QUESTION:** ISN'T EVERY CODE SYSTEMATIC ?

EXAMPLE: $\mathbb{F}_3^3$

$$G = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 2 \end{pmatrix} \qquad \mathcal{C} = \{ (0\ 0\ 0), (1\ 1\ 2), \vdots \}$$

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

OBS: **NOT** A SYSTEMATIC CODE!

OBS 2: BUT THERE EXISTS A CODE THAT IS:

- EXACTLY OF THE SAME QUALITY
- BUT IT IS SYSTEMATIC.

$$\tilde{G} = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \end{pmatrix} \quad \mathcal{C} = \{(000), (010)\ (020), (101), (111), \dots\}$$

$$\tilde{G}^{(1)} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \nearrow$$

Example (Systematic Form)

Here an example where we have to swap columns.

Let

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

The steps towards the reduced echelon form are

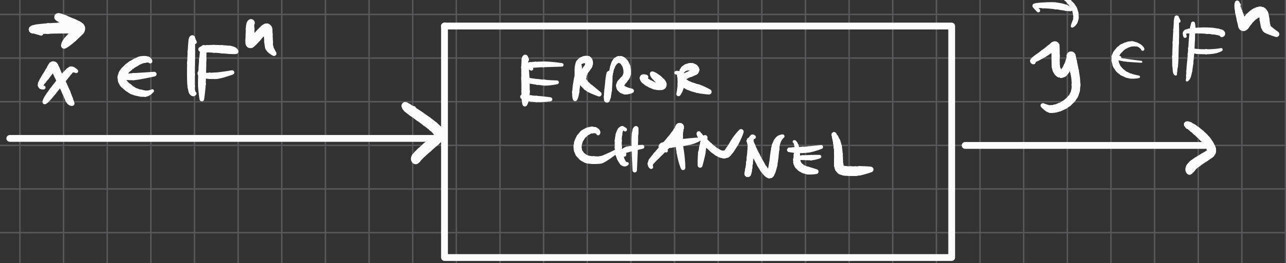
$$\begin{pmatrix} 1 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix} \quad \sim \quad \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

By swapping the second and third columns, we obtain the following generator matrix of a different (but equivalent) code.

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

DECODING

→ ERROR CHANNEL



$$\rightsquigarrow \quad \vec{y} = \vec{x} \oplus \vec{e}$$

component-wise addition in $\mathbb{F}$.

$\vec{e}$: error pattern

$\vec{e} = \vec{y} - \vec{x}$

FIRST QUESTION:

IS $\vec{y} \in \mathbb{F}^n$ A CODEWORD ?

RECALL:

$$S = \{ \vec{y} : \vec{y} H^T = \vec{0} \}$$

$$= \left\{ \vec{y} : \begin{aligned} &y_1 H_{11} + y_2 H_{12} + \dots + y_n H_{1n} = 0 \\ &\vdots \\ &y_1 H_{m,1} + y_2 H_{m,2} + \dots + y_n H_{m,n} = 0 \end{aligned} \right\}$$

DECODING

How to decode?

Decoding is about deciding the information word from the channel output.

If the channel output $\vec{y}$ is a codeword, then we assume that it equals the channel input.

In this case decoding is about inverting the encoding map. This is trivial if the generator matrix is in systematic form. (We read out the first k symbols of $\vec{y}$.)

But how to know if the channel output is a codeword?

We use the fact that a linear block code, like every subspace of a vector space, can be defined by a system of homogeneous linear equations.

The channel output is a codeword iff it satisfies those equations.

EXAMPLE

$$\begin{cases} c_4 = 3c_1 + 3c_2 \\ c_5 = 2c_1 + 4c_2 \end{cases} \Rightarrow \begin{cases} -3c_1 - 3c_2 + c_4 = 0 \\ -2c_1 - 4c_2 + c_5 = 0 \end{cases} \Rightarrow \begin{cases} 2c_1 + 2c_2 + c_4 = 0 \\ 3c_1 + c_2 + c_5 = 0 \end{cases}$$

Therefore, $\vec{y} \in \mathcal{C}$ iff

$$\begin{cases} 2y_1 + 2y_2 + y_4 & = 0 \\ 3y_1 + y_2 + y_5 & = 0 \end{cases}$$

i.e., iff

$$(y_1, y_2, y_3, y_4, y_5) \underbrace{\begin{pmatrix} 2 & 3 \\ 2 & 1 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}}_{H^\mathsf{T}} = \vec{0}.$$

Parity-Check Matrix

A **parity-check matrix** H for a linear (n, k) code is an $(n-k) \times n$ matrix that contains the coefficients of a system of homogeneous linear equations that defines the code.

Example

$$(y_1, y_2, y_3, y_4, y_5)\underbrace{\begin{pmatrix} 2 & 3 \\ 2 & 1 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}}_{H^{\mathsf{T}}} = \vec{0} \qquad \Leftrightarrow \qquad \begin{cases} 2y_1 + 2y_2 + y_4 & = 0 \\ 3y_1 + y_2 + y_5 & = 0 \end{cases}$$

Theorem (Textbook Theorem 13.1)

If $G = (I_k, P)$, where P is a $k \times (n-k)$ matrix, is a generator matrix (in systematic form) of a linear (n, k) block code, then

$$H = \big(-P^{\mathsf{T}}, I_{n-k}\big)$$

is a parity-check matrix of the same code.

PROOF:

(1) IF G IS A GENERATOR MATRIX, THEN ANY CODEWORD CAN BE WRITTEN AS:

$$\vec{c} = \vec{u} G$$ FOR SOME $\vec{u} \in \mathbb{F}^k$

(2) IF H IS A PARITY CHECK MATRIX. THEN ANY CODEWORD MUST SATISFY

$$\vec{c} H^T = \vec{0}$$

$$\vec{u} G H^T = \vec{0}$$

$$G H^T = 0$$

$$\begin{pmatrix} I_k & \vdots & P \end{pmatrix} \begin{pmatrix} -P \\ I_{n-k} \end{pmatrix} = P - P$$

$$= 0$$

$\Rightarrow$ we need $\vec{u}GH^T = \vec{0}$ FOR <u>ALL</u> $\vec{u} \in \mathbb{F}^k$

$$\Updownarrow$$

$\mathbb{F}^{k\times(n-k)} \ni GH^T$ must be the all-zero matrix

FOR THE CASE OF <u>SYSTEMATIC</u>:

$$\left(I_k \;\vdots\; P \right) \begin{pmatrix} -P \\ \cdots \\ I_{n-k} \end{pmatrix}$$

$$= -P + P = 0. \quad ✓$$

Proof:

$H = \big(-P^{\mathsf{T}}, I_{n-k} \big)$ has rank $(n-k)$, hence it defines a system of equations, the solution of which is a subspace of $\mathbb{F}^n$ of dimension k.

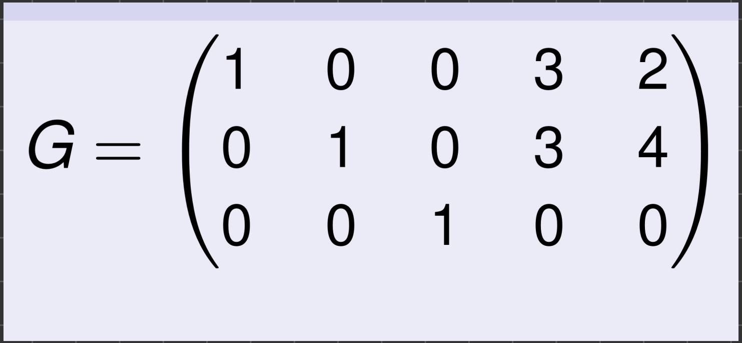
We want to show that $\vec{u}GH^T = \vec{0}$ for all information vectors $\vec{u}$.

This is true iff GH^T is the zero matrix (of size $k \times (n-k)$).

$$GH^{\mathsf{T}} = (I_k, P) \begin{pmatrix} -P \\ I_{n-k} \end{pmatrix} = -P + P = 0.$$

□

EXAMPLE:



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

$n = 5$

$k = 3$

↳ $n - k = 2$

$$H = \begin{array}{cc} & 5 \\ 2 & \begin{pmatrix} -3 & -3 & 0 & 1 & 0 \\ -2 & -4 & 0 & 0 & 1 \end{pmatrix} \end{array}$$

BINARY CODE OF ALL SEQUENCES WITH EVEN # OF ONES.

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

$$n, \; k = n-1$$

$$G = \left(\begin{array}{cccc|c} 1 & & 0 & & -1 \\ & 1 & & & -1 \\ & & 1 & & \vdots \\ 0 & & & \ddots & \\ & & & 1 & -1 \end{array}\right)$$

Example

$G = \begin{pmatrix} 1 & 0 & 0 & 3 & 2 \\ 0 & 1 & 0 & 3 & 4 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix}$ is the generator matrix of a $(5, 3)$ code over $\mathbb{F}_5$.

$$H = \begin{pmatrix} -3 & -3 & 0 & 1 & 0 \\ -2 & -4 & 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 2 & 0 & 1 & 0 \\ 3 & 1 & 0 & 0 & 1 \end{pmatrix}$$

is a corresponding parity-check matrix.

ADDITIONAL TRICK . . .

RECALL EXAMPLE IN $\mathbb{F}_3^3$

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

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

$$\downarrow$$

$$\tilde{G}^{(1)} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

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

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

LAST WEEK

LINEAR CODE:

$$\mathcal{C} = \{ \vec{c} : \vec{c} = \vec{u}G, \ \vec{u} \in \mathbb{F}_q^k \}$$

$$= \{ \vec{c} : \vec{c} H^T = \vec{0}, \ \vec{0} \in \mathbb{F}_q^{n-k} \}$$

SYSTEMATIC:

$$G = \begin{pmatrix} I_k & P \end{pmatrix} \qquad H = \begin{pmatrix} -P^T & I_{n-k} \end{pmatrix}$$

Syndrome

Definition

Let H be the $(n-k) \times n$ parity-check matrix of a linear block code $\mathcal{C} \subset \mathbb{F}^n$ and let $\vec{y} \in \mathbb{F}^n$.

The **syndrome** of $\vec{y}$ is the vector

$$\vec{s} = \vec{y} H^\mathsf{T}.$$

By definition,

$$\vec{y} \in \mathcal{C} \quad \Longleftrightarrow \quad \vec{s} = \vec{0}.$$

$$\vec{s} = \vec{y} H^T = (\vec{x} + \vec{e}) H^T$$
$$= \vec{x} H^T + \vec{e} H^T$$

HAMMING CODES

$\mathbb{F}_2^{15}$

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

$$\vec{y} = \begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 \end{pmatrix}$$

$$0 \quad 0 \quad 0 \quad 1 \quad 1 \quad 1 \quad 1 \quad 0 \quad 0 \quad 0 \quad 0 \quad 1 \quad 1 \quad 1 \quad 1$$

$$\vec{s} = \vec{y} H^T = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 0 \end{pmatrix}$$

$$\vec{y} = \vec{c} + \vec{e}$$

IF WE CHOOSE

$$\vec{e} = (0\ 0\ 0\ 0\ 0\ 0\ \underset{\substack{\uparrow \\ \text{seventh position}}}{1}\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0)$$

THEN

$$\vec{c} =$$

$$(1\ 1\ 0\ 1\ 0\ 0\ 1\ 1\ 0\ 0\ 1\ 0\ 1\ 1\ 0)$$

IS THE DECODED CODEWORD.

NOW SUPPOSE THAT

$$\vec{y} = \underset{\substack{\uparrow \\ \text{SOME CODEWORD}}}{\vec{c}} + \vec{e}_i$$

WHERE:

$$\vec{e}_i = (0\ 0\ \ldots\ \underset{\substack{\uparrow \\ i\text{th POSITION}}}{1}\ 0\ 0\ \ldots\ 0)$$

<u>EX</u>: $\vec{e}_2 = (0\ 1\ 0\ \ldots\ 0)$

$$\Rightarrow \quad \vec{s} = (\vec{c} + \vec{e_i}) H^T$$

$$= \vec{e_i} H^T$$

$$= \vec{h_i}$$: the ith COLUMN OF THE PARITY CHECK MATRIX H.

Example (Hamming codes)

For every integer $m \geq 2$, there exists a binary Hamming code of parameters

$$n = 2^m - 1,$$
$$k = n - m.$$

The parity-check matrix is the $m \times n$ matrix whose columns consist of all non-zero vectors of length m.

Hamming codes are easy to encode and to decode.

$$d_{min} \leq n - k + 1 = m + 1$$

Example (Cont.)

For instance, let $m = 3$.

A valid parity check matrix is

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

where, for convenience, the ith column is the binary representation of i.

The block-length is $n = 2^m - 1 = 7$.

The rank of H is m, hence the code dimension is $k = n - m = 4$.

$\vec{c} = (1, 1, 1, 0, \ldots, 0)$ is a codeword because $\vec{c}H^{\mathsf{T}} = 0$.

Hence $d_{min} \leq 3$.

We show that $d_{min} = 3$ by showing how to correct all error patterns of weight 1.

Example (Cont.)

Let $\vec{y} = \vec{c} + \vec{e}$, be the channel output, where $\vec{c} \in \mathcal{C}$, $\vec{e} \in \mathbb{F}_2^n$, and $w(\vec{e}) = 1$.

$\vec{s} = \vec{y}H^{\mathsf{T}} = \vec{c}H^{\mathsf{T}} + \vec{e}H^{\mathsf{T}} = \vec{e}H^{\mathsf{T}}$.

$\vec{s} = \vec{e}H^{\mathsf{T}}$ is the binary representation of the position of the error.

Hence we can correct the error.

MINIMUM DISTANCE OF HAMMING CODE

1) THEY CAN CORRECT ONE ERROR

$$\Rightarrow d_{min} \geq 3$$

2) $\vec{c} = (1\ 1\ 1\ 0\ 0\ 0 \ldots 0)$ IS A CODEWORD

$$\Rightarrow d_{min} = 3 .$$

$\Rightarrow$ FOR $m \geq 3$, <u>NOT</u> MDS.

SYNDROME DECODING

SPECIAL CASE OF SINGLE ERROR:

OBS: IF ALL COLUMNS OF H ARE DIFFERENT, THEN YOUR CODE CAN CORRECT ANY SINGLE ERROR.

Exercise

For the $(7,4)$ Hamming code,

1. find a parity-check matrix of the form $H = \big(-P^{\mathsf{T}}, I_3\big)$.
2. find the corresponding generator matrix $G = \big(I_4, P\big)$.

Solution

1. By moving to the far right the first, second, and fourth columns of the original parity-check matrix we obtain

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

which has the desired form. Notice that the result is a different code (we have reordered the components) — still a Hamming code.

2.

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

Exercise

For the binary $(4, 1)$ repetition code,

1. find a generator matrix in systematic form.
2. find a parity-check matrix.

SOLUTION

1. The code consists of the two codewords $(0, 0, 0, 0)$ and $(1, 1, 1, 1)$. There is only one basis of this code, hence there is only one generator matrix

$$G = (1, 1, 1, 1).$$

2. Since the generator matrix is of the form (I_1, P), a corresponding parity-check matrix is of the form $(-P^T, I_3)$, namely

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

Theorem

Let H be any parity check matrix of a linear code. The minimum distance of the code is the smallest positive integer d such that there are d columns of H that are linearly dependent.

PROOF:

$$H = \begin{pmatrix} | & | & | & & | \\ \vec{h}_1^T & \vec{h}_2^T & \vec{h}_3^T & \cdots & \vec{h}_n^T \\ | & | & | & & | \end{pmatrix}$$

$$\vec{c} = (c_1, c_2, c_3, \ldots, c_n)$$

$$\vec{c}\, H^T = (c_1, c_2, \ldots, c_n) \begin{pmatrix} - & \vec{h}_1 & - \\ - & \vec{h}_2 & - \\ & \vdots & \\ - & \vec{h}_n & - \end{pmatrix}$$

$$= c_1 \vec{h}_1 + c_2 \vec{h}_2 + \ldots + c_n \vec{h}_i$$

$$\overset{!}{=} \vec{0}$$ IF $\vec{c}$ IS A CODEWORD.

Proof:

For a linear code, the minimum distance is the smallest weight of a non-zero codeword. Let $\vec{c} \neq 0$ be a codeword of smallest weight d. The fact that $\vec{c}H^T = 0$ proves that H has d linearly dependent columns.

We need to argue that fewer than d columns of H are not linearly dependent. Suppose that H has $t < d$ linearly dependent columns. Then we could find a non-zero codeword $\vec{c}$ of weight smaller than d such that $\vec{c}H^T = 0$. This is a contradiction. □

$$\vec{c}\, H^T = \vec{c} \begin{pmatrix} | & | & & | \\ \vec{h}_1^T & \vec{h}_2^T & \cdots & \vec{h}_n^T \\ | & | & & | \end{pmatrix}^T$$

$$= c_1 \vec{h}_1^T + c_2 \vec{h}_2^T + \cdots + c_n \vec{h}_n^T$$

$$\overset{!}{=} \vec{0}$$

Example

The following is a parity check matrix for a Hamming code of parameters $n = 7$, $k = 4$.

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

Clearly no two columns are linearly dependent.

Column 1, 2, and 3 are linearly dependent.

Hence $d_{min} = 3$.

SYNDROME DECODING

GENERAL CASE

Standard Array: background Material

Equivalence Relation (review):

- $\mathcal{G}$ a set;
- $\sim$ an equivalence relation on $\mathcal{G}$;
- $[a]$ the equivalence class of $a \in \mathcal{G}$.

Key property that we will use: An equivalence relation on a set partitions the set into disjoint equivalence classes.

Example

- $\mathcal{G}$ is the set of all students in Switzerland
- $a \sim b$ if a and b attend the same university
- $[a]$ the subset of $\mathcal{G}$ that contains all the students that attend the same university as a

Note: As in the above example, equivalence classes need not have the same size.

Special Case: Group-Theoretic Construction

When $\mathcal{G}$ forms a commutative group $(\mathcal{G}, \star)$ and $(\mathcal{H}, \star)$ is a subgroup, there is a natural choice for $\sim$ defined as follows:

$a \sim b$ if there exists an $h \in \mathcal{H}$ such that $b = a \star h$.

Equivalently:
$a \sim b$ if $a^{-1} \star b \in \mathcal{H}$.

Claim: The above $\sim$ is indeed an equivalence relation:

Proof:

- (reflexive) $a \sim a$:
  true because $a^{-1} \star a \in \mathcal{H}$

- (symmetric) if $a \sim b$ then $b \sim a$:
  true because if $a^{-1} \star b = h \in \mathcal{H}$ then $b^{-1} \star a = h^{-1} \in \mathcal{H}$

- (transitive) if $a \sim b$ and $b \sim c$ then $a \sim c$:
  true because if $a^{-1} \star b = h_1 \in \mathcal{H}$ and $b^{-1} \star c = h_2 \in \mathcal{H}$, then $\mathcal{H}$ contains also $h_1 \star h_2$ which is $a^{-1} \star b \star b^{-1} \star c = a^{-1} \star c$

□

Since in this case an equivalence class has the form

$$[a] = \{a \star h : h \in \mathcal{H}\},$$

it makes sense to write

$$[a] = a \star \mathcal{H}.$$

For instance, if $\star$ is the addition, then $[a]$ is $\mathcal{H}$ translated by a.

COSET

THE EQUIVALENCE CLASS

$$[a] = \{ a \star h : h \in \mathcal{H} \}$$
$$= a \star \mathcal{H}$$

IS CALLED THE <u>COSET</u>.

OF $\mathcal{H}$ W.R.T. $a \in \mathcal{G}$

Example

Let $(\mathcal{G}, +) = (\mathbb{Z}/10\mathbb{Z}, +)$ and let $\mathcal{H} = \{0, 5\}$.

Then $(\mathcal{H}, +)$ is a subgroup of $(\mathcal{G}, +)$, and the equivalence classes are:

$$
\begin{aligned}
[0] &= \mathcal{H} = \{0, 5\} \\
[1] &= 1 + \mathcal{H} = \{1, 6\} \\
[2] &= 2 + \mathcal{H} = \{2, 7\} \\
[3] &= 3 + \mathcal{H} = \{3, 8\} \\
[4] &= 4 + \mathcal{H} = \{4, 9\}.
\end{aligned}
$$

$$[5] = [0]$$

$$(\mathcal{G}, \star) \rightsquigarrow (\mathcal{H}, \star)$$

THM :

1) ALL COSETS HAVE CARDINALITY $|\mathcal{H}|$.

2) EVERY $g \in \mathcal{G}$ IS IN **EXACTLY ONE** COSET.

3) THERE ARE $\dfrac{|\mathcal{G}|}{|\mathcal{H}|}$ COSETS.

In the group theoretic language, $[a]$ is called the coset of $\mathcal{H}$ with respect to a.

Claim: All cosets of $\mathcal{H}$ have the same cardinality $\mathsf{card}(\mathcal{H})$.

Proof:

$$h_1, h_2 \in \mathcal{H} \text{ s.t. } h_1 \neq h_2 \implies a \star h_1 \neq a \star h_2.$$

Hence $a \star \mathcal{H}$ has the same cardinality as $\mathcal{H}$. □

$$(G, \star) \rightsquigarrow (H, \star)$$

$$(\mathbb{F}^n, +) \rightsquigarrow (\mathcal{C}, +)$$

LINEAR CODE OVER $\mathbb{F}^n$.

$$(G, \star) \rightsquigarrow (H, \star)$$

$$(\mathbb{F}^n, +) \rightsquigarrow (\mathcal{C}, +)$$

EX: $(\mathbb{F}_2^3, +) \rightsquigarrow (\mathcal{C} \quad , +)$

$$\mathcal{C} = \{000, 111\}$$

Here is the group and subgroup of interest to us:

- The group $(\mathcal{G}, \star)$ is $(\mathbb{F}^n, +)$ for some finite field $\mathbb{F}$ and positive integer n;
- the subset $\mathcal{H}$ is a linear code $\mathcal{C} \subset \mathbb{F}^n$;
- then if $x, y \in \mathbb{F}^n$, $x \sim y$ iff $-x + y \in \mathcal{C}$;
- (equivalently, $x \sim y$ iff $y = x + c$ for some $c \in \mathcal{C}$).

EXAMPLE $(\mathbb{F}_2^3, +)$

$\mathcal{C} = \{000, \; 111\}$

$001 + \mathcal{C} = [001] = \{001, 110\}$

$000 + \mathcal{C} = [000] = \{000, 111\}$

$111 + \mathcal{C} = [111] = \{111, 000\}$

$110 + \mathcal{C} = [110] = \{110, 001\}$

$100 + \mathcal{C} = [100) = \{100, 011\}$

$\vdots$

"STANDARD ARRAY"

LEADER t	$t + \mathcal{C}$	
$t_0 = 000$	000	111
$t_1 = 001$	001	110
$t_2 = 010$	010	101
$t_3 = 100$	100	011

"STANDARD ARRAY"

LEADER t	$t + c$	
$t_0 = 000$	000	111
$t_1 = 001$	001	110
$t_2 = 010$	010	101
$t_3 = 100$	100	011
	DECODING REGION D_0	DECODING REGION D_1

"STANDARD ARRAY"

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

LEADER t	$t + \mathcal{C}$		
$t_0 = 000$	000	111	$\vec{s} = (0,0)$
$t_1 = 001$	001	110	$\vec{s} = (0,1)$
$t_2 = 010$	010	101	$\vec{s} = (1,0)$
$t_3 = 100$	100	011	$\vec{s} = (1,1)$
	DECODING REGION D_0	DECODING REGION D_1	

YESTERDAY

- A LINEAR CODE $\mathcal{C}$ PARTITIONS $\mathbb{F}^n$ INTO $\boxed{\text{COSETS}}$

- $y_1, y_2 \in \mathbb{F}^n$ ARE IN THE SAME COSET IFF $y_1 - y_2 \in \mathcal{C}$.

$\mathcal{C} = \{000, 111\} \in \mathbb{F}_2^3$ $\quad H = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}$

↓

$\vec{t}$	coset of $\mathcal{C}$ with respect to $\vec{t}$			
000	$\mathcal{C}$	000	111	$\vec{s} = (0,0)$
001	$001 + \mathcal{C}$	001	110	$\vec{s} = (0,1)$
010	$010 + \mathcal{C}$	010	101	$\vec{s} = (1,0)$
100	$100 + \mathcal{C}$	100	011	$\vec{s} = (1,1)$

$\vec{t}$	coset of $\mathcal{C}$ with respect to $\vec{t}$			
000	$\mathcal{C}$	000	111	$\vec{s} = (0,0)$
011	$011 + \mathcal{C}$	011	100	$\vec{s} = (1,1)$
101	$101 + \mathcal{C}$	101	010	$\vec{s} = (1,0)$
001	$001 + \mathcal{C}$	001	110	$\vec{s} = (0,1)$

COSET DECODER

$\vec{t}$	coset of $\mathcal{C}$ with respect to $\vec{t}$		
000	$\mathcal{C}$	000	111
001	$001 + \mathcal{C}$	001	110
010	$010 + \mathcal{C}$	010	101
100	$100 + \mathcal{C}$	100	011

$\vec{t}$	coset of $\mathcal{C}$ with respect to $\vec{t}$		
000	$\mathcal{C}$	000	111
011	$011 + \mathcal{C}$	011	100
101	$101 + \mathcal{C}$	101	010
001	$001 + \mathcal{C}$	001	110

$\mathbb{F}_2^5$ $\quad C = \text{span}\{(1\,1\,1\,1\,1)\}$

	00000	11111
10000	10000	01111

$\mathbb{F}_2^5$ $\quad C = \text{span}\{(1\,1\,1\,1\,1)\}$

	0 0 0 0 0	1 1 1 1 1
1 0 0 0 0		
0 1 0 0 0		
0 0 1 0 0		
0 0 0 1 0		
0 0 0 0 1		
1 1 0 0 0		
1 0 1 0 0		
1 0 0 1 0		
1 0 0 0 1		
0 1 1 0 0		
0 1 0 1 0		
0 1 0 0 1		
0 0 1 1 0		
0 0 1 0 1		
0 0 0 1 1		

$\mathbb{F}_3^5$. $\quad C = \text{span}\{(1\ 1\ 1\ 1\ 1)\}$

00000 $\quad$ 11111 $\quad$ 22222

10000
01000
00100
00010
00001

20000
02000
00200
00020
00002

$\vdots$

EXAMPLE $(\mathbb{F}_2^4, +)$

$G = \{ 0000, 0011, 1100, 1111 \}$

0000	0000	0011	1100	1111
1000	1000	1011	0100	0111
0010	0010	0001	1110	1101
1010	1010	1001	0110	0101

STANDARD ARRAY

The **Standard Array** is an array that has the elements of $\mathcal{C}$ in the top row, starting with the 0 codeword, and each row forms a coset of $\mathcal{C}$. Each element of $\mathbb{F}^n$ shows up exactly once in the standard array.

$$\begin{array}{|ccccc|l}
c_0 = 0 & c_1 & c_2 & \ldots & c_{M-1} & \leftarrow [\mathcal{C}] \\
t_1 & t_1 + c_1 & t_1 + c_2 & \ldots & t_1 + c_{M-1} & \leftarrow [t_1] \\
t_2 & t_2 + c_1 & t_2 + c_2 & \ldots & t_2 + c_{M-1} & \leftarrow [t_2] \\
\vdots & \vdots & \vdots & & \vdots & \vdots \\
t_{L-1} & t_{L-1} + c_1 & t_{L-1} + c_2 & \ldots & t_{L-1} + c_{M-1} & \leftarrow [t_{L-1}]
\end{array}$$

where for each $j = 1, \ldots, L-1$, t_j is such that

$$t_j \notin \left(\mathcal{C} \bigcup_{k=1}^{j-1} [t_k] \right).$$

Later we will choose the coset leaders more carefully.

Decoding Regions

Suppose that $\mathsf{card}(\mathcal{C}) = M$.

Think of the decoder as being specified by M decoding regions $\mathcal{D}_0, \ldots, \mathcal{D}_{M-1}$ that partition $\mathbb{F}^n$:

$$\mathcal{D}_i \bigcap \mathcal{D}_j = \emptyset \text{ if } i \neq j;$$

$$\bigcup_{i=0}^{M-1} \mathcal{D}_i = \mathbb{F}^n.$$

Upon observing $y \in \mathbb{F}^n$, the decoder finds the i such that $y \in \mathcal{D}_i$, and declares

$$\hat{c} = c_i.$$

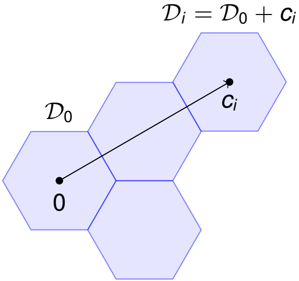
THE COSET DECODER : HOW TO DECODE WITH THE STANDARD ARRAY

We let $\mathcal{D}_i$ be the ith column of the standard array.

$$\begin{array}{ccccc}
c_0 = 0 & c_1 & c_2 & \ldots & c_{M-1} \\
t_1 & t_1 + c_1 & t_1 + c_2 & \ldots & t_1 + c_{M-1} \\
t_2 & t_2 + c_1 & t_2 + c_2 & \ldots & t_2 + c_{M-1} \\
\vdots & \vdots & \vdots & & \vdots \\
t_{L-1} & t_{L-1} + c_1 & t_{L-1} + c_2 & \ldots & t_{L-1} + c_{M-1} \\
\hline
\uparrow & \uparrow & \uparrow & \ldots & \uparrow \\
\mathcal{D}_0 & \mathcal{D}_1 & \mathcal{D}_2 & & \mathcal{D}_{M-1}
\end{array}$$

Note that $\mathcal{D}_i = c_i + \mathcal{D}_0$.

Geometrical Interpretation:



The union of all the $\mathcal{D}_i$ is $\mathbb{F}^n$. Hence every $y \in \mathbb{F}^n$ is in exactly one decoding region.

THE COSET DECODER : HOW TO IMPLEMENT

To find the codeword associated to a channel output y, we could find y in the standard array and read out the entry on top of the same column.

Storing the whole standard array is impractical (often impossible for large codes).

The first column describes the geometry of all decoding regions. We should be able to leverage on that.

THEOREM:

$\vec{y}_1$ AND $\vec{y}_2$ HAVE THE SAME SYNDROME IF AND ONLY IF THEY ARE IN THE SAME COSET.

PROOF:

(1) SAME SYNDROME ⇒ SAME COSET

$$\vec{y}_2 H^T = \vec{y}_2 H^T \Rightarrow (\vec{y}_1 - \vec{y}_2) H^T = \vec{0}$$

$\Rightarrow (\vec{y}_1 - \vec{y}_2)$ IS A CODEWORD.

$\overset{(*)}{\Rightarrow}$ $\vec{y}_1$ AND $\vec{y}_2$ IN SAME COSET.

NOTE FOR THE LAST IMPLICATION (†):

THIS IS BY CONSTRUCTION:

SUPPOSE y_1 IS IN SOME COSET.

THEN, $[y_1] = y_1 + \mathcal{C}$.

SO, y_1 PLUS EVERY CODEWORD.

HENCE, IF $y_2 = y_1 +$ SOME CODEWORD,

THEN $[y_2] = [y_1]$.

(2) SAME COSET ⇒ SAME SYNDROME

$$\left.\begin{aligned}\vec{y}_1 &= \vec{t} + \vec{c}_i \\ \vec{y}_2 &= \vec{t} + \vec{c}_j\end{aligned}\right\}$$ for some $\vec{c}_i, \vec{c}_j \in \mathcal{C}$

$\hookrightarrow \vec{s}_1 = \vec{y}_1 H^T = (\vec{t} + \vec{c}_i) H^T$

$= \vec{t} H^T + \underbrace{\vec{c}_i H^T}_{= \vec{0}} = \vec{t} H^T$

$\vec{S}_2 = \vec{y}_2 H^T = \dots = \vec{t} H^T$

$\square$

Claim: In the standard array, each element of a row has the same syndrome as the coset leader.

Proof:

- the elements of $[t_i]$ have the form $t_i + c$ for some $c \in \mathcal{C}$
- the syndrome of such an element is

$$(t_i + c)H^T = t_iH^T + cH^T = t_iH^T$$

which is the syndrome of the coset leader t_i

□

Claim: The syndrome uniquely identifies the coset leader.

Proof:

Let t_i and t_j be coset leaders.

Suppose that $t_i H^T = t_j H^T$.

Then $(t_i - t_j)H^T = 0$.

Hence $t_i - t_j = c_k \in \mathcal{C}$.

It follows that t_i and t_j are in the same coset.

Since both are coset leaders, $t_i = t_j$.

□

In the previous slide, we have proved that the map

$$\begin{aligned} \mathcal{D}_0 &\to \mathbb{F}^{n-k} \\ t &\mapsto tH^T \end{aligned}$$

is one-to-one.

We use the pigeonhole principle to prove that it is also onto, hence it is a bijection. (We will not use this fact.)

Let $\mathbb{F} = \mathbb{F}_q$.

The standard array places the q^n elements of $\mathbb{F}^n$ into $\mathsf{card}(\mathcal{C}) = q^k$ columns and $\mathsf{card}(\mathcal{D}_0) = \frac{q^n}{q^k} = q^{n-k}$ rows.

The cardinality of $\mathbb{F}^{n-k}$ is also q^{n-k}. □

Hence the **coset decoder** can be implemented as follows:

1. we precompute and store the coset leaders and the corresponding syndrome;
2. to decode y, we compute its syndrom $s = yH^T$;
3. s encodes the row of y;
4. we use the lookup table to determine the corresponding coset leader, say, t_i;
5. t_i and y uniquely determine the column of y, namely $y = t_i + c_j$;
6. hence $c_j = y - t_i$;
7. the decoder declares that the transmitted codeword is $\hat{c} = y - t_i$.

To find the information word $\hat{u}$ that corresponds to $\hat{c}$ we solve the linear system

$$\hat{u}G = \hat{c}.$$

If G is in systematic form, then $\hat{u}$ consists of the first k components of $\hat{c}$.

CHOOSING COSET LEADERS

<u>RULE "MD"</u>: IN EVERY COSET, THE LEADER IS THE MINIMUM-WEIGHT SEQUENCE.

THEOREM: UNDER RULE "MD", THE COSET DECODER IS A MINIMUM DISTANCE DECODER.

PROOF: SUPPOSE $\vec{y}$ IS IN ROW i, COLUMN j OF THE STANDARD ARRAY.

$$\begin{array}{cccccc}
 & j=0 & j=1 & j=2 \; \cdots & j & -- \\
i=0 & \vec{c}_0 & \vec{c}_1 & \vec{c}_2 & \vec{c}_j & \\
i=1 & t_1+\vec{c}_0 & & & \vdots & \\
\vdots & & & & & \\
i & t_i+\vec{c}_0 & ------ & ------ & \vec{y} & = \vec{c}_j+\vec{t}_i
\end{array}$$

$$\begin{array}{cccc} & j=0 & j=1 & j=2 \\ i=0 & \vec{c}_0 & \vec{c}_1 & \vec{c}_2 \\ i=1 & t_1 + \vec{c}_0 & & \\ \vdots & & & \end{array}$$

HENCE : $\vec{y} = \vec{c}_j + \vec{t}_i$

- THE COSET DECODER WILL DECODE: $\vec{y} \longrightarrow \vec{c}_j$

- HENCE $d(\vec{y}, \vec{c}_j) = w(\vec{t}_i)$.

$$
\begin{aligned}
d(\vec{y}, \vec{c}_j) &= w(\vec{y} - \vec{c}_j) \\
&= w(\vec{c}_j + \vec{t}_i - c_j) \\
&= w(\vec{t}_i)
\end{aligned}
$$

NOW CONSIDER

$$\begin{aligned} d(\vec{y}, \vec{c}_k) &= w(\vec{y} - \vec{c}_k) \\ &= w(\vec{c}_j + \vec{t}_i - \vec{c}_k) \\ &= w(\underbrace{(\vec{c}_j - \vec{c}_k) + \vec{t}_i}_{\text{WHICH COSET IS THIS IN?}}) \end{aligned}$$

$$\boxed{\text{NOW CONSIDER}}$$

$$d(\vec{y}, \vec{c}_k) = w(\vec{c}_j + \vec{t}_i - \vec{c}_k)$$

$$= w(\underbrace{\vec{c}_j - \vec{c}_k}_{\in \mathcal{C}} + \vec{t}_i)$$

<u>QUESTION</u>: IN WHICH <u>ROW</u> IS

$$\vec{c}_j - \vec{c}_k + \vec{t}_i \quad ?$$

→ IN ROW i.

NOW CONSIDER

$$d(\vec{y}, \vec{c}_k) = w(\vec{c}_j + \vec{t}_i - \vec{c}_k)$$

$$= w(\underbrace{\underbrace{\vec{c}_j - \vec{c}_k}_{\in \mathcal{C}} + \vec{t}_i}_{\text{IN ROW } i})$$

$$\geq w(\vec{f}_i)$$

SINCE BY RULE "MD", $\vec{t}_i$ IS THE LOWEST-WEIGHT SEQUENCE IN ROW i.

$\square$

Choosing Coset Leaders

There are many ways to choose the coset leaders $t_1, t_2, \cdots$.

In fact, every element of every row of the standard array can be chosen as the coset leader.

Theorem

In every row of the standard array:

- *select the coset leader to be one of the minimum-weight vectors in that row.*

Then the coset decoder is a minimum-distance decoder.

Proof:

Let y be in the ith row and jth column of the standard array, i.e., $y \in [t_i]$ and $y \in \mathcal{D}_j$.

We want to show that $d(y, c_j) \leq d(y, c_k)$ for every k.

On the LHS we have $d(y, c_j) = d(t_i + c_j, c_j) = w(t_i)$.

On the RHS we have $d(y, c_k) = d(t_i + c_j, c_k) = w(t_i + c_j - c_k) = w(t_i + c_l)$ for some l.

t_i and $t_i + c_l$ are in the same coset, and t_i is the coset leader. By choice, $w(t_i) \leq w(t_i + c_l)$. □

Example (Standard array for a Toy-Code)

Let $\mathcal{C} = \{\underbrace{(0,0,0)}_{c_0}, \underbrace{(1,1,1)}_{c_1}\}$

Standard Array:

$$\begin{array}{rr|l}
c_0 = (0,0,0) & c_1 = (1,1,1) & \leftarrow \mathcal{C} \\
t_1 = (0,0,1) & t_1 + c_1 = (1,1,0) & \leftarrow t_1 + \mathcal{C} \\
t_2 = (0,1,0) & t_2 + c_1 = (1,0,1) & \leftarrow t_2 + \mathcal{C} \\
t_3 = \underbrace{(1,0,0)}_{\mathcal{D}_0} & t_3 + c_1 = \underbrace{(0,1,1)}_{\substack{\mathcal{D}_1 = \\ \mathcal{D}_0 + c_1}} & \leftarrow t_3 + \mathcal{C}
\end{array}$$

Example (Decoding for the Toy-Code)

Transposed parity-check matrix and corresponding syndrome lookup table:

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

t	s
$(0,0,0)$	$(0,0)$
$(0,0,1)$	$(0,1)$
$(0,1,0)$	$(1,0)$
$(1,0,0)$	$(1,1)$

If $y = (1,0,1)$ is received, the syndrome is $s = yH^T = (1,0)$, the coset leader is $t = (0,1,0)$, and the decoded codeword is $\hat{c} = y - t = (1,1,1)$.

Example ($(6,3)$ binary linear block code)

The code is defined by the following parity-check matrix

$$H = \left(\begin{array}{ccc|ccc} 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 \end{array}\right).$$

No two columns are linearly dependent. The first three columns are linearly dependent. Hence $d_{min} = 3$.

H has the form (P, I). The generator matrix G has the form $(I, -P^T)$:

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

Example (cont.)

The standard array has $2^k = 2^3 = 8$ columns and $2^{n-k} = 2^3 = 8$ rows.

The top row of the standard array is the code, and the left column consists of the elements of $\mathcal{D}_0$.

The code can correct all the errors of weight 1. Hence all the weight-1 words are in $\mathcal{D}_0$.

This gives:

000000	001110	010101	011011	100011	101101	110110	111000
000001							
000010							
000100							
001000							
010000							
100000							
t_7							

Example (cont.)

The next step is to fill in the rows for which we know the coset leader:

000000	001110	010101	011011	100011	101101	110110	111000
000001	001111	010100	011010	100010	101100	110111	111001
000010	001100	010111	011001	100001	101111	110100	111010
000100	001010	010001	011111	100111	101001	110010	111100
001000	000110	011101	010011	101011	100101	111110	110000
010000	011110	000101	001011	110011	111101	100110	101000
100000	101110	110101	111011	000011	001101	010110	011000
t_7							

As t_7 we choose a weight-2 word that has not yet appeared in any row, i.e., anything except 001100, 001010, 000110, 010100, 010001, 000101, 100010, 100001, 000011, 110000, 101000, 011000.

We choose $t_7 = 100100$.

Example (cont.)

000000	001110	010101	011011	100011	101101	110110	111000
000001	001111	010100	011010	100010	101100	110111	111001
000010	001100	010111	011001	100001	101111	110100	111010
000100	001010	010001	011111	100111	101001	110010	111100
001000	000110	011101	010011	101011	100101	111110	110000
010000	011110	000101	001011	110011	111101	100110	101000
100000	101110	110101	111011	000011	001101	010110	011000
100100	101010	110001	111111	000111	001001	010010	011100

The code will correct all the weight-1 channel-error patterns and the weight-2 channel-error pattern 100100. All the other channel-error patterns will lead to a decoding error.

EXAMPLE (CONT.)

The syndrome lookup table, i.e., the table that associates the coset leader t_i to the syndrome $s_i = t_i H^T$ is:

t_i	s_i
000000	000
000001	001
000010	010
000100	100
001000	110
010000	101
100000	011
100100	111

The code will correct all the weight-1 channel errors and the weight-2 channel error 100100. All the other error patterns will lead to a decoding error.

Example (cont.)

If $y = 011000$ is received, the decoder determines its syndrome

$$s = yH^T = 011,$$

and the corresponding coset leader

$$t = 100000.$$

t_i	s_i
000000	000
000001	001
000010	010
000100	100
001000	110
010000	101
100000	011
100100	111

The decoded word is

$$\hat{c} = y - t = 111000,$$

which is indeed a codeword. □

Disclaimer:

The procedure that we have described requires to store $|\mathbb{F}|^{(n-k)}$ coset leaders and the corresponding syndrome.

While the approach is theoretically appealing, a lookup table of that size is prohibitive for most codes of practical interest.

Example (CD Rom)

The information on a CD Rom is encoded via two codes.

One code has parameters $|\mathbb{F}| = 2^8$, $n = 32$ and $k = 28$.

For it, there are $|\mathbb{F}|^{(n-k)} = \left(2^8\right)^4 = 4.29 \times 10^9$ coset leaders.

A coset leader is $8 \times 32 = 256$ bits long.

A syndrome is $8 \times 4 = 32$ bits long.

This requires more than 10^{12} bits — far more than the capacity of a CD Rom.

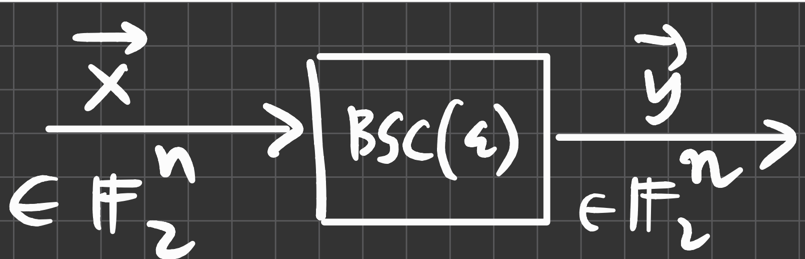
Error Probability

There are many ways to choose $\mathcal{D}_0$.

In fact, every element of every row of the standard array can be chosen as the coset leader.

Next, we learn how to choose the coset leaders so as to minimize the decoding error-probability.

BINARY SYMMETRIC CHANNEL

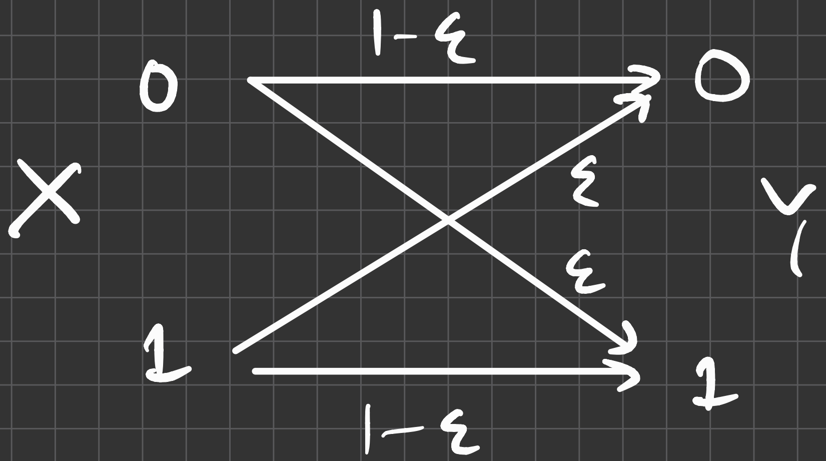


- EVERY BIT IS FLIPPED WITH PROBABILITY ε (AND UNTOUCHED WITH PROBABILITY $1-\varepsilon$), EVERY BIT INDEPENDENTLY.

- $\varepsilon \leq 1/2$

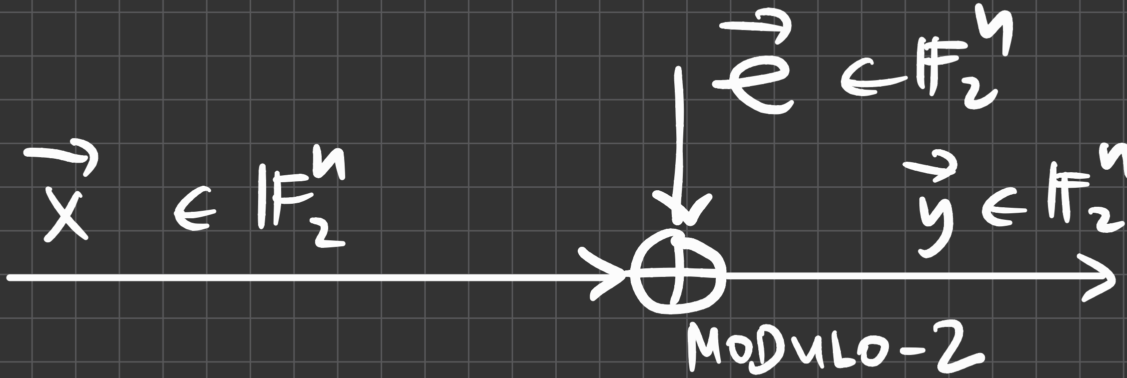
BINARY SYMMETRIC CHANNEL

MOST PEOPLE'S FAVORITE PICTURE:



BINARY SYMMETRIC CHANNEL

YET ANOTHER PICTURE:


$\vec{e} \in \mathbb{F}_2^n$
$\vec{x} \in \mathbb{F}_2^n$
$\vec{y} \in \mathbb{F}_2^n$
$\oplus$
MODULO-2

WHERE $\vec{e}$ IS A STRING OF INDEPENDENT BINARIES WITH

$$p(e_i = 1) = \varepsilon.$$

BINARY SYMMETRIC CHANNEL

WHERE $\vec{e}$ IS A STRING OF INDEPENDENT BINARIES WITH

$$p(e_i = 1) = \varepsilon.$$

HENCE, WHAT IS

$$p(\vec{e} = (0, 1, 0, 0, 1, 0, 1))$$

$$= p(e_1 = 0)\, p(e_2 = 1)\, p(e_3 = 0) \ldots$$

$$= (1-\varepsilon) \quad \varepsilon \quad (1-\varepsilon) \ldots$$

$$= (1-\varepsilon)^{\#\text{ of zeros}} \; \varepsilon^{\#\text{ of ones}}$$

$$= (1-\varepsilon)^{4} \; \varepsilon^{3}$$

$\#$ of ones in sequence $\vec{e}$

$$= w(\vec{e})$$

$\Rightarrow$ $\#$ of zeros — '' —

$$= n - w(\vec{e})$$

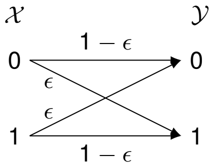
$$\boxed{p(\vec{e}) = (1-\varepsilon)^{n-w(\vec{e})} \; \varepsilon^{w(\vec{e})}}$$

$$= (1-\varepsilon)^n \left(\frac{\varepsilon}{1-\varepsilon}\right)^{w(\vec{e})}$$

IF: $\varepsilon \leq 1/2$

THEN: $\dfrac{\varepsilon}{1-\varepsilon} \leq 1$

Suppose that the channel is the following binary symmetric channel with input alphabet $\mathcal{X} = \{0, 1\}$ and output alphabet $\mathcal{Y} = \{0, 1\}$.



The probability of the error pattern $e \in \mathbb{F}^n$ is

$$\epsilon^{w(e)}(1-\epsilon)^{n-w(e)} = \left(\frac{\epsilon}{1-\epsilon}\right)^{w(e)} (1-\epsilon)^n.$$

We assume $\epsilon < 1/2$. Then $\frac{\epsilon}{1-\epsilon} < 1$, and the above expression is a decreasing function of $w(e)$.

PROBABILITY OF CORRECT DECODING

↳ START BY ASSUMING THAT $\vec{c}_0 = (0, 0, 0 \ldots 0)$ WAS TRANSMITTED.

$$P_c(0) = p(\text{error pattern was "good"})$$

$= p($ error pattern is one from the Column below $\vec{c}_0)$

$$= p(\vec{e} \in D_0)$$

$$= \sum_{\vec{e} \in D_0} p(\vec{e})$$

$$= \sum_{\vec{e} \in D_0} (1-\varepsilon)^n \left(\frac{\varepsilon}{1-\varepsilon}\right)^{w(\vec{e})}$$

Let $P_C(c_i)$ be the probability that the decoder decodes correctly, given that $c_i \in \mathcal{C}$ was transmitted.

When $c_0 \in \mathcal{C}$ is transmitted, the decoder makes the correct decision whenever $y \in \mathcal{D}_0$. But when c_0 is transmitted, the event $y \in \mathcal{D}_0$ is the same as the event $e \in \mathcal{D}_0$. Hence,

$$P_C(0) = \sum_{e \in \mathcal{D}_0} \epsilon^{w(e)}(1-\epsilon)^{n-w(e)} = \sum_{j=0}^{L-1} \epsilon^{w(t_j)}(1-\epsilon)^{n-w(t_j)},$$

where we have defined $t_0 = c_0 = 0$.

In conclusion, we maximize $P_C(0)$ if the coset leaders have the smallest possible weights.

EXAMPLE (CONT.)

The probability of error P_E is $1 - P_C$, where P_C is the probability that the error pattern e is in $\mathcal{D}_0$.

For the binary symmetric channel, this probability is

$$P_C = (1-\epsilon)^6 + 6(1-\epsilon)^5\epsilon + (1-\epsilon)^4\epsilon^2.$$

t_i
000000
000001
000010
000100
001000
010000
100000
100100

For $\epsilon = 0.1$, $P_C = 0.8923$.

(The probability that the channel output equals the channel input is $(1-\epsilon)^6$. This is 0.531 when $\epsilon = 0.1$.)

NOW, WHAT IF $\vec{c}_i$ IS THE TRANSMITTED CODEWORD?

$$P_C(i) = \mathbb{P}(\vec{y} \in D_i)$$

$$= \mathbb{P}(\vec{c}_i + \vec{e} \in D_i)$$

$$= \mathbb{P}(\vec{c}_i + \vec{e} \in c_i + D_0)$$

$$= \mathbb{P}(\vec{e} \in D_0)$$

We have assumed for simplicity a binary symmetric channel, but the same idea applies to nonbinary channels (with input and output alphabet $\mathbb{F}$) for which the probability of an error pattern $e \in \mathbb{F}^n$ is decreasing with $w(e)$.

$P_C(c_i)$ is the probability that $y \in \mathcal{D}_i$ when c_i is transmitted.

This is the probability that $\underbrace{c_i + e}_{y} \in \underbrace{c_i + \mathcal{D}_0}_{\mathcal{D}_i}$.

It is the same as the probability that $e \in \mathcal{D}_0$. But this is $P_C(0)$.

Hence, for all i, $P_C(c_i) = P_C(0)$.

Hence, the unconditional probability of correct decoding is $P_C = P_C(0)$.

We conclude that the error probability is minimized when the coset leaders have the smallest weight.

This can be achieved by reordering each row of the standard array as follows.

Suppose that, in row j we have

$$w(t_j) > w(t_j + c_i).$$

By making $t_j + c_i$ the new coset leader of the jth row, the following happens:

1. the elements of that row are permuted;
2. the term of $P_C(0)$ that corresponds to the jth row increases whereas the other terms of $P_C(0)$ are unaffected.

To Summarize:

- A linear code $\mathcal{C}$ is a subspace of a vector space $\mathbb{F}^n$.
- The code partitions $\mathbb{F}^n$ into cosets.
- We cannot chose the cosets, but we can choose the coset leaders.
- To maximize $P_C(0)$, we let the leader of each coset be one of the smallest-weight elements of the coset.
- The cost leaders form $\mathcal{D}_0$, and $\mathcal{D}_i = \boldsymbol{c}_i + \mathcal{D}_0$.
- The error probability is the same and equal to $P_C(0)$, no matter which codeword is transmitted.