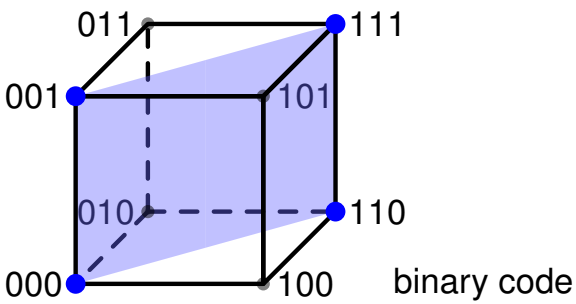
Weeks 12&13: Vector Spaces and Linear Codes (Textbook Chapter 13)

Prof. Michael Gastpar

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

EPFL

Spring Semester 2025



Outline

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

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

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

IF $\vec{v}_1, \vec{v}_2, \vec{v}_3$ LIN. INDEP., THEN $\dim(S) = 3$

$= \{ \vec{s} : s_1 h_{11} + s_2 h_{12} + \dots + s_n h_{1n} = 0$

$\vdots$

$s_1 h_{m1} + s_2 h_{m2} + \dots + s_n h_{mn} = 0 \}$

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

- EXCEPT THAT THE # OF
  POINTS IS FINITE:

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

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 SUBSPACE = 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 SUBSPACE ?

2) FULL CHECK: FIND BASIS!

$k=3 \Rightarrow$ 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 SUBSPACE ?

3) FIND A SET OF EQUATIONS !

$m + k = n$

$k = 3$, $n = 7$

$\Rightarrow$ 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?

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

We have

$$\vec{c}_4 = \vec{c}_1 + \vec{c}_2$$
$$\vec{c}_5 = \vec{c}_1 + \vec{c}_3$$
$$\vec{c}_6 = \vec{c}_2 + \vec{c}_3$$
$$\vec{c}_7 = \vec{c}_1 + \vec{c}_2 + \vec{c}_3$$

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

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 $\dim(\mathcal{C}) = 3$.

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

We have seen that a k-dimensional subspace of $\mathbb{F}^n$ has cardinality $[\mathsf{card}(\mathbb{F})]^k$. (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:

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

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

OBS: $d(u_i, v_i) = w(u_i - v_i)$

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:

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

$$= \min_{\substack{\vec{u}, \vec{v} \in \mathcal{C}, \vec{u} \neq \vec{v}}} w(\underbrace{\vec{u} - \vec{v}}_{\in \mathcal{C} \text{ BY LINEARITY}})$$

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

PROOF PART 2 :

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

PROOF:

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

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

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

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:

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

$$= \min_{\vec{u},\vec{v}\in\mathcal{C};\vec{u}\neq\vec{v}} w(\vec{u}-\vec{v})$$

$$\geq \min_{\vec{c}\in\mathcal{C};\vec{c}\neq\vec{0}} w(\vec{c}) \qquad \text{(reason: } \mathcal{C} \text{ is a vector space, so } \vec{u}-\vec{v}\in\mathcal{C}\text{).}$$

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

$$\geq \min_{\vec{c},\vec{v}\in\mathcal{C};\vec{c}\neq\vec{v}} d(\vec{c},\vec{v}) \qquad \text{(reason: we have equality with } \vec{v}=0\text{)}$$

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

□

EXERCISE

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

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

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

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

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

SOLUTION

- Compute the weight of each non-zero codeword:

$$w_1 = w(0011100) = 3$$
$$w_2 = w(0111011) = 5$$
$$w_3 = w(1110100) = 4$$
$$w_4 = w(0100111) = 4$$
$$w_5 = w(1101000) = 3$$
$$w_6 = w(1001111) = 5$$
$$w_7 = w(1010011) = 4$$

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

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

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

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

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

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

$\dim(C) = n - 1 = k$

$d_{min}(C) = 2$

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.,

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

(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

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

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

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

$\Rightarrow$ MDS code (!)

$\Rightarrow$ HOW MANY EQUATIONS? $m = n-1$.

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

REPETITION CODE

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

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

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

$\Rightarrow$ 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).

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

$\downarrow$

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

NOT LINEAR BUT IT'S MDS.

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

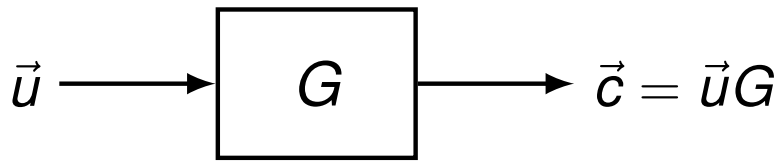
$(\vec{c}_1, \vec{c}_2, \vec{c}_3)$ form a basis for $\mathcal{C}$. Therefore

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

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

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

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:

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

Exercise

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

Solution

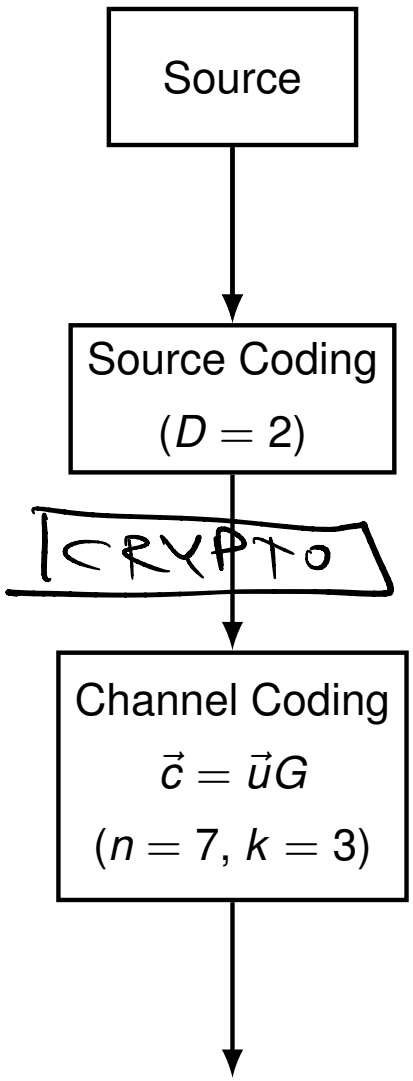
It is the number of lists that form a basis. A q-ary linear code of dimension k has q^k codewords and the number of bases is

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

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

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

$$\begin{cases} 000 \\ 011 \\ 111 \\ 100 \\ 101 \end{cases} \xrightarrow{G} \begin{matrix} 0000000 \\ 1001111 \\ 1010011 \\ 0011100 \\ 1101000 \end{matrix}$$

0000000100111110100110011100 1101000

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 $u_1 u_2 u_3 \to c_1 c_2 c_3 c_4 c_5 c_6 c_7$.

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:

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

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

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

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:

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

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

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

EXAMPLE:

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

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

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

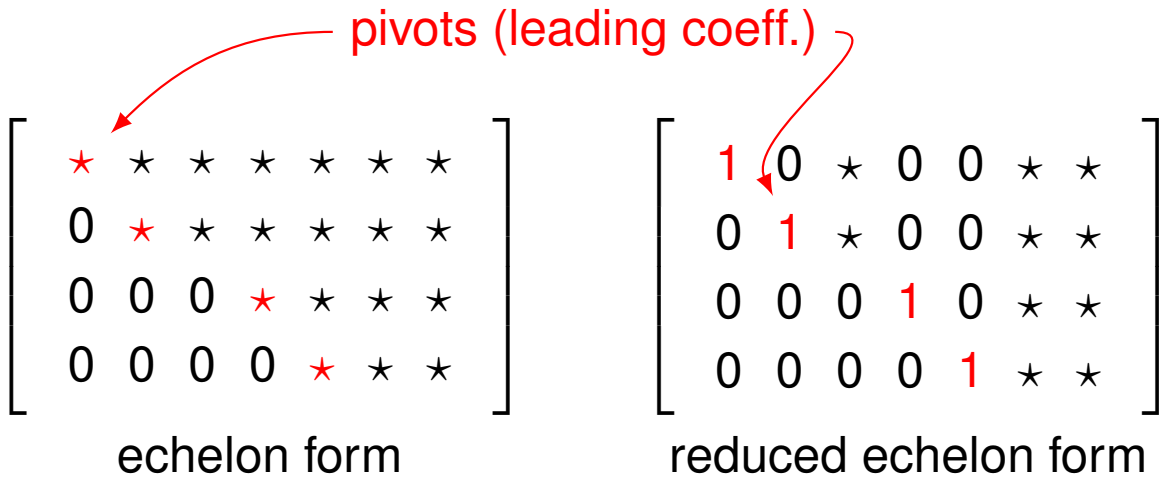
$$G^{(3)} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 \end{pmatrix}$$

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

1. Find a basis $\{\vec{c}_1, \ldots, \vec{c}_k\}$ 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.



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 $\mathbb{F}_5$:

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

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

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

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

$$c_1 = u_1$$
$$c_2 = u_2$$
$$c_3 = u_3$$
$$c_4 = 3u_1 + 3u_2$$
$$c_5 = 2u_1 + 4u_2$$

YESTERDAY

- DEF: (n, k) LINEAR BLOCK CODE

  $\Leftrightarrow$

  k-DIM SUBSPACE OF $\mathbb{F}^n$

- HAS EXACTLY $\left(\text{CARD}(\mathbb{F})\right)^k$ CODEWORDS

- d_{min} = minimum weight of any non-zero codeword

- GENERATOR MATRIX G : EVERY CODEWORD CAN BE WRITTEN AS

$$\vec{c} = \vec{u} G$$

$$\uparrow \in \mathbb{F}^k$$

- ALSO GIVES ENCODING MAP
- PARTICULARLY 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: $\mathbb{F}_3^3$ $C = \{$

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

$(0\ 0\ 0)$,
$(1\ 1\ 2)$,
$\vdots$
$\}$

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

OBS: NOT A SYSTEMATIC CODE!

OBS 2: BUT THERE EXISTS A CODE THAT IS:

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

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

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

$C =$
$\{(000),$
(010)
$(020),$
$(101),$
$(111),$
$\ldots\}_3$

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

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

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:

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

$$= \{ \vec{y} : y_1 H_{11} + y_2 H_{12} + \cdots + y_n H_{1n} = 0$$

$$\vdots$$

$$y_1 H_{m,1} + y_2 H_{m,2} + \cdots + y_n H_{m,n} = 0 \}$$

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

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

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

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

is a parity-check matrix of the same code.

PROOF:

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

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

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

FOR ALL $\vec{u} \in \mathbb{F}^k$

$\Updownarrow$

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

FOR THE CASE OF SYSTEMATIC:

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

$$= -P + P = 0. \checkmark$$

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:

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

$n = 5$

$k = 3$

$\hookrightarrow n - k = 2$

$$H = \begin{matrix} \\ 2 \end{matrix} \overbrace{\begin{pmatrix} -3 & -3 & 0 & 1 & 0 \\ -2 & -4 & 0 & 0 & 1 \end{pmatrix}}^{5}$$

BINARY CODE OF ALL SEQUENCES WITH EVEN # OF ONES.

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

$n,\ k = n-1$

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

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 $\mathbb{F}_3^3$

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

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

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

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

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

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,

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

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