Math 5251 Minimum distance & linear codes (Chap. 12)

Recall Shannon's Noisy Coding Theorem said we could find q-ary codes $\mathcal{C} \subset \Sigma^*$ (so $q = |\Sigma|$) whose words all have the same length n (called the block length of $\mathcal{C}$) having high q-ary rate $\frac{\log_q(m)}{n}$ where $m = |\mathcal{C}|$

and probability of error in decoding $\to 0$ as $n \to \infty$, by picking the m code words of $\mathcal{C}$ randomly.

This makes it tough to do minimum distance decoding, that is, decode a received $y = (y_1, \ldots, y_n)$ as $x = (x_1, \ldots, x_n)$ where x is any word in $\mathcal{C}$ that minimizes the Hamming distance $d(x,y) := |\{i : x_i \neq y_i\}|$.

Our ability to detect/correct errors this way is controlled by ...

DEF'N: The minimum distance of code $\mathcal{C}$ is

$$d = d(\mathcal{C}) := \min\left\{ d(x,y) : \begin{matrix} x,y \in \mathcal{C} \\ x \neq y \end{matrix} \right\}$$

Call $\mathcal{C}$ an (n,m,d) q-ary code

if n = blocklength of its words

$m = |\mathcal{C}|$

$d = d(\mathcal{C})$

$q = |\Sigma|$

PROPOSITION: An (n,m,d) q-ary code can

(i) detect up to $d-1$ errors

(ii) correct up to $\left\lfloor \frac{d-1}{2} \right\rfloor$ errors via min. distance decoding

greatest integer $\leq \frac{d-1}{2}$

It will be easy to prove, but first let's see...

EXAMPLES

(1) $\mathcal{C} = \{000, 110, 101, 101\}$

$\subset (\mathbb{F}_2)^3$ = words of length 3 using $\Sigma = \mathbb{F}_2 = \{0,1\}$

is a $(\underset{\substack{\| \\ \text{length}}}{\overset{\|}{3}}{}_{n},\ \underset{\substack{\| \\ |\mathcal{C}|}}{\overset{\|}{4}}{}_{m},\ \underset{\substack{\| \\ d(\mathcal{C})}}{\overset{\|}{2}}{}_{d})$ 2-ary code (binary)

that can detect 1 $(=d-1)$ bit errors, but cannot correct any errors at all (Why?)

(and $0 = \lfloor \frac{d-1}{2} \rfloor = \lfloor \frac{1}{2} \rfloor$)

Note that it is a parity check code (Why?)

---

(2) This 3-fold repetition code

$\mathcal{C}_3 = \{000, 111, 222, 333, 444\} \subset (\mathbb{F}_5)^3$

is a $(3, 5, 3)$ 5-ary code that can detect up to $\underset{=d-1}{2}$ errors, correct $\underset{=\lfloor \frac{d-1}{2} \rfloor}{1}$ error. (Why?)

The 4-fold version of the repetition code

$\mathcal{C}_4 = \{0000, 1111, 2222, 3333, 4444\} \subset (\mathbb{F}_5)^4$

is $(4, 5, 4)$ 5-ary, detecting up to $\underset{=d-1}{3}$ errors

still correcting only $\underset{=\lfloor\frac{d-1}{2}\rfloor}{1}$ error

The 7-fold version

$\mathcal{C}_7 = \{0000000, 1111111, 2222222, 3333333, 4444444\} \subset (\mathbb{F}_5)^7$

is $(7, 5, 7)$ 5-ary, detecting up to $\underset{=d-1}{6}$ errors

correcting up to $\underset{=\lfloor\frac{d-1}{2}\rfloor}{3}$ errors.

---

proof of PROPOSITION: Assume $d(\mathcal{C}) = d$.

Then any sent word $x \in \mathcal{C}$ corrupted by noise to a received word y with $\leq d-1$ letters different will have $d(x,y) \leq d-1 < d(\mathcal{C})$, so $y \notin \mathcal{C}$, and recipient will detect this.

If the received y has $\leq \lfloor\frac{d-1}{2}\rfloor$ letters different from the sent x, then x is the unique word in $\mathbb{C}$ with $d(x,y) \leq \lfloor \frac{d-1}{2} \rfloor$, else $\exists\, x' \in \mathbb{C}$ with $x' \neq x$ and $d(x',y) \leq \lfloor \frac{d-1}{2} \rfloor$,

$$\text{so} \quad d(x,x') \leq d(x,y) + \underbrace{d(y,x')}_{\shortparallel}$$

$$= d(x,y) + \overbrace{d(x',y)}$$

symmetry of HAMMING DISTANCE

TRIANGLE INEQUALITY holds for Hamming distance (why?)

$$\leq \lfloor \tfrac{d-1}{2} \rfloor + \lfloor \tfrac{d-1}{2} \rfloor$$

$$\leq \frac{d-1}{2} + \frac{d-1}{2} \leq d-1 < d$$

contradiction $d = d(\mathbb{C})$ ▨

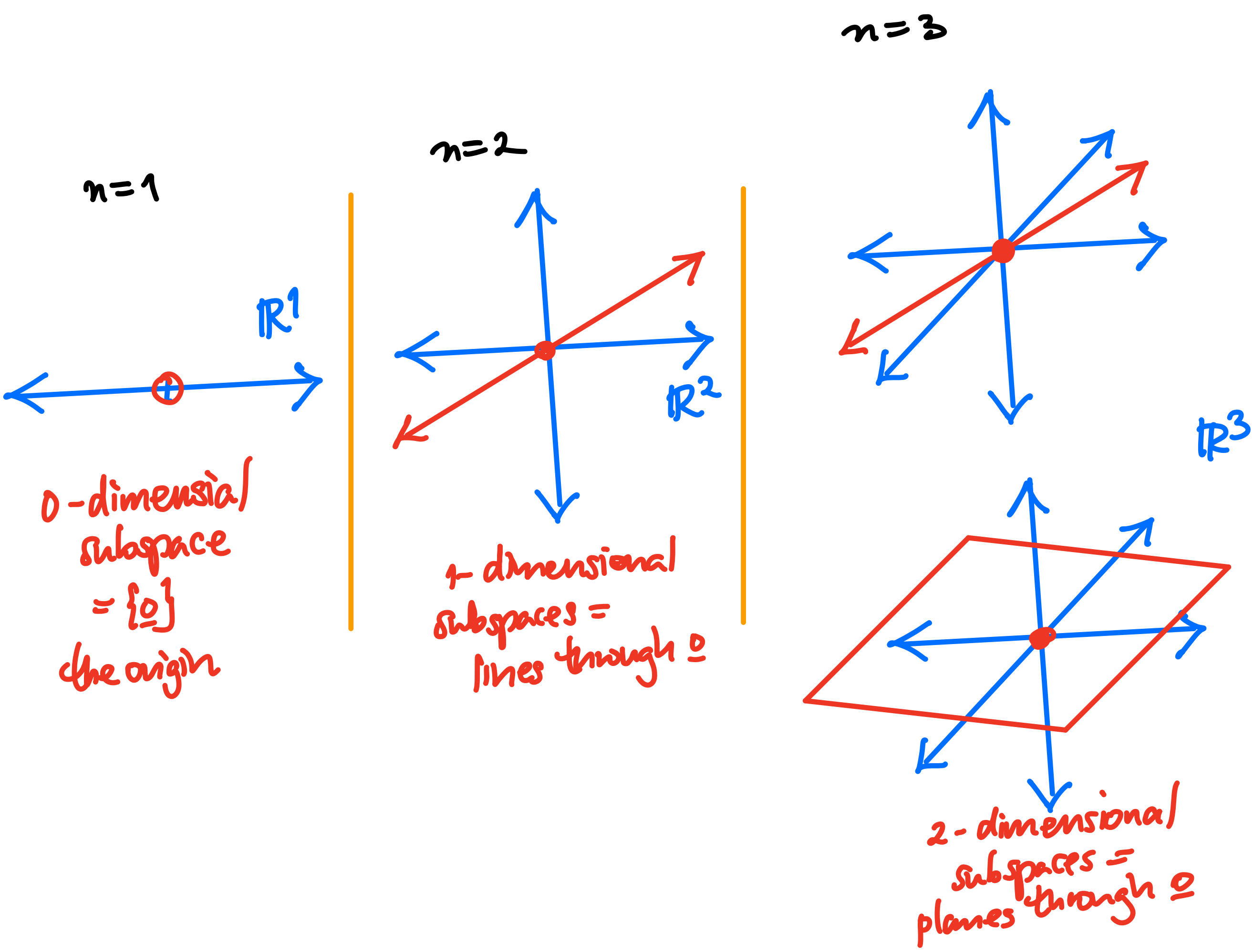
Linear Codes

Computing $d(\mathbb{C})$ and doing min. distance decoding turn out to be much easier when we pick $\mathbb{C} \subset (\mathbb{F}_q)^n$ where $\mathbb{F}_q$ is a field with q elements and $\mathbb{C}$ is a k-dimensional linear subspace inside $(\mathbb{F}_q)^n$.

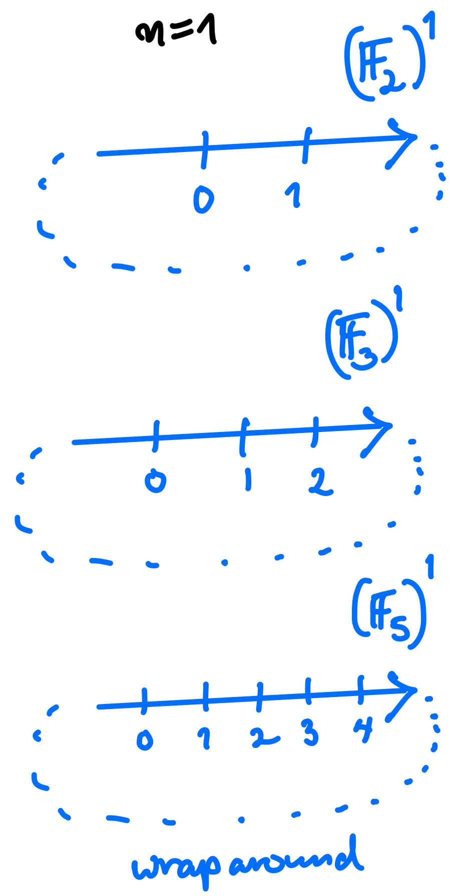
NOTATION: Such a k-dimensional subspace $\mathcal{C} \subset (\mathbb{F}_q)^n$ is called an $[n, k, d]$ $\mathbb{F}_q$-linear code if $d = d(\mathcal{C})$, (and it will turn out that $m = |\mathcal{C}| = q^k$, so it is an (n, q^k, d) q-ary code in the previous notation).

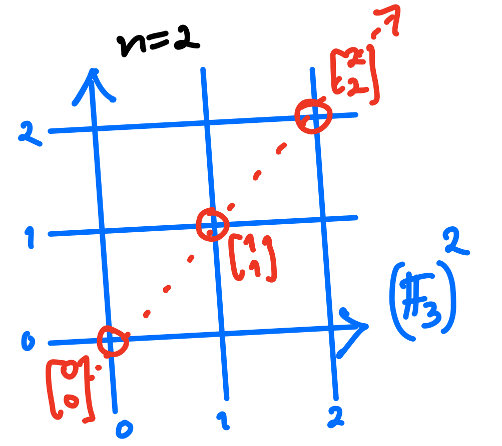
What does this mean ?!

Recall linear subspaces $\mathcal{C} \subset \mathbb{R}^n$ for small n:



We can similarly try to visualize linear subspaces in $(\mathbb{F}_q)^n$ for small n:



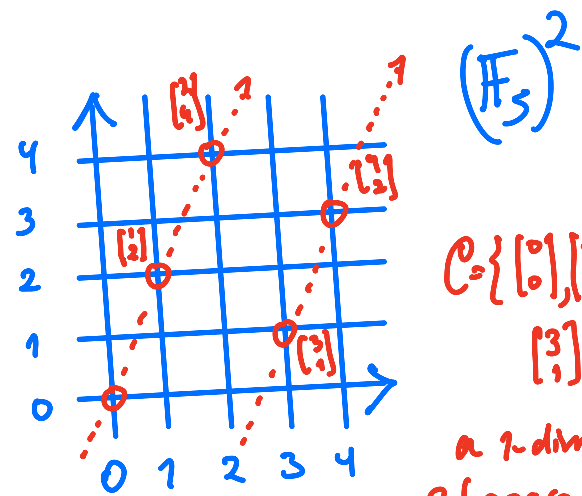


$\mathcal{C} = \left\{ \begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}1\\1\end{bmatrix}, \begin{bmatrix}2\\2\end{bmatrix} \right\}$

= 3-ary 2-fold repetition code

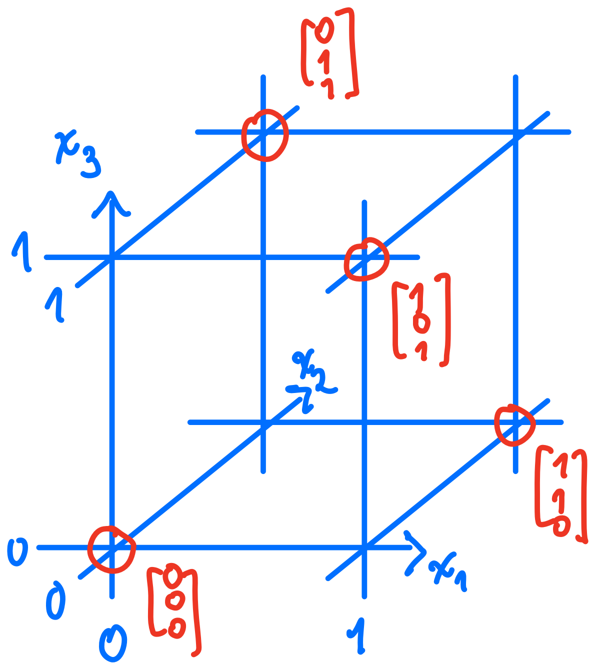
is a 1-dimensional subspace,

a line through $\underline{0}$ in $(\mathbb{F}_3)^2$



$\mathcal{C} = \left\{ \begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}1\\2\end{bmatrix}, \begin{bmatrix}2\\4\end{bmatrix}, \begin{bmatrix}3\\1\end{bmatrix}, \begin{bmatrix}4\\3\end{bmatrix} \right\}$

a 1-dimensional subspace in $(\mathbb{F}_5)^2$



$$C = \left\{ \begin{bmatrix}0\\0\\0\end{bmatrix}, \begin{bmatrix}1\\1\\0\end{bmatrix}, \begin{bmatrix}1\\0\\1\end{bmatrix}, \begin{bmatrix}0\\1\\1\end{bmatrix} \right\}$$

our 2-ary parity check code is a 2-dimensional subspace of $(\mathbb{F}_2)^3$.

It's a plane through $\underline{0}$ with equation $x_1+x_2+x_3=0$ in $(\mathbb{F}_2)^3$

Review of linear algebra and vector spaces over a field (§§ 12.5, 12.6, 12.7, A.1, A.2)

DEF'N: A vector space V over a field $\mathbb{F}$ ← scalars

is a set V with 2 operations with
"vectors"

operations vector addition $+$

$$V \times V \longrightarrow V$$
$$(v, w) \longmapsto v+w$$

and scalar multiplication

$$\mathbb{F} \times V \longrightarrow V$$
$$(c, v) \longmapsto cv$$

satisfying some reasonable axioms that we're used to from $V = \mathbb{R}^n$ and $\mathbb{F} = \mathbb{R}$

e.g. $+$ is
- commutative — $v + w = w + v$
- associative — $(u+v) + w = u + (v+w)$
- has an identity zero vector $\underline{0}$ — $\underline{0} + v = v$
- has inverses $-v$ — $(-v) + v = \underline{0}$

Scalar mult. and $+$ distribute over each other
- $c(v+w) = cv + cw$
- $(c + c')v = cv + c'v$

Lastly, $1 \in \mathbb{F}$ has $1 \cdot v = v$ $\forall v \in V$

A (linear) subspace $W \subseteq V$ is just a nonempty subset closed under addition, $v \mapsto -v$, and scalar mult. (so W is itself an $\mathbb{F}$-vector space)

An $\mathbb{F}_q$-linear code is just a subspace $\mathcal{C} \subseteq (\mathbb{F}_q)^n$ where $(\mathbb{F}_q)^n = \left\{ \text{column vectors } \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} : x_i \in \mathbb{F}_q \right\}$

a finite field with q elements

with usual $+$ and scalar mult:

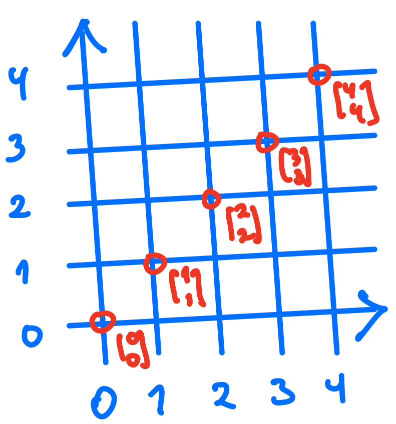
$$\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_1+y_1 \\ \vdots \\ x_n+y_n \end{bmatrix}$$

$$c\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} cx_1 \\ \vdots \\ cx_n \end{bmatrix}$$

EXAMPLES

(1) For p a prime,
the p-ary n-fold repetition code $\mathcal{C} \subset (\mathbb{F}_p)^n$
is the line through $\underline{0}$ consisting of all
$\mathbb{F}_p$- scalar multiples of $\begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$, that is

$$\mathcal{C} = \left\{ \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 2 \\ \vdots \\ 2 \end{bmatrix}, \dots, \begin{bmatrix} p-1 \\ p-1 \\ \vdots \\ p-1 \end{bmatrix} \right\} = \left\{ c\begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} : c \in \mathbb{F}_p \right\}$$



$\mathcal{C}$

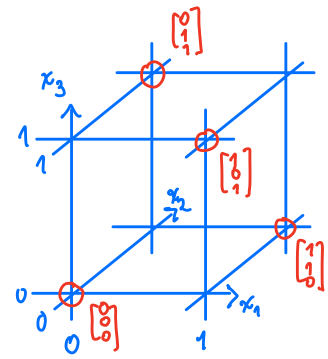
for $p=5$
$n=2$

(2) The binary (2-ary) single parity check code of length n is

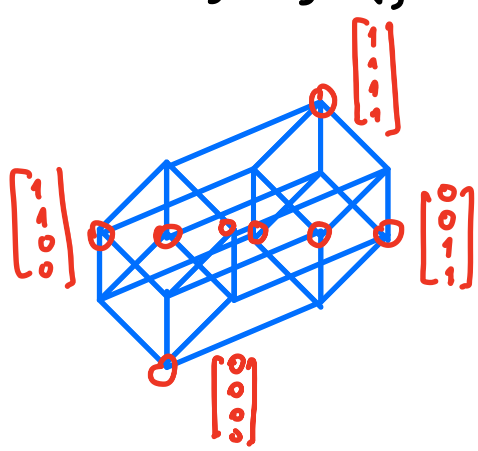
$$\mathcal{C} = \left\{ \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} : x_i \in \mathbb{F}_2,\ x_1 + x_2 + \ldots + x_n = 0 \text{ in } \underset{\{0,1\}}{\underset{\parallel}{\mathbb{F}_2}} \right\} \subset (\mathbb{F}_2)^n$$

i.e. evenly many x_i are 1's

$n=3$: $\mathcal{C} = \left\{ \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \right\}$



$n=4$: $\mathcal{C} = \left\{ \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \right\}$



Q: Why is the single parity check code $\mathcal{C} \subset \mathbb{F}_2^n$ always a subspace of $\mathbb{F}_2^n$?

Spanning, linear independence, bases, dimension

DEF'N: For a subspace $W \subset V$ a vector space over $\mathbb{F}$, say $w_1, \dots, w_m \in W$ **span** W if every $w \in W$ can be written $w = c_1 w_1 + \dots + c_m w_m = \sum_{i=1}^{m} c_i w_i$ for some $c_i \in \mathbb{F}$

EXAMPLES

(1) The n-fold p-ary **repetition code** $\mathcal{C} \subset (\mathbb{F}_p)^n$ is spanned by $\left.\begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}\right\}$ n times

(or by any $\begin{bmatrix} c \\ c \\ \vdots \\ c \end{bmatrix}$ with $c \in \mathbb{F}_p^\times = \mathbb{F}_p - \{0\}$; Why?)

(2) The single **parity check code** of length 3

$$\mathcal{C} = \left\{ \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \right\} \subset (\mathbb{F}_2)^3$$

is spanned by $\left\{ \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \right\}$

since $\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} = 0 \cdot \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + 0 \cdot \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$

$\begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} = 1 \cdot \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + 1 \cdot \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$

DEF'N: A generator matrix G for a linear code $\mathcal{C}$ is any matrix whose rows span $\mathcal{C}$,

row vectors

that is $\mathcal{C}$ is the row space of G.

EXAMPLES

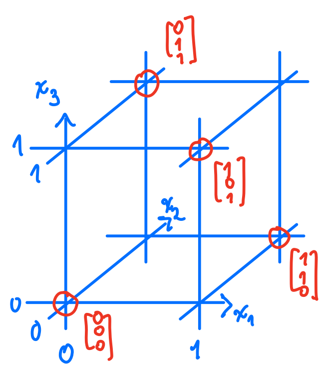
(1) p-ary n-fold repetition code $\mathcal{C} \subset (\mathbb{F}_p)^n$
has generator matrix

$$G = \underbrace{[1\ 1\ \cdots\ 1]}_{n \text{ entries}}$$

---

(2) parity check code $\mathcal{C} = \left\{ \begin{bmatrix}0\\0\\0\end{bmatrix}, \begin{bmatrix}1\\1\\0\end{bmatrix}, \begin{bmatrix}1\\0\\1\end{bmatrix}, \begin{bmatrix}0\\1\\1\end{bmatrix} \right\} \subset (\mathbb{F}_2)^3$
has generator matrix

$$G = \begin{bmatrix}1 & 1 & 0\\ 1 & 0 & 1\end{bmatrix} \quad \left(\text{or } G = \begin{bmatrix}1 & 1 & 0\\ 0 & 1 & 1\end{bmatrix} \text{ or } G = \begin{bmatrix}1 & 0 & 1\\ 0 & 1 & 1\end{bmatrix}\right)$$



DEF'N: Say $v_1, \ldots, v_m \in V$ a vector space over $\mathbb{F}$ are linearly dependent if $\exists\, c_1, \ldots, c_m \in \mathbb{F}$ not all 0 with $c_1 v_1 + \ldots + c_m v_m = \underline{0}$.

Otherwise, if $c_1 v_1 + \ldots + c_m v_m = \underline{0}$ forces $c_1 = \ldots = c_m = 0$, say $v_1, \ldots, v_m$ are linearly independent.

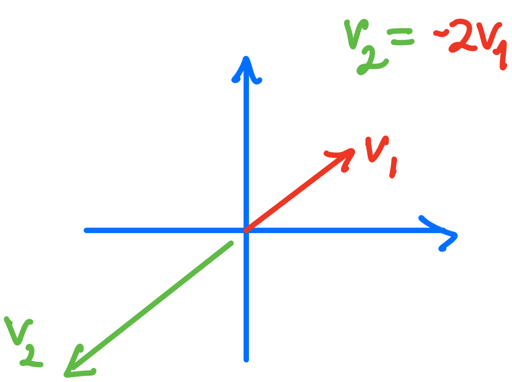
EXAMPLES

$\underline{0}$ is always lin. dependent

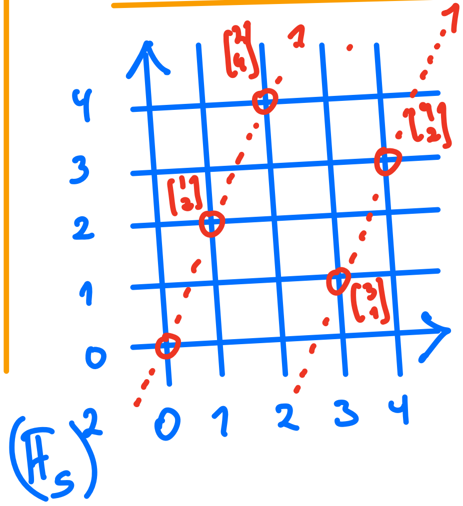
$v \neq \underline{0}$ is always lin. indep.

v_1, v_2 are lin. dependent $\iff$ $v_2 = c v_1$ for some $c \in \mathbb{F}$ or $v_1 = c v_2$

i.e. they lie on a common line through $\underline{0}$



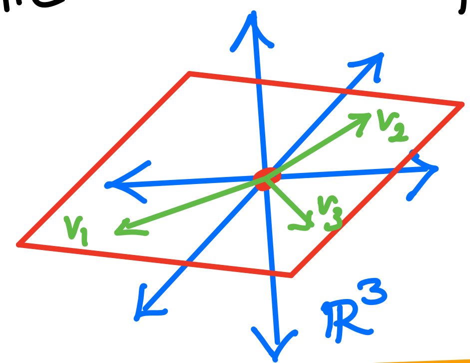
v_1, v_2 are dependent



$\begin{bmatrix}3\\1\end{bmatrix} = 3\begin{bmatrix}1\\2\end{bmatrix}$ so $\begin{bmatrix}3\\1\end{bmatrix}, \begin{bmatrix}1\\2\end{bmatrix}$ are dependent in $(\mathbb{F}_5)^2$

v_1, v_2, v_3 are lin. dependent $\iff$

they lie on a common plane through $\underline{0}$



v_1, v_2, v_3 dependent

DEF'N: Say $w_1, \ldots, w_k$ are a **basis** for $W \subset V$ if they are **lin. indep.** and **span** W.

EXAMPLES

(1) $\begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$ is a basis for the p-ary repetition code $C \subset (\mathbb{F}_p)^n$

(2) The parity check code $\mathcal{C} = \left\{ \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \right\} \subset (\mathbb{F}_2)^3$

has bases $\left\{ \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \right\}$, $\left\{ \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \right\}$, $\left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \right\}$

SOME LINEAR ALGEBRA FACTS

(familiar when $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$, but work over all fields $\mathbb{F}$)

- Every lin. indep. set in W is contained in a basis for W.
- Every spanning set for W contains a basis for W.
- Every basis $v_1, \ldots, v_n$ for W has the same size n called the dimension $n = \dim_{\mathbb{F}}(W) = \dim(W)$
- $\dim_{\mathbb{F}}(\mathbb{F}^n) = n$ since $e_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, e_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \ldots, e_n = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}$ are a basis for it
- W_1 a subspace of $W_2 \Rightarrow \dim(W_1) \leq \dim(W_2)$

EXAMPLE $\mathbb{F}^2$ only has subspaces of dimensions $0, 1, 2$