I. First examples

Example (French social security number)

A French social security number has the following format: $s\ yy\ mm\ dd\ iii\ oooo\ kk$, where:
- s: 1 for male, 2 for female
- yy: year of birth
- mm: month of birth
- dd: department of birth
- iii and ooo: Insee number and registering order
- kk: a security key to be able to identify errors in the values above.

Example (Repetition encoding)

$b \in \mathbb{F}_2 \mapsto (b, ..., b) \in \mathbb{F}_2^n$:
- detects any error pattern of $< n$ errors.
- corrects up to $\lfloor \frac{n-1}{2} \rfloor$ errors by majority voting.
- but not efficient because we transmit many bits.

Example (Parity encoding)

$(b_1, ..., b_{n-1}) \mapsto \left(b_1, ..., b_{n-1}, \sum_{i=1}^{n-1} b_i\right)$:
- detects only one error.
- does not correct.

II. Error correcting codes

II.1. Definitions

Definition (Linear code)

A *linear code* is a subspace $\mathcal{C} \subseteq \mathbb{F}_2^n$.

Remarks:
- In the next lectures, $\mathbb{F}_2$ might be replaced by $\mathbb{F}_q$ $(q > 2)$.
- Anne C. will use a bit *non-linear codes* (*i.e.* $\mathcal{C}$ is an arbitrary subset of $\mathbb{F}_2^n$).

II.2. Parameters

Definition (Hamming distance)

The *Hamming distance* between $x, y \in \mathbb{F}_2^n$ is $d_H(x, y) = |\{i \mid x_i \neq y_i\}|$.

The *Hamming weight* of $x \in \mathbb{F}_2^n$ is $w_H(x) = d_H(x, 0)$.

A code $\mathcal{C} \in \mathbb{F}_2^n$ is associated to 3 fundamental parameters:
- its length n
- its dimension $k = \dim_{\mathbb{F}_2}(\mathcal{C}) = \log_2|\ \mathcal{C}\ |$ (for non-linear codes)
- its minimal distance $d = d_{\min}\mathcal{C} = \min_{\substack{x,y\in\mathcal{C} \\ x\neq y}} \{d_H(x, y)\}$

Equivalently, if $\mathcal{C}$ is linear, $d = d_{\min}(\mathcal{C}) = \min\limits_{\substack{x \in \mathcal{C} \\ x \neq 0}} \{w_H(x)\}$.

Example (Repetition code)

$\{(0...0), (1...1)\} \subseteq \mathbb{F}_2^n$ with parameters:
- n
- $k = 1$
- $d = n$

Example (Parity code)

$\{c \in \mathbb{F}_2^n \mid w_H(c) \text{ is even}\}$ with parameters:
- n
- $k = n - 1$
- $d = 2$

Exercise. Show that this is a linear space.
Let $x, y \in \mathcal{C}$, we want to prove that $w_H(x + y)$ is even too, *i.e.* $x + y$ has an even number of 1's.
x and y both have an even number of 1's because they belong in $\mathcal{C}$.
- We can remove 1's where x and y agree (all indexes i such that $x_i = y_i = 1$), because they lead to 0's. We're left with p indexes in x that will add up to a 0 in y leading to a 1, and $p + 2k (k \in \mathbb{Z})$ indexes from y in a similar fashion. Thus, there are $2p + 2k$ 1's in $x + y$.

Intuitively $\frac{k}{n}$ **is a measure of efficiency** and $\frac{d}{n}$ **of ability to correct**.

Notations.
- We usually denote parameters of $\mathcal{C} \subseteq \mathbb{F}_2^n$ as $[n, k, d]_q$ or $[n, k]_q$ if d is unknown.
- We denote:
  - $R := \frac{k}{n}$ the *rate of the code*
  - $\delta := \frac{d}{n}$ the *relative distance*

There is a tradeoff between R and δ.

Having a δ close to 1 is a good criterion to indicate that we might be able to correct, but it is not sufficient by itself.

II.3. How to represent a linear code?

II.3.1. Using generator matrices

Definition (Generator matrix)

A *generator matrix* $G \in \mathbb{F}_2^{l \times n}$ is a matrix whose rows span $\mathcal{C}$ as a vector space ($l \geq k$), *i.e.* $\mathcal{C} = \left\{mG \mid m \in \mathbb{F}_2^l\right\}$.

Remark. $\substack{\mathbb{F}_2^l \to \mathbb{F}_2^n \\ m \mapsto mG}$ is an encoding map (take $l = k$).

Note that in coding theory, vector are rows.

II.3.2. Parity-check matrices

Definition (Parity-check matrix)

A *parity-check matrix* (*p.c.m.*) $H \in \mathbb{F}_2^{l \times n}$ $(l \leq n - k)$ is a matrix whose right kernel is $\mathcal{C}$, *i.e.* $\mathcal{C} = \{y \in \mathbb{F}_2^n \mid Hy^T = 0\}$.

II.3.3. Examples of such matrices

Example (Repetition code)

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

Example (Parity code)

- G: take H above.
- H: take G above.

There will be a lecture on this duality.

III. The Hamming code

III.1. Further properties of the minimal distance

Prop (Disjoint balls)

Let $\mathcal{C} \subseteq \mathbb{F}_2^n$ be a code with minimum distance d.

Then, the sets $B\left(c, \left\lfloor \frac{d-1}{2} \right\rfloor\right)$ when c ranges over $\mathcal{C}$ are pairwise disjoint.

Proof. Exercise or see the official lecture notes.

Prop (Linearly linked columns of p.c.m.)

Let $\mathcal{C} \subseteq \mathbb{F}_2^n$ be a code with parity-check matrix H.

Then, d is the smallest number of linearly linked columns of H.

Proof. Same as above.

III.2. Definition

Definition (Hamming code)

The *Hamming code* is the code in $\mathbb{F}_2^7$ with p.c.m.

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

Prop (Hamming code parameters)

The Hamming code is $[7, 4, 3]_2$.

Proof.

- dimension $= 4$, indeed, $\mathrm{rk}(H) = 3$ so $\dim(\ker(H)) = 7 - \mathrm{rk}(H)$ (by rank nullity theorem).
- minimum distance:
  - $d \leq 3$: $y = \begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 0 & 0 \end{pmatrix}$ is in the code
  - $d > 1$: since no zero column in H
  - $d > 2$: no two equal columns (because we're in $\mathbb{F}_2$)

(Fun?) fact: the Hamming code **corrects one error**:

Suppose we receive $y = c + e$ with $c \in \ker(H)$ and $w_{H(e)} = 1$, ie $e = \begin{pmatrix} 0 & \dots & \underbrace{1}_{i-\text{th position}} & 0 & \dots 0 \end{pmatrix}$

Compute $Hy^T = \underbrace{Hc^T}_{0} + \underbrace{He^T}_{i\text{-th column of } H}$ then return $y + e_i$.

III.3. Comparison

- **Hamming code** has rate $R = \frac{4}{7}$ that corrects a $\frac{1}{7}$ error ratio.
- **Repetition code** has rate $R = \frac{1}{7}$ and corrects a $\frac{3}{7}$ error ratio.

Hamming code yields a better tradeoff.