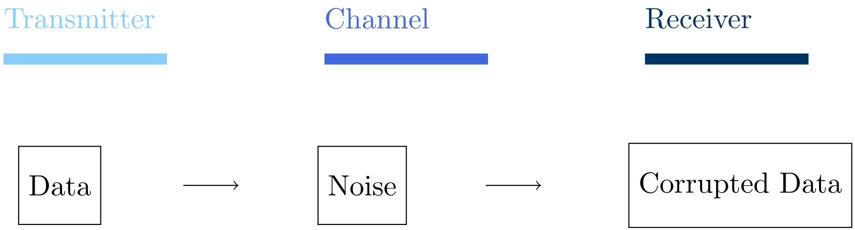
Coding Theory: Linear Codes

What is Coding Theory?

Communication channels

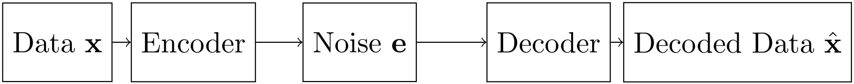


A Generic Communication Channel

Transmitter

Channel

Receiver



$$\mathbf{x} = (x_1, \ldots, x_k) \mapsto \underbrace{\mathbf{c} = (c_1, \ldots, c_n)}_{\text{codeword, } n \geq k} \longrightarrow \mathbf{c} + \mathbf{e} \to \hat{\mathbf{x}} = (\hat{x}_1, \ldots, \hat{x}_k)$$

Encoding

Given an alphabet A, a map that sends k data symbols $(x_1, \ldots, x_k) \in A^k$ to $n \geq k$ encoded symbols $(c_1, \ldots, c_n) \in A^n$. The encoded vector $\mathbf{c} = (c_1, \ldots, c_n)$ is called a codeword.

An encoding for $k = 1$: $(x_1) \mapsto (x_1, \ldots, x_1)$, that is $c_1 = c_2 = \ldots = c_n$.

If the alphabet A is $A = \{0, 1\}$, then $(0) \mapsto (0, \ldots, 0)$ and $(1) \mapsto (1, \ldots, 1)$.

Modulo 2

$+$	0	1
0	0	1
1	1	0

$\cdot$	0	1
0	0	0
1	0	1

Modulo 3

$+$	0	1	2
0	0	1	2
1	1	2	0
2	2	0	1

$\cdot$	0	1	2
0	0	0	0
1	0	1	2
2	0	2	1

Modulo p (a prime): take the remainder of the Euclidean division by p.

Modulo 4

$+$	0	1	2	3
0	0	1	2	3
1	1	2	3	0
2	2	3	0	1
3	3	0	1	2

$\cdot$	0	1	2	3
0	0	0	0	0
1	0	1	2	3
2	0	2	0	2
3	0	3	2	1

Modulo m (m not a prime): what is the difference?

So $2 \cdot 2 \equiv 0 \pmod{4}$. What does it change?

So $2 \cdot 2 \equiv 0 \pmod 4$. What does it change?

Actually a lot...

- If $2x = 0$, it is not true that $x = 0$, it could also be that $x = 2$.
- This also shows that a polynomial of degree 1 can have two solutions...
- and that two numbers different from 0 once multiplied together actually give 0.
- Also $2x = 1$ does not have a solution.

Finite field $\mathbb{F}_p$

For p a prime, the set of integers modulo p represented by $\{0, 1 \ldots, p-1\}$ is a finite field, denoted by $\mathbb{F}_p$.

$\mathbb{F}_p$ is finite means $|\mathbb{F}_p| = p < \infty$.

Informally, that $\mathbb{F}_p$ is a field means that computations work as usual, namely we can add, subtract, multiply, in a commutative manner, and divide as long as it is not by 0.

Inverse in $\mathbb{F}_p$

For x a non-zero element in $\mathbb{F}_p$, its (multiplicative) inverse is the element in $\mathbb{F}_p$ denoted by x^{-1} which satisfies that $x \cdot x^{-1} = x^{-1} \cdot x = 1$.

The inverse of 3 modulo 5:

Inverse in $\mathbb{F}_p$

For x a non-zero element in $\mathbb{F}_p$, its (multiplicative) inverse is the element in $\mathbb{F}_p$ denoted by x^{-1} which satisfies that $x \cdot x^{-1} = x^{-1} \cdot x = 1$.

The inverse of 3 modulo 5: 2 is the inverse of 3 since $2 \cdot 3 \equiv 1 \pmod 5$.

Find two elements that are their own inverse modulo 7:

Inverse in $\mathbb{F}_p$

For x a non-zero element in $\mathbb{F}_p$, its (multiplicative) inverse is the element in $\mathbb{F}_p$ denoted by x^{-1} which satisfies that $x \cdot x^{-1} = x^{-1} \cdot x = 1$.

The inverse of 3 modulo 5: 2 is the inverse of 3 since $2 \cdot 3 \equiv 1 \pmod 5$.

Find two elements that are their own inverse modulo 7: 1 is its own inverse since $1 \cdot 1 \equiv 1 \pmod 7$, but so is 6 since $6 \cdot 6 = 36 \equiv 1 \pmod 7$.

Exercise. Prove that if m is not a prime integer, then integers modulo m cannot form a finite field.

Exercise. Prove that if m is not a prime integer, then integers modulo m cannot form a finite field.

If m is not a prime, then m is a composite number, that is $m = ab$ for a, b some integers which are not zero. Then $ab \equiv 0 \mod m$.

Exercise. Prove that if m is not a prime integer, then integers modulo m cannot form a finite field.

If m is not a prime, then m is a composite number, that is $m = ab$ for a, b some integers which are not zero. Then $ab \equiv 0 \mod m$. Now a cannot be invertible, if it were, consider a^{-1} and multiply $ab \equiv 0 \mod m$ by a^{-1} to get $b \equiv 0 \mod m$, a contradiction.

Suppose there exists an element ω which is a zero of $X^2 + X + 1 \pmod 2$. Then $\omega \neq 0, 1$,

$$\omega^2 = \omega + 1 \pmod 2, \; \omega^3 = \omega(\omega + 1) = \omega^2 + \omega = 1 \pmod 2.$$

Suppose there exists an element ω which is a zero of $X^2 + X + 1 \pmod 2$. Then $\omega \neq 0, 1$,

$$\omega^2 = \omega + 1 \pmod 2,\ \omega^3 = \omega(\omega+1) = \omega^2 + \omega = 1 \pmod 2.$$

$\mathbb{F}_4$

$+$	0	1	ω	ω^2
0	0	1	ω	ω^2
1	1	0	ω^2	ω
ω	ω	ω^2	0	1
ω^2	ω^2	ω	1	0

$\cdot$	0	1	ω	ω^2
0	0	0	0	0
1	0	1	ω	ω^2
ω	0	ω	ω^2	1
ω^2	0	ω^2	1	ω

We will denote by $\mathbb{F}_q$ a finite field with q elements.

So far, we know $\mathbb{F}_p$ and $\mathbb{F}_4$, we will know more later, but in the meantime, we will use the notation.

Linear Encoding

Given a finite field $\mathbb{F}_q$, a linear map that sends k data symbols $(x_1, \ldots, x_k) \in \mathbb{F}_q^k$ to $n \geq k$ encoded symbols $(c_1, \ldots, c_n) \in \mathbb{F}_q^n$.

Since we work over $\mathbb{F}_q$:

- Codewords belong to $\mathbb{F}_q^n$, which is a vector space.
- Since the encoding is a linear map, by definition (1) the sum of two codewords is again a codeword, and (2) a multiple of a codeword is again a codeword.

Linear (n, k) code

Given a finite field $\mathbb{F}_q$, a set of codewords $\{(c_1, \ldots, c_n),\ c_i \in \mathbb{F}_q\} \in \mathbb{F}_q^n$ is said to form a linear code (or codebook) $\mathcal{C}$ if (1) the sum of two codewords is again a codeword, and (2) a multiple of a codeword is again a codeword.

The whole zero codeword $\mathbf{0} \in \mathcal{C}$.

If $\mathbf{c} \in \mathcal{C}$, so is $-\mathbf{c}$.

$\mathcal{C}$ forms a linear subspace of $\mathbb{F}_q^n$, it thus has a dimension, namely k, and we call n the length.

An (n, k) linear code $\mathcal{C}$ over $\mathbb{F}_q$ contains q^k codewords.

We write $|\mathcal{C}| = q^k$.

Let $\mathbf{b}_1, \ldots, \mathbf{b}_k$ be a basis for $\mathcal{C}$. Then codewords are obtained as every possible linear combination:

$$x_1\mathbf{b}_1 + \ldots + x_k\mathbf{b}_k,$$

there are q possible values for each x_i, $i = 1, \ldots, k$. $\square$

Linear algebra

For V, W finite-dimensional vector spaces, with a basis for each space, a linear map can be represented by a matrix in the given bases.

Linear codes

Given $\mathbb{F}_q^k, \mathbb{F}_q^n$, fix a basis for each space, a linear encoding is represented by a generator matrix.

Generator matrix

Given a finite field $\mathbb{F}_q$, a generator matrix G for an (n, k) linear code $\mathcal{C}$ is a $k \times n$ matrix, which contains as rows the basis vectors of $\mathcal{C}$.

There are many generator matrices.

There is a unique generator matrix of the form $G = [\mathbf{I}_k | A]$ where $\mathbf{I}_k$ is the identity matrix. The code is said to be in systematic form.

$$\underbrace{(x_1, \ldots, x_k)}_{\text{information data}} \underbrace{\begin{bmatrix} & a_{11} & a_{1,n-k} \\ \mathbf{I}_k & \vdots & \vdots \\ & a_{k,1} & a_{k,n-k} \end{bmatrix}}_{\text{generator matrix } G} = \underbrace{(x_1, \ldots, x_k, c_{k+1}, \ldots, c_n)}_{\text{codeword}}$$

The $(n, 1)$ repetition code

- Dimension: $k = 1$.
- Length: n.
- Encoding: $(x_1) \mapsto (x_1, \ldots, x_1) \in \mathbb{F}_q^n$.

The $(n, 1)$ repetition code

- Dimension: $k = 1$.
- Length: n.
- Encoding: $(x_1) \mapsto (x_1, \ldots, x_1) \in \mathbb{F}_q^n$.

$$(x_1)[1, \ldots, 1] = (x_1 \ldots, x_1).$$

The $(n, n-1)$ single parity check code

- Dimension: $k = n - 1$.
- Length: n
- Encoding: $(x_1, \ldots, x_k) \mapsto (x_1, \ldots, x_k, \sum_{i=1}^{k} x_i) \in \mathbb{F}_q^n$

The $(n, n-1)$ single parity check code

- Dimension: $k = n - 1$.
- Length: n
- Encoding: $(x_1, \ldots, x_k) \mapsto (x_1, \ldots, x_k, \sum_{i=1}^{k} x_i) \in \mathbb{F}_q^n$

$$(x_1, \ldots, x_k) \begin{bmatrix} & 1 \\ \mathbf{I}_k & \vdots \\ & 1 \end{bmatrix} = (x_1 \ldots, x_k, x_1 + \ldots + x_k).$$

Exercise. Consider the following codebook of length $n = 4$ over $\mathbb{F}_2$: $\{(1,0,0,0),(0,1,0,1),(1,1,0,1),(0,0,0,1)\}$. Is this code a linear code? If so, provide a generator matrix.

Exercise. Consider the following codebook of length $n = 4$ over $\mathbb{F}_2$: $\{(1,0,0,0),(0,1,0,1),(1,1,0,1),(0,0,0,1)\}$. Is this code a linear code? If so, provide a generator matrix.

The code cannot be linear, because $(0,0,0,0)$ does not belong to the code.

Exercise. Consider the following codebook of length $n = 4$ over $\mathbb{F}_2$: $\{(1,0,0,0),(0,1,0,1),(1,1,0,1),(0,0,0,1)\}$. Is this code a linear code? If so, provide a generator matrix.

The code cannot be linear, because $(0,0,0,0)$ does not belong to the code.

The code cannot be linear, because $(1,1,0,1)+(0,0,0,1)$ does not belong to the code.

Exercise. Consider the codebook of length $n = 5$ over $\mathbb{F}_2$ containing: $(0, 0, 0, 0, 0)$,$(1, 0, 0, 1, 0)$, $(0, 1, 0, 1, 1)$, $(1, 1, 0, 0, 1)$, $(0, 0, 1, 0, 1)$, $(1, 0, 1, 1, 1)$, $(0, 1, 1, 1, 0)$, $(1, 1, 1, 0, 0)$. Is this code a linear code? If so, provide a generator matrix.

Exercise. Consider the codebook of length $n = 5$ over $\mathbb{F}_2$ containing: $(0,0,0,0,0)$,$(1,0,0,1,0)$, $(0,1,0,1,1)$, $(1,1,0,0,1)$, $(0,0,1,0,1)$, $(1,0,1,1,1)$, $(0,1,1,1,0)$, $(1,1,1,0,0)$. Is this code a linear code? If so, provide a generator matrix.

The code is linear. The first three coefficients run through every possible vectors in $\mathbb{F}_2^3$, namely $(0,0,0)$,$(1,0,0)$, $(0,1,0)$, $(1,1,0)$, $(0,0,1)$, $(1,0,1)$, $(0,1,1)$, $(1,1,1)$.

Exercise. Consider the codebook of length $n = 5$ over $\mathbb{F}_2$ containing: $(0,0,0,0,0)$,$(1,0,0,1,0)$, $(0,1,0,1,1)$, $(1,1,0,0,1)$, $(0,0,1,0,1)$, $(1,0,1,1,1)$, $(0,1,1,1,0)$, $(1,1,1,0,0)$. Is this code a linear code? If so, provide a generator matrix.

The code is linear. The first three coefficients run through every possible vectors in $\mathbb{F}_2^3$, namely $(0,0,0)$,$(1,0,0)$, $(0,1,0)$, $(1,1,0)$, $(0,0,1)$, $(1,0,1)$, $(0,1,1)$, $(1,1,1)$. We next show that there is a generator matrix (which is enough to conclude the code is linear):

$$(x_1, x_2, x_3) \begin{bmatrix} 1 & 0 & 0 & a_{11} & a_{12} \\ 0 & 1 & 0 & a_{21} & a_{22} \\ 0 & 0 & 1 & a_{31} & a_{32} \end{bmatrix}.$$

Exercise. Consider the codebook of length $n = 5$ over $\mathbb{F}_2$ containing: $(0,0,0,0,0)$,$(1,0,0,1,0)$, $(0,1,0,1,1)$, $(1,1,0,0,1)$, $(0,0,1,0,1)$, $(1,0,1,1,1)$, $(0,1,1,1,0)$, $(1,1,1,0,0)$. Is this code a linear code? If so, provide a generator matrix.

$$(1,0,0)\begin{bmatrix}1 & 0 & 0 & a_{11} & a_{12}\\ 0 & 1 & 0 & a_{21} & a_{22}\\ 0 & 0 & 1 & a_{31} & a_{32}\end{bmatrix} = (1,0,0,a_{11},a_{12})$$

$$(x_1,x_2,x_3)\begin{bmatrix}1 & 0 & 0 & 1 & 0\\ 0 & 1 & 0 & 1 & 1\\ 0 & 0 & 1 & 0 & 1\end{bmatrix}.$$

Linear algebra

A subspace W of a vector space V is the kernel of some linear transformation (the projection onto W).

Linear codes

Given an (n, k) linear code over $\mathbb{F}_q$, there exists an $(n-k) \times n$ matrix H such that

$$\mathcal{C} = \{\mathbf{x} \in \mathbb{F}_q^n,\ H\mathbf{x}^T = \mathbf{0}\},$$

called a parity check matrix.

If $G = [\mathbf{I}_k | A]$ is a generator matrix for the (n, k) code $\mathcal{C}$, then $H = [-A^T | \mathbf{I}_{n-k}]$.

We have

$$G^T = \begin{bmatrix} \mathbf{I}_k \\ A^T \end{bmatrix}$$

and
$HG^T = -A^T + A^T = \mathbf{0}$.

If $\mathbf{c} \in \mathcal{C}$, $\mathbf{c} = \mathbf{x}G$ and $H\mathbf{c}^T = HG^T\mathbf{x}^T$.

If $G = [\mathbf{I}_k|A]$ is a generator matrix for the (n, k) code $\mathcal{C}$, then $H = [-A^T|\mathbf{I}_{n-k}]$.

We have

$$G^T = \begin{bmatrix} \mathbf{I}_k \\ A^T \end{bmatrix}$$

and
$HG^T = -A^T + A^T = \mathbf{0}$.

If $\mathbf{c} \in \mathcal{C}$, $\mathbf{c} = \mathbf{x}G$ and $H\mathbf{c}^T = HG^T\mathbf{x}^T$.
Thus $\mathcal{C}$ is contained in the kernel of the linear map $\mathbf{v} \mapsto H\mathbf{v}^T$. As H has rank $n-k$, this map has a kernel of dimension k, which is the dimension of $\mathcal{C}$. □

The $(n,1)$ repetition code

Generator matrix in systematic form:

$$[1, \underbrace{1 \ldots, 1}_{A}] = [\mathbf{I}_k | A].$$

Parity check matrix in systematic form:

The $(n, 1)$ repetition code

Generator matrix in systematic form:

$$[1, \underbrace{1 \ldots, 1}_{A}] = [\mathbf{I}_k | A].$$

Parity check matrix in systematic form:

$$[-A^T | \mathbf{I}_{n-k}] = \begin{bmatrix} -1 & \\ \vdots & \mathbf{I}_{n-k} \\ -1 & \end{bmatrix}$$

Data - Encoder - Channel - Decoder

(n, k) linear code

Generator matrix

Parity check matrix

$\mathbb{F}_p$, $\mathbb{F}_4$