4F5: Advanced Communications and Coding

Coding Handout 2: Mathematical Background and Linear Codes

Jossy Sayir

Signal Processing and Communications Lab
Department of Engineering
University of Cambridge
jossy.sayir@eng.cam.ac.uk

Michaelmas 2014

Channel Coding

Definition

1. A set of M m-ary messages of length K, i.e., $M = m^K$.
2. An encoding function that assigns a q-ary codeword $X_1 \dots X_N$ of length N to each message.
3. A decoding function that assigns a codeword or message to every possible received sequence $Y_1 \dots Y_N$

We call the set $\mathcal{C}$ of codewords a *code*. If $m = q$, we speak of an (N, K) q-ary code.

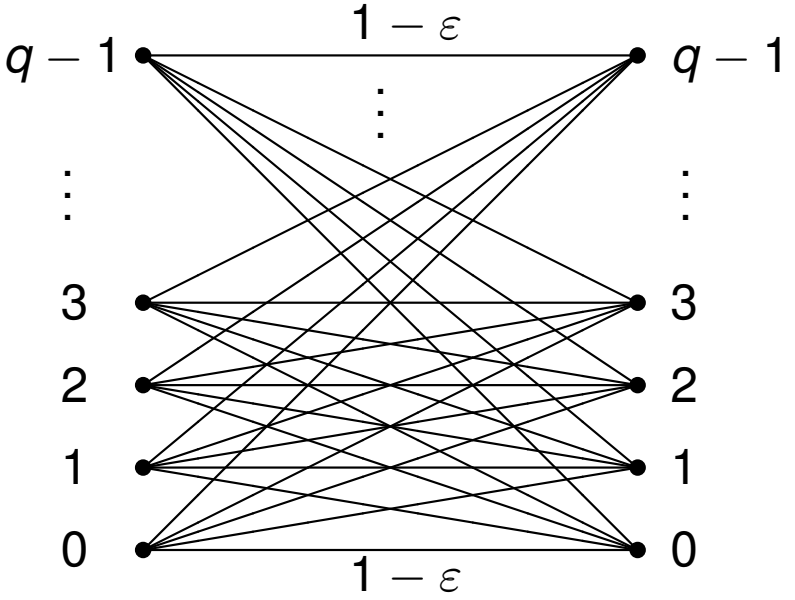
Maximum Likelihood (ML) Decoding

- Assumes a uniform prior for the messages
- Pick the most probable codeword given the observation, i.e., for any received sequence $Y_1 \dots Y_N = y_1 \dots y_N$,

$$\hat{X}_1 \dots \hat{X}_N = \arg\max_{x_1 \dots x_N \in \mathcal{C}} P(y_1 \dots y_N | x_1 \dots x_N)$$

$$= \arg\max_{x_1 \dots x_N \in \mathcal{C}} \prod_{i=0}^{N} P(y_i | x_i)$$

q-ary symmetric channel



ML decoding for the q-ary symmetric channel

- Channel definition: $P_{Y|X}(x|x) = 1-\varepsilon$ while $P_{Y|X}(y|x) = \frac{\varepsilon}{q-1}$ for $x \neq y$.
- ML rule

$$\hat{X}_1 \ldots \hat{X}_N = \arg\max_{x_1 \ldots x_N \in \mathcal{C}} (1-\varepsilon)^{N-d(x_1 \ldots x_N, y_1 \ldots y_N)} \left(\frac{\varepsilon}{q-1}\right)^{d(x_1 \ldots x_N, y_1 \ldots y_N)}$$

where $d(x_1 \ldots x_N, y_1 \ldots y_N)$, the number of positions where $x_1 \ldots x_N$ differs from $y_1 \ldots y_N$ is called the Hamming distance between $x_1 \ldots x_N$ and $y_1 \ldots y_N$

Error Correction Coding

Minimum Distance

- Let $d_{\min}$ be the minimum Hamming distance between any two codewords in a code $\mathcal{C}$
- The code can correct any received sequence with t or less errors if and only if

$$2t < d_{\min}$$

- Singleton Bound: $d_{\min} \leq N - K + 1$ for any (N, K) q-ary code. A code that achieves the Singleton bound with equality is called Maximum Distance Separable (MDS)
- It may appear that $d_{\min}$ is the most important design parameter for a channel code
- However, remember that channel coding is only equivalent to error correction coding for q-ary symmetric channels

Proof of the Singleton Bound: an (N, K) q-ary code has q^K codewords where any two codewords differ in at least $d_{\min}$ positions, hence if we remove the first $d_{\min} - 1$ components of each codeword, we obtain q^K distinct q-ary sequences of length $N - d_{\min} + 1$, but since there are only $q^{N-d_{\min}+1}$ possible sequences of that length, it follows that $q^K \leq q^{N-d_{\min}+1}$ or $K \leq N - d_{\min} + 1$.

Algebraic Systems

Single Operation

A single operation *algebraic system* $\langle S, \star \rangle$ consists of a set S and an operation $\star$ on pairs of elements of S

Conditions	**Statement**	$\langle \mathbf{S}, \star \rangle$
(i) Closure	for any $a, b \in S$, $c = a \star b \in S$	
(ii) Associative law	for any $a, b, c \in S$, $(a \star b) \star c = a \star (b \star c)$	Semi-group
(iii) Neutral element	there exists $e \in S$ such that for any $a \in S$, $e \star a = a \star e = a$. e is the neutral element of S	Monoid
(iv) Inverse	for any $a \in S$, there exists $b \in S$ such that $a \star b = b \star a = e$	Group
(v) Commutativity	for any $a, b \in S$, $a \star b = b \star a$	Abelian group

Algebraic Systems

Two Operations

A double operation algebraic system $\langle S, +, \cdot \rangle$ consists of a set S and two operations $+$ and $\cdot$ on pairs of elements of S

$\langle \mathbf{S}, + \rangle$	$\langle \mathbf{S}, \cdot \rangle$	**Conditions**	$\langle \mathbf{S}, +, \cdot \rangle$
Ab. Group	Monoid with $e_\cdot \neq e_+$	Distributive law, for any $a, b, c \in S$, $a \cdot (b + c) = (a \cdot b) + (a \cdot c)$ and $(a + b) \cdot c = (a \cdot c) + (b \cdot c)$	Ring
Ab. Group	Monoid with $e_\cdot \neq e_+$	$\langle S \setminus \{e_+\}, \cdot \rangle$ is an Ab. Group	**Field**

Algebraic Systems

Examples

- the real numbers $\mathbb{R}$ and complex numbers $\mathbb{C}$ are both fields, because they are closed, operations are associative, commutative and distributive, they have the neutral elements 0 for addition and 1 for multiplication, every element x has an additive inverse $-x$ and every element except 0 has a multiplicative inverse $1/x$
- the set of integers $\mathbb{Z}$ is not a field but a ring because elements do not have a multiplicative inverse, e.g., $1/2$ is not an integer. The set of non-negative integers $\mathbb{N}$ is not a ring because elements don't have additive inverses, e.g., $-1 \notin \mathbb{N}$
- $\langle \mathbb{Z}_q, + \rangle$, the set $\mathbb{Z}_q = \{0, 1, \ldots, q-1\}$ with addition modulo q is an Abelian group with neutral element 0, and every element a has an inverse $q - a$ denoted $-a$
- $\langle \mathbb{Z}_q, \cdot \rangle$, the same set with multiplication modulo q is a monoid because 0 has no inverse. 1 is its neutral element.

Algebraic Systems

Examples

- $\langle \mathbb{Z}_q, +, \cdot \rangle$ the set $\mathbb{Z}_q$ with addition and multiplication modulo q is a ring, because
  1. every element a has an additive inverse $b = q - a$ because $a + b = a + q - a = q \mod q = 0$, but
  2. some elements besides 0 may have no multiplicative inverses, e.g., for $q = a \cdot b$ with $1 < a < q$ and $1 < b < q$, $a \cdot b = 0$ in $\mathbb{Z}_q$, so if there existed an inverse a^{-1} of a in $\mathbb{Z}_q$, we would have

$$\begin{aligned} b &= 1 \cdot b = (a^{-1} \cdot a) \cdot b \\ &= a^{-1} \cdot (a \cdot b) = a^{-1} \cdot 0 \\ &= a^{-1} \cdot (1 + (-1)) \\ &= a^{-1} + (-a^{-1}) = 0 \end{aligned}$$

  where we have used associativity and the existence of additive inverses

- if q is prime, then $\mathbb{Z}_q$ is the Galois field GF(q), i.e., every element except 0 has a multiplicative inverse

Évariste Galois, 1811-32

Order of a group, order of an element

- The order of a group $\langle \boldsymbol{S}, \star \rangle$ is the number of elements in the set $\boldsymbol{S}$.
- For any $a \in \boldsymbol{S}$, we define

$$a^k \stackrel{\text{def}}{=} \underbrace{a \star a \star \ldots \star a}_{k \text{ times}}$$

- The order of an element a in a group $\langle \boldsymbol{S}, \star \rangle$ is the minimum $k > 0$ such that $a^k = e$. The order of e is 1 by convention.
- For any $a \in \boldsymbol{S}$, $\{a^0, a^1, a^2, \ldots, a^{\text{ord}\, a - 1}$ is a sub-group of $\boldsymbol{S}$

Lagrange's Theorem

The order of a sub-group of a finite group $\langle \boldsymbol{S}, \star \rangle$ divides the order of the group $\boldsymbol{S}$.

Corollary: the order of any element in a group divides the order of the group.

J.L. Lagrange, 1736-1813

Lagrange's Theorem

Proof

- Let $\boldsymbol{S}' \subset \boldsymbol{S}$ be a sub-group of $\langle \boldsymbol{S}, \star \rangle$, i.e., $e \in \boldsymbol{S}'$ and for any $a, b \in \boldsymbol{S}'$, $a \star b \in \boldsymbol{S}'$
- If ord $\boldsymbol{S}' <$ ord $\boldsymbol{S}$, then pick an element $c \in \boldsymbol{S}$ not in $\boldsymbol{S}'$. For any distinct $a, b \in \boldsymbol{S}'$, $c \star a \neq c \star b$. Hence, the set $c\boldsymbol{S}' = \{c \star a \text{ for any } a \in \boldsymbol{S}'\}$ has ord $\boldsymbol{S}'$ elements
- Next, pick any element not in $\boldsymbol{S}' \cup c\boldsymbol{S}'$ and repeat the above procedure. This can be repeated as long as there are elements left not in the union of the sets, adding ord $\boldsymbol{S}'$ elements to the union at every step.
- The procedure stops after k steps when there are no elements left outside the union, at which point the union has $(k+1)$ ord $\boldsymbol{S}' =$ ord $\boldsymbol{S}$ elements, which proves the theorem.

Extension fields

- it can be shown that a field with a finite number q of elements can only exist if $q = p^m$ where p is prime and $m \geq 1$. Fields with $m = 1$ are called base fields, whereas other fields ($m > 1$) are called extension fields

Construction of an extension field of size $q = p^m$

- the set GF(q) is the set of m-tuples over GF(p) whose elements $\boldsymbol{a} \in$ GF(q) can be represented either in vector form as $\boldsymbol{a} = (a_0, a_1, \ldots, a_{m-1})$ with $a_i \in$ GF(p) for $i = 0 \ldots m-1$, or alternatively in polynomial form

$$\boldsymbol{a}(X) = a_0 + a_1 X + a_2 X^2 + \ldots a_{m-1} X^{m-1}$$

- addition is defined as component-wise addition or polynomial addition, i.e., if $\boldsymbol{a}, \boldsymbol{b} \in$ GF(q), $\boldsymbol{a}(X) + \boldsymbol{b}(X) = \sum_{i=0}^{m-1} (a_i + b_i) X^i$ where the sums of coefficients are in GF(p)
- multiplication requires an irreducible polynomial $\pi(X)$ of degree m with coefficients in GF(p), i.e., a polynomial that cannot be factorised into a product of polynomials of degree smaller than m. Multiplication of $\boldsymbol{a}$ and $\boldsymbol{b} \in$ GF(q) is defined as the product of the associated polynomials modulo $\pi(X)$

Extension fields

Example

- GF(8) with $\pi(X) = 1 + X + X^3$
- example addition

$$\begin{aligned}(1,1,0) + (1,0,1) &= (1+X) + (1+X^2) \\ &= X + X^2 = (0,1,1)\end{aligned}$$

- example multiplication

$$\begin{aligned}(1,1,0) \cdot (1,0,1) &= (1+X) \cdot (1+X^2) \\ &= 1 + X + X^2 + X^3 \mod (1 + X + X^3) \\ &= X^2 = (0,0,1)\end{aligned}$$

Companion Matrices

- Pick $\pi(X)$ so that X has order $q-1$ in GF($q = p^m$) (it can be shown that there exists such a $\pi(X)$ for any $q = p^m$.)
- We can pre-compute the powers of X in a list and represent them in a matrix $\boldsymbol{M}$, for example taking GF(8) with $\pi(X) = 1 + X + X^3$,

$$\boldsymbol{M} \stackrel{\text{def}}{=} \begin{bmatrix} X^0 \\ X^1 \\ X^2 \\ X^3 \\ X^4 \\ X^5 \\ X^6 \end{bmatrix} = \begin{bmatrix} 1 \\ X \\ X^2 \\ 1+X \\ X+X^2 \\ 1+X+X^2 \\ 1+X^2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix}$$

- Let $a(X) = \sum_{i=0}^{m-1} a_i X^i$ and $b(X) = \sum_{i=0}^{m-1} b_i X^i = X^k$ for some k, then we have

$$a(X) \cdot b(X) = \sum_{i=0}^{m-1} a_i X^i X^k = \sum_{i=0}^{m-1} a_i \boldsymbol{m}_{k+i}$$

where $\boldsymbol{m}_n$ is the $(n \mod q-1)$-th row of the precomputed matrix $\boldsymbol{M}$

Companion Matrices

We conclude:

- There is a one-to-one correspondance between field elements X^k and matrices $\boldsymbol{M}_k$ consisting of m cyclic-consecutive rows of $\boldsymbol{M}$ starting at row k.
- Any product in an extension field GF(p^m) can be computed as a matrix-vector product in the vector space GF(p)^m.
- $\boldsymbol{M}_k$ is called the companion matrix of X^k in GF(q).

Companion Matrices

Example

- GF(8) with $\pi(X) = 1 + X + X^3$
- the precomputed product matrix is

$$\boldsymbol{M} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix}$$

- example multiplication

$$\begin{aligned}(1,1,0) \cdot (1,0,1) &= (1,1,0) \begin{bmatrix} 1 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \\ &= (0,0,1)\end{aligned}$$

Linear Block Codes

Definitions

- A q-ary linear code $\mathcal{C}$ of length N and dimension K is a K-dimensional subspace of the vector space $\mathcal{F}^N$ based on a field $\mathcal{F}$ of additive order q, e.g., $\mathcal{F} = \mathrm{GF}(q)$.
- A linear code is a set of codewords forming a vector space
- A generator or encoder matrix $\boldsymbol{G}$ for $\mathcal{C}$ is a basis of the vector space and can be used for mapping information vectors to code vectors
- The rate of the code is $R = K/N$

Systematicity

Systematic Encoders

- There are many bases for vector space, and hence many generator matrices for the same code. Each generator matrix will yield a different encoding or mapping of information sequences to codewords.
- The encoding is said to be systematic if the information sequence is part of the codeword, i.e.,

$$\boldsymbol{c} = [\boldsymbol{u}, \boldsymbol{p}]$$

  where $\boldsymbol{p}$ is called the parity sequence of the codeword.
- The systematic encoder matrix of a code is of the form

$$\boldsymbol{G} = [\boldsymbol{I}_K, \boldsymbol{P}]$$

  where $\boldsymbol{I}_K$ is the $K \times K$ identity matrix, and $\boldsymbol{P}$ is a $K \times (N - K)$ matrix (the example $\boldsymbol{G}$ in the previous slide is a systematic encoder matrix)

Parity-Check Matrices

Definitions

- Every vector subspace has an orthogonal complement, and so a linear code $\mathcal{C}$ of dimension K in $\mathcal{F}^N$ has an orthogonal complement code $\mathcal{C}^\perp$ of dimension $N-K$, called the dual code of $\mathcal{C}$
- For any $\boldsymbol{a} \in \mathcal{C}$ and any $\boldsymbol{b} \in \mathcal{C}^\perp$, we have $\boldsymbol{a} \cdot \boldsymbol{b}^T = 0$.
- If $\boldsymbol{H}$ is any generator matrix of $\mathcal{C}^\perp$, then for any $\boldsymbol{c} \in \mathcal{C}$, $\boldsymbol{c}\boldsymbol{H}^T = 0$ and any vector $\boldsymbol{c} \in \mathcal{F}^N$ satisfying $\boldsymbol{c}\boldsymbol{H}^T = 0$ is a codeword in $\mathcal{C}$. A generator matrix of the dual code is called a parity check matrix of $\mathcal{C}$. It is an $(N-K) \times N$ matrix.
- One parity-check matrix of $\mathcal{C}$ can easily be recovered from the systematic generator matrix of $\mathcal{C}$ as follows

$$\boldsymbol{G} = [\boldsymbol{I}_K, \boldsymbol{P}] \quad \boldsymbol{H} = [-\boldsymbol{P}^T, \boldsymbol{I}_{N-K}]$$

Proof: let $\boldsymbol{A} = \boldsymbol{G}\boldsymbol{H}^T$, for any i and j,

$$a_{ij} = \sum_{k=1}^{K} g_{ik} h_{jk} = 1 h_{ji} + g_{i(j+K)} 1 = -p_{ij} + p_{ij} = 0$$

Weight Distance Equivalence

Hamming Weight

- The Hamming weight $w(\boldsymbol{x})$ of a q-ary sequence $\boldsymbol{x}$ is the number of positions where $\boldsymbol{x}$ differs from zero. Remember that the Hamming distance $d(\boldsymbol{x}, \boldsymbol{y})$ defined previously is the number of positions where $\boldsymbol{x}$ differs from $\boldsymbol{y}$
- Let $\boldsymbol{x}$ and $\boldsymbol{y}$ be two codewords in a linear code $\mathcal{C}$. Since the code is linear, $\boldsymbol{z} = \boldsymbol{x} - \boldsymbol{y}$ is also in $\mathcal{C}$. The codeword $\boldsymbol{z}$ has non-zero components precisely where $\boldsymbol{x}$ differs from $\boldsymbol{y}$. Hence,
$$d(\boldsymbol{x}, \boldsymbol{y}) = w(\boldsymbol{x} - \boldsymbol{y})$$
- When looking at other codewords from any given codeword, a linear code "looks" the same, i.e., there will be the same number of codewords at any distance from any given codeword
- For any linear code $\mathcal{C}$, the minimum distance d_{min} is the minimum weight of any non-zero codeword w_{min}

Codes over extension fields

Base/extension equivalence

- We have seen that multiplication over an extension field can be viewed as vector matrix multiplication with companion matrices
- For a code defined over an extension field, the parity-check matrix $\boldsymbol{H}_{q^m}$ can be replaced by a parity-check matrix $\boldsymbol{H}_q$ over the base field where every element in $\boldsymbol{H}_{q^m}$ is replaced by a companion matrix in $\boldsymbol{H}_q$, and the code vector over GF(q^m) is written as a concatenation of elements from GF(q), i.e.,

$$\boldsymbol{H}_q = \left[\begin{array}{ccc|ccc|c|ccc} h_{11}^{11} & \dots & h_{11}^{1m} & h_{12}^{11} & \dots & h_{12}^{1m} & \dots & h_{1N}^{11} & \dots & h_{1N}^{1m} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & & \vdots & \ddots & \vdots \\ h_{11}^{m1} & \dots & h_{11}^{mm} & h_{12}^{m1} & \dots & h_{12}^{mm} & \dots & h_{1N}^{m1} & \dots & h_{1N}^{mm} \\ \hline h_{21}^{11} & \dots & h_{21}^{1m} & h_{22}^{11} & \dots & h_{22}^{1m} & \dots & h_{2N}^{11} & \dots & h_{2N}^{1m} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & & \vdots & \ddots & \vdots \\ h_{21}^{m1} & \dots & h_{21}^{mm} & h_{22}^{m1} & \dots & h_{22}^{mm} & \dots & h_{2N}^{m1} & \dots & h_{2N}^{mm} \\ \hline \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{array}\right]$$

Codes over extension fields

Base/Extension equivalence (continued)

- The same applies to the generator matrix
- Any linear code over an extension field is in reality a linear code over the base field!
- In terms of code properties (distance, etc.) there is no advantage to operating over the extension field. For example, a code over GF(256) is in reality a binary code.
- However, there are specific code and decoder constructions over extension fields that take advantage of operations on those fields. Hence, extension field can be a neat way of constructing practical coding systems over base fields.
- We will see one such method (Reed Solomon codes)