Digital Communications
SSY125, Lecture 9

Linear Block Codes
(Chapter 8)

Alexandre Graell i Amat
alexandre.graell@chalmers.se
https://sites.google.com/site/agraellamat

November 20, 2019

CHALMERS

Generator Matrix

$\mathcal{C}(n,k)$ Linear Block Code

A k-dimensional subspace of the n-dimensional vector space of all binary vectors of length n, i.e., $\mathcal{C} \subset \{0,1\}^n$.

- We can find k linearly independent vectors $\boldsymbol{g}_1, \ldots, \boldsymbol{g}_k \in \{0,1\}^n$ that span $\mathcal{C} \longrightarrow$ Every codeword in $\mathcal{C}$ is a linear combination of $\boldsymbol{g}_1, \ldots, \boldsymbol{g}_k$.
- The codeword $\boldsymbol{c} = (c_1, \ldots, c_n)$ for the message $\boldsymbol{u} = (u_1, \ldots, u_k)$ can be expressed as

$$\boldsymbol{c} = \sum_{i=1}^{k} u_i \boldsymbol{g}_i = u_1 \boldsymbol{g}_1 + \ldots u_k \boldsymbol{g}_k.$$

- Can be rewritten in matrix form as

$$\boldsymbol{c} = \boldsymbol{u}\boldsymbol{G},$$

  where $\boldsymbol{G}$ is a $k \times n$ binary matrix with rows $\boldsymbol{g}_1, \ldots, \boldsymbol{g}_k$.
- $\boldsymbol{G}$ spans (i.e., generates) the code $\mathcal{C} \longrightarrow$ generator matrix of the code.

Generator Matrix

- Both the code as well as the encoder are completely specified by the generator matrix $\boldsymbol{G}$.
- Formally,
$$\mathcal{C} \triangleq \{\boldsymbol{c} : \boldsymbol{c} = \boldsymbol{u}\boldsymbol{G}\}.$$
- Several bases generate the same subspace $\longrightarrow$ several generator matrices generate the same code $\mathcal{C}$.

Equivalent Encoders

Two encoders (equivalently two generator matrices) that generate the same code are called equivalent encoders.

Generator Matrix

- From basic linear algebra, any linear operation on the basis vectors leads to another basis that generates the same subspace $\longrightarrow$ can always find a generator matrix in the form

$$\boldsymbol{G}_{\mathsf{s}} = (\boldsymbol{I}_k \quad \boldsymbol{P}),$$

  where $\boldsymbol{I}_k$ is a $k \times k$ identity matrix and $\boldsymbol{P}$ is a $k \times (n-k)$ matrix.
- $\boldsymbol{G}_{\mathsf{s}}$ is called a systematic generator matrix and the resulting code is a systematic code.
- For a systematic code,

$$\boldsymbol{c} = (c_1 \ldots, c_n) = \boldsymbol{u}\boldsymbol{G}_{\mathsf{s}} = (u_1, , \ldots, u_k, p_1, \ldots p_{n-k}) = (\boldsymbol{u} \ \boldsymbol{p}).$$

Systematic Code

A code that contains the information word $\boldsymbol{u}$ as verbatim copy in $\boldsymbol{c}$.

Linear Block Codes

Example: $(3,1)$ Repetition Code

- Encoder:

$$0 \to (0,0,0) \qquad 1 \to (1,1,1) \qquad \boxtimes$$

- The corresponding generator matrix is

$$\boldsymbol{G} = \begin{pmatrix} 1 & 1 & 1 \end{pmatrix}.$$

- The code is systematic, $R_{\mathsf{c}} = 1/3$ and $d_{\min} = 3$.

Linear Block Codes

Example: $(3,2)$ Parity-Check Code

$$\mathcal{C}_{\text{check}}(n=3,k=2)=\{(0,0,0),(0,1,1),(1,0,1),(1,1,0)\}\,.$$

- c_3 is a parity-check of the first two code bits, i.e., $c_3 = c_1 + c_2$.
- Encoder:

$$\begin{aligned} \boldsymbol{u}=(0,0) \quad &\rightarrow \quad \boldsymbol{c}_1=(0,0,0)\\ \boldsymbol{u}=(0,1) \quad &\rightarrow \quad \boldsymbol{c}_2=(0,1,1)\\ \boldsymbol{u}=(1,0) \quad &\rightarrow \quad \boldsymbol{c}_3=(1,0,1)\\ \boldsymbol{u}=(1,1) \quad &\rightarrow \quad \boldsymbol{c}_4=(1,1,0). \end{aligned}$$

- Generator matrix:

$$\boldsymbol{G}=\begin{pmatrix} 1 & 0 & 1\\ 0 & 1 & 1 \end{pmatrix}. \quad \boxtimes$$

- The code is systematic, $R_{\mathsf{c}} = 2/3$ and $d_{\min}(\mathcal{C}_{\text{check}}) = 2$.

Linear Block Codes

Example: $(7, 4)$ Hamming Code

- Hamming codes have parameters $n = 2^r - 1$, $k = 2^r - r - 1$ with $r \geq 3$, and $d_{\min} = 3$.

$$\mathcal{C}_{\text{Hamm}}(n = 7, k = 4) =$$
$$\{(0,0,0,0,0,0,0), (0,0,0,1,0,1,1), (0,0,1,0,1,1,0), (0,0,1,1,1,0,1),$$
$$(0,1,0,0,1,1,1), (0,1,0,1,1,0,0), (0,1,1,0,0,0,1), (0,1,1,1,0,1,0),$$
$$(1,0,0,0,1,0,1), (1,0,0,1,1,1,0), (1,0,1,0,0,1,1), (1,0,1,1,0,0,0),$$
$$(1,1,0,0,0,1,0), (1,1,0,1,0,0,1), (1,1,1,0,1,0,0), (1,1,1,1,1,1,1)\}.$$

- Generator matrix in systematic form:

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

Parity-Check Matrix

- The null (or dual) space of $\mathcal{C}$,

$$\mathcal{C}_{\perp} = \{\tilde{\boldsymbol{c}} : \langle \tilde{\boldsymbol{c}}, \boldsymbol{c} \rangle = 0 \text{ for all } \boldsymbol{c} \in \mathcal{C}\},$$

  is an $(n-k)$-dimensional subspace of $\{0,1\}^n$.
- $\mathcal{C}_{\perp}$ is a binary $(n, n-k)$ linear block code, referred to as dual code of $\mathcal{C}$.
- Let $\boldsymbol{H}$ be a generator matrix of $\mathcal{C}_{\perp}$, of dimensions $(n-k) \times n$ (the rows of $\boldsymbol{H}$ form a basis that spans $\mathcal{C}_{\perp}$).
- Since every codeword $\boldsymbol{c} \in \mathcal{C}$ is orthogonal to every codeword $\tilde{\boldsymbol{c}} \in \mathcal{C}_{\perp}$,

$$\boldsymbol{c}\boldsymbol{H}^{\mathsf{T}} = \boldsymbol{0}_{n-k},$$

  and

$$\boldsymbol{G}\boldsymbol{H}^{\mathsf{T}} = \boldsymbol{0}_{k \times (n-k)}.$$

- $\boldsymbol{c}\boldsymbol{H}^{\mathsf{T}} = \boldsymbol{0}_{n-k}$ if and only if $\boldsymbol{c} \in \mathcal{C} \longrightarrow$ We can define a code $\mathcal{C}$ through the generator matrix of its dual code $\mathcal{C}_{\perp}$.

Parity-Check Matrix

- Formally,

$$\mathcal{C} \triangleq \{\boldsymbol{c} : \boldsymbol{c} = \boldsymbol{u}\boldsymbol{G}\}.$$

- Since $\boldsymbol{c} \in \mathcal{C}$ if an only if $\boldsymbol{c}\boldsymbol{H}^{\mathsf{T}} = \mathbf{0}$, $\mathcal{C}$ is also defined as the null space of $\boldsymbol{H}$,

$$\mathcal{C} \triangleq \{\boldsymbol{c} : \boldsymbol{c}\boldsymbol{H}^{\mathsf{T}} = \mathbf{0}\}.$$

- A linear block code is uniquely specified by $\boldsymbol{G}$ and $\boldsymbol{H}$.
- Usually, the generator matrix is used for encoding, while the decoding is based on $\boldsymbol{H}$.

Interpretation of the Parity-Check Matrix

- $\boldsymbol{c}\boldsymbol{H}^{\mathsf{T}} = \mathbf{0}$ forms a system of $n-k$ linearly independent equations:

$$\begin{array}{ccccccccc}
h_{11}c_1 & + & h_{12}c_2 & + & \dots & + & h_{1n}c_n & = & 0 \\
h_{21}c_1 & + & h_{22}c_2 & + & \dots & + & h_{2n}c_n & = & 0 \\
\vdots & & \vdots & & \vdots & & \vdots & & \vdots \\
h_{(n-k)1}c_1 & + & h_{(n-k)2}c_2 & + & \dots & + & h_{(n-k)n}c_n & = & 0
\end{array}$$

  where h_{ij} is the entry of the matrix $\boldsymbol{H}$ at row i and column j.

- Parity-check equations: The equations that each codeword must satisfy $\longrightarrow$ $\boldsymbol{H}$ is known as the parity-check matrix of the code $\mathcal{C}$.
- $\boldsymbol{H}$ is not unique!
- If $\boldsymbol{G}$ is in systematic form, $\boldsymbol{G}_{\mathsf{s}} = (\boldsymbol{I}_k \;\; \boldsymbol{P})$, its corresponding parity-check matrix in systematic form is given by

$$\boldsymbol{H}_{\mathsf{s}} = \left(\boldsymbol{P}^{\mathsf{T}} \;\; \boldsymbol{I}_{n-k}\right).$$

Linear Block Codes

Example: $(3,2)$ Parity-check code

- $(3,2)$ Parity-check code with systematic generator matrix

$$\boldsymbol{G}_{\mathsf{s}} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix}.$$

- The corresponding parity-check matrix in systematic form is

$$\boldsymbol{H}_{\mathsf{s}} = \begin{pmatrix} \boldsymbol{P}^{\mathsf{T}} & \boldsymbol{I}_{n-k} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 \end{pmatrix}. \qquad \boxtimes$$

- The $(3,1)$ repetition code is the dual code of the $(3,2)$ parity-check code! (and vice versa).

Decoding of Linear Block Codes over the BSC

- Transmission over the BSC using an (n, k) linear block code $\mathcal{C}$ with parity-check matrix $\boldsymbol{H}$.
- Received vector:

$$\bar{\boldsymbol{y}} = \boldsymbol{c} + \boldsymbol{e},$$

- $\boldsymbol{e} = (e_1, \ldots, e_n)$ is the error vector or error pattern, with $e_i = 1$ if transmission error occurs for position i.
- ML decoding:

$$\begin{aligned}\hat{\boldsymbol{c}} &= \arg\min_{\boldsymbol{c}\in\mathcal{C}} d_{\mathsf{H}}(\boldsymbol{c}, \bar{\boldsymbol{y}}) \\ &= \arg\min_{\boldsymbol{c}\in\mathcal{C}} w_{\mathsf{H}}(\boldsymbol{c} + \bar{\boldsymbol{y}}) \\ &= \arg \min_{\substack{\boldsymbol{e} \\ \bar{\boldsymbol{y}}+\boldsymbol{e}=\boldsymbol{c}\in\mathcal{C}}} w_{\mathsf{H}}(\boldsymbol{e}).\end{aligned}$$

- ML decoding: finding the error pattern $\boldsymbol{e}$ with smallest Hamming weight that we need to add to $\bar{\boldsymbol{y}}$ to obtain a valid codeword!

Syndrome-Based Decoding

- Using $\boldsymbol{H}$,

$$\begin{aligned}\bar{\boldsymbol{y}}\boldsymbol{H}^{\mathsf{T}} &= (\boldsymbol{c}+\boldsymbol{e})\,\boldsymbol{H}^{\mathsf{T}} \\ &= \boldsymbol{e}\boldsymbol{H}^{\mathsf{T}}.\end{aligned}$$

- We define

$$\boldsymbol{s} \triangleq \bar{\boldsymbol{y}}\boldsymbol{H}^{\mathsf{T}} = \boldsymbol{e}\boldsymbol{H}^{\mathsf{T}},$$

- $\boldsymbol{s} = (s_1, \ldots, s_{n-k})$, of length $n-k$, is called the syndrome.
- There are 2^{n-k} possible syndromes $\boldsymbol{s} = (s_1, \ldots, s_{n-k})$, i.e., all binary vectors of length $n-k$.
- Property: $\bar{\boldsymbol{y}} \in \mathcal{C}$ if and only if $\boldsymbol{s} = \mathbf{0}_{(n-k)}$.
- If $\boldsymbol{s} = \mathbf{0}_{(n-k)}$ $(\bar{\boldsymbol{y}} \in \mathcal{C})$, the most-likely transmitted codeword is $\boldsymbol{c} = \bar{\boldsymbol{y}}$.

Syndrome-Based Decoding

- $\boldsymbol{s} = \mathbf{0}_{(n-k)}$ if:
  - (i) $\boldsymbol{e} = \mathbf{0}_{(n)}$, i.e., the channel introduces no errors.
  - (ii) $\boldsymbol{e}$ is such that $\bar{\boldsymbol{y}} = \boldsymbol{c} + \boldsymbol{e} \in \mathcal{C}$ but $\bar{\boldsymbol{y}} \neq \boldsymbol{c}$ $\longrightarrow$ The decoder will decide erroneously that $\bar{\boldsymbol{y}}$ was transmitted. (undetectable error pattern)
- There are $2^k - 1$ undetectable error patterns.
- Since there are 2^{n-k} possible syndromes $\boldsymbol{s} = (s_1, \ldots, s_{n-k})$, there are

$$\frac{2^n}{2^{n-k}} = 2^k$$

received vectors $\bar{\boldsymbol{y}}$ (equivalently error patterns) that generate the same syndrome.

Syndrome-Based Decoding

- For a given syndrome $\boldsymbol{s}$ what's the most likely transmitted codeword?
- Equivalently, what is the most likely error pattern? The solution to $\boldsymbol{s} = \boldsymbol{e}\boldsymbol{H}^{\mathsf{T}}$ with smallest Hamming weight!

ML decoding for the BSC

For a received vector $\bar{\boldsymbol{y}}$ that generates the syndrome $\boldsymbol{s}$, among all possible 2^k error patterns that generate $\boldsymbol{s}$ find the one with smallest Hamming weight, $\boldsymbol{e}_{\mathsf{min}}(\boldsymbol{s})$ and decode onto

$$\hat{\boldsymbol{c}} = \bar{\boldsymbol{y}} + \boldsymbol{e}_{\mathsf{min}}(\boldsymbol{s}).$$

Syndrome-Based Decoding

- Can be implemented efficiently (for reasonable $n-k$) using a decoding table that associates to each syndrome $\boldsymbol{s}$ the error pattern with smallest weight that generates it, $\boldsymbol{e}_{\mathsf{min}}(\boldsymbol{s})$.
- Decoding:
  1. Compute the syndrome of $\bar{\boldsymbol{y}}$,
  $$\boldsymbol{s} = \bar{\boldsymbol{y}}\boldsymbol{H}^{\mathsf{T}}.$$
  2. Find in the decoding table the error pattern $\boldsymbol{e}_{\mathsf{min}}(\boldsymbol{s})$.
  3. Decode $\bar{\boldsymbol{y}}$ onto
  $$\hat{\boldsymbol{c}} = \bar{\boldsymbol{y}} + \boldsymbol{e}_{\mathsf{min}}(\boldsymbol{s}).$$

Syndrome-Based Decoding

Example: $(7,4)$ Hamming Code

$(7,4)$ Hamming Code with parity-check matrix

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

There are $2^3 = 8$ possible syndromes. The decoding table is

$\boldsymbol{s}$	$\boldsymbol{e}_{\text{min}}$
(0,0,0)	(0,0,0,0,0,0,0)
(0,0,1)	(0,0,0,0,0,0,1)
(0,1,0)	(0,0,0,0,0,1,0)
(0,1,1)	(0,0,0,1,0,0,0)
(1,0,0)	(0,0,0,0,1,0,0)
(1,0,1)	(1,0,0,0,0,0,0)
(1,1,0)	(0,0,1,0,0,0,0)
(1,1,1)	(0,1,0,0,0,0,0)

Syndrome-Based Decoding

Example: $(7,4)$ Hamming Code

Assume $\bar{\boldsymbol{y}} = (1,1,0,0,0,0,0)$.

1. Compute $\boldsymbol{s} = \bar{\boldsymbol{y}}\boldsymbol{H}^{\mathsf{T}} = (0,1,0)$.
2. We find $\boldsymbol{e}_{\mathsf{min}}((0,1,0)) = (0,0,0,0,0,1,0)$.
3. ML decision: $\hat{\boldsymbol{c}} = \bar{\boldsymbol{y}} + \boldsymbol{e}_{\mathsf{min}}((0,1,0)) = (1,1,0,0,0,1,0)$ (minimizes the distance $d_{\min}(\boldsymbol{c}, \bar{\boldsymbol{y}})$).

The error pattern $\boldsymbol{e} = (1,1,0,0,0,0,0)$ also generates the syndrome $\boldsymbol{s} = (0,1,0)$. Therefore, if the actual error introduced by the channel was $\boldsymbol{e} = (1,1,0,0,0,0,0)$, we would have decoded incorrectly onto $\hat{\boldsymbol{c}} = (1,1,0,0,0,1,0)$!

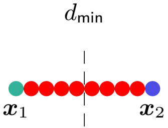
Error Correction Capability

Theorem (Error correction capability)

For transmission over the BSC, a block code with minimum Hamming distance $d_{\min}$ can correct all error patterns with

$$t = \left\lfloor \frac{d_{\min} - 1}{2} \right\rfloor$$

or fewer errors.



Error Detection Capability

Theorem (Error detection capability)

For transmission over the BSC, a block code with minimum Hamming distance $d_{\min}$ can detect all error patterns with

$$d = d_{\min} - 1$$

or fewer errors.