8 Lineare Codes

8 Lineare Codes

- Laut dem zweiten Shannonsches Codierungstheorem, erreichen zufällige Codes für große Codewortlängen n die Kanalkapazität.
- Ein zufälliger Code würde es erforderlich machen, die $2^l = 2^{Rn}$ Abbildungen der Codierung zu tabellieren $\Rightarrow$ unpraktibel für große n.
- Für begrenzte n sind Codes mit definierten Eigenschaften wünschenswert.

$\Rightarrow$ Systematik in der Codierung erforderlich.

Definition 8.1 (Lineare Codes)

Bei einem linearen Code C erfolgt die Codierung

$$\mathcal{C}(\boldsymbol{a}) = \boldsymbol{a}\boldsymbol{G} \tag{8.1}$$

durch die Multiplikation mit einer Generatormatrix $\boldsymbol{G} \in \{0,1\}^{l \times n}$, wobei alle Rechenoperationen modulo 2 erfolgen. Die Zeilen von $\boldsymbol{G}$ sind linear unabhängig.

Hinweis: Da $\boldsymbol{a}$ ein Zeilenvektor ist, ist die Matrix der rechte Faktor.

Beim Rechnen modulo 2 entspricht die Multiplikation der UND-, die Addition und Subtraktion der XOR-Verknüpfung.

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

$+$	0	1
0	0	1
1	1	0

$-$	0	1
0	0	1
1	1	0

Beispiel 8.1

Der Wiederholungscode

$$0 \mapsto 000, \qquad 1 \mapsto 111$$

aus Beispiel 7.1 ist ein linearer Code mit der Generatormatrix

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

Beispiel 8.2

Der Parity-Check-Code

$$00 \mapsto 000, \qquad 01 \mapsto 011, \qquad 10 \mapsto 101, \qquad 11 \mapsto 110$$

aus Beispiel 7.2 ist ein linearer Code mit der Generatormatrix

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

Die l unhängigen Zeilen der Generatormatrix bilden eine Basis des l-dimensionalen Coderaums.

Satz 8.1

Der minimale Abstand $\mathrm{d}(C)$ *eines linearen Codes* C *ist gegeben durch*

$$\mathrm{d}(C) = \min_{x \in C} \mathrm{d}(\boldsymbol{x}, \mathbf{0}). \tag{8.2}$$

Beweis.

Da $\mathbf{0} \in C$, gilt offensichtlich $\mathrm{d}(\boldsymbol{x}, \mathbf{0}) \geq \mathrm{d}(C)$.
Im allgemeinen gilt

$$\mathrm{d}(C) = \min_{\substack{\boldsymbol{x}', \boldsymbol{x}'' \in C \\ \boldsymbol{x}' \neq \boldsymbol{x}''}} \mathrm{d}(\boldsymbol{x}', \boldsymbol{x}''). \tag{8.3}$$

Seien nun $\boldsymbol{x}'$ und $\boldsymbol{x}''$ Codewörter mit $\mathrm{d}(C) = \mathrm{d}(\boldsymbol{x}', \boldsymbol{x}'')$, dann ist aufgrund der Linearität auch $\boldsymbol{x} = \boldsymbol{x}' - \boldsymbol{x}'' \in C$, und $\mathrm{d}(\boldsymbol{x}, \mathbf{0}) = \mathrm{d}(\boldsymbol{x}', \boldsymbol{x}'') = \mathrm{d}(C)$. □

Um den minimalen Abstand zu bestimmen brauchen also nicht $2^l \cdot (2^l - 1)$ Codewortpaare, sondern nur $2^l - 1$ Codewörter betrachtet werden.

8.1 Kontrollmatrizen

Ob ein Wort $\boldsymbol{y}$ in einem linearen Code C enthalten ist, lässt sich mit $n - l$ linearen Gleichungen prüfen.

Beispiel 8.3

Ein Wort $\boldsymbol{y} = y_1 y_2 y_3$ ist genau dann im Wiederholungscode $\{000, 111\}$ enthalten, wenn

$$\begin{array}{lllll} y_1 = y_2 & \Leftrightarrow & y_1 - y_2 = 0 & \Leftrightarrow & y_1 + y_2 = 0 \\ y_2 = y_3 & \Leftrightarrow & y_2 - y_3 = 0 & \Leftrightarrow & y_2 + y_3 = 0 \end{array}$$

bzw.

$$\boldsymbol{y} \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix}^T = \mathbf{0}.$$

Beispiel 8.4

Alle Codewörter des Parity-Check-Codes $\{000, 011, 101, 110\}$ aus Beispiel 7.2 haben eine gerade Anzahl Einsen, also

$$y_1 + y_2 + y_3 = 0$$

bzw.

$$\boldsymbol{y} \begin{pmatrix} 1 & 1 & 1 \end{pmatrix}^T = \boldsymbol{0}.$$

Definition 8.2 (Kontrollmatrix)

Eine Matrix $\boldsymbol{H} \in \{0,1\}^{(n-l)\times n}$ mit $n - l$ unabhängigen Zeilen, ist eine Kontrollmatrix des linearen Codes C, wenn

$$\boldsymbol{x}\boldsymbol{H}^T = \boldsymbol{0} \quad \Leftrightarrow \quad \boldsymbol{x} \in C, \tag{8.4}$$

also der Coderaum C gleich dem Nullraum von $\boldsymbol{H}$ ist.

Satz 8.2

Sei $\boldsymbol{G} \in \{0,1\}^{l\times n}$ die Generatormatrix eines linearen Codes C, so ist $\boldsymbol{H} \in \{0,1\}^{(n-l)\times n}$ genau dann eine Prüfmatrix von C, wenn die Zeilen von $\boldsymbol{H}$ linear unabhängig sind und

$$\boldsymbol{G}\boldsymbol{H}^T = \boldsymbol{0}. \tag{8.5}$$

Beweis.

Da die Zeilen von $\boldsymbol{H}$ linear unabhängig sind, hat der Nullraum $n - (n - l) = l$ Dimensionen.
Die l Zeilen von $\boldsymbol{G}$ spannen den l-dimensionalen Coderaum C auf.
Da jede Zeile von $\boldsymbol{G}$ im Nullraum von $\boldsymbol{H}$ liegt, liegt der gesamt Coderaum C im Nullraum von $\boldsymbol{H}$, und auf Grund der gleichen Dimensionalität, müssen C und der Nullraum von $\boldsymbol{H}$ gleich sein. □

Beispiel 8.5 (Wiederholungscode)

$$\begin{pmatrix} 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix}^T = \begin{pmatrix} 1+1+0 & 0+1+1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \end{pmatrix}$$

Beispiel 8.6 (Parity-Check-Code)

$$\begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 \end{pmatrix}^T = \begin{pmatrix} 1+0+1 \\ 0+1+1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Satz 8.3

Sei $\boldsymbol{H}$ eine Kontrollmatrix des linearen Codes C, so ist der minimale Abstand $\mathrm{d}(C)$ gleich der kleinsten Anzahl abhängiger Spalten von $\boldsymbol{H}$.

Beweis.

Seien $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n$ die Spalten von $\boldsymbol{H}$. Dann lässt sich

$$\boldsymbol{x}\boldsymbol{H}^T = \sum_{i=1}^{n} x_i \boldsymbol{s}_i^T \tag{8.6}$$

schreiben, für ein Codewort $\boldsymbol{x}$ muss also

$$\sum_{i=1}^{n} x_i \boldsymbol{s}_i = \boldsymbol{0} \tag{8.7}$$

gelten, was genau die Bedingung linearer Abhängigkeit der $\boldsymbol{s}_i$ ist, für die $x_i \neq 0$.
Die kleinste Anzahl linear abhängiger Spalten von $\boldsymbol{H}$ ist also die geringste Anzahl von gesetzen Bits in einem Codewort von C (mit Ausnahme von $\boldsymbol{0}$).Nach Satz 8.1 ist dies aber der minimale Abstand $\mathrm{d}(C)$. □

Bemerkungen:

- Nur wenn $\boldsymbol{H}$ eine Nullspalte enthält, ist $\mathrm{d}(C) = 1$.
- Nur wenn $\boldsymbol{H}$ mindestens zwei identische Spalten enthält, ist $\mathrm{d}(C) = 2$.

Beispiel 8.7

Die Prüfmatrix

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

des Wiederholungscodes besitzt keine gleichen Spalten; die Summe aller drei Spalten ist der Nullvektor. Also ist $\mathrm{d}(C) = 3$.

Beispiel 8.8

Die Prüfmatrix

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

des Parity-Check-Codes besitzt (nur) paarweise gleiche Spalten, also ist $\mathrm{d}(C) = 2$.

8.2 Äquivalenz von Codes, systematische Codes

- Werden Zeilen einer Generatormatrix vertauscht, ändert dies zwar die Codierung $\mathcal{C}$, nicht aber den Code C (die Basis bleibt gleich).
- Wird eine Zeile einer Generatormatrix durch die Summe dieser und einer anderen Zeile ersetzt, ändert dies zwar die Codierung $\mathcal{C}$, nicht aber den Code C (die Basis ändert sich, spannt aber denselben Raum auf).
- Werden Spalten einer Generatormatrix vertauscht, ändert dies nicht nur die Codierung, sondern auch den Code. Es werden jedoch lediglich die Stellen der Codewörter (für alle gleich) vertauscht, die wesentlichen Eigenschaften des Codes bleiben dabei unverändert.

Definition 8.3 (Äquivalenz von Codes)

Zwei lineare Codes $\mathcal{C}$ und $\mathcal{C}'$ sind äquivalent, wenn sich ihre Generatormatrizen durch obige Operationen ineinander überführen lassen.

Definition 8.4

Ein linearer Code heißt systematisch, wenn die Generatormatrix die Form

$$\boldsymbol{G} = (\boldsymbol{I} \quad \boldsymbol{P}) \tag{8.8}$$

hat, wobei $\boldsymbol{I}$ die (hier l-dimensionale) Einheitsmatrix bezeichnet.

Die ersten l Bits eines Codeworts sind also gerade die Nutzdatenbits.

Satz 8.4

Zu jedem linearen Code gibt es einen äquivalenten systematischen Code.

Beweisidee.
Anwendung des Gauß'schen Eliminationsverfahrens mit Zeilen- und Spaltenpivotisierung erlaubt geeignete Umformung der Generatormatrix. □

Beispiel 8.9

Ein Code sei durch die Generatormatrix

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

gegeben. Wie lautet ein äquivalenter systematischer Code?

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

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

Die Generatormatrix

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

beschreibt also einen äquivalenten systematischen Code.

Satz 8.5

Für einen systematischen linearen Code mit der Generatormatrix

$$\boldsymbol{G} = \begin{pmatrix} \boldsymbol{I} & \boldsymbol{P} \end{pmatrix} \tag{8.9}$$

ist $\boldsymbol{H} = \begin{pmatrix} \boldsymbol{P}^T & \boldsymbol{I} \end{pmatrix}$ *eine Kontrollmatrix.*

Beachte: Für $\boldsymbol{G}$ ist $\boldsymbol{I}$ l-dimensional, für $\boldsymbol{H}$ $(n-l)$-dimensional.

Beweis.

$$\boldsymbol{G}\boldsymbol{H}^T = \begin{pmatrix} \boldsymbol{I} & \boldsymbol{P} \end{pmatrix} \begin{pmatrix} \boldsymbol{P} \\ \boldsymbol{I} \end{pmatrix} = \boldsymbol{I}\boldsymbol{P} + \boldsymbol{P}\boldsymbol{I} = \boldsymbol{P} + \boldsymbol{P} = \boldsymbol{0} \tag{8.10}$$

□

Beispiel 8.10

Der allgemeine Wiederholungscode, der jedes Nutzdatenbit n mal wiederholt, hat die Generatormatrix

$$\boldsymbol{G} = \begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \end{pmatrix} \qquad 1 \text{ Zeile, } n \text{ Spalten,}$$

also

$$\boldsymbol{P} = \begin{pmatrix} 1 & 1 & \cdots & 1 \end{pmatrix} \qquad 1 \text{ Zeile, } n-1 \text{ Spalten.}$$

Damit ergibt sich die Kontrollmatrix

$$\boldsymbol{H} = \begin{pmatrix} 1 & 1 & 0 & \cdots & 0 \\ 1 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & \cdots & 1 \end{pmatrix} \qquad n-1 \text{ Zeilen, } n \text{ Spalten.}$$

Jede Auswahl von $n-1$ (oder weniger) Spalten ist linear unabhängig, also $\mathrm{d}(C) = n$.
Da aber $R = \frac{1}{n}$ geht für den Wiederholungscode zwar die Wahrscheinlichkeit einer falschen Decodierung für große n gegen Null, aber auch die Coderate.

Beispiel 8.11

Der allgemeine Parity-Check-Code, der jeweils l Nutzdatenbits zu Codewörtern der Länge $n = l + 1$ mit gerader Anzahl von Einsen erweitert, hat die Generatormatrix

$$\boldsymbol{G} = \begin{pmatrix} 1 & 0 & \cdots & 0 & 1 \\ 0 & 1 & \cdots & 0 & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 1 \end{pmatrix} \qquad n-1 \text{ Zeilen, } n \text{ Spalten.}$$

also

$$\boldsymbol{P} = \begin{pmatrix} 1 & 1 & \cdots & 1 \end{pmatrix}^T \qquad n-1 \text{ Zeilen, } 1 \text{ Spalte.}$$

Damit ergibt sich die Kontrollmatrix

$$\boldsymbol{H} = \begin{pmatrix} 1 & 1 & \cdots & 1 & 1 \end{pmatrix} \qquad 1 \text{ Zeile, } n \text{ Spalten,}$$

Es gibt (mindestens) zwei gleiche Spalten, also $\mathrm{d}(C) = 2$.
Da keine Fehler korrigiert werden können, geht die Wahrscheinlichkeit einer falschen Decodierung für große n gegen Eins, aber auch die Coderate $R = \frac{n-1}{n}$.

8.3 Hamming-Codes

Hamming-Codes sind eine Familie von perfekten Codes, die $t = 1$ Fehler korrigieren können.
Zur Erinnerung: Für perfekte Codes gilt (vgl. Satz 7.3)

$$2^{n-l} = \sum_{i=0}^{t} \binom{n}{i}. \tag{8.11}$$

Hier $t = 1$, also

$$2^{n-l} = \sum_{i=0}^{1} \binom{n}{i} = 1 + n, \tag{8.12}$$

bzw. mit der Anzahl $k = n - l$ der Prüfbits

$$2^k = n + 1. \tag{8.13}$$

k	2	3	4	5	6	7	8	⋯
n	3	7	15	31	63	127	255	⋯
l	1	4	11	26	57	120	247	⋯

Für den Entwurf von Hamming-Codes ist zu berücksichtigen:

- Für vorgegebene Anzahl k an Prüfbits haben die Codewörter die Länge $n = 2^k - 1$.
- Die Kontrollmatrix hat also $n = 2^k - 1$ Spalten und $n - l = k$ Zeilen.
- Der Code soll $t = 1$ Fehler korrigieren können, daher muss $\mathrm{d}(C) \geq 3$.
- Die Spalten der Kontrollmatrix müssen also von $\mathbf{0}$ verschieden sein (sonst $\mathrm{d}(C) = 1$) und paarweise voneinander verschieden sein (sonst $\mathrm{d}(C) = 2$).

Jede Spalte der Kontrollmatrix besteht aus k Bits, es gibt also 2^k Möglichkeiten für eine Spalte, von denen jedoch eine, der Nullvektor entfällt. Die verbleibenden $2^k - 1$ Möglichkeiten bilden genau die $2^k - 1$ verschiedenen Spalten.
Die Reihenfolge der Spalten ist belibig; zweckmäßigerweise wird ein systematischer Code gebildet.

Beispiel 8.12 (Hamming-Code für $k = 2$)

Für $k = 2$ lautet eine mögliche Kontrollmatrix

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

Die Generatormatrix lautet damit

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

Beispiel 8.13 (Hamming-Code für $k = 3$)

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

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

Beispiel 8.14 (Hamming-Code für $k = 4$)

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

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

8.4 Zusammenfassung

- Lineare Codes erlauben eine Codierung durch Multiplikation mit einer Generatormatrix.
- Die Überprüfung, ob ein Wort in einem Code enthalten ist, kann durch Multiplikation mit einer Kontrollmatrix erfolgen.
- Für systematische Codes lässt sich leicht eine Kontrollmatrix zu einer Generatormatrix finden und umgekehrt.
- Hamming-Codes sind eine Klasse von perfekten 1-Fehler-korrigierenden Codes mit leicht zu konstruierenden Kontrollmatrizen.

Aber wie decodiert man, wenn man Fehler nicht nur erkennen, sondern auch korrigieren möchte?