

Miniprojekt 3: Fejlkorligerende køder $\mapsto$ Fejlkorrigerende koder

*Denne note er skrevet med udgangspunkt i [1, p. 240-243, 249].
Et videre studium kan eksempelvis tage udgangspunkt i [2].
Eventuelle kommentarer kan sendes til olav@math.aau.dk*

I kurset “Lineær Algebra” arbejder vi normalt med matricer, hvor elementerne er reelle tal. Denne note behandler tilfældet, hvor elementerne i matricerne er tal i talstrukturen $\mathbb{Z}_2 = \{0, 1\}$. Regnereglerne i $\mathbb{Z}_2$ er som følger:

$$\begin{array}{c|cc} + & 0 & 1 \\ \hline 0 & 0 & 1 \\ 1 & 1 & 0 \end{array} \quad \begin{array}{c|cc} \cdot & 0 & 1 \\ \hline 0 & 0 & 0 \\ 1 & 0 & 1 \end{array}$$

Bemærk, at $1 + 1 = 0 = 1 - 1$. Derfor er plus og minus det samme i $\mathbb{Z}_2$.

Eksempel 1 *Betragt vektorene*

$$\vec{u} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix} \quad \vec{v} = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 1 \end{bmatrix}$$

i $\mathbb{Z}_2^4$. Vi har $\vec{u} \cdot \vec{v} = 1 \cdot 1 + 0 \cdot 1 + 0 \cdot 0 + 1 \cdot 1 = 1 + 0 + 0 + 1 = 1 + 1 = 0$.

I Matlab udregnes prikproduktet sådan her:

```
>> u=[1 0 0 1], v=[1 1 0 1]
```

```
u =
```

```
1      0      0      1
```

```
v =
```

```
1      1      0      1
```

```
>> mod(u*transpose(v),2)
```

```
ans =      0
```

De fleste resultater, I kender for matricer over $\mathbb{R}$, gælder også for matricer over $\mathbb{Z}_2$. Men ikke alle...

Opgave 1 Udregn ved hjælp af Matlab $\vec{u} \cdot \vec{u}$ og $\vec{v} \cdot \vec{v}$ i $\mathbb{Z}_2$. Kan du få øje på et fænomen, som aldrig ville finde sted, hvis du havde regnet med reelle tal?

Betragt situationen, hvor vi vil kommunikere over en støjfyldt kanal (tænk på et langt ikke-afskærmet datakabel). Vi ønsker at sende en følge af 0'er og 1'er svarende til indholdet af en fil. At kanalen er støjfyldt betyder, at nogle af 0'erne bliver lavet om til 1'er, inden de når frem til modtageren. Tilsvarende bliver nogle af 1'erne lavet om til 0'er. En simpel måde, hvorpå vi kan beskytte vores datatransmission, er hvis afsenderen sender tre kopier af hvert enkelt binært symbol. Istedet for 0 sendes altså 000, og istedet for 1 sendes 111. De tilsvarende vektorer

$$\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

kaldes kodeord. Antag nu, at modtageren i den anden ende af informationsskanalen modtager 110. Denne vil nu regne med, at der nok er sket fejl i tredje position, hvorfor den sendte besked antages at være 1. Tilsvarende vil man dekode 001 til 000, og dermed vil man antage, at den sendte besked er 0. Bemærk, at vi kan rette én fejl, men ikke to eller tre. Lad os bringe ovenstående på matrix form. Beskederne 0 og 1 identificeres med vektorerne $[0]$ og $[1]$, og som nævnt ovenfor skrives kodeordene som søjlevektorer af længde 3.

Mængden af kodeord kaldes en kode.

Indkodningen ovenfor af beskeder til kodeord svarer til den lineære transformation

$$T : \mathbb{Z}_2 \rightarrow \mathbb{Z}_2^3$$

som repræsenteres af matricen

$$G = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

Altså er indkodningen givet ved $G\vec{m}$, hvor $\vec{m}$ tilhører $\mathbb{Z}_2 = \{0, 1\}$. Matricen G kaldes en generatormatrix.

Opgave 2 Indkod i Matlab beskeden 1. Altså udregn $G\vec{m}$. Husk, Matlabkoden for $\vec{m}$ er

```
>> m=[1]
```

```
m =
```

```
1
```

Hvis vi vil tjekke, om

$$\vec{c} = \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} \in \mathbb{Z}_2^3$$

er et kodeord, skal vi undersøge, om alle tre komponenter er ens. Dette svarer til at tjekke, om

$$\begin{cases} c_1 = c_2 \\ c_1 = c_3 \end{cases}$$

holder, hvilket er det samme som at tjekke, om

$$\begin{cases} c_1 + c_2 = 0 \\ c_1 + c_3 = 0 \end{cases}$$

holder (husk $+$ og $-$ er det samme), hvilket igen er det samme som at tjekke om

$$\begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

holder. Matricen

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

kaldes for en paritetstjekmatrix.

Opgave 3 Lad $\vec{u} = [0 \ 0 \ 0]^T$ og $\vec{v} = [0 \ 1 \ 0]^T$. Udregn i Matlab $H\vec{u}$ og $H\vec{v}$ og konkluder hvilket af ordene, der er et kodeord.

Da $\vec{c} = [c_1, c_2, c_3]^T$ altså er et kodeord præcis hvis $H\vec{c} = \vec{0}$, ja da er koden simpelthen nulrummet af H .

Eksempel 2 Koden $\{[0 \ 0 \ 0]^T, [1 \ 1 \ 1]^T\}$ har dimension 1. Fra SIF side 47 øverst plus SIF side 255 nederst har vi følgende resultat. Givet en matrix, da er dimensionen af nulrummet lig med antallet af pivot-frie søjler. Da koden er nulrummet af H , så må H altså have præcis en pivot-fri søjle. Dette tjekkes med Matlabkommandoerne:

```
>> H=[1 1 0; 1 0 1]
```

```
H =
```

```

1      1      0
1      0      1
```

```
>> H(2,:)=mod(H(2,:)-H(1,:),2)
```

```
H =
```

```

1      1      0
0      1      1
```

Antag, at $\vec{c} = [1 \ 1 \ 1]^T$ er afsendt, men at $\vec{r} = [1 \ 0 \ 1]^T$ er modtaget. Vi har $\vec{r} = \vec{c} + \vec{e}$, hvor $\vec{e} = [0 \ 1 \ 0]^T$ er fejlvektoren. I stedet for den tidligere beskrevne dekodningsmetode (den, hvor vi ser hvilket symbol 0 eller 1, der forekommer flest gange) dekoder vi nu vha. paritetstjekmatricen. Det vil hjælpe os til at demonstrere et snedigt generelt princip. Vi udregner

$$H\vec{r} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}. \quad (1)$$

Da dette er forskelligt fra $[0 \ 0]^T$, er der sket fejl. Bemærk, at $H\vec{r} = H(\vec{c} + \vec{e})$, hvor $\vec{e}$ er fejlvektoren beskrevet ovenfor. Men $H\vec{c} = \vec{0}$, og derfor har vi $H\vec{e} = H\vec{r} = [1 \ 0]^T$. Læg mærke til, at hver ikke-nul vektor i $\mathbb{Z}_2^2$ forekommer præcis én gang som søjle i H , samt at højre side af (1) svarer til anden søjle i H . Vi konkluderer, at hvis der kun er sket én fejl, da er det i position 2. Altså $\vec{e} = [0 \ 1 \ 0]^T$, og dermed

$$\vec{c} = \vec{r} - \vec{e} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} - \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

Den afsendte besked er altså lig 1.

Opgave 4 Du modtager $\vec{r} = [0 \ 1 \ 1]^T$. Udregn i Matlab $H\vec{r}$. Brug metoden ovenfor til at rette fejl. Gentag med $\vec{r} = [1 \ 0 \ 0]^T$.

Vi generaliserer nu ovenstående konstruktion af koder.

Definition 1 Et k -dimensionalt underum af $\mathbb{Z}_2^n$ kaldes en $[n, k]$ -kode. Givet en $[n, k]$ -kode C , da kaldes en $n \times k$ matrix G for en tilhørende generatormatrix, når søjlerne i G er en basis for C . Tilsvarende kaldes en $(n - k) \times n$ matrix H for en paritetstjekmatrix for C , når C er nulrummet hørende til H .

En $[n, k]$ -kode kan altså bruges til at indkode beskeder af længde k . En besked $\vec{m} \in \mathbb{Z}_2^k$ identificeres med et kodeord $\vec{c} \in \mathbb{Z}_2^n$ ved $\vec{c} = G\vec{m}$.

Definition 2 Lad $k < n$. En $[n, k]$ kode kaldes systematisk, hvis den har en generatormatrix af formen

$$G = \begin{bmatrix} I_k \\ A \end{bmatrix}.$$

(Her er A selvfølgelig en $(n - k) \times k$ matrix).

Bemærkning 1 Givet et kodeord i en systematisk kode, da optræder den indkodede besked som de første k komponenter.

Opgave 5 Indtast i Matlab en binær 7×3 matrix G som i Definition 2. Indkod forskellige beskeder (af længde 3) og tjek at Bemærkning 1 holder. Husk, "mod" kommandoen.

Sætning 1 Lad C være en systematisk $[n, k]$ -kode med generatormatrix

$$G = \begin{bmatrix} I_k \\ A \end{bmatrix}.$$

Så er $H = [A \ I_{n-k}]$ en paritetstjekmatrix for C (dvs. $\vec{c} \in C$, hvis og kun hvis $H\vec{c} = \vec{0}$).

Bevis: Lad $\vec{c} \in C$ være et vilkårligt kodeord, dvs. $\vec{c}$ er af formen $\vec{c} = G\vec{m}$. Men så er

$$\begin{aligned} H\vec{c} &= HG\vec{m} = [A \ I_{n-k}] \begin{bmatrix} I_k \\ A \end{bmatrix} \vec{m} \\ &= (A + A)\vec{m} = O\vec{m} = \vec{0}. \end{aligned}$$

Vi har nu vist, at C er indeholdt i nulrummet hørende til H . Men såvel nulrummet hørende til H som C har dimension k , så C må være lig med nulrummet. $\square$

Eksempel 3 *Koden med paritetstjekmatrix*

$$H = \begin{bmatrix} 1 & 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 & 1 \end{bmatrix} = \left[\begin{bmatrix} 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \right] = [A \ I_3]$$

kaldes en binær Hammingkode. Ifølge Definition 2 og Sætning 1 er den systematisk med generatormatrix

$$G = \begin{bmatrix} I_4 \\ A \end{bmatrix}.$$

Der er således tale om en $[7,4]$ -kode. Bemærk, at søjlerne i H består af samtlige ikke-nul vektorer. Hver af disse forekommer præcis én gang. Beskeden $\vec{m} = [1 \ 0 \ 1 \ 0]^T$ indkodes nu til

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \\ 1 \\ 0 \\ 1 \end{bmatrix}.$$

Der sker nu fejl under transmissionen, således modtages

$$\vec{r} = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}.$$

Modtageren kender ej det afsendte ord, men leder efter $\vec{c} \in C$, således at der gælder $\vec{r} = \vec{c} + \vec{e}$, hvor $\vec{e}$ kun har én ikke-nul komponent. Der regnes

$$H\vec{r} = \begin{bmatrix} 1 & 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}.$$

Bemærk, at $H\vec{c} = \vec{0}$ (jævnfør definitionen af H), og derfor har vi $H\vec{r} = H(\vec{c} + \vec{e}) = H\vec{e}$. Hvis der kun er én ikke-nul komponent i $\vec{e}$, da må det være den femte, da $[1\ 0\ 0]^T$ er den femte søjle i H . Modtageren dekode derfor korrekt til

$$\vec{c} = \vec{r} + \vec{e} = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \\ 1 \\ 0 \\ 1 \end{bmatrix}.$$

Vi har i Eksempel 3 argumenteret for, at man kan bruge følgende dekodningsalgoritme for Hammingkoden:

1. Modtageren udregner $H\vec{r}$.
2. Hvis $H\vec{r} = \vec{0}$, da konkluderes det, at der nok ikke er sket fejl.
3. Hvis $H\vec{r} \neq \vec{0}$, da undersøges hvilket søjlenummer j , som svarer hertil. Fejlen antages at være sket i position j , og der dekodes i overensstemmelse hermed.

Opgave 6 Betragt den binære Hammingkode i Eksempel 3. Indkod ved hjælp af Matlab beskeden $\vec{m} = [1\ 0\ 1\ 1]^T$.

Opgave 7 Betragt den binære Hammingkode i Eksempel 3. Du modtager ordet $\vec{r} = [1\ 1\ 0\ 0\ 1\ 1\ 0]^T$. Benyt Matlab og metoden beskrevet i eksemplet til at dekode.

Opgave 8 Betragt den binære Hammingkode fra Eksempel 3. Bring ved hjælp af rækkeoperationer H på række-echelonform (trappeform). Find antal pivotfrie søjler og bestem dimensionen af koden. Udover kommandoerne fra Eksempel får du muligvis brug for noget i stil med følgende kommando, som bytter om på række 1 og række 3:

```
>> H([1 3], :) = H([3 1], :)
```

References

- [1] David Poole, “Linear Algebra - A Modern Introduction,” Sec. ed., Thomson, 2006

- [2] J. Justesen og T. Høholdt, “ A Course In Error-Correcting Codes,” European Mathematical Society, *Textbooks in Mathematics*, 2004