LinAlgDat

Hamming code

De reelle tal $\mathbb{R}$ er et eksempel på et såkaldt *legeme*. Vi skal ikke komme ind på den præcise definition af et legeme, men det indebærer blandt andet, at $\mathbb{R}$ er udstyret med addition $+$ og multiplikation $\cdot$, som opfylder nogle naturlige regneregler. En anden vigtig egenskab er, at ethvert ikke-nul element $x \in \mathbb{R}$ har en multiplikativ invers $x^{-1} \in \mathbb{R}$, som opfylder $xx^{-1} = 1$. De komplekse tal $\mathbb{C}$ er også et legeme. Et vigtigt eksempel fra datalogi er *legemet med to elementer*, som skrives $\mathbb{F}_2 = \{0,1\}$ ($0 =$ FALSE og $1 =$ TRUE), hvori addition og multiplikation er defineret som følger (også kaldet *regning modulo* 2):

$+$	0	1
0	0	1
1	1	0

Addition

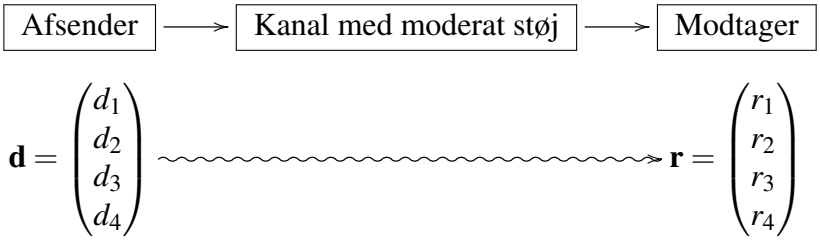
$\cdot$	0	1
0	0	0
1	0	1

Multiplikation

Meget af det vi lærer i LinAlgDat, som fx matrixregning og Gauss-Jordan eliminering, fungerer problemfrit hvis man erstatter de reelle tal $\mathbb{R}$ med et vilkårligt andet legeme, som fx $\mathbb{C}$ eller $\mathbb{F}_2$.

Nytten af matricer med indgange fra legemet $\mathbb{F}_2$ illustreres af fænomenet *Hamming code*:

En *afsender* sender en 4-bit databesked $\mathbf{d} = (d_1, d_2, d_3, d_4)$, hvor $d_i \in \mathbb{F}_2 = \{0,1\}$, til en *modtager*. Beskeden går gennem en *kanal med moderat støj*, hvilket betyder, at den besked $\mathbf{r} = (r_1, r_2, r_3, r_4)$ som modtageren får højst afviger én bit fra den oprindeligt afsendte besked $\mathbf{d}$.



Hvis fx modtageren får beskeden $\mathbf{r} = (1,0,0,1)$, så må den oprindeligt afsendte besked have været en blandt følgende:

$$\mathbf{d} = (1,0,0,1),\ (0,0,0,1),\ (1,1,0,1),\ (1,0,1,1),\ (1,0,0,0),$$

men det er som udgangspunkt umuligt at vide hvilken af ovenstående beskeder der er den korrekte. Hamming code løser dette problem på en elegant måde. Den grundlæggende idé er følgende:

I stedet for at afsende 4-bit beskeden $\mathbf{d} = (d_1, d_2, d_3, d_4)$, så afsender man en (kodet) 7-bit besked

$$\tilde{\mathbf{d}} = (\tilde{d}_1, \tilde{d}_2, \tilde{d}_3, \tilde{d}_4, \tilde{d}_5, \tilde{d}_6, \tilde{d}_7) := (p_1, p_2, d_1, p_3, d_2, d_3, d_4)\,.$$

Her er p_1, p_2, p_3 de såkaldte *paritets-bit*, som er snedigt konstrueret ud fra de oprindelige data-bit d_1, d_2, d_3, d_4 og som omtales nærmere nedenfor. Af tekniske grunde er de tre paritets-bit indsat på koordinatnumrene $1 = 2^0$, $2 = 2^1$ og $4 = 2^2$ i vektoren $\tilde{\mathbf{d}}$ mens de oprindelige data-bit udfylder de resterende koordinatnumre 3, 5, 6 og 7 i vektoren $\tilde{\mathbf{d}}$.

Modtageren får altså en 7-bit besked $\tilde{\mathbf{r}} = (\tilde{r}_1, \tilde{r}_2, \tilde{r}_3, \tilde{r}_4, \tilde{r}_5, \tilde{r}_6, \tilde{r}_7)$ som højst afviger én bit fra $\tilde{\mathbf{d}}$. Det smarte ved hele konstruktionen er, at uanset hvilken (om nogen) af de 7 bit i vektoren $\tilde{\mathbf{d}}$ der ændres, så er det faktisk muligt at genskabe den oprindeligt afsendte 4-bit besked $\mathbf{d} = (d_1, d_2, d_3, d_4)$.

Matematikken bag Hamming code er baseret på følgende to matricer (som har størelser hhv. 7×4 og 3×7) med indgange fra $\mathbb{F}_2 = \{0, 1\}$:

$$\mathbf{G} = \begin{pmatrix} 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \qquad \text{og} \qquad \mathbf{H} = \begin{pmatrix} 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix}.$$

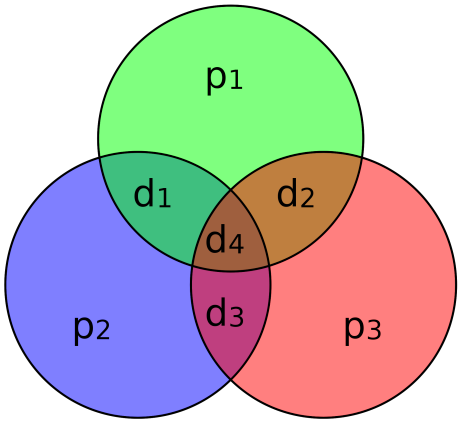
Matricen $\mathbf{G}$ kaldes *kodningsmatricen* og bruges til at skabe beskeden $\tilde{\mathbf{d}}$ ud fra $\mathbf{d}$; man definerer:

$$\tilde{\mathbf{d}} := \mathbf{G}\mathbf{d} = \begin{pmatrix} 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} d_1 \\ d_2 \\ d_3 \\ d_4 \end{pmatrix} = \begin{pmatrix} d_1 + d_2 + d_4 \\ d_1 + d_3 + d_4 \\ d_1 \\ d_2 + d_3 + d_4 \\ d_2 \\ d_3 \\ d_4 \end{pmatrix}.$$

Mao. så er de tre paritets-bit p_1, p_2 og p_3 i vektoren $\tilde{\mathbf{d}} = (p_1, p_2, d_1, p_3, d_2, d_3, d_4)$ givet ved:

$$\begin{aligned} p_1 &= d_1 + d_2 + d_4 \\ p_2 &= d_1 + d_3 + d_4 \\ p_3 &= d_2 + d_3 + d_4 \,. \end{aligned}$$

Husk at regning med bit foregår i legemet $\mathbb{F}_2 = \{0, 1\}$. Systemet i hvilke data-bit d_1, d_2, d_3, d_4 der bidrager til hvilke paritets-bit p_1, p_2, p_3 illustreres af følgende figur (et såkaldt Venn-diagram):



Matricen $\mathbf{H}$ kaldes *afkodningsmatricen* og er lavet sådan, at den i'te søjle i $\mathbf{H}$ giver tallet i i det bineære talsystem, altså:

i	i'te søjle i $\mathbf{H}$	Bineær fremstilling af tallet i
1	1 0 0	$1\cdot 2^0+0\cdot 2^1+0\cdot 2^2=1$
2	0 1 0	$0\cdot 2^0+1\cdot 2^1+0\cdot 2^2=2$
3	1 1 0	$1\cdot 2^0+1\cdot 2^1+0\cdot 2^2=3$
4	0 0 1	$0\cdot 2^0+0\cdot 2^1+1\cdot 2^2=4$
5	1 0 1	$1\cdot 2^0+0\cdot 2^1+1\cdot 2^2=5$
6	0 1 1	$0\cdot 2^0+1\cdot 2^1+1\cdot 2^2=6$
7	1 1 1	$1\cdot 2^0+1\cdot 2^1+1\cdot 2^2=7$

Bemærk, at pga. den specielle addition og multiplikation i legemet $\mathbb{F}_2=\{0,1\}$ har man:

$$\mathbf{HG}=\begin{pmatrix}1&0&1&0&1&0&1\\0&1&1&0&0&1&1\\0&0&0&1&1&1&1\end{pmatrix}\begin{pmatrix}1&1&0&1\\1&0&1&1\\1&0&0&0\\0&1&1&1\\0&1&0&0\\0&0&1&0\\0&0&0&1\end{pmatrix}=\begin{pmatrix}0&0&0&0\\0&0&0&0\\0&0&0&0\end{pmatrix}=\mathbf{O};$$

her er $\mathbf{O}$ altså nulmatricen af størrelse 3×4 med indgange fra legemet $\mathbb{F}_2=\{0,1\}$.

Hamming code fungerer nu således: Afsenderen har kodningsmatricen $\mathbf{G}$ og den 4-bit besked $\mathbf{d}$ han ønsker at sende; modtageren har afkodningsmatricen $\mathbf{H}$ og hvad end han måtte modtage fra afsenderen.

I stedet for at afsende den ønskede 4-bit besked $\mathbf{d}$, så afsendes 7-bit beskeden $\tilde{\mathbf{d}}=\mathbf{Gd}$; altså $\mathbf{d}$ kodet med matricen $\mathbf{G}$. Hvis man fx ønsker at sende $\mathbf{d}=(1,0,0,1)$, så afsender man i stedet for

$$\tilde{\mathbf{d}}=\mathbf{Gd}=\begin{pmatrix}1&1&0&1\\1&0&1&1\\1&0&0&0\\0&1&1&1\\0&1&0&0\\0&0&1&0\\0&0&0&1\end{pmatrix}\begin{pmatrix}1\\0\\0\\1\end{pmatrix}=\begin{pmatrix}0\\0\\1\\1\\0\\0\\1\end{pmatrix}. \qquad (*)$$

Modtageren kender som udgangspunkt hverken $\mathbf{d}$ eller $\tilde{\mathbf{d}}$; han har jo kun den modtagne besked $\tilde{\mathbf{r}}=(\tilde{r}_1,\tilde{r}_2,\tilde{r}_3,\tilde{r}_4,\tilde{r}_5,\tilde{r}_6,\tilde{r}_7)$ og afkodningsmatricen $\mathbf{H}$. Følgende resultat fortæller hvordan den oprindelige besked $\mathbf{d}$ rekonstrueres ud fra disse informationer.

Sætning. *Modtageren udregner vektoren* $\mathbf{z} := \mathbf{H}\tilde{\mathbf{r}}$*; husk at alle udregninger foregår i* $\mathbb{F}_2 = \{0,1\}$*. Lad* $\mathbf{z} = (z_0, z_1, z_2)$ *være koordinaterne i* $\mathbf{z}$ *og sæt* $j := z_0 2^0 + z_1 2^1 + z_2 2^2 \in \{0,\ldots,7\}$*.*

- *Hvis* $j \in \{0,1,2,4\}$ *så er den oprindelige besked* $\mathbf{d} = (\tilde{r}_3, \tilde{r}_5, \tilde{r}_6, \tilde{r}_7)$*.*
- *Hvis* $j = 3$ *så er den oprindelige besked* $\mathbf{d} = (\tilde{r}_3 + 1, \tilde{r}_5, \tilde{r}_6, \tilde{r}_7)$*.*
- *Hvis* $j = 5$ *så er den oprindelige besked* $\mathbf{d} = (\tilde{r}_3, \tilde{r}_5 + 1, \tilde{r}_6, \tilde{r}_7)$*.*
- *Hvis* $j = 6$ *så er den oprindelige besked* $\mathbf{d} = (\tilde{r}_3, \tilde{r}_5, \tilde{r}_6 + 1, \tilde{r}_7)$*.*
- *Hvis* $j = 7$ *så er den oprindelige besked* $\mathbf{d} = (\tilde{r}_3, \tilde{r}_5, \tilde{r}_6, \tilde{r}_7 + 1)$*.*

Bevis. Fordi transmissionen foregår gennem en kanal med moderat støj vides det, at den modtagne besked $\tilde{\mathbf{r}}$ højst afviger én bit fra den afsendte $\tilde{\mathbf{d}} = \mathbf{G}\mathbf{d}$. Man har altså $\tilde{\mathbf{r}} = \tilde{\mathbf{d}}$ i det tilfælde hvor ingen bit er ændret, og $\tilde{\mathbf{r}} = \tilde{\mathbf{d}} + \mathbf{e}_i$ i det tilfælde at netop én bit (den i'te bit) er ændret. Her er $\mathbf{e}_i$ vektoren med 1 på den i'te koordinat og 0 på de øvrige koordinater, fx er $\mathbf{e}_3 = (0,0,1,0,0,0,0)$.

I det tilfælde hvor $\tilde{\mathbf{r}} = \tilde{\mathbf{d}}$, så bliver $\mathbf{z}$ nul-vektoren idet jo:

$$\mathbf{z} = \mathbf{H}\tilde{\mathbf{r}} = \mathbf{H}\tilde{\mathbf{d}} = \mathbf{H}\mathbf{G}\mathbf{d} = \mathbf{O}\mathbf{d} = \mathbf{0}.$$

Altså er $j = 0\cdot 2^0 + 0\cdot 2^1 + 0\cdot 2^2 = 0$. Da $\tilde{\mathbf{d}} = (p_1, p_2, d_1, p_3, d_2, d_3, d_4)$ er lig med $\tilde{\mathbf{r}} = (\tilde{r}_1, \tilde{r}_2, \tilde{r}_3, \tilde{r}_4, \tilde{r}_5, \tilde{r}_6, \tilde{r}_7)$ følger det, at $\mathbf{d} = (d_1, d_2, d_3, d_4) = (\tilde{r}_3, \tilde{r}_5, \tilde{r}_6, \tilde{r}_7)$, som påstået i sætningen.

I det tilfælde hvor $\tilde{\mathbf{r}} = \tilde{\mathbf{d}} + \mathbf{e}_i$ for et eller andet indeks $i = 1,\ldots,7$ (der jo som udgangspunkt er ukendt for modtageren), så bliver vektoren $\mathbf{z} = (z_0, z_1, z_2)$ den i'te søjle i $\mathbf{H}$ fordi:

$$\mathbf{z} = \mathbf{H}\tilde{\mathbf{r}} = \mathbf{H}(\tilde{\mathbf{d}} + \mathbf{e}_i) = \mathbf{H}\tilde{\mathbf{d}} + \mathbf{H}\mathbf{e}_i = \mathbf{0} + \mathbf{H}\mathbf{e}_i = \mathbf{H}\mathbf{e}_i\,.$$

Den måde hvorpå $\mathbf{H}$ er konstrueret giver derfor, at $i = z_0 2^0 + z_1 2^1 + z_2 2^2 = j$ (mao. det a priori ukendte indeks i viser sig altså alligevel at være "kendt" for modtageren, idet det åbenbart netop er det tal j, altså $2^0 + z_1 2^1 + z_2 2^2$, som modtageren beregnede ud fra vektoren $\mathbf{z}$). De i sætningen angivne formler for $\mathbf{d}$ følger nu ved at gennemgå de 7 forskellige muligheder for j. Hvis fx $j = 1$ så er altså $\tilde{\mathbf{r}} = (\tilde{r}_1, \tilde{r}_2, \tilde{r}_3, \tilde{r}_4, \tilde{r}_5, \tilde{r}_6, \tilde{r}_7)$ lig med $\tilde{\mathbf{d}} + \mathbf{e}_1 = (p_1 + 1, p_2, d_1, p_3, d_2, d_3, d_4)$ og det følger, at

$$\mathbf{d} = (d_1, d_2, d_3, d_4) = (\tilde{r}_3, \tilde{r}_5, \tilde{r}_6, \tilde{r}_7)\,.$$

Hvis fx $j = 3$ så er $\tilde{\mathbf{r}} = (\tilde{r}_1, \tilde{r}_2, \tilde{r}_3, \tilde{r}_4, \tilde{r}_5, \tilde{r}_6, \tilde{r}_7)$ lig med $\tilde{\mathbf{d}} + \mathbf{e}_3 = (p_1, p_2, d_1 + 1, p_3, d_2, d_3, d_4)$ og det følger, at

$$\mathbf{d} = (d_1, d_2, d_3, d_4) = (\tilde{r}_3 + 1, \tilde{r}_5, \tilde{r}_6, \tilde{r}_7)\,,$$

som påstået. ///

Lad os se hvordan Hamming code og sætningen ovenfor virker i praksis: Hvis afsenderen ønsker at sende beskeden $\mathbf{d} = (1,0,0,1)$ så skal han i stedet for sende

$$\tilde{\mathbf{d}} = \mathbf{G}\mathbf{d} = (0,0,1,1,0,0,1)\,,$$

som udregnet i $(*)$ ovenfor. Fordi transmissionen foregår gennem en kanal med moderat støj vil den modtagne besked $\tilde{\mathbf{r}}$ højst afvige én bit fra $\tilde{\mathbf{d}}$.

Lad os sige at modtageren får beskeden

$$\tilde{\mathbf{r}} = (\tilde{r}_1, \tilde{r}_2, \tilde{r}_3, \tilde{r}_4, \tilde{r}_5, \tilde{r}_6, \tilde{r}_7) = (0,0,1,1,0,0,1)\,.$$

(Kanalen med moderat støj ændrede i dette tilfælde ingen bit i $\tilde{\mathbf{d}}$, men det ved modtageren jo som udgangspunkt ikke noget om.) Han udregner nu vektoren

$$\mathbf{z} = \mathbf{H}\tilde{\mathbf{r}} = \begin{pmatrix} 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

Eftersom $\mathbf{z} = (0,0,0)$ er $j = 0 \cdot 2^0 + 0 \cdot 2^1 + 0 \cdot 2^2 = 0$, så modtageren konkluderer ifølge sætningen ovenfor, at den oprindelige databesked $\mathbf{d}$ er $(\tilde{r}_3, \tilde{r}_5, \tilde{r}_6, \tilde{r}_7) = (1,0,0,1)$, hvilket jo er korrekt!

Lad os sige at modtageren får beskeden

$$\tilde{\mathbf{r}} = (\tilde{r}_1, \tilde{r}_2, \tilde{r}_3, \tilde{r}_4, \tilde{r}_5, \tilde{r}_6, \tilde{r}_7) = (0,0,1,1,1,0,1)\,.$$

(Kanalen med moderat støj ændrede i dette tilfælde den 5'te bit i $\tilde{\mathbf{d}}$, men det ved modtageren jo som udgangspunkt ikke noget om.) Han udregner nu vektoren

$$\mathbf{z} = \mathbf{H}\tilde{\mathbf{r}} = \begin{pmatrix} 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 1 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}.$$

Eftersom $\mathbf{z} = (1,0,1)$ er $j = 1 \cdot 2^0 + 0 \cdot 2^1 + 1 \cdot 2^2 = 5$, så modtageren konkluderer ifølge sætningen ovenfor, at den oprindelige databesked $\mathbf{d}$ er $(\tilde{r}_3, \tilde{r}_5 + 1, \tilde{r}_6, \tilde{r}_7) = (1, 1+1, 0, 1) = (1,0,0,1)$, hvilket jo er korrekt!

Henrik Holm (holm@math.ku.dk)
Henrik Laurberg Pedersen (henrikp@math.ku.dk)