LUNDS
UNIVERSITET

Matematikcentrum
Matematik, NF

ÄNDLIGA KROPPAR

OCH

FELRÄTTANDE KODER

Karl Gustav Andersson

Lund 1996

RINGAR OCH KROPPAR

1. Inledande definitioner och exempel

I detta förberedande avsnitt skall vi diskutera de grundläggande algebraiska operationerna *addition* och *multiplikation* från en abstrakt utgångspunkt. Vi skall betrakta mängder A på vilka två operationer är definierade på ett sådant sätt att till varje par av element a och b i A hör två nya element $a+b$ och $a \cdot b$ i A, kallade *summan* respektive *produkten* av a och b. Beträffande addition skall vi förutsätta att följande fyra axiom är uppfyllda

(A1) $$a + (b + c) = (a + b) + c$$

(A2) $$a + b = b + a$$

(A3) det finns ett element $0 \in A$ sådant att

$$a + 0 = a \quad \text{för alla } a \in A$$

(A4) för varje $a \in A$ finns ett element $-a \in A$ sådant att

$$a + (-a) = 0$$

Dessa axiom garanterar att *subtraktion* är väldefinierad i A. Man kontrollerar lätt att (A1)–(A4) medför att ekvationen $a + x = b$ i A har den entydiga lösningen $x = b + (-a)$. I fortsättningen skriver vi $b - a$ för $b + (-a)$.

Motsvarande axiom för multiplikation är

(M1) $$a \cdot (b \cdot c) = (a \cdot b) \cdot c$$

(M2) $$a \cdot b = b \cdot a$$

(M3) det finns ett element $1 \in A$ sådant att

$$a \cdot 1 = a \quad \text{för alla } a \in A$$

(M4) för varje $a \neq 0$ i A finns ett element $a^{-1} \in A$ sådant att

$$a \cdot a^{-1} = 1$$

Ibland kommer vi bara att förutsätta att vissa av axiomen för multiplikation är uppfyllda. Om samtliga gäller så följer, precis som för subtraktion ovan, att *division* är väldefinierad i A, dvs att ekvationen $ax = b$ med $a \neq 0$ har den entydiga lösningen $x = a^{-1}b$.

Slutligen antar vi alltid att de distributiva lagarna gäller i A:

(D) $$a \cdot (b+c) = a \cdot b + a \cdot c \quad \text{och} \quad (a+b) \cdot c = a \cdot c + b \cdot c$$

Definition. En *ring* A är en mängd på vilken addition och multiplikation är definierade så att samtliga lagar för addition (A1)–(A4) gäller och dessutom (M1) och (D) är uppfyllda. Om även (M2) gäller, kallas A en *kommutativ ring* och om (M3) gäller säger man att ringen har *etta*. En ring som innehåller minst två element och i vilken samtliga lagar för multiplikation (M1)–(M4) är uppfyllda, kallas en *kropp*.

Exempel 1. De rationella talen Q, de reella talen R och de komplexa talen C med de vanliga definitionerna av addition och multiplikation är exempel på kroppar. Heltalen Z bildar en kommutativ ring, men är ingen kropp eftersom (M4) inte gäller i Z.

Exempel 2. Mängden $M_2(R)$ av 2×2 matriser med reella element är en ring. Här är 0 nollmatrisen och 1 enhetsmatrisen. I $M_2(R)$ gäller inte den kommutativa lagen (M2). Inte heller (M4) är uppfylld, ty det finns matriser skilda från nollmatrisen som ej är inverterbara. Exempelvis är

$$\begin{pmatrix} 1 & -2 \\ -2 & 4 \end{pmatrix} \begin{pmatrix} 4 & -2 \\ 2 & -1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

och härav följer att ingen av matriserna i vänsterledet är inverterbar.

Definition. Två element $a \neq 0$ och $b \neq 0$ i en ring kallas *nolldelare* om $a \cdot b = 0$.

Exempel 3. Matriserna

$$\begin{pmatrix} 1 & -2 \\ -2 & 4 \end{pmatrix} \text{ och } \begin{pmatrix} 4 & -2 \\ 2 & -1 \end{pmatrix}$$

i Exempel 2 är nolldelare i ringen $M_2(R)$.

Vi skall nu lite närmare diskutera en familj av ringar som spelar en viktig roll i fortsättningen. Låt $n \geq 2$ vara ett givet heltal. Man säger att två heltal a och b är *kongruenta modulo* n om talet $a - b$ är delbart med n. Mer kortfattat skriver man $a \equiv b \pmod{n}$. T ex är $13 \equiv 4 \pmod{3}$. Beteckna med $[a]$ klassen av heltal som är kongruenta med a modulo n. Man kan då definiera addition och multiplikation av sådana kongruensklasser genom

$$[a] + [b] = [a + b] \qquad \text{och} \qquad [a] \cdot [b] = [a \cdot b] .$$

Här måste man dock verifiera att definitionerna inte beror på valet av representanter för kongruensklasserna. Antag alltså att $a \equiv a_1 \pmod{n}$ och $b \equiv b_1 \pmod{n}$. Då är $a_1 = a + kn$ och $b_1 = b + ln$ för några heltal k och l. Härav följer att

$$a_1 + b_1 = a + b + (k + l)n \qquad \text{och} \qquad a_1 b_1 = ab + (al + bk + kln)n,$$

så $a_1 + b_1$ är kongruent med $a + b$ och $a_1 b_1$ med ab modulo n. Beteckna med Z_n mängden av kongruensklasser modulo n, dvs $Z_n = \{[0], [1], [2], \ldots, [n-1]\}$. Man kontrollerar lätt att med ovanstående definitioner av addition och multiplikation blir Z_n en kommutativ ring.

Exempel 4. I Z_{11} är

$$[5] + [9] = [14] = [3] \quad \text{och} \quad [5] \cdot [9] = [45] = [1]$$

och i Z_{12} är

$$[4] + [9] = [13] = [1] \quad \text{och} \quad [4] \cdot [9] = [36] = [0].$$

Av exemplet framgår att [5] är multiplikativ invers till [9] i ringen Z_{11}. Följande sats ger ett kriterium för att ett element i ringen Z_n har en multiplikativ invers.

Sats 1. *Låt $[a] \in Z_n$ vara skild från [0]. Då finns ett element $[b]$ i Z_n sådant att $[a][b] = [1]$ om och endast om a och n är relativt prima, dvs inte har någon icke-trivial gemensam delare.*

Bevis. Antag först att a och n har en gemensam delare $d \geq 2$. Då är $a = kd$ och $n = ld$ för några heltal k och l med $0 < l < n$. Härav följer att $[l][a] = [lkd] = [kn] = [0]$. Därmed kan det inte finnas någon multiplikativ invers $[b]$ till $[a]$, ty i så fall skulle man ha

$$[l] = [l][1] = [l][a][b] = [0][b] = [0].$$

Om, omvänt, a och n är relativt prima så ger Euklides algoritm att det finns heltal b och c så att $1 = ab + nc$ och alltså är $[1] = [a][b]$.

Exempel 5. För att avgöra om [235] har en multiplikativ invers i Z_{567}, tillämpar vi Euklides algoritm

$$\begin{aligned} 567 &= 2 \cdot 235 + 97 \\ 235 &= 2 \cdot 97 + 41 \\ 97 &= 2 \cdot 41 + 15 \\ 41 &= 3 \cdot 15 - 4 \\ 15 &= 4 \cdot 4 - 1 \end{aligned}$$

Alltså är 567 och 235 relativt prima och genom att gå baklänges i räkningarna får man

$$1 = 4 \cdot 4 - 15 = 4 \cdot (3 \cdot 15 - 41) - 15 = 11 \cdot 15 - 4 \cdot 41 = \cdots = 63 \cdot 567 - 152 \cdot 235.$$

Den multiplikativa inversen till [235] är således $[-152] = [415]$.

Om $n = p$ är ett primtal, så har naturligtvis inget av talen $1, 2, \ldots, p-1$ någon icke-trivial delare gemensam med p. Alla klasser $[1], [2], \ldots, [p-1]$ skilda från $[0]$ i Z_p har således multiplikativ invers och följaktligen är Z_p en kropp. Om n ej är ett primtal, så gäller att $n = kl$ för några heltal $k, l \geq 2$. Ingen av klasserna $[k]$ och $[l]$ har då invers i Z_n och alltså är Z_n ingen kropp. Vi sammanfattar:

Sats 2. *Ringen Z_n är en kropp om och endast n är ett primtal.*

Avslutningsvis skall vi också definiera begreppet *isomorfi* mellan ringar. Låt A_1 och A_2 vara två ringar och antag att det finns en bijektiv funktion f från A_1 till A_2 sådan att

$$f(a+b) = f(a) + f(b) \qquad \text{och} \qquad f(a \cdot b) = f(a) \cdot f(b)$$

för alla element a och b i A_1. Man säger då att ringarna A_1 och A_2 är isomorfa och att funktionen f är en isomorfi från A_1 till A_2. Två ringar som är isomorfa är egentligen bara olika representationer av samma ring. En isomorfi svarar mot att byta beteckningar på elementen. Alla räkningar i den ena ringen speglas fullständigt av motsvarande räkningar i den andra ringen.

Exempel 6. Låt M vara ringen av alla 2×2 matriser av formen

$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix}$$

där a och b är reella tal och operationerna är vanlig matrisaddition och matrismultiplikation. Avbildningen

$$M \ni \begin{pmatrix} a & -b \\ b & a \end{pmatrix} \longrightarrow a + ib \in C$$

definierar då en isomorfi mellan M och ringen av komplexa tal. Läsaren uppmanas att kontrollera detta.

Övningar

1. Visa att följande räknelagar gäller i en godtycklig ring:
(a) $0 \cdot a = a \cdot 0 = 0$ (*Ledning*: $0 \cdot a + 0 \cdot a = 0 \cdot a$)
(b) $(-a)b = a(-b) = -ab$
(c) $(-a)(-b) = ab$

2. Visa att en kropp inte har några nolldelare.

3. Visa att om a inte är en nolldelare i ringen A, så gäller annuleringslagen

$$ax = ay \Rightarrow x = y$$

för x och y i A.

4. Låt M vara mängden av alla matriser

$$\begin{pmatrix} a & 2b \\ -b & a \end{pmatrix}$$

där a och b är heltal. Visa att, med vanlig addition och multiplikation av matriser, bildar M en kommutativ ring med etta. Har M några nolldelare?

5. Beteckna med $Q[\sqrt{2}]$ mängden av alla tal $a+b\sqrt{2}$, där a och b är rationella tal. Visa att $Q[\sqrt{2}]$, med vanlig addition och multiplikation av reella tal, är en kropp.

6. Låt $Z[i]$ beteckna mängden av Gaussiska heltal $a+ib$, där a och b är heltal. Visa att $Z[i]$, med vanlig addition och multiplikation av komplexa tal, är en kommutativ ring med etta. För vilka element $u \in Z[i]$ finns en multiplikativ invers v, dvs ett element v sådant att $uv = 1$?

7. Visa att en ring A är kommutativ om och endast om

$$(a+b)^2 = a^2 + 2ab + b^2$$

för alla a och b i A.

8. Avgör om determinanten

$$\begin{vmatrix} 325 & 131 & 340 \\ 142 & 177 & 875 \\ 214 & 122 & 961 \end{vmatrix}$$

är ett udda eller jämnt tal.

9. Lös i Z_{23} ekvationerna

$$[17]\cdot x = [5] \quad \text{och} \quad [12]\cdot x = [7]\,.$$

10. Undersök om [121] och [212] är inverterbara i Z_{9999}. Bestäm i förekommande fall inversen.

11. Betrakta elementen [39], [41], [46] och [51] i Z_{221}.
(a) Vilka av dessa är nolldelare?
(b) Vilka har en multiplikativ invers? Beräkna inversen.

12. Lös följande ekvationssystem

$$\text{(a)} \quad \begin{cases} 4x+7y \equiv 3 \pmod{11} \\ 8x+5y \equiv 9 \pmod{11} \end{cases} \qquad \text{(b)} \quad \begin{cases} 4x+7y \equiv 5 \pmod{13} \\ 7x+5y \equiv 8 \pmod{13} \end{cases}$$

13. Bestäm siffrorna x och y så att följande (decimal)tal blir delbara med 11

$$2x653874 \quad , \quad 37y64943252\,.$$

Ledning: $10^n \equiv (-1)^n \pmod{11}$.

14. Låt A vara en *ändlig* kommutativ ring med etta. Visa att om $a \in A$ inte är en nolldelare, så har a en multiplikativ invers.
Ledning: Betrakta avbildningen $x \to ax$, $x \in A$.

15. (a) Låt a vara ett element skilt från noll i en kropp K. Visa att om $a^{-1} = a$, så är $a = 1$ eller $a = -1$.
(b) Visa *Wilsons sats* som säger att för varje primtal p är

$$(p-1)! \equiv -1 \pmod{p}\,.$$

2. Räkning med kongruenser

Låt K vara en ändlig kropp med q element och sätt $K^* = \{x \in K \ ; \ x \neq 0\}$. Ordna elementen i K^* i en följd $x_1, x_2, \ldots, x_{q-1}$. För varje fixt $a \in K^*$ genomlöper då även elementen ax_i hela K^* när i går från 1 till $q-1$, ty om $ax_i = ax_j$ så ger multiplikation med a^{-1} att $x_i = x_j$. Vi har därför att

$$\prod_{i=1}^{q-1}(ax_i) = \prod_{i=1}^{q-1} x_i \, .$$

Bryter man ut a ur faktorerna på vänstra sidan och dividerar med $\prod_{i=1}^{q-1} x_i$, får man $a^{q-1} = 1$ och vi har därmed bevisat följande sats

Sats 3. *I en ändlig kropp K med q element gäller ekvationen*

$$a^{q-1} = 1$$

för varje $a \neq 0$ i K.

Specialisering till fallet då $K = Z_p$, för något primtal p, ger följande resultat av Pierre de Fermat från 1640:

Fermats lilla sats. *Om p är ett primtal och a ett heltal som ej delas av p, så är*

$$a^{p-1} \equiv 1 \pmod{p}.$$

Exempel 7. Vi vill beräkna den minsta positiva resten då 3^{350} divideras med 17. Eftersom 17 är ett primtal, ger Fermats sats att $3^{16} \equiv 1 \pmod{17}$. Alltså är

$$3^{350} = 3^{21 \cdot 16 + 14} \equiv 3^{14} \pmod{17}.$$

Fortsatt räkning modulo 17 ger

$$3^{14} = 9^7 = 9 \cdot 81^3 \equiv 9 \cdot (-4)^3 = 9 \cdot (-4) \cdot 16 \equiv 9 \cdot (-4) \cdot (-1) = 36 \equiv 2.$$

Den sökta resten är således 2.

Alternativt kan man visa att $3^{14} \equiv 2$ genom att observera att $3^{14} \cdot 3^2 = 3^{16} \equiv 1$. Härav följer att $[3^{14}] = [9]^{-1} = [2]$, ty $2 \cdot 9 = 18 \equiv 1$.

Nästa sats är en generalisering av Fermats sats.

Sats 4. *Låt p och q vara olika primtal och m ett positivt heltal. Då är*

$$a^{m(p-1)(q-1)+1} \equiv a \pmod{pq}$$

för varje heltal a.

Bevis. Om p inte delar a, ger Fermats sats att

$$a^{p-1} \equiv 1 \pmod{p}.$$

Alltså är

$$a^{m(p-1)(q-1)} \equiv 1 \pmod{p}.$$

Multiplikation med a ger

$$a^{m(p-1)(q-1)+1} \equiv a \pmod{p}.$$

Denna likhet gäller naturligtvis också då p delar a, ty då är $a \equiv 0 \pmod{p}$. På samma sätt ser man att

$$a^{m(p-1)(q-1)+1} \equiv a \pmod{q}.$$

Eftersom differensen $a^{m(p-1)(q-1)+1} - a$ delas av både p och q, måste den delas av $p \cdot q$ och satsen är bevisad.

Exempel 8. Sats 4 har en intressant tillämpning inom kryptologin. Antag att en mottagare, t ex en bank, tar emot meddelanden från ett stort antal olika avsändare och inte vill att innehållet skall kunna läsas av obehöriga. Då måste meddelandena krypteras. Det innebär att avsändarna måste ha tillgång till en krypteringsnyckel. En metod för att åstadkomma detta är att använda system med *öppen nyckel*. Dylika system bygger på idén att det finns funktioner som är lätta att beräkna, men för vilka inversen i praktiken är mycket svår att bestämma utan tillgång till extra information. Följande metod (RSA-systemet) föreslogs 1978 av Rivest, Shamir och Adelman.

Välj två stora[1] olika primtal p och q och sätt $n = pq$. Välj också ett stort tal d som saknar gemensam faktor med $(p-1)(q-1)$. Enligt Sats 1 i föregående avsnitt har d i ringen $Z_{(p-1)(q-1)}$ en multiplikativ invers e som kan beräknas med Euklides algoritm. Talen n och e publiceras öppet tillsammans med anvisningar för hur de skall användas för kryptering. Talen p, q och d håller mottagaren hemliga.

Antag att alla meddelanden har formen av ett eller flera heltal mellan 1 och n. En avsändare som vill sända ett sådant tal M krypterar det genom att beräkna $C \equiv M^e \pmod{n}$. När C når mottagaren, beräknar denne det entydiga tal D mellan 1 och n som uppfyller $D \equiv C^d \pmod{n}$. Då följer av Sats 4 att $D = M$. Eftersom e är den multiplikativa inversen till d i ringen $Z_{(p-1)(q-1)}$, har man nämligen att $ed = m(p-1)(q-1)+1$ för något heltal m. Alltså är

$$D \equiv C^d \equiv M^{ed} = M^{m(p-1)(q-1)+1} \equiv M \pmod{n}.$$

Frågan är nu om det är möjligt för en utomstående person att med hjälp av den öppna informationen e och n forcera ett krypterat meddelande. För att klara detta inom rimlig tid, behövs säkert primtalen p och q. Dessa kan bestämmas genom att

[1] Med stora tal menas här tal med hundratals siffror.

faktorisera n. I praktiken lär detta emellertid, med dagens teknik och de mycket stora tal det gäller, i allmänhet vara en hopplös uppgift.

Nästa exempel handlar om problemet att finna en gemensam lösning till flera olika kongruenser.

Exempel 9. I en omkring tvåtusen år gammal bok av den kinesiske författaren Sun-Tsu förekommer följande uppgift:

"Det finns ett obekant tal som vid division med 3 ger resten 2, med 5 resten 3 och med 7 resten 2. Vilket är talet?"

Det gäller, med andra ord, att finna ett tal x som samtidigt uppfyller de tre kongruenserna

$$\begin{aligned} x &\equiv 2 \pmod 3 \\ x &\equiv 3 \pmod 5 \\ x &\equiv 2 \pmod 7 \end{aligned}$$

Den metod som Sun-Tsu angav för att lösa problemet ger en allmän sats som vi först formulerar.

Kinesiska restsatsen. *Antag att talen $n_1, n_2, \ldots, n_k$ är parvis relativt prima. Då har kongruenserna*

$$\begin{aligned} x &\equiv a_1 \pmod{n_1} \\ x &\equiv a_2 \pmod{n_2} \\ &\cdots \\ x &\equiv a_k \pmod{n_k} \end{aligned}$$

en entydig lösning modulo $n = n_1 n_2 \cdots n_k$.

Bevis. Sätt

$$N_i = \frac{n}{n_i} = \prod_{j \neq i} n_j.$$

Då är talen N_i och n_i relativt prima för varje i. Alltså finns heltal s_i och t_i så att

$$s_i N_i + t_i n_i = 1.$$

Sätt

$$x = \sum_{j=1}^{k} a_j s_j N_j = a_1 s_1 N_1 + \cdots + a_k s_k N_k.$$

Man har $s_i N_i \equiv 1 \pmod{n_i}$ och $N_j \equiv 0 \pmod{n_i}$ då $j \neq i$. Följaktligen är

$$x \equiv a_i \pmod{n_i} \quad , \quad i = 1, \ldots, k.$$

Det återstår att visa att lösningen x är entydigt bestämd modulo n. Antag att $\tilde{x}$ vore en annan lösning. Då skulle man ha $x \equiv \tilde{x} \pmod{n_i}$ för alla i. Eftersom talen n_i är parvis relativt prima, följer härav att $x \equiv \tilde{x} \pmod{n}$ och satsen är bevisad.

Exempel 9 (fortsättning). Här är $n_1 = 3$, $n_2 = 5$, $n_3 = 7$ och $N_1 = 35$, $N_2 = 21$, $N_3 = 15$. Man har

$$\begin{aligned} 2 \cdot 35 - 23 \cdot 3 &= 1 \\ 1 \cdot 21 - \ 4 \cdot 5 &= 1 \\ 1 \cdot 15 - \ 2 \cdot 7 &= 1 \end{aligned}$$

Alltså är

$$x = 2 \cdot 2 \cdot 35 + 3 \cdot 1 \cdot 21 + 2 \cdot 1 \cdot 15 = 233$$

en lösning till problemet. Den minsta positiva lösningen är

$$233 - 2n = 233 - 210 = 23.$$

Den kinesiska restsatsen kan också ges en lite mer abstrakt formulering. Om $A_1, \ldots, A_k$ är k stycken ringar, så kan man bilda en ny ring som betecknas $A_1 \times \cdots \times A_k$ och som består av alla k-tipler $(a_1, \ldots, a_k)$ där $a_i \in A_i$. Addition och multiplikation i den nya ringen definieras genom

$$\begin{aligned} (a_1, \ldots, a_k) + (b_1, \ldots, b_k) &= (a_1 + b_1, \ldots, a_k + b_k) \\ (a_1, \ldots, a_k) \cdot (b_1, \ldots, b_k) &= (a_1 \cdot b_1, \ldots, a_k \cdot b_k). \end{aligned}$$

Antag nu att $n = n_1 n_2 \ldots n_k$ där talen n_i är parvis relativt prima. Då säger den kinesiska restsatsen att, för givna heltal $a_1, \ldots, a_k$ med $0 \leq a_i < n_i$, finns precis ett heltal a med $0 \leq a < n$ sådant att

$$a \equiv a_i \pmod{n_i} \quad , \quad i = 1, \ldots, k.$$

Man kontrollerar lätt att avbildningen som tar a på $(a_1, \ldots, a_k)$ är en isomorfi mellan Z_n och $Z_{n_1} \times \cdots \times Z_{n_k}$.

Exempel 10. Låt $n = 1001 = 7 \cdot 11 \cdot 13$ och betrakta de två elementen [778] och [431] i Z_{1001}. Man har

$$\begin{aligned} 778 &\equiv 1 \pmod{7} & 431 &\equiv 4 \pmod{7} \\ 778 &\equiv 8 \pmod{11} & 431 &\equiv 2 \pmod{11} \\ 778 &\equiv 11 \pmod{13} & 431 &\equiv 2 \pmod{13} \end{aligned}$$

I stället för att beräkna produkten $778 \cdot 431$ modulo 1001, kan man alltså beräkna produkten

$$(1, 8, 11) \cdot (4, 2, 2) = (4, 16, 22) \equiv (4, 5, 9)$$

i ringen $Z_7 \times Z_{11} \times Z_{13}$ och sedan, som i beviset för den kinesiska restsatsen, bestämma motsvarande element i Z_{1001}. Vid omfattande räkningar med stora tal kan denna typ av aritmetik ibland vara användbar.

Övningar

1. Bestäm den multiplikativa inversen till [45] i Z_{101}. Bestäm också talet x mellan 1 och 100 så att

$$45^{99} \equiv x \pmod{101}.$$

2. Ange, i vart och ett av följande fall, det minsta icke-negativa heltalet x som satisfierar

$$x \equiv 3^{5000} \pmod{13}, \qquad x \equiv 3^{100} \pmod{101}$$
$$x \equiv 3^{40} \pmod{23}, \qquad x \equiv 2^{1000} \pmod{7}.$$

3. Visa att

$$p^{q-1} + q^{p-1} \equiv 1 \pmod{pq}$$

om p och q är två skilda primtal.

4. Låt $p_1, p_2, \ldots, p_k$ vara olika primtal och r ett positivt heltal som delas av $p_i - 1$ för $i = 1, \ldots, k$. Visa att

$$a^{r+1} \equiv a \pmod{p_1 \cdot p_2 \cdots p_k}$$

för alla heltal a.

5. Visa att för alla heltal n är
(a) $n^7 \equiv n \pmod{42}$ (b) $n^{13} \equiv n \pmod{2730}$.
Ledning: Använd resultatet i föregående övning.

6. Bestäm det minsta positiva heltalet M, sådant att

$$M^{49} \equiv 21 \pmod{209}.$$

7. Visa att om p är ett primtal och m ett positivt heltal, så är

$$a^{(p-1)p^{m-1}} \equiv 1 \pmod{p^m}$$

för alla heltal a som inte är delbara med p.
Ledning: Kopiera beviset för Sats 3 med K^* lika med mängden av inverterbara element i Z_{p^m}.

8. Visa att för alla udda heltal k är
(a) $k^4 \equiv 1 \pmod{16}$ (b) $k^{2^n} \equiv 1 \pmod{2^{n+2}}$ då $n \geq 2$.

9. Bestäm alla heltal x sådana att

$$\begin{cases} x \equiv 1 \pmod{3} \\ x \equiv 3 \pmod{7} \\ x \equiv 7 \pmod{16}. \end{cases}$$

10. Finn det minsta positiva heltalet x som uppfyller

$$\begin{cases} 2x \equiv 9 \pmod{11} \\ 7x \equiv 2 \pmod{19} \end{cases}.$$

11. Verifiera att

$$\begin{cases} 95 \equiv 3 \pmod{23} \\ 95 \equiv 2 \pmod{31} \end{cases}$$

och använd detta till att beräkna $95^{36} \pmod{713}$.

3. Vektorrum

Definition. Ett *vektorrum* (eller ett *lineärt rum*) över en kropp K är en mängd V, som innehåller ett element betecknat $\underline{0}$, och på vilken man för varje par $u, v \in V$ och varje $\alpha \in K$ har definierat summan $u + v \in V$ och produkten $\alpha u \in V$ så att följande räknelagar är uppfyllda

(i) $u + (v + w) = (u + v) + w$

(ii) $u + v = v + u$

(iii) $\alpha(\beta u) = (\alpha\beta)u$

(iv) $1u = u$

(v) $0u = \underline{0}$

(vi) $\alpha(u + v) = \alpha u + \alpha v$

(vii) $(\alpha + \beta)u = \alpha u + \beta u$.

Anmärkning 1. Det följer lätt av dessa lagar att samtliga axiom för addition (A1)–(A4) i avsnitt 1 är uppfyllda i ett vektorrum. Av (iv), (v) och (vii) får man

$$u + \underline{0} = 1u + 0u = (1 + 0)u = 1u = u$$

så (A3) gäller. Axiomet (A4) verifieras på följande sätt

$$u + (-1)u = 1u + (-1)u = (1 + (-1))u = 0u = \underline{0}.$$

Anmärkning 2. Elementen i ett vektorrum kallas ofta *vektorer*. I (v) har vi satt ett streck under nollan i högerledet för att markera att den är en vektor. Fortsättningsvis betecknas även nollvektorn med 0.

Den grundläggande teorin ser likadan ut för ett vektorrum V över en allmän kropp K som i specialfallet då $K = R$. Ett antal vektorer $u_1, \ldots, u_l$ i V kallas *lineärt beroende* om det finns $\alpha_1, \ldots, \alpha_l \in K$, ej alla noll, så att

$$\alpha_1 u_1 + \cdots + \alpha_l u_l = 0.$$

Man säger att $u_1, \ldots, u_l$ är *lineärt oberoende* om de ej är lineärt beroende. Vektorerna $u_1, \ldots, u_l$ *genererar* vektorrummet V om varje vektor $u \in V$ är en *lineärkombination* av $u_1, \ldots, u_l$, dvs om

$$u = \alpha_1 u_1 + \cdots + \alpha_l u_l$$

för några $\alpha_1, \ldots, \alpha_l \in K$. En *bas* för V är en uppsättning vektorer $e_1, \ldots, e_n$ som är lineärt oberoende och som genererar V. Detta är ekvivalent med att varje vektor $u \in V$ entydigt kan skrivas

$$u = \alpha_1 e_1 + \cdots + \alpha_n e_n$$

med $\alpha_1, \ldots, \alpha_n \in K$. Koefficienterna $\alpha_1, \ldots, \alpha_n$ kallas *koordinaterna* för vektorn u i basen $e_1, \ldots, e_n$. Två olika baser för ett vektorrum består alltid av lika många vektorer och ett vektorrum säges ha *dimensionen* n om det har en bas av n vektorer. Om ett vektorrum V genereras av ett ändligt antal vektorer $v_1, \ldots, v_m$, så kan man alltid välja ut en bas bland dessa. Om vektorerna $v_1, \ldots, v_m$ är lineärt oberoende, bildar de en bas. I annat fall är någon av dem, t ex v_m, en lineärkombination av de övriga. Då genereras V av $v_1, \ldots, v_{m-1}$. Så kan man fortsätta och eliminera vektorer tills man har fått fram en lineärt oberoende mängd som genererar V.

Exempel 11. $K^n = \{(\alpha_1, \ldots, \alpha_n)\ ;\ \alpha_i \in K\}$, med komponentvis addition och multiplikation med element från K, är ett vektorrum över K. Varje vektorrum V av dimension n kan identifieras med K^n genom att välja en bas i V.

Exempel 12. Låt $\mathbf{k}$ vara en *delkropp* i en större kropp K. Detta innebär att $\mathbf{k} \subset K$ och att $\mathbf{k}$ själv är en kropp med samma operationer som i den stora kroppen K. För att så skall vara fallet krävs att $\mathbf{k}$ innehåller minst två element, att operationerna addition och multiplikation tillämpade på element i $\mathbf{k}$ inte för utanför $\mathbf{k}$ samt att $-\alpha$ och α^{-1} tillhör $\mathbf{k}$ för varje $\alpha \neq 0$ i $\mathbf{k}$. I denna situation kan man uppfatta K som ett vektorrum över delkroppen $\mathbf{k}$. Att axiomen (i)–(vii) för vektorrum är uppfyllda följer av räknelagarna i K. Om K är en ändlig kropp så genereras K naturligtvis av ändligt många vektorer, betraktad som vektorrum över $\mathbf{k}$. Det finns alltså en bas $e_1, \ldots, e_n$ av element i K sådan att varje $u \in K$ entydigt kan skrivas

$$u = \alpha_1 e_1 + \cdots + \alpha_n e_n$$

med $\alpha_1, \ldots, \alpha_n \in \mathbf{k}$. Som vektorrum över $\mathbf{k}$ har K då dimensionen n. Om p är antalet element i $\mathbf{k}$ så följer att antalet element i K är p^n, ty varje koordinat α_i kan ha p olika värden.

Vi skall senare återkomma mera utförligt till vektorrum över ändliga kroppar i samband med teorin för felrättande koder. Här skall vi bara visa hur man, med hjälp av Exempel 12, kan övertyga sig om att antalet element i en ändlig kropp måste vara en primtalspotens.

Låt alltså K vara en ändlig kropp och beteckna som vanligt med 1 ettan i K. Betrakta summorna

$$1\ ,\ 1+1\ ,\ 1+1+1\ ,\ \ldots\ ,\ m1\ ,\ \ldots$$

där $m1$ betyder summan av m stycken ettor. Eftersom K är ändlig, måste det finnas positiva heltal $r < s$ sådana att $r1 = s1$. Om $m = s - r$, gäller då att $m1 = 0$. Det minsta positiva heltalet p sådant att $p1 = 0$ kallas *karakteristiken* för kroppen K. Karakteristiken p måste vara ett primtal, ty om p vore produkten av två heltal p_1 och p_2 större än 1 skulle man ha

$$(p_1 1) \cdot (p_2 1) = p1 = 0$$

och därmed att $p_1 1 = 0$ eller $p_2 1 = 0$. Detta strider mot att p är det *minsta* positiva heltalet med $p1 = 0$. Sätt nu

$$\mathbf{k} = \{m1\ ;\ m \in Z\} = \{\ 0\ ,\ 1\ ,\ 1+1\ ,\ \ldots\ ,\ (p-1)1\ \}\ .$$

Man kontrollerar lätt att $\mathbf{k}$ är en delkropp i kroppen K och att avbildningen $m \longrightarrow m1$ ger en isomorfi mellan Z_p och $\mathbf{k}$. Eftersom $\mathbf{k}$ har p element, följer av Exempel 12 att kroppen K har p^n element för något positivt heltal n. Vi formulerar detta som en särskild sats.

Sats 5. *För varje ändlig kropp K finns ett primtal p och ett positivt heltal n sådant att antalet element i K är lika med p^n. Primtalet p är karakteristiken för kroppen.*

Anmärkning. Även för oändliga kroppar kan man definiera begreppet karakteristik, men här finns två fall. Antingen finns ett minsta positivt heltal p sådant att $p1 = 0$, och detta kallas då kroppens karakteristik, eller också är elementen $m1$ alltid olika för olika heltal m. I det senare fallet säger man att kroppens karakteristik är 0. Exempel på kroppar med karakteristik noll är Q, R och C.

Övningar

1. Låt V vara ett vektorrum över en kropp K. En delmängd U av V kallas ett *underrum* i V om

$$u, v \in U \Rightarrow \alpha u + \beta v \in U \quad , \quad \text{för alla } \alpha, \beta \in K\ .$$

(a) Kontrollera att varje underrum U i V är ett vektorrum med samma operationer som i V.
(b) Låt K vara kroppen Z_3 och U det underrum i K^4 som genereras av vektorerna

$$(0,1,2,1)\ ,\ (1,0,2,2)\ ,\ (1,2,0,1)\ .$$

Bestäm dimensionen för U och ange en bas i U.

2. Låt K vara en kropp med karakteristiken $p \neq 0$.
(a) Visa att $pa = 0$ för alla $a \in K$.
(b) Visa att

$$(a+b)^p = a^p + b^p$$

för alla $a, b \in K$.
Ledning: Visa först att för $0 < k < p$ är binomialkoefficienterna $\binom{p}{k}$ delbara med p.

3. (a) Visa att i en kropp med karaktcristikcn $p \neq 0$ är

$$(a_1 + a_2 + \cdots + a_l)^p = a_1^p + a_2^p + \cdots + a_l^p \ .$$

(b) Bevisa Fermats lilla sats genom att välja alla $a_i = 1$ i (a).

4. Polynomringar

Enligt Sats 5 måste en ändlig kropp ha p^n element, för något primtal p och något positivt heltal n. Hittills har vi bara mött kropparna Z_p för vilka $n = 1$. För att kunna konstruera kroppar där $n > 1$, behöver vi först diskutera polynom med koefficienter i en ändlig kropp.

Ett polynom med koefficienter i en kropp K är ett uttryck

$$f(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0 \tag{1}$$

där alla $a_i \in K$. Strängt taget är ett polynom bara en ändlig följd $a_0, a_1, \ldots, a_n$ av element i K och bokstaven x skall uppfattas som en *formell* symbol. Värdet $f(\alpha)$ av polynomet (1) i $\alpha \in K$ är

$$a_n \alpha^n + a_{n-1} \alpha^{n-1} + \cdots + a_1 \alpha + a_0 \in K \ .$$

Exempel 13. Betrakta polynomen

$$f(x) = x^3 + 1 \quad \text{och} \quad g(x) = x^4 + x^2 + x + 1$$

med koefficienter i Z_2 (observera att man inte skriver ut termer som har koefficienten 0). Värdena av f och g är lika för alla $\alpha \in Z_2$, dvs för $\alpha = 0$ eller 1, men de skall ändå uppfattas som *olika* polynom.

Om $a_n \neq 0$ i (1) säger man att polynomet $f(x)$ har graden n och om $a_n = 1$ kallas polynomet *moniskt*. Mängden av alla polynom med koefficienter i kroppen K betecknas $K[x]$. Addition och multiplikation av polynom definieras på vanligt sätt, som då koefficienterna ligger i R eller C. Likaså fungerar bevisen för divisionsalgoritmen, faktorsatsen och Euklides algoritm oförändrade för en allmän kropp. Divisionsalgoritmen säger att om f och g är polynom med grad $f \geq$ grad g, så finns polynom q och r sådana att

$$f(x) = q(x)g(x) + r(x) \ ,$$

där antingen $r(x)$ är nollpolynomet eller grad $r <$ grad g. Om r är nollpolynomet, säger man att g delar f och skriver $g|f$. Innebörden av faktorsatsen är att $f(\alpha) = 0$ om och endast om $(x - \alpha)$ delar $f(x)$. Euklides algoritm, slutligen, ger en metod för att bestämma den största gemensamma delaren till två polynom f och g. Att h är den största gemensamma delaren till f och g betyder att h delar både f och g samt att

varje annat polynom som delar både f och g måste dela h. Den största gemensamma delaren är inte helt entydigt bestämd, men två olika största gemensamma delare h_1 och h_2 kan bara skilja sig med en multiplikativ konstant. Man måste nämligen ha att h_1 delar h_2 och h_2 delar h_1. Detta är bara möjligt om $h_1 = ah_2$ för något $a \in K$. Om man kräver att den största gemensamma delaren till f och g skall vara ett moniskt polynom, är den entydigt bestämd och betecknas (f, g).

Exempel 14. Vi skall illustrera Euklides algoritm genom att beräkna den största gemensamma delaren till följande två polynom i $Z_3[x]$:

$$f(x) = x^5 + 2x^3 + x^2 + 2 \qquad g(x) = x^4 + 2x^3 + 2x^2 + 2x + 1\,.$$

Observera att, eftersom koefficienterna ligger i Z_3, så gäller i räkningarna t ex att $4 \equiv 1$ och att $2 \equiv -1$. (I fortsättningen sätter vi i allmänhet inte klammer runt elementen i Z_n).

$$\begin{aligned} x^5 + 2x^3 + x^2 + 2 &= (x+1)(x^4 + 2x^3 + 2x^2 + 2x + 1) + (x^3 + 1) \\ x^4 + 2x^3 + 2x^2 + 2x + 1 &= (x+2)(x^3 + 1) + (2x^2 + x + 2) \\ x^3 + 1 &= (2x + 2)(2x^2 + x + 2) \end{aligned}$$

Den sista icke-försvinnande resten $2x^2 + x + 2$ är en största gemensam delare till f och g. Motsvarande moniska polynom fås genom att multiplicera med $2^{-1} = 2$. Detta ger $(f, g) = x^2 + 2x + 1$.

Definition. Ett polynom $s(x)$ i $K[x]$ av grad $n \geq 1$ kallas *irreducibelt* om det inte har någon äkta delare, dvs om det inte finns något polynom $g(x)$ som delar $s(x)$ och för vilket $1 \leq \text{grad } g < n$. Irreducibla polynom kallas även *primpolynom*.

Exempel 15. Polynomet $f(x) = x^3 + 2x + 1$ är irreducibelt i $Z_3[x]$. För att kontrollera detta observerar vi att, eftersom gradtalet är 3, måste polynomet ha en förstagradsfaktor om det inte är ett primpolynom. Då skulle f ha ett nollställe i Z_3, men så är inte fallet ty $f(0) = 1$, $f(1) = 1$ och $f(-1) = 1$.

Vi skall bevisa att varje moniskt polynom i $K[x]$ kan skrivas som en produkt av moniska primpolynom och att denna produktframställning är entydig, så när som på ordningsföljden mellan primpolynomen. För detta ändamål behöver vi en hjälpsats.

Lemma. *Antag att f, g och h är tre polynom i $K[x]$ sådana att $f(x)$ delar produkten $g(x)h(x)$. Om f och g är relativt prima, dvs $(f, g) = 1$, så måste f dela h.*

Bevis. Eftersom $(f, g) = 1$, så kan man genom att gå baklänges i Euklides algoritm visa att det finns polynom $c(x)$ och $d(x)$ sådana att

$$1 = c(x)f(x) + d(x)g(x)\,.$$

Alltså är

$$h(x) = c(x)f(x)h(x) + d(x)g(x)h(x)\,.$$

Båda termerna på högra sidan delas av f. Därför delas även vänsterledet av f.

Sats 6. *Låt K vara en kropp och $f(x)$ ett moniskt polynom med koefficienter i K. Då finns ett antal olika moniska primpolynom $s_1(x), \ldots, s_l(x)$ i $K[x]$ och tillhörande positiva heltal $m_1, \ldots, m_l$ sådana att*

$$f(x) = s_1(x)^{m_1} \cdots s_l(x)^{m_l} .$$

Primpolynomen s_i och heltalen m_i är entydigt bestämda så när som på ordningsföljden.

Bevis. Att f kan skrivas som en produkt av primpolynom visas med induktion över graden av f. Då f har graden 1 är saken klar. Antag att graden av f är n och att påståendet är bevisat för polynom av lägre gradtal. Om f är ett primpolynom är vi klara. I annat fall är $f(x) = g_1(x)g_2(x)$ för några polynom g_1 och g_2 av lägre grad än n. Enligt induktionsantagandet kan dessa skrivas som en produkt av primpolynom. Därmed är existensen av primfaktoruppdelningen bevisad.

Det återstår att visa entydigheten. Antag att man hade två primfaktoruppdelningar av $f(x)$. För dessa skulle då gälla att

$$s_1(x)^{m_1} \cdots s_l(x)^{m_l} = t_1(x)^{n_1} \cdots t_j(x)^{n_j} . \tag{2}$$

Betrakta först $t_1(x)$. Vi skall visa att $t_1(x)$ är lika med något av polynomen $s_i(x)$ på vänstra sidan. Eftersom s_1 och t_1 är moniska primpolynom, gäller antingen att $s_1 = t_1$ eller att s_1 och t_1 är relativt prima. Om $s_1 = t_1$ är vi klara. Annars är $s_1(x)^{m_1}$ och $t_1(x)$ relativt prima. Enligt lemmat måste då $t_1(x)$ dela produkten

$$s_2(x)^{m_2} \cdots s_l(x)^{m_l} .$$

Så kan man fortsätta. Antingen är $t_1 = s_2$ eller också delar $t_1(x)$ produkten

$$s_3(x)^{m_3} \cdots s_l(x)^{m_l} .$$

Förr eller senare inträffar att $t_1(x) = s_i(x)$. Vi kan då förkorta med $t_1(x)$ på båda sidor i (2) och upprepa proceduren. När alla faktorer $t_i(x)$ har tagit slut på högra sidan, måste också faktorerna $s_i(x)$ på vänstra sidan ha tagit slut, ty annars skulle man ha en produkt av polynom $s_i(x)$ som var lika med 1, vilket är orimligt. Därmed är entydigheten av faktoruppdelningen bevisad.

För en given kropp K bildar mängden $K[x]$, försedd med operationerna polynomaddition och polynommultiplikation, en ring. Som bör ha framgått av det ovanstående, finns det stora likheter mellan $K[x]$ och ringen Z av heltal. I båda ringarna gäller divisionsalgoritmen och Euklides algoritm och dessutom har man entydig primfaktoruppdelning både i Z och i $K[x]$. Primtalen i Z motsvaras av primpolynomen i $K[x]$. Vi skall nu i $K[x]$ kopiera konstruktionen av ringarna Z_n. Låt $s(x)$ vara ett givet polynom med koefficienter i K, som ej är nollpolynomet. Två polynom $f(x)$ och $g(x)$ i $K[x]$ kallas *kongruenta* modulo $s(x)$ om polynomet $f(x) - g(x)$ är delbart med $s(x)$. Mer kortfattat skriver man $f \equiv g \pmod{s}$. Beteckna med $[f(x)]$ klassen

av polynom som är kongruenta med $f(x)$ modulo $s(x)$. Man definierar addition och multiplikation av sådana kongruensklasser genom

$$[f(x)] + [g(x)] = [f(x) + g(x)] \quad \text{och} \quad [f(x)] \cdot [g(x)] = [f(x)g(x)] \, .$$

På samma sätt som för heltal kontrollerar man att dessa definitioner är oberoende av valet av representanter för kongruensklasserna. Beteckna med

$$K[x]/(s(x))$$

mängden av kongruensklasser modulo $s(x)$. Man kontrollerar lätt att med ovanstående definitioner av addition och multiplikation blir $K[x]/(s(x))$ en kommutativ ring.

Exempel 16. I ringen $Z_5[x]/(x^3+1)$ är

$$\begin{aligned}[x^2 + 2x + 1] \cdot [x^2 + x + 2] &= [x^4 + 3x^3 + 5x^2 + 5x + 2] \\ &= [x^4 + 3x^3 + 2] = [(x+3)(x^3 + 1 - 1) + 2] \\ &= [(x+3)(-1) + 2] = [-x - 1] = [4x + 4] \, .\end{aligned}$$

Observera att x^3 alltid kan ersättas med -1, eftersom vi räknar modulo $x^3 + 1$.

I analogi med ringarna Z_n kan man visa att $K[x]/(s(x))$ är en kropp om och endast om $s(x)$ är ett primpolynom. Om $s(x)$ ej är ett primpolynom, är $s(x) = s_1(x)s_2(x)$ för några polynom s_1 och s_2 av högre grad än noll. Då blir $[s_1(x)][s_2(x)] = 0$, så $K[x]/(s(x))$ har nolldelare och är därför ingen kropp. Om $s(x)$ är ett primpolynom är $(f, s) = 1$ för varje polynom $f(x)$ av lägre gradtal än s och skilt från nollpolynomet. Euklides algoritm ger att det finns polynom $c(x)$ och $d(x)$ sådana att

$$1 = c(x)f(x) + d(x)s(x) \, .$$

Härav följer att $[1] = [c(x)][f(x)]$, så $[c(x)]$ är invers till $[f(x)]$. Eftersom, enligt divisionsalgoritmen, varje kongruensklass kan representeras av ett polynom $f(x)$ av lägre grad än $s(x)$, så har varje element skilt från noll i $K[x]/(s(x))$ en multiplikativ invers och man har en kropp.

Exempel 17. Polynomet $x^2 + 1$ är irreducibelt i ringen $R[x]$ av polynom med reella koefficienter. Därför är

$$R[x]/(x^2 + 1)$$

en kropp. Varje kongruensklass representeras av ett förstagradspolynom och, om man utnyttjar att $[x^2 + 1] = 0$, får man att

$$[a + bx][c + dx] = [(ac - bd) + (ad + bc)x]$$

Med hjälp härav kontrollerar man lätt att $R[x]/(x^2 + 1)$ är isomorf med kroppen C av komplexa tal.

Övningar

1. Låt $f(x)$ vara polynomet $x^{214} + 3x^{152} + 2x^{47} + 2$ i $Z_5[x]$. Beräkna värdet $f(3)$ i Z_5 .

2. Visa att om $f(x)$ är ett polynom av grad n med koefficienter i en kropp K, så har f högst n nollställen i K .

3. Bestäm den största gemensamma delaren (f, g) till följande polynom i $Z_2[x]$:
(a) $f(x) = x^7 + 1 \quad , \quad g(x) = x^5 + x^3 + x + 1$
(b) $f(x) = x^5 + x + 1 \quad , \quad g(x) = x^6 + x^5 + x^4 + x + 1$.

4. Finn den största gemensamma delaren $h = (f, g)$ till polynomen $f(x) = x^{17}+1$ och $g(x) = x^7+1$ i $Z_2[x]$ och bestäm två polynom $c(x)$ och $d(x)$ sådana att

$$h(x) = c(x)f(x) + d(x)g(x) .$$

5. (a) Visa att det bara finns ett irreducibelt andragradspolynom i $Z_2[x]$.
(b) Avgör om polynomet $x^5 + x^4 + 1$ är irreducibelt i $Z_2[x]$.

6. Bestäm samtliga moniska irreducibla andragradspolynom i $Z_3[x]$.

7. Skriv följande polynom i $Z_3[x]$ som produkter av primpolynom:
(a) $x^5 + x^4 + x^3 + x - 1$
(b) $x^4 + 2x^2 + 2x + 2$
(c) $x^4 + 1$
(d) $x^8 + 2$.

8. Hur många nolldelare finns det i ringen $Z_5[x]/(x^3 + 1)$?

9. (a) Visa att i en ändlig kropp är produkten av samtliga element skilda från noll alltid lika med -1.
Ledning: Använd Sats 3 och sambandet mellan rötter och koefficienter.
(b) Visa att för varje primtal p är

$$(p - 1)! = -1 \pmod{p} .$$

(jämför med övning 14 i avsnitt 1).

10. Låt K vara en kropp med q element, där $q = 2m + 1$ är udda. Visa att $x \in K$ är kvadraten på något element skilt från noll i K om och endast om $x^m = 1$.
Ledning: Visa först att $a^2 = b^2$ medför att $a = b$ eller $a = -b$ och använd sedan övning 2 ovan.

11. Visa att i en kropp med ett jämnt antal element är varje element kvadraten på ett och endast ett element.

Exempel 18. Vi skall bestämma alla irreducibla polynom av grad ≤ 4 i $Z_2[x]$. Det finns bara två polynom av första graden, nämligen

$$x \quad \text{och} \quad x+1\,.$$

Dessa är naturligtvis irreducibla. Då graden är 2 eller 3 är ett polynom irreducibelt om och endast om det saknar nollställen i Z_2. Man kontrollerar lätt att nollställen saknas precis då polynomet har ett udda antal termer och den "konstanta" termen är 1. Detta ger följande primpolynom av grad 2 och 3 :

$$x^2+x+1$$

$$x^3+x^2+1 \quad \text{och} \quad x^3+x+1\,.$$

För att ett polynom av grad 4 skall vara irreducibelt krävs dels att det inte har någon förstagradsfaktor, dvs att det saknar nollställe i Z_2, och dels att det inte är produkten av irreducibla andragradspolynom. Det senare villkoret eliminerar endast $(x^2+x+1)^2 = x^4+x^2+1$, ty det finns bara ett primpolynom av grad 2. De återstående fjärdegradspolynomen som saknar nollställe i Z_2 är

$$x^4+x^3+1\,,\; x^4+x+1 \quad \text{och} \quad x^4+x^3+x^2+x+1\,.$$

Detta är samtliga primpolynom av grad 4 i $Z_2[x]$.

Om $s(x)$ är något av de irreducibla fjärdegradspolynomen i Exempel 18, så är $Z_2[x]/(s(x))$ en kropp med $2^4 = 16$ element. Varje kongruensklass representeras nämligen av ett entydigt polynom av grad ≤ 3 och i detta kan var och en av de fyra koefficienterna väljas på två sätt, som 0 eller 1. De irreducibla polynomen av grad 2 och 3 ger kroppar med $2^2 = 4$ och $2^3 = 8$ element. I nästa avsnitt skall vi visa att för varje primtal p och varje positivt heltal n finns ett irreducibelt polynom i $Z_p[x]$ av grad n. Följaktligen finns för varje p och n en kropp med p^n element. Vi skall också visa att två ändliga kroppar med samma antal element alltid är isomorfa. Så när som på isomorfier finns det alltså för varje primtal p och positivt heltal n precis en kropp med p^n element. Denna kallas, efter den franske matematikern Evariste Galois (1811-1832), för Galoiskroppen av ordning p^n och betecknas $GF(p^n)$ (där GF står för engelskans Galois Field). I detta avsnitt skall vi ge exempel på hur man räknar i ändliga kroppar.

Exempel 19. För att beräkna den multiplikativa inversen till $[x^2+1]$ i kroppen $Z_2[x]/(x^3+x^2+1)$ tillämpar vi Euklides algoritm:

$$x^3+x^2+1 = (x+1)(x^2+1)+x$$
$$x^2+1 = x\cdot x+1\,.$$

Alltså är (observera att $+ = -$ i Z_2)

$$\begin{aligned} 1 &= (x^2+1) + x \cdot x = (x^2+1) + x((x^3+x^2+1) + (x+1)(x^2+1)) \\ &= (x^2+x+1)(x^2+1) + x(x^3+x^2+1)\,. \end{aligned}$$

Detta ger att $[x^2+1]^{-1} = [x^2+x+1]$.

Vi övergår nu till att diskutera räkning med potenser. Om a är ett element skilt från noll i en ändlig kropp, så måste någon potens av a vara lika med 1. T ex vet vi från Sats 3 i avsnitt 2 att $a^{q-1} = 1$, där q är antalet element i kroppen.

Definition. *Ordningen* av ett element $a \neq 0$ i en ändlig kropp är det minsta positiva heltalet m sådant att $a^m = 1$. Man betecknar ordningen av a med $o(a)$.

Exempel 20. Vi beräknar ordningen av [10] i kroppen Z_{73}:

$$\begin{aligned} 10^2 &= 100 \equiv 27 \\ 10^3 &\equiv 270 \equiv -22 \\ 10^4 &\equiv -220 \equiv -1\,. \end{aligned}$$

Härav följer att $10^5 \equiv -10$, $10^6 \equiv -27$, $10^7 \equiv 22$ och $10^8 \equiv 1$. Ordningen av [10] är alltså lika med 8.

I kroppen Z_{73} gäller, enligt Fermats lilla sats, alltid att $a^{72} = 1$ om $a \neq 0$. Följande lemma visar att det inte är någon tillfällighet att ordningen 8 i Exempel 20 delar 72.

Lemma 1. *Låt $a \neq 0$ vara ett element i en ändlig kropp. Om $a^n = 1$, för något positivt heltal n, så måste ordningen av a dela n.*

Bevis. Antag motsatsen. Om m är ordningen av a, så finns då heltal q och r, med $0 < r < m$, så att

$$n = qm + r\,.$$

Härav följer att

$$1 = a^n = (a^m)^q \cdot a^r = a^r\,.$$

Detta strider mot att $m = o(a)$, ty $0 < r < m$.

Nästa lemma ger en metod för att konstruera element av hög ordning i en kropp.

Lemma 2. *Antag att elementen a_1 och a_2 i en ändlig kropp har ordningarna m_1 respektive m_2 samt att m_1 och m_2 är relativt prima. Då har $a = a_1a_2$ ordningen m_1m_2.*

Bevis. Antag att $a^k = 1$. Då är

$$1 = a^{km_1} = a_1^{km_1} \cdot a_2^{km_1} = a_2^{km_1} .$$

Enligt Lemma 1 måste m_2 dela km_1. Eftersom $(m_1, m_2) = 1$, delar m_2 talet k. På samma sätt ser man att m_1 delar k. Alltså delas k av $m_1 m_2$, ty m_1 och m_2 är relativt prima. Ordningen av a är således minst $m_1 m_2$. Att den är exakt $m_1 m_2$ följer av att

$$a^{m_1 m_2} = (a_1^{m_1})^{m_2} \cdot (a_2^{m_2})^{m_1} = 1 .$$

Exempel 21. I Z_{73} är

$$8^2 = 64 \equiv -9$$
$$8^3 \equiv -72 \equiv 1$$

så ordningen av [8] är 3. Enligt Exempel 20 och Lemma 2 har alltså [80] = [7] ordningen $8 \cdot 3 = 24$.

Innan vi formulerar huvudresultatet i detta avsnitt, behöver vi ännu en hjälpsats.

Lemma 3. *Låt a och b vara element av ordning m respektive n i en ändlig kropp K och antag att m inte delar n. Då finns ett element av högre ordning än n i K.*

Bevis. Om m inte delar n, så finns en primtalspotens p^k som delar m men inte n. Då är $m = m'p^k$ och $n = n'p^l$, där $0 \leq l < k$ och n' ej delas av p. Alltså är $(p^k, n') = 1$ och $a^{m'} \cdot b^{p^l}$ har ordning $p^k n' > n$, enligt Lemma 2.

Sats 7. *I en ändlig kropp K med q element finns alltid ett element av ordning $q-1$.*

Bevis. Låt $b \neq 0$ vara ett element i K vars ordning är större eller lika med ordningen av alla andra element i K. Sätt $n = o(b)$. Enligt Lemma 3 måste ordningen av varje annat element i K dela n, ty annars fanns ett element av högre ordning än n. Detta innebär att samtliga element skilda från noll i K uppfyller ekvationen

$$x^n = 1 .$$

Polynomet $x^n - 1$ har således $q - 1$ olika nollställen. Enligt faktorsatsen måste man därför ha att $n \geq q-1$. Å andra sidan ger Sats 3 att ordningen aldrig kan vara större än $q - 1$. Alltså är $n = q - 1$ och satsen är bevisad.

Definition. Låt K vara en kropp med q element. Ett element av ordning $q - 1$ i K kallas ett *primitivt element*.

Exempel 22. Vi skall visa att [3] är ett primitivt element i Z_{101}. Eftersom ordningen av [3] måste dela $100 = 2^2 \cdot 5^2$, räcker det att kontrollera exponenterna 2, 4, 5, 10, 20, 25 och 50 :

$$3^2 = 9$$
$$3^4 = 81 \equiv -20$$
$$3^5 \equiv -60$$
$$3^{10} \equiv 3600 \equiv -36$$
$$3^{20} \equiv 1296 \equiv -17$$
$$3^{25} \equiv 1020 \equiv 10$$
$$3^{50} \equiv 100 \equiv -1$$

Det minsta positiva heltalet m för vilket $3^m \equiv 1$ är alltså 100.

För ett primitivt element a i en kropp K med q element är potenserna

$$a^0\,,\ a^1\,,\ a^2\,,\ \ldots\,,\ a^{q-2}$$

alla olika. I annat fall skulle man ha $a^j = a^k$ för några heltal $j < k$ mellan 0 och $q-2$. Då vore $a^{k-j} = 1$, vilket strider mot att ordningen av a är $q-1$. För varje $b \neq 0$ i K finns alltså ett entydigt bestämt heltal j med $0 \leq j \leq q-2$ sådant att $b = a^j$. Man kallar j för *index* av b och skriver $j = \mathrm{ind}(b)$. Index kallas också den *diskreta logaritmen* av b med avseende på det primitiva elementet a. Med hjälp av index kan man förenkla beräkningen av produkter och kvoter i ändliga kroppar. Om kroppen har q element gäller nämligen att

$$\mathrm{ind}(b_1 \cdot b_2) \equiv \mathrm{ind}(b_1) + \mathrm{ind}(b_2) \pmod{q-1}$$
$$\mathrm{ind}(b_1 \cdot b_2^{-1}) \equiv \mathrm{ind}(b_1) - \mathrm{ind}(b_2) \pmod{q-1}\,.$$

Exempel 23. Enligt Exempel 18 är polynomet $x^4 + x^3 + 1$ irreducibelt i $Z_2[x]$. Kroppen

$$K = Z_2[x]/(x^4 + x^3 + 1)$$

har $2^4 = 16$ element. Varje element i K kan beskrivas med en sträng av fyra binära siffror som anger koefficienterna i det polynom av grad ≤ 3 som ger motsvarande kongruensklass. T ex betecknar 1011 klassen $[x^3 + x + 1]$. Klassen $[x]$ är ett primitivt element i K och med hjälp av detta kan man göra en tabell över samtliga element i K^* :

index	0	1	2	3	4	5	6	7
element	0001	0010	0100	1000	1001	1011	1111	0111
index	8	9	10	11	12	13	14	
element	1110	0101	1010	1101	0011	0110	1100	

Exempelvis går beräkningen av elementet med index 5 till på följande sätt

$$[x^5] = [x \cdot x^4] = [x \cdot (x^3 + 1)] = [x^4 + x]$$
$$= [(x^3 + 1) + x] = [x^3 + x + 1] .$$

Vi illustrerar hur tabellen kan användas genom att beräkna

$$(1111) \cdot (1101)^{-1} .$$

Index för detta element är

$$6 - 11 = -5 \equiv 10 \pmod{15}$$

Alltså är

$$(1111) \cdot (1101)^{-1} = (1010) .$$

Övningar

1. Bestäm samtliga irreducibla femtegradspolynom i $Z_2[x]$.

2. Visa att $Z_3[x]/(x^3 + x^2 + 2)$ är en kropp med 27 element och beräkna i denna den multiplikativa inversen till $[x + 2]$.

3. Visa att $Z_{11}[x]/(x^2 + x + 4)$ är en kropp och beräkna inversen till $[3x + 2]$. Hur många element har kroppen?

4. (a) Bestäm ordningen av elementen [3] och [4] i Z_{37}.
(b) Ange ett primitivt element i Z_{37}.

5. Bestäm ett primitivt element i Z_{73}.

6. (a) Visa att $L = Z_2[x]/(x^3 + x + 1)$ är en kropp.
(b) Visa att $[x]$ är ett primitivt element och beräkna, som i Exempel 23, en indextabell i L.
(c) Beräkna $[x^2 + 1] \cdot [x^2 + x + 1]^{-1}$.

7. Använd tabellen i Exempel 23 till att beräkna följande uttryck
(a) $(1001) \cdot ((1011)^2 + (0011)^{-2})$
(b) $((1010)^2 + (0101)^3) \cdot ((0001) + (1101)^2)^{-1}$.

6. Existens och entydighet av $GF(p^n)$

För att visa att det finns en kropp med p^n element skall vi visa att för varje primtal p och varje positivt heltal n finns ett irreducibelt polynom av grad n i $Z_p[x]$. Vi konstaterar först att det totala antalet moniska polynom

$$f(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$$

med koefficienter i Z_p är lika med p^n. Enligt Sats 6 kan varje sådant polynom entydigt, så när som på ordningsföljden, skrivas som en produkt

$$f(x) = s_1(x)^{m_1} \cdots s_l(x)^{m_l} , \tag{3}$$

där $s_1(x), \ldots, s_l(x)$ är moniska primpolynom i $Z_p[x]$. Om d_i är graden av $s_i(x)$, så är

$$n = m_1 d_1 + \cdots + m_l d_l . \tag{4}$$

Antalet moniska polynom av grad n i $Z_p[x]$ är lika med antalet sätt att, som i (3), skriva moniska polynom av grad n som produkter av primpolynom. Om I_d betecknar antalet moniska primpolynom av grad d, så är alltså enligt (4) det totala antalet moniska polynom av grad n i $Z_p[x]$ lika med koefficienten för t^n i produkten

$$(1 + t + t^2 + \cdots)^{I_1}(1 + t^2 + t^4 + \cdots)^{I_2}(1 + t^3 + t^6 \cdots)^{I_3} \cdots .$$

Eftersom vi vet att denna koefficient är lika med p^n, gäller att

$$\prod_d \left(\frac{1}{1 - t^d}\right)^{I_d} = \frac{1}{1 - pt} .$$

Logaritmering ger

$$\sum_d -I_d \left(\ln(1 - t^d)\right) = -\ln(1 - pt)$$

och om man Taylorutvecklar båda sidorna får man

$$I_1(t + \frac{t^2}{2} + \frac{t^3}{3} + \cdots) + I_2(t^2 + \frac{t^4}{2} + \frac{t^6}{3} + \cdots) + I_3(t^3 + \frac{t^6}{2} + \frac{t^9}{3} + \cdots) + \cdots = \\ = pt + \frac{p^2t^2}{2} + \frac{p^3t^3}{3} + \cdots .$$

Jämförelse av koefficienterna för t^n ger

$$\sum_{d|n} I_d \cdot \frac{d}{n} = \frac{p^n}{n} .$$

Observera att på vänstra sidan förekommer endast termer där d delar n. Multiplikation med n ger följande sats:

Sats 8. *Om I_d är antalet moniska irreducibla polynom av grad d i $Z_p[x]$, så är*

$$\sum_{d|n} dI_d = p^n .$$

Exempel 24. För $p = 2$ och $n = 6$ får man

$$I_1 + 2I_2 + 3I_3 + 6I_6 = 2^6 = 64 .$$

Enligt Exempel 18 är $I_1 = 2$, $I_2 = 1$ och $I_3 = 3$, så $I_6 = 7$.

Med hjälp av Sats 8 kan man successivt bestämma talen I_d, men för att i ett steg kunna visa att I_d alltid är större än noll skall vi använda *Möbius inversionsformel* som bevisas i ett appendix. Möbius funktion $\mu(n)$ är definierad för positiva heltal n och antar endast tre värden 0, 1, och -1. Den ges av

$$\mu(n) = \begin{cases} 1 & \text{om } n = 1 \\ (-1)^k & \text{om } n \text{ är produkten av } k \textit{ olika} \text{ primtal} \\ 0 & \text{annars.} \end{cases}$$

Sats 12 i appendix tillämpad på formeln i Sats 8 ger

$$nI_n = \sum_{d|n} \mu(d) p^{n/d} .$$

Högerledet innehåller en lägsta potens av p. Om den lägsta potensen är p^m, så är

$$\frac{nI_n}{p^m} = \pm 1 + (\text{ett antal } p\text{-potenser med koefficienter } \pm 1).$$

Alltså är

$$\frac{nI_n}{p^m} = \pm 1 \pmod{p}$$

och speciellt är $nI_n \neq 0$.

Sats 9. *För varje primtal p och varje positivt heltal n finns ett irreducibelt polynom av grad n i $Z_p[x]$.*

Av Sats 9 följer att det existerar en kropp med p^n element. Vi övergår nu till att bevisa att det, så när som på isomorfier, bara finns en sådan kropp.

Låt K vara en godtycklig ändlig kropp med karakteristik p. Då innehåller K delkroppen

$$\mathrm{k} = \{\, 0\,,\, 1\,,\, \ldots\,,\, (p-1)1 \,\}$$

som är isomorf med Z_p. Om $m1 \in \mathrm{k}$ och $\beta \in K$, så är $(m1) \cdot \beta = m\beta$. Vi kan därför uppfatta K som ett vektorrum över kroppen Z_p. Eftersom K är ändlig, har detta vektorrum ändlig dimension. För varje $\alpha \in K$ finns därför ett positivt heltal d sådant att potenserna

$$\alpha^0, \alpha^1, \alpha^2, \ldots, \alpha^d$$

blir lineärt beroende, dvs det finns $a_0, a_1, \ldots, a_d \in Z_p$ ej alla noll så att

$$a_0 1 + a_1\alpha + a_2\alpha^2 + \cdots + a_d\alpha^d = 0 .$$

Låt d vara det minsta heltalet för vilket detta inträffar och sätt $s(x) = a_0 + a_1 x + \cdots + a_d x^d$. Då har $s(x)$ minimalt gradtal bland de icke-triviala polynom i $Z_p[x]$ som har α som nollställe. Tar man $a_d = 1$, vilket alltid är möjligt, så är $s(x)$ entydigt bestämt och kallas *minimalpolynomet* till α. Minimalpolynomet är irreducibelt i $Z_p[x]$ ty om $s(x)$ vore lika med en produkt $s_1(x)s_2(x)$ med faktorer av lägre grad än

d, så skulle antingen s_1 eller s_2 ha α som nollställe och detta strider mot att $s(x)$ är minimalpolynomet till α.

Sats 10. *Låt K vara en ändlig kropp med karakteristik p och α ett element i K. Om L är den minsta delkroppen i K som innehåller α och om $s(x)$ är minimalpolynomet till α, så är L isomorf med kroppen $Z_p[x]/(s(x))$.*

Bevis. Sätt

$$L = \{f(\alpha)\ ;\ f \in Z_p[x]\}\,.$$

Varje delkropp i K som innehåller α måste omfatta L, ty en sådan kropp innehåller alla potenser av α och alla lineärkombinationer av sådana potenser. Vi skall visa att L är isomorf med kroppen $Z_p[x]/(s(x))$. Härav följer att L själv är en kropp och därmed den minsta delkroppen i K som innehåller α. Betrakta avbildningen

$$Z_p[x]/(s(x)) \ni [f(x)] \longrightarrow f(\alpha) \in L\,.$$

Den är väldefinierad ty om f och g tillhör samma kongruensklass, dvs om $f(x) = g(x) + h(x)s(x)$ för något polynom h, så är

$$f(\alpha) = g(\alpha) + h(\alpha)s(\alpha) = g(\alpha)\,.$$

Av definitionen följer omedelbart att $[f(x)] + [g(x)]$ avbildas på $f(\alpha) + g(\alpha)$ och $[f(x)] \cdot [g(x)]$ på $f(\alpha)g(\alpha)$. Det återstår att visa att visa att avbildningen är bijektiv. Att den är surjektiv är klart. För att visa att den är injektiv observerar vi först att, om minimalpolynomet $s(x)$ har grad d, så räcker det att betrakta polynom $f(x)$ av lägre grad än d. Varje kongruensklass i $Z_p[x]/(s(x))$ representeras ju av ett sådant polynom. Antag att $f(\alpha) = g(\alpha)$ för två olika polynom av lägre grad än d. Då är α ett nollställe till $f - g$, vilket strider mot att $s(x)$ är minimalpolynomet till α. Alltså är avbildningen injektiv och satsen bevisad.

Följdsats. *Låt K vara en kropp med p^n element och $s(x)$ ett moniskt primpolynom i $Z_p[x]$ som har ett nollställe α i K. Då är $s(x)$ minimalpolynomet till α och graden av s delar n.*

Bevis. Elementet α är ett nollställe både till $s(x)$ och till sitt minimalpolynom $t(x)$. Därför är α ett nollställe till den största gemensamma delaren (s,t). Eftersom s och t är irreducibla, måste man ha $s = (s,t) = t$. Om $s(x)$ har graden d och L är den minsta delkroppen som innehåller α, så ger Sats 10 att L har p^d element. Eftersom K kan uppfattas som ett vektorrum över L, så är

$$|K| = |L|^m$$

för något positivt heltal m, där $|K|$ och $|L|$ betecknar antalet element i K respektive L. Alltså är

$$p^n = p^{dm}$$

och härav följer att d delar n.

Vi har nu alla hjälpmedel som behövs för att visa att två ändliga kroppar med samma antal element måste vara isomorfa. Låt K vara en godtycklig kropp med $q = p^n$ element. Enligt Sats 3 är då varje element i K ett nollställe till polynomet $x^q - x$. Vi har multiplicerat ekvationen i Sats 3 med x för att också få med $x = 0$. Enligt Sats 6 kan $x^q - x$ skrivas som en produkt av primpolynom i $Z_p[x]$:

$$x^q - x = \prod_i s_i(x) \,. \tag{5}$$

Här är summan av gradtalen av polynomen s_i lika med q. Eftersom $x^q - x$ har q olika nollställen i K, måste alla primpolynomen i högerledet vara olika och varje polynom s_i ha lika många nollställen i K som sitt gradtal. Följdsatsen ovan ger att gradtalen av polynomen s_i delar n. Betrakta nu formeln i Sats 8. Den visar att summan av gradtalen av *samtliga* primpolynom i $Z_p[x]$ vars gradtal delar n är lika med p^n. Därför måste produkten i högerledet av (5) innehålla alla primpolynom vars gradtal delar n. Speciellt måste, enligt Sats 9, högerledet i (5) innehålla ett primpolynom av grad n. Detta är minimalpolynomet till vart och ett av sina n nollställen i K. Låt α vara ett sådant nollställe. Då ger Sats 10 att den minsta delkroppen i K som innehåller α är isomorf med kroppen $Z_p[x]/(s(x))$ och följaktligen innehåller p^n element. Hela kroppen K är således isomorf med $Z_p[x]/(s(x))$. Vi har därmed bevisat följande sats.

Sats 11. *Låt $s(x)$ vara ett primpolynom av grad n i $Z_p[x]$. Då är varje kropp med p^n element isomorf med $Z_p[x]/(s(x))$.*

Anmärkning. Speciellt har vi visat att om s_1 och s_2 är två olika primpolynom av grad n i $Z_p[x]$ så är kropparna $Z_p[x]/(s_1(x))$ och $Z_p[x]/(s_2(x))$ isomorfa.

Appendix. Möbius inversionsformel

Vi påminner om att Möbius funktion $\mu(n)$ definieras, för positiva heltal n, som 0 om n innehåller någon multipel primfaktor och som $(-1)^k$ om n är en produkt av k olika primtal. Speciellt är $\mu(1) = 1$.

Lemma.

$$\sum_{d|n} \mu(d) = \begin{cases} 1 & \text{för } n = 1 \\ 0 & \text{för } n > 1 \,. \end{cases}$$

Bevis. Då $n = 1$ är summan lika med $\mu(1) = 1$. Om $n > 1$ och $n = p_1^{m_1} \cdots p_r^{m_r}$ är primfaktoruppdelningen av n, sätter vi $n^* = p_1 \cdots p_r$. Då är

$$\sum_{d|n} \mu(d) = \sum_{d|n^*} \mu(d) = 1 - r + \cdots + (-1)^k \binom{r}{k} + \cdots + (-1)^r \binom{r}{r} = (1-1)^r = 0 \,.$$

Binomialkoefficienterna $\binom{r}{k}$ anger hur många olika tal d som är produkter av k primfaktorer valda bland $p_1, \ldots, p_r$.

Sats 12 (Möbius inversionsformel). *Låt $f(n)$ och $g(n)$ vara definierade för positiva heltal n och antag att*

$$f(n) = \sum_{d|n} g(d)$$

för alla n. Då är

$$g(n) = \sum_{d|n} \mu(d) f\left(\frac{n}{d}\right) .$$

Bevis. Eftersom

$$f\left(\frac{n}{d}\right) = \sum_{d'|\frac{n}{d}} g(d') ,$$

så är

$$\sum_{d|n} \mu(d) f\left(\frac{n}{d}\right) = \sum_{d|n} \mu(d) \sum_{d'|\frac{n}{d}} g(d') = \sum_{d'|n} g(d') \sum_{d|\frac{n}{d'}} \mu(d) = g(n) .$$

I den sista likheten användes lemmat, som ger att

$$\sum_{d|\frac{n}{d'}} \mu(d) = \begin{cases} 1 & \text{då } d' = n \\ 0 & \text{då } d' < n . \end{cases}$$

FELRÄTTANDE KODER

1. Inledning

Vid all slags överföring och lagring av information finns det risk för att fel uppkommer. För att öka möjligheten att upptäcka och eventuellt rätta sådana fel, kan man tillföra den informationsbärande texten en viss redundans, t ex i form av kontrollsiffror. Vi ger två enkla exempel.

Exempel 1. Antag att en avsändare vill vidarebefordra en text uppdelad på ett antal sexställiga binära ord. Varje sådant ord består av sex siffror som var och en är 0 eller 1. För att öka möjligheten för mottagaren att upptäcka eventuella fel som sker vid överföringen kan avsändaren till varje ord lägga en sjunde binär siffra så att det i varje sjusiffrigt ord alltid finns ett jämnt antal ettor. Om mottagaren registrerar ett ord med ett udda antal ettor, så vet han att ett fel har uppstått och kan eventuellt be avsändaren att upprepa meddelandet.

Exempel 2. Om mottagaren i Exempel 1 inte har möjlighet att begära repetition, kan avsändaren gå tillväga på ett annat sätt. I stället för att lägga till en sjunde siffra kan han sända varje sexställigt ord tre gånger i följd. Om de tre orden inte längre är lika när de når mottagaren vet denne att ett fel har inträffat och han kan försöka rätta detta genom att på varje plats välja den siffra som finns på motsvarande plats i minst två av orden. Naturligtvis kan han inte vara helt säker på att felet har rättats, men om sannolikheten är liten för att mer än ett fel uppkommer är chansen god.

En nackdel med metoden i Exempel 2 är att i jämförelse med den ursprungliga texten tar varje meddelande försett med den felrättande anordningen tre gånger så lång tid att sända. Det är uppenbarligen en angelägen uppgift att försöka finna effektivare metoder och detta är syftet med teorin för felrättande koder. Denna tog sin början med arbeten av Shannon, Golay och Hamming i slutet av 1940-talet och har sedan dess utvecklats kraftigt med användning av alltmer sofistikerade matematiska metoder. I synnerhet spelar ändliga kroppar en framträdande roll.

En förutsättning för att kunna skriva text är att man har ett *alfabet*. Ett sådant är en ändlig mängd K av symboler som kallas *bokstäver*. I kodningsteorin är K vanligen en ändlig kropp och vi antar alltid att så är fallet. Då $K = Z_2$, som i exemplen ovan, talar man om en *binär kod*. Ett *ord* är en ändlig följd $x_1x_2 \ldots x_m$ av bokstäver. Vi skall enbart syssla med s.k. *blockkoder*. I en sådan har alla ord samma längd m och de kan därför uppfattas som vektorer i vektorrummet K^m. När så är lämpligt,

skriver vi orden som vektorer $x = (x_1, \ldots, x_m)$. En *kodningsfunktion* E är en injektiv avbildning

$$K^m \xrightarrow{E} K^n$$

från K^m in i ett vektorrum K^n av högre dimension. Bildmängden $C = E(K^m)$ kallas en *kod*. För att ge goda möjligheter att upptäcka och rätta fel är det önskvärt att koden C ligger glest i K^n, så att sannolikheten är liten för att ett felaktigt mottaget ord är ett nytt kodord.

Definition. Hammingavståndet $d(x, y)$ mellan två vektorer $x = (x_1, \ldots, x_n)$ och $y = (y_1, \ldots, y_n)$ i K^n definieras som antalet koordinater där $x_i \neq y_i$.

Exempel 3. I Z_2^5 är $d(10111, 11001) = 3$ och i Z_3^4 är $d(1122, 1220) = 2$.

Anmärkning. Om det är lika sannolikt att en felaktigt mottagen bokstav är vilken som helst av de övriga bokstäverna i alfabetet, är Hammingavståndet ett naturligt mått på storleken av ett fel. I vissa situationer kan andra mått vara lämpligare, men vi kommer endast att använda Hammingavståndet.

Definition. Separationen $d(C)$ för en kod C i K^n definieras som det minsta möjliga avståndet mellan två olika ord i koden, dvs

$$d(C) = \min\{d(x, y)\,;\; x, y \in C\,,\; x \neq y\}\,.$$

Sats 1. *(i) En kod C kan avslöja upp till k fel i varje ord om $d(C) \geq k + 1$.*
(ii) En kod C kan rätta upp till k fel i varje ord om $d(C) \geq 2k + 1$.

Anmärkning. Innebörden av *(ii)* är att om $d(C) \geq 2k + 1$ så finns för varje ord som innehåller högst k fel ett entydigt närmaste kodord. Det förutsätts att det felaktiga ordet rättas genom att välja det närmaste kodordet. I praktiken är det av stort intresse att finna effektiva algoritmer för att rätta fel och existensen av sådana algoritmer kan vara ett viktigt argument för att välja en viss kod. Vi kommer dock i det följande huvudsakligen att diskutera hur man kan konstruera koder med god separation och inte närmare gå in på algoritmer för felkorrigering.

Bevis för Sats 1. Om $d(C) \geq k + 1$, så skiljer sig två olika kodord alltid på minst $k + 1$ platser. Ett ord som mottages med minst en och högst k bokstäver fel kan alltså inte vara ett kodord och avslöjas därför som felaktigt.

För att bevisa *(ii)* antar vi att x är ett mottaget ord som skiljer sig från ett kodord y på högst k platser. Om $d(C) \geq 2k + 1$ kan det inte finnas något annat kodord z som skiljer sig från x på högst k platser, ty då skulle man ha $d(y, z) \leq 2k$. Man kan alltså rätta x till y.

Om man vill konstruera en kod $C = E(K^m)$ i K^n med given separation $\sigma = d(C)$, så finns det naturligtvis en begränsning på hur stort m kan väljas. Vi skall nu ge en teoretisk uppskattning av det största möjliga värdet på m.

Definition. För varje heltal $r \geq 0$ definieras sfären $S(x,r)$ med radie r och centrum i $x \in K^n$ genom

$$S(x,r) = \{y \in K^n \,;\, d(x,y) \leq r\}\,.$$

Lemma. *Om K har q element, så innehåller sfären $S(x,r)$ exakt*

$$\binom{n}{0} + \binom{n}{1}(q-1) + \binom{n}{2}(q-1)^2 + \cdots + \binom{n}{r}(q-1)^r$$

ord.

Bevis. Om $0 \leq j \leq r$, så är antalet ord som avviker från x på j ställen lika med $\binom{n}{j}(q-1)^j$.

Sats 2. *Antag att K har q element och att koden C i K^n innehåller M ord och har separation $2k+1$. Då är*

$$M\left[\binom{n}{0} + \binom{n}{1}(q-1) + \cdots + \binom{n}{k}(q-1)^k\right] \leq q^n\,. \tag{1}$$

Bevis. Sfärer med radie k och centrum i *olika* kodord i C kan inte skära varandra, ty $d(C) = 2k+1$. Eftersom antalet element i K^n är q^n, följer satsen av lemmat.

Anmärkning 1. Om $C = E(K^m)$, så är $M = q^m$.

Anmärkning 2. Olikheten (1) kallas *sfärpacknings-* eller *Hammingbegränsningen.* En kod för vilken man har likhet i (1) kallas en *perfekt kod.* För en sådan finns för varje ord x i K^n precis ett kodord på avstånd högst k från x.

Övningar

1. I exemplen 1 och 2 definieras två kodningsfunktioner från Z_2^6 in i Z_2^7 respektive Z_2^{18}. Bestäm separationen för motsvarande koder. Jämför resultaten med Sats 1.

2. Låt $\sigma > 0$ vara ett *udda* heltal och C en kod i Z_2^n med M ord och separation σ. Visa att det finns en kod $\widehat{C}$ i Z_2^{n+1} med M ord och separation $\sigma + 1$.
Ledning: Jämför med Exempel 1.

3. Konstruera en kod i Z_2^8 med 4 ord och separation 5.

4. Visa att det inte finns någon kod i Z_2^{12} med 2^7 ord och separation 5.

2. Linjära koder och generatormatriser

Definition. En kod C i K^n kallas *linjär* om den är ett underrum i K^n. Om dimensionen av C är m kallas den en $[n, m]$ kod.

Anmärkning. Att C är ett underrum i K^n innebär att alla linjärkombinationer av vektorer i C också ligger i C. Då är C själv ett vektorrum med samma operationer som i K^n och därmed är dimensionen av C väldefinierad.

De flesta felrättande koder som används i praktiken är linjära eller kan enkelt erhållas från linjära koder. En stor fördel med linjära koder är att det är mycket enklare att bestämma separationen för en sådan än för en allmän kod.

Definition. Med vikten $w(x)$ av ett kodord $x = (x_1, \ldots, x_n)$ i K^n menas antalet koordinater i x som är skilda från noll. Vikten $w(C)$ av en linjär kod C i K^n definieras genom

$$w(C) = \min\{w(x)\,;\, x \in C\,,\, x \neq 0\}\,.$$

Sats 3. *För en linjär kod C är separationen $d(C)$ lika med vikten $w(C)$.*

Bevis. En linjär kod som innehåller två ord x och y innehåller även $x - y$. Hammingavståndet $d(x, y)$ är då lika med vikten $w(x - y)$ och härav följer satsen.

Anmärkning. Vill man bestämma separationen för en allmän kod som innehåller M kodord, måste man i princip beräkna $M(M-1)/2$ olika Hammingavstånd, ett för varje par av ord i koden. För en linjär kod räcker det att beräkna vikten av $M - 1$ kodord.

Definition. En *generatormatris* för en linjär $[n, m]$ kod C i K^n är en $m \times n$ matris G, med element ur K, vars rader bildar en bas för underrummet C.

Exempel 4. Betrakta följande 3×7 matris med element ur $K = Z_3$

$$G = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 2 & 2 & 1 & 1 & 2 \\ 2 & 1 & 2 & 1 & 2 & 1 & 2 \end{bmatrix}.$$

Genom att dra första raden från den andra och addera första raden till den tredje, får man matrisen

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 2 & 0 & 2 & 0 & 2 & 0 \end{bmatrix}.$$

Multiplikation av tredje raden med 2 ger

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

Dras slutligen både andra och tredje raden från den första, får man

$$\tilde{G} = \begin{bmatrix} 1 & 0 & 0 & 2 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 \end{bmatrix}.$$

Raderna i $\tilde{G}$ genererar samma underrum i K^7 som raderna i G, ty man kan skriva raderna i den ena matrisen som lineärkombinationer av raderna i den andra. De två matriserna G och $\tilde{G}$ är alltså generatormatriser för samma kod C i K^7. Vi observerar nu att de tre första kolonnerna i $\tilde{G}$ är kolonnerna i enhetsmatrisen av ordning tre. Om man skiftar plats på andra och tredje kolonnen övergår $\tilde{G}$ i

$$\begin{bmatrix} 1 & 0 & 0 & 2 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 & 1 & 0 \end{bmatrix}.$$

Denna matris genererar en kod C' i K^7 som fås från koden C genom att kasta om bokstäverna på plats 2 och 3 för alla ord i C.

Definition. Två koder C och C' i K^n kallas *ekvivalenta* om det finns en permutation π av talen $1, \ldots, n$ sådan att

$$C' = \{x_{\pi(1)} x_{\pi(2)} \ldots x_{\pi(n)} \; ; \; x_1 x_2 \ldots x_n \in C\}.$$

Anmärkning. För två ekvivalenta koder C och C' gäller att $d(C) = d(C')$.

På liknande sätt som i Exempel 4 kan man visa följande sats.

Sats 4. *Varje lineär $[n, m]$ kod C är ekvivalent med en kod som har en generatormatris av formen*

$$[I_m \mid A]$$

där I_m är enhetsmatrisen av ordning m och A är en $m \times (n-m)$ matris.

Definition. När generatormatrisen för en lineär kod ser ut som i Sats 4 säger man att den är på *normalform.*

Låt $G = [I_m \mid A]$ vara generatormatrisen på normalform för en lineär $[n, m]$ kod C i K^n. Om elementen i K^m och K^n uppfattas som radmatriser, så ger avbildningen

$$K^m \ni x \longrightarrow xG \in K^n$$

en naturlig kodningsfunktion. De m första bokstäverna i ordet xG ges av ordet x i K^m och de återstående $n - m$ bokstäverna (kontrollsiffrorna) av xA.

Övningar

1. Bestäm generatormatriser för koderna i exemplen 1 och 2.

2. Låt C vara en binär lineär kod med generatormatrisen

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

Räkna upp samtliga kodord i C och bestäm separationen för C.

3. Matrisen

$$\begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 2 \end{bmatrix}$$

är generatormatris för en lineär kod C i Z_3^4. Bestäm alla kodord i C samt separationen $d(C)$. Visa också att C är en perfekt kod.

4. Låt C vara en binär lineär kod med generatormatrisen

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

Finn en generatormatris på normalform för C.

5. Bevisa Sats 4 genom att visa att varje $m \times n$ matris G, med element ur en kropp K och lineärt oberoende rader, kan överföras på en matris av formen $[I_m | A]$ med upprepad användning av följande operationer:
(i) multiplikation av en rad med ett element ur K
(ii) addition av en rad till en annan
(iii) skifta plats på två kolonner.
Ledning: Använd induktion över antalet rader i G.

3. Kontrollmatriser och syndromavkodning

Definition. Skalärprodukten $< x, y >$ av två vektorer $x = (x_1, \ldots, x_n)$ och $y = (y_1, \ldots, y_n)$ i K^n definieras av

$$< x, y > = x_1 y_1 + \cdots + x_n y_n .$$

Definition. Den *duala* koden $C^\perp$ till en lineär kod C i K^n är

$$C^\perp = \{ y \in K^n \, ; \, < x, y > = 0 \text{ för alla } x \in C \} .$$

Anmärkning. Liksom för underrum i R^n är det lätt att visa att om koden C i K^n har dimensionen m, så har den duala koden $C^\perp$ dimensionen $n - m$. För vektorrum K^n över en ändlig kropp K gäller emellertid i allmänhet *inte* att varje vektor i K^n entydigt kan skrivas som en summa av en vektor i C och en vektor i $C^\perp$. Det kan rent av inträffa att $C^\perp = C$. Man säger då att koden är *självdual.*

Exempel 5. I matrisen

$$G = \begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 2 \end{bmatrix}$$

är skalärprodukten av första raden med sig själv lika med 3, skalärprodukten av andra raden med sig själv lika med 6 och skalärprodukten av de två raderna lika med 3. Samtliga skalärprodukter är alltså noll modulo 3. Den $[4, 2]$ kod över Z_3 som har generatormatrisen G är följaktligen självdual.

Definition. En generatormatris för den duala koden till en kod C kallas en *kontollmatris* för C.

Ett ord $x \in K^n$ tillhör en kod C om och endast om x har skalärprodukt noll med raderna i en kontrollmatris för C. Man kan således med hjälp av en kontrollmatris enkelt avgöra om man har ett kodord eller ej.

Om G är en generatormatris för en $[n, m]$ kod C och H är en kontrollmatris för C, så är G en $m \times n$ matris och H en $(n - m) \times n$ matris. Villkoret för att H skall vara en kontrollmatris kan skrivas

$$G \cdot H^t = 0\,, \tag{2}$$

där H^t är den transponerade matrisen till H. Innebörden av (2) är nämligen att raderna i G har skalärprodukt noll med raderna i H. Antag nu att generatormatrisen G är på normalform $[I_m \mid A]$, där A är en $m \times (n - m)$ matris. Om man väljer $H = [-A^t \mid I_{n-m}]$, så verifierar man lätt att villkoret (2) är uppfyllt. Vi formulerar detta som en särskild sats.

Sats 5. *Om en lineär $[n, m]$ kod C har generatormatrisen $[I_m \mid A]$, så har den kontrollmatrisen $[-A^t \mid I_{n-m}]$.*

Anmärkning. Om kroppen K är Z_2, så är $-A^t = A^t$ och man kan ta $[A^t \mid I_{n-m}]$ som kontrollmatris.

Exempel 6. Den binära $[5, 2]$ kod som har generatormatrisen

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

har kontrollmatrisen

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

Vi skall nu beskriva hur en mottagare kan använda sig av en kontrollmatris H till en lineär kod C för att rätta eventuella fel som uppstått vid överföring av information med hjälp av koden C. Först kontrollerar man om ett mottaget ord $x \in K^n$ uppfyller villkoret $xH^t = 0$. Om så är fallet, är x ortogonal mot raderna i H och därmed ett

kodord. Man får då anta att inget fel har skett och att x är lika med det avsända kodordet. Om däremot $xH^t \neq 0$, så finns ett fel. För att rätta detta, kan man betrakta mängden av alla ord y i K^n för vilka $yH^t = xH^t$. Denna mängd kallas *biklassen* (eller multiplanet) hörande till *syndromet* xH^t. I biklassen som hör till xH^t väljer man ett ord $\bar{y}$ med minimal vikt, dvs med minsta möjliga Hammingavstånd till origo. Eftersom $\bar{y}H^t = xH^t$ är $x - \bar{y}$ ett kodord och det finns inget annat kodord närmare x, ty $\bar{y}$ har minimal vikt. Det är därför rimligt att rätta x till $x - \bar{y}$. Ordet $\bar{y}$ kallas en *klassledare* hörande till syndromet xH^t.

Exempel 7. För koden i Exempel 6 har man följande lista över klassledare med tillhörande syndrom:

klassledare	00000	10000	01000	00100	00010	00001	11000	10010
syndrom	000	101	011	100	010	001	110	111

Syndromet 000 svarar mot biklassen av kodord. De följande fem syndromen hör till biklasser bestående av ord som endast avviker på en plats från ett kodord. Klassledarna är här entydigt bestämda eftersom olika ord med vikten 1 ger olika syndrom. Detta beror på att kolonnerna i kontrollmatrisen H alla är olika. Syndromet för ett ord som är 1 på plats j och noll för övrigt är den j:te raden i H^t. De två sista klassledarna är ej entydigt bestämda av sina syndrom. T ex ger även 01100 syndromet 111. Här kan mottagaren agera på flera sätt. En möjlighet är att han bestämmer sig för en av de tänkbara klassledarna och använder den för felkorrigering. Andra alternativ är att be avsändaren att repetera meddelandet eller att helt enkelt ignorera det aktuella ordet.

Låt oss tillämpa listan på de tre mottagna orden 11111, 01110 och 01101. För det första ordet är syndromet 001. Motsvarande klassledare är 00001 och det rättade ordet blir 11110. För 01110 är syndromet 101 och klassledaren 10000. Det rättade ordet 01110 − 10000 blir även i detta fall 11110. För ordet 01101 är syndromet 110 och därför måste minst två bokstäver vara fel. Om mottagaren väljer klassledaren i listan, blir det rättade ordet 10101.

Vi avslutar detta avsnitt med en sats som beskriver hur man kan se på en kontrollmatris hur stor separation motsvarande kod har.

Sats 6. *En lineär kod C med kontrollmatrisen H har separationen σ om och endast om det finns σ kolonner i H som är lineärt beroende och dessutom $\sigma - 1$ kolonner i H alltid är lineärt oberoende.*

Bevis. Att σ kolonner i H är lineärt beroende betyder att det finns ett ord x med vikt högst σ sådant att $xH^t = 0$. Ett dylikt ord kan aldrig ha lägre vikt än σ, ty $\sigma - 1$ kolonner i H är alltid lineärt oberoende. Alltså är $w(C) = \sigma$ och satsen följer av Sats 3 i föregående avsnitt.

Övningar

1. Ange en kontrollmatris för koden i Exempel 1.

2. Visa att för en linear $[n, m]$ kod C har den duala koden $C^{\perp}$ dimensionen $n - m$.
Ledning: Använd Sats 4.

3. Matriserna

$$\begin{bmatrix} 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 \end{bmatrix} \quad \text{och} \quad \begin{bmatrix} 1 & 2 & 4 & 0 & 3 \\ 0 & 2 & 1 & 4 & 1 \\ 2 & 0 & 3 & 1 & 4 \end{bmatrix}$$

är generatormatriser för två linear koder C_1 och C_2 i Z_2^5 respektive Z_5^5. Bestäm kontrollmatriser för C_1 och C_2. Vad är separationen för de två koderna?

4. Betrakta den linear kod i Z_2^6 som har generatormatrisen

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

(a) Vilka av följande ord är kodord

$$111001\,,\ 010100\,,\ 101100\,,\ 110111\,,\ 100001\,?$$

(b) Vilka av orden i (a) kan avkodas entydigt? Avkoda dessa!

5. Låt C vara en binär kod med generatormatrisen

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

Avkoda i C, så gott det går, följande ord

$$1101011\,,\ 0110111\,,\ 0111000\,.$$

6. Bestäm separationen för den linear kod i Z_3^8 som har kontrollmatrisen

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 2 & 1 & 2 \\ 0 & 0 & 1 & 0 & 1 & 1 & 2 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 1 \end{bmatrix}.$$

7. Låt C vara den kod i Z_5^6 som har generatormatrisen

$$\begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 2 & 3 \\ 0 & 0 & 1 & 1 & 3 & 4 \end{bmatrix}.$$

Visa att $d(C) = 4$.

Exempel 8. Matrisen

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

är kontrollmatris för en binär $[7,4]$ kod som består av alla ord i Z_2^7 sådana att $xH^t = 0$. De sju kolonnerna i H är alla olika och utgör *samtliga* kolonner skilda från noll i Z_2^3. Varje syndrom skilt från noll i Z_2^3 har därför en entydig klassledare med vikten 1. Till exempel ger $\bar{y} = 0001000$ syndromet $\bar{y}H^t = 011$ som motsvarar den fjärde kolonnen i H. Varje ord x i Z_2^7 som inte är ett kodord kan alltså rättas till ett sådant genom att bara ändra en siffra i x. Vilken siffra som skall ändras bestäms av vilken kolonn i H som svarar mot syndromet xH^t.

Koder med egenskaper som i detta exempel har ett särskilt namn.

Definition. En lineär $[n,m]$ kod över Z_2 som har en kontrollmatris vars kolonner alla är olika och utgör samtliga kolonner skilda från noll i Z_2^{n-m} kallas en binär *Hammingkod.*

Anmärkning 1. Hammingkoder kan bara förekomma för vissa värden på parametrarna m och n. Om $n - m = r$, så är antalet vektorer skilda från nollvektorn i Z_2^{n-m} lika med $2^r - 1$. För en binär Hammingkod är alltså $n = 2^r - 1$ och $m = n - r = 2^r - 1 - r$ för något positivt heltal r. Exempel 8 svarar mot $r = 3$.

Anmärkning 2. På samma sätt som i Exempel 8 inser man att för en godtycklig binär Hammingkod gäller att varje ord i Z_2^n antingen är ett kodord eller ligger på Hammingavståndet 1 från ett entydigt bestämt kodord. Detta innebär att sfärerna med radie 1 och centrum i kodord täcker hela Z_2^n och att två olika sfärer aldrig skär varandra. Varje binär Hammingkod är därför en perfekt kod.

Exempel 9. Låt C vara den $[10,8]$ kod över kroppen Z_{11} som definieras av $xH^t = 0$, där

$$H = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \end{bmatrix}.$$

Observera att kontrollmatrisen H inte har normalformen $[-A^t \mid I_2]$. Om man så önskar är det lätt att transformera den till normalform, men för fortsättningen är det bekvämast att använda det givna uttrycket. Lägg också märke till att alla räkningar sker i Z_{11}, så $x \in Z_{11}^{10}$ är ett kodord om och endast om

$$\begin{cases} x_1 + x_2 + \cdots + x_{10} = 0 & (\text{mod } 11) \\ x_1 + 2x_2 + \cdots + 10x_{10} = 0 & (\text{mod } 11). \end{cases}$$

Antag nu att vid överföring av ett kodord $z = (z_1, \ldots, z_{10})$ *ett enda fel* av storlek e har uppstått på plats k så att det mottagna ordet är $x = (z_1, \ldots, z_k + e, \ldots, z_{10})$. Då blir syndromet xH^t lika med (e, ke). Härur kan man direkt bestämma felets storlek och också den plats k som det har inträffat på, genom att dividera den andra komponenten med den första. Om t ex $x = 0610271355$, så blir $xH^t = (8, 6)$. Eftersom $8^{-1} = 7$ i Z_{11} är $6 \cdot 8^{-1} \equiv 42 \equiv 9 \pmod{11}$. Om endast ett fel har uppstått i x, så är det alltså på plats 9 och siffran där skall ändras till $5 - 8 \equiv 8$.

Om man inte vill utnyttja "siffran 10" i kodorden, kan man i C stryka alla ord som innehåller 10 någonstans. Med hjälp av principen om inklusion-exklusion kan man beräkna att det ändå blir kvar 82644629 kodord. Man skulle alltså kunna distribuera så många tiosiffriga telefonnummer och garantera att rätt abonnent nås även om en siffra är felslagen.

Som förberedelse för nästa exempel skall vi beskriva en metod med vars hjälp man utifrån två givna koder kan konstruera en ny kod.

Sats 7. *Låt K vara en ändlig kropp och C_1, C_2 två lineära koder i K^n av dimension m_1 respektive m_2. Då är*

$$C = \{(x, x + y) \in K^{2n} \,;\, x \in C_1 \text{ och } y \in C_2\}$$

en lineär $[2n, m_1 + m_2]$ kod. Om σ_1 är separationen av C_1 och σ_2 separationen av C_2, så har C separationen

$$\sigma = min\,(2\sigma_1, \sigma_2)\,.$$

Bevis. Vi överlåter åt läsaren att kontrollera att C blir en lineär kod av dimensionen $m_1 + m_2$. För att bestämma separationen för C, gäller det att uppskatta den minsta möjliga vikten av ett ord skilt från noll i C. Om $y = 0$, är $w(x, x) = 2w(x) \geq 2\sigma_1$ och likhet inträffar för något $x \neq 0$ i C_1. Om $y \neq 0$, så är $w(x, x + y) \geq w(y) \geq \sigma_2$ och likhet inträffar för $x = 0$ och något $y \in C_2$. Alltså är separationen av C lika med $\min\,(2\sigma_1, \sigma_2)$.

Exempel 10. Med upprepad användning av Sats 7 skall vi konstruera en kod som bl a användes av Mariner 9 för att sända bilder av planeten Mars tillbaka till Jorden.

Låt C_1 vara den binära [4,3] kod som består av alla ord $x = x_1x_2x_3x_4$ i Z_2^4 för vilka

$$x_1 + x_2 + x_3 + x_4 = 0 \pmod{2}\,.$$

Koden C_1 bildas alltså av de ord som har ett jämnt antal ettor. Ett ord som inte är noll måste innehålla minst två ettor, så separationen för C_1 är 2. Som C_2 tar vi den kod som bara består av de två orden 0000 och 1111. Koden C_2 har dimensionen 1 och separationen 4. Tillämpas konstruktionen i Sats 7 på C_1 och C_2, får man en binär $[8, 3 + 1]$ kod med separationen 4. Kalla denna kod för C_1' och välj C_2' som den kod i Z_2^8 som bara innehåller de två ord för vilka antingen alla siffror är 0 eller alla siffror 1. Sats 7 tillämpad på C_1' och C_2' ger en [16,5] kod med separationen 8. Kalla

denna C_1'' och tag C_2'' som den kod i Z_2^{16} som består av de två ord som har alla siffror lika. Ännu en tillämpning av Sats 7 resulterar i en [32,6] kod med separationen 16. Det var denna kod som användes av Mariner 9. Eftersom separationen är 16, ger Sats 1 att 15 fel upptäcks och 7 fel rättas i varje ord med 32 bokstäver. För detta krävs här $32-6 = 26$ kontrollsiffror. Mariner-koden tillhör en allmän klass som kallas *Reed-Muller* koder.

Det sista exemplet i detta avsnitt är en klassisk kod som konstruerades av M. J. E. Golay 1949.

Exempel 11. Låt C vara den [12,6] kod över Z_3 som har generatormatrisen

$$G = [I_6 \mid A] = \left[\begin{array}{cccccc|cccccc} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 2 & 2 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 2 & 2 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 2 & 1 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 2 & 2 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 2 & 2 & 1 & 0 \end{array}\right].$$

De fem sista siffrorna i de fem sista raderna erhålles genom att cykliskt permutera vektorn 01221. Man kontrollerar lätt att raderna i G har skalärprodukt noll med varandra (observera att $2 = -1$ i Z_3). Koden C är alltså självdual. Speciellt är $< x, x >= 0$ för varje ord x i C. Eftersom bokstäverna i x är 0 eller ± 1, innebär detta att vikten $w(x)$ måste vara delbar med 3. Vi skall visa att det inte finns något ord i C med vikten 3. Ett sådant ord måste vara av typen $(3\,|\,0)$, $(2\,|\,1)$, $(1\,|\,2)$ eller $(0\,|\,3)$, där siffrorna till vänster och höger om strecket anger hur många av de sex första respektive sex sista siffrorna i ordet som är skilda från noll. Eftersom koden är självdual, måste varje kodord ha skalärprodukt noll med raderna i generatormatrisen G. Detta är omöjligt för ord av typerna $(3\,|\,0)$ och $(2\,|\,1)$. Å andra sidan måste varje kodord vara en lineärkombination av raderna i G. Detta är omöjligt för typerna $(1\,|\,2)$ och $(0\,|\,3)$. Alltså är den minsta vikten av ett ord skilt från noll i C lika med 6, vilket därför är kodens separation. Stryker man nu i generatormatrisen den första kolonnen i A, får man en [11,6] kod som kallas *Golaykoden* över Z_3 och betecknas $\mathcal{G}_{11}$. Genom att ta bort en bokstav i ett ord kan man högst minska vikten med ett, så $\mathcal{G}_{11}$ har separationen 5 och rättar därför 2 fel.

Det visar sig att $\mathcal{G}_{11}$ är en perfekt kod. För att kontrollera detta måste man visa att likhet gäller i olikheten (1) i Sats 2. För $\mathcal{G}_{11}$ är $M = 3^6$, $n = 11$, $k = 2$ och $q = 3$, så det gäller att verifiera att

$$3^6 \cdot \left[\binom{11}{0} + \binom{11}{1} \cdot 2 + \binom{11}{2} \cdot 2^2\right] = 3^{11}.$$

Detta lämnas åt läsaren.

Anmärkning. Golay konstruerade 1949 också en perfekt binär [23,12] kod med separation 7 som betecknas $\mathcal{G}_{23}$. Man kan bevisa att Golays koder är de enda perfekta koderna över en ändlig kropp som innehåller mer än två kodord och korrigerar mer än ett fel. Mer precist uttryckt måste varje annan sådan kod väsentligen vara ekvivalent med $\mathcal{G}_{11}$ eller $\mathcal{G}_{23}$.

Övningar

1. Ange en kontrollmatris för en binär [15,11] Hammingkod.

2. Låt K vara en ändlig kropp och C en $[n, m]$ kod i K^n med separation 3. Om C har en kontrollmatris H sådan att *varje* vektor i K^{n-m} kan fås genom att multiplicera någon kolonn i H med ett element ur K, så kallas C en Hammingkod över K.
(a) Visa att varje sådan Hammingkod är perfekt.
(b) Bestäm en kontrollmatris för en [8,6] Hammingkod över Z_7.
(c) Ange en kontrollmatris för en [13,10] Hammingkod över Z_3.
(d) För vilka värden på n och m finns det en $[n, m]$ Hammingkod över Z_p?

3. Avkoda, med avseende på koden i Exempel 9,det mottagna ordet 0617960587 under förutsättning att högst en siffra är fel.

4. Låt H vara kontrollmatrisen i Exempel 9. Vad kan man dra för slutsats om en siffra, men inte båda, är noll i syndromet xH^t för ett mottaget ord x?

5. Beskriv en generatormatris för koden C i Sats 7, om G_1 och G_2 är generatormatriser för koderna C_1 och C_2. Ange också en generatormatris för koden C_1' i Exempel 10.

6. Visa att i en binär självdual kod måste varje kodord ha en vikt som är delbar med 2.

7. Låt C vara en binär kod med generatormatrisen

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

(a) Visa att C är självdual.
(b) Använd resultatet i Övning 6 till att beräkna separationen $d(C)$.

5. Vandermondematriser och Reed-Solomon koder

Avslutningsvis skall vi beskriva en typ av koder som har hög kapacitet för felrättning och som bland annat har haft betydelse för utvecklingen av den moderna CD-tekniken.

Enligt Sats 6 i avsnitt 3 har en lineär kod med kontrollmatris H separationen σ eller mer om *varje* uppsättning av $\sigma - 1$ kolonner i H är lineärt oberoende. Vi visar först hur man på ett enkelt sätt kan konstruera matriser med ett föreskrivet minsta antal lineärt beroende kolonner.

Låt K vara en ändlig kropp och $\beta_0, \beta_1, \ldots, \beta_d$ olika element i K. Enligt faktorsatsen måste ett polynom $c(x)$ i $K[x]$ av grad d eller lägre och med nollställen i $\beta_0, \beta_1, \ldots, \beta_d$ vara nollpolynomet. Om

$$c(x) = c_0 + c_1 x + \cdots + c_d x^d ,$$

så innebär detta att ekvationssystemet

$$\begin{bmatrix} 1 & \beta_0 & \beta_0^2 & \ldots & \beta_0^d \\ 1 & \beta_1 & \beta_1^2 & \ldots & \beta_1^d \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & \beta_d & \beta_d^2 & \ldots & \beta_d^d \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \\ \vdots \\ c_d \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

bara har den triviala lösningen $c_0 = c_1 = \cdots = c_d = 0$. Koefficientmatrisen är alltså inverterbar, så kolonnerna i den transponerade matrisen

$$\begin{bmatrix} 1 & 1 & \ldots & 1 \\ \beta_0 & \beta_1 & \ldots & \beta_d \\ \beta_0^2 & \beta_1^2 & \ldots & \beta_d^2 \\ \vdots & \vdots & & \vdots \\ \beta_0^d & \beta_1^d & \ldots & \beta_d^d \end{bmatrix} \tag{3}$$

är lineärt oberoende. En matris med detta utseende kallas en *Vandermondematris*.

Låt nu n vara ett heltal som är större än d och låt $\alpha_0, \alpha_1, \ldots, \alpha_n$ vara olika element i kroppen K. Då är varje uppsättning av $d+1$ kolonner i matrisen

$$\begin{bmatrix} 1 & 1 & \ldots & \ldots & 1 \\ \alpha_0 & \alpha_1 & \ldots & \ldots & \alpha_n \\ \alpha_0^2 & \alpha_1^2 & \ldots & \ldots & \alpha_n^2 \\ \vdots & \vdots & & & \vdots \\ \alpha_0^d & \alpha_1^d & \ldots & \ldots & \alpha_n^d \end{bmatrix} \tag{4}$$

lineärt oberoende, ty de bildar en Vandermondematris. Enligt Sats 6 är alltså varje matris av formen (4) kontrollmatris för en lineär kod i K^{n+1} med separationen $d+2$.

Exempel 12. Betrakta den lineära [10,6] kod över Z_{11} som definieras av kontrollmatrisen

$$H = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ 1 & 2^2 & 3^2 & 4^2 & 5^2 & 6^2 & 7^2 & 8^2 & 9^2 & 10^2 \\ 1 & 2^3 & 3^3 & 4^3 & 5^3 & 6^3 & 7^3 & 8^3 & 9^3 & 10^3 \end{bmatrix}$$

där alla potenser beräknas i Z_{11}. Enligt vad vi just har visat är alltid fyra godtyckliga kolonner i H lineärt oberoende, så motsvarande kod har separationen 5. Observera att kolonnerna ligger i ett vektorrum av dimension fyra, så fler än fyra kolonner är alltid lineärt beroende.

Att separationen är 5 innebär, enligt Sats 1, att koden rättar två fel. Detta är en förbättring i förhållande till koden i Exempel 9, som ju bara rättar ett fel. Priset för detta är att antalet kodord i Z_{11}^{10} nu endast är 11^6 jämfört med 11^8 i Exempel 9.

Koden i Exempel 12 är en så kallad *Reed-Solomon* kod. Mer allmänt så används denna benämning på varje kod över en ändlig kropp K som har en kontrollmatris av formen (4) där $\alpha_0, \alpha_1, \ldots, \alpha_n$ är *samtliga* element skilda från noll i K. Om K har q element, innebär detta att $n = q - 2$. Vanligen räknar man då upp elementen $\alpha_0, \alpha_1, \ldots, \alpha_n$ genom att välja ett primitivt element $\alpha \in K$ och sätta $\alpha_i = \alpha^i$. Då får kontrollmatrisen (4) utseendet

$$\begin{bmatrix} 1 & 1 & 1 & \ldots & 1 \\ 1 & \alpha & \alpha^2 & \ldots & \alpha^{q-2} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & \alpha^d & \alpha^{2d} & \ldots & \alpha^{(q-2)d} \end{bmatrix}$$

Eftersom $\alpha^{q-1} = 1$ i K, räcker det naturligtvis att beräkna exponenterna modulo $q - 1$.

Anmärkning. Om $d = 2k - 1$ i kontrollmatrisen (4), så blir separationen $2k + 1$ och motsvarande kod rättar k fel. I de flesta tillämpningarna är $K = GF(2^m)$ och varje "bokstav" i K kan då skrivas med m binära symboler, 0 eller 1. Om man betraktar en sammanhängande följd av $(k-1)m+1$ binära symboler i ett ord, så kan dessa inte beröra fler än k bokstäver i $GF(2^m)$. En enstaka "kaskad" av längd $\leq (k-1)m+1$ av binära fel rättas alltså. Att Reed-Solomon koder används inom CD-tekniken beror bland annat på att de kan rätta sådana kaskader av fel. Detta utnyttjas för att vid avspelningen av en skiva eliminera störningar från damm, fingeravtryck, mindre repor och dylikt.

Exempel 13. Tag $K = GF(2^6)$ och $k = 5$. Eftersom K har 64 element, har orden i varje Reed-Solomon kod över K längden 63 om bokstäverna är element i K. Detta svarar mot binära ord av längden $6 \cdot 63 = 378$. Med $k = 5$ rättas enstaka kaskader av binära fel av längd högst $(k-1)m + 1 = 25$. Kontrollmatrisen (4) har i detta fall $d + 1 = 2k = 10$ rader, så koden har dimensionen $63 - 10 = 53$ som vektorrum över K. Den innehåller alltså $(2^6)^{53} = 2^{318}$ ord.

Övningar

1. Konstruera en lineär [8,4] kod över Z_{17} med separation 5.

2. Ange på lämpligt sätt en kontrollmatris för en Reed-Solomon kod över $K = GF(2^3)$ som rättar två fel i K.