# TALSYSTEM OCH RESTARITMETIKER

**Juliusz Brzezinski**

**MATEMATISKA INSTITUTIONEN**
**CHALMERS TEKNISKA HÖGSKOLA**
**GÖTEBORGS UNIVERSITET**
**GÖTEBORG 2002**

# FÖRORD

Detta häfte handlar om talsystem, restaritmetiker och polynomringar i anslutning till kursen "MAL 200". Först visar vi hur och varför man definierar olika typer av tal. Därefter kommer vi att bekanta oss med andra algebraiska system som har mycket gemensamt med talen. I avsnitt **2** diskuteras restaritmetiker som gör det möjligt att visa flera mycket intressanta egenskaper hos heltalen. I avsnitt **3** utvidgar vi våra kunskaper om polynom med koefficienter i olika talområden.

Om du har några kommentarer, upptäcker några tryckfel eller har förslag till förbättringar av texten skicka gärna e-mail till **jub at math.chalmers.se** (Julius Brzezinski).

# INNEHÅLL

# AVSNITT 1

# TALBEGREPPET

Med all säkerhet har Du redan mött olika typer av tal: naturliga, hela, rationella, reella och komplexa. Vad är det som skiljer olika talmängder? Finns det andra typer av tal? Vad menas egentligen med ett tal? Vi skall försöka svara på dessa frågor genom att analysera olika egenskaper hos olika talmängder. Men svaren är inte alltid enkla, och riktigt tillfredsställande svar kräver ibland djupare kunskaper som först är tillgängliga i senare kurser.

Vi skall beteckna med:

$\mathbb{N}$ de naturliga talen,

$\mathbb{Z}$ de hela talen,

$\mathbb{Q}$ de rationella talen,

$\mathbb{R}$ de reella talen,

$\mathbb{C}$ de komplexa talen.

Vi har $\mathbb{N} = \{1, 2, 3, ...\}$, $\mathbb{Z} = \{0, \pm 1, \pm 2, \pm 3, ...\}$, $\mathbb{Q} = \{\frac{m}{n} : m, n \in \mathbb{Z},\ n \neq 0\}$. Det är inte lika lätt att beskriva alla reella och komplexa tal. Vi skall försöka göra det i detta avsnitt och visa hur och varför man definierar olika typer av tal.

Alla tal kan adderas och multipliceras. Detta betyder att om a och $b$ är två tal så kan man bilda deras summa $a + b$ och deras produkt $ab$. Det är mycket viktigt att om $X$ betecknar något av talområdena ovan så

$$a, b \in X \Rightarrow a + b,\ ab \in X, \tag{1}$$

dvs summa och produkt av två naturliga tal är ett naturligt tal och samma gäller för alla andra talmängder $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$. Hur är det med de två andra räknesätten – subtraktion och division? Om man kräver att

$$a, b \in X \Rightarrow a - b \in X, \tag{2}$$

så är det inte möjligt att välja $X = \mathbb{N}$, ty trots att t ex $2, 3 \in \mathbb{N}$ så $2 - 3 = -1 \notin \mathbb{N}$. Däremot kan $X$ vara lika med $\mathbb{Z}, \mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$. Hur är det med

$$a, b \in X \Rightarrow \frac{a}{b} \in X? \tag{3}$$

Först och främst måste man tillägga att $b \neq 0$ (varför?). Det är klart att $\mathbb{N}$ och $\mathbb{Z}$ saknar egenskapen (3) ty t ex $2, 3 \in \mathbb{Z}$, men $\frac{2}{3} \notin \mathbb{Z}$. Alla andra talområden $\mathbb{Q}$, $\mathbb{R}$ och $\mathbb{C}$ uppfyller villkoret (3) (med $b \neq 0$). Man säger att $\mathbb{Q}$, $\mathbb{R}$ och $\mathbb{C}$ är slutna med avseende på de fyra räknesätten. $\mathbb{Z}$ är inte sluten med avseende på division, och $\mathbb{N}$ är inte sluten med avseende på subtraktion och division. Det visar sig att just slutenheten med avseende på olika operationer (här de fyra räknesätten) har en stor betydelse när det gäller skillnader mellan olika talområden. Av den anledningen har man infört följande begrepp:

**(1.1) Definition.** Man säger att en talmängd $K$ är en **talkropp** om $1 \in K$ och $K$ är sluten m a p de fyra räknesätten dvs om $a, b \in K$ så $a \pm b$, $ab \in K$, och i fall $b \neq 0$, $\frac{a}{b} \in K$. □

Som exempel kan vi nämna kroppen av de rationella talen $\mathbb{Q}$, de reella talen $\mathbb{R}$ och de komplexa talen $\mathbb{C}$. Finns det andra talkroppar? Svaret är att det finns många fler t o m oändligt många. Innan vi konstruerar andra talkroppar låt oss tänka en stund på $\mathbb{N}$ och $\mathbb{Z}$ som inte är kroppar men ändå måste anses som mycket viktiga talmängder. Heltalen är den enklaste talmängd som kallas för ring:

**(1.2) Definition.** Man säger att en talmängd $R$ är en **talring** om $1 \in R$ och $R$ är sluten m a p addition, subtraktion och multiplikation dvs om $a, b \in R$ så $a \pm b$, $ab \in R$. □

Heltalen $\mathbb{Z}$ är en talring. Det är också klart att varje talkropp är en talring. $\mathbb{N}$ är inte en talring.

Hur kan man konstruera talringar och talkroppar? Vi visar en enkel sats som är ett specialfall av en mycket allmän konstruktion av talringar och talkroppar (den allmänna konstruktionen behandlas i fortsättningskurser i algebra).

**(1.3) Sats.** *Låt $R$ vara en talring och låt $\alpha$ vara ett tal sådant att $\alpha \notin R$ men $\alpha^2 \in R$. Då bildar alla tal*

$$a + b\alpha, \text{ där } \ a, b \in R,$$

*en talring som betecknas med $R[\alpha]$. Om $R$ är en kropp så är också $R[\alpha]$ en kropp.*

Innan vi bevisar satsen låt oss titta på några intressanta exempel:

**(1.4) Exempel.** (a) Låt $R = \mathbb{Z}$ och låt $\alpha = \sqrt{2}$. Då har vi $\sqrt{2} \notin \mathbb{Z}$ och $(\sqrt{2})^2 = 2 \in \mathbb{Z}$. Satsen säger att talen:

$$a + b\sqrt{2}, \quad \text{där} \quad a, b \in \mathbb{Z},$$

bildar en ring. Om vi i stället för $\mathbb{Z}$ väljer $R = \mathbb{Q}$ får vi att talen

$$a + b\sqrt{2}, \ \text{ där } \quad a, b \in \mathbb{Q},$$

bildar en kropp. Detta betyder bl a att kvoten av två tal $a + b\sqrt{2}$ och $c + d\sqrt{2} \neq 0$ med $c, d \in \mathbb{Q}$ måste kunna skrivas som $e + f\sqrt{2}$, där $e, f \in \mathbb{Q}$. Låt oss pröva:

$$\frac{1 + \sqrt{2}}{3 + 2\sqrt{2}} = \frac{(1 + \sqrt{2})(3 - 2\sqrt{2})}{(3 + 2\sqrt{2})(3 - 2\sqrt{2})} = -1 + \sqrt{2}.$$

Det här kan inte vara någon överraskning – det finns många liknande exempel i grundskolans läroböcker !

(b) I stället för $\alpha = \sqrt{2}$ kan man välja $\alpha = \sqrt{a}$, där $a$ är ett godtyckligt heltal sådant att $\sqrt{a} \notin \mathbb{Q}$. På så sätt får vi oändligt många ringar $\mathbb{Z}[\sqrt{a}\,]$ och kroppar $\mathbb{Q}[\sqrt{a}\,]$. Är de verkligen olika? Det är ganska lätt att visa att för olika primtal $p$ är kropparna $\mathbb{Q}[\sqrt{p}\,]$ olika (se övning 5). Alltså existerar oändligt många olika kroppar därför att primtalen bildar en oändlig mängd.

(c) En mycket intressant ring får man då man väljer $R = \mathbb{Z}$ och $\alpha = i$. Vi har $i^2 = -1 \in \mathbb{Z}$. Enligt satsen bildar talen

$$a + bi, \ \text{ där } \ a, b \in \mathbb{Z},$$

en ring. Tal av denna typ kallas Gaussiska heltal *. De spelar en viktig roll i algebraisk talteori. □

Låt oss nu bevisa satsen:

**Bevis av (1.3)**: Låt $x = a + b\alpha$, $y = c + d\alpha \in R[\alpha]$. Vi vill visa att $R[\alpha]$ är en ring dvs att $x \pm y$, $xy \in R[\alpha]$. Vi har

$$x \pm y = (a + b\alpha) \pm (c + d\alpha) = (a \pm c) + (b \pm d)\alpha \in R[\alpha]$$

samt

$$xy = (a + b\alpha)(c + d\alpha) = (ac + bd\alpha^2) + (ad + bc)\alpha \in R[\alpha].$$

Om $R$ är en kropp, vill vi visa att $x, y \in R[\alpha]$ och $y \neq 0$ ger $x/y \in R[\alpha]$. Detta är lite svårare. Här har vi:

$$\frac{x}{y} = \frac{a + b\alpha}{c + d\alpha} = \frac{(a + b\alpha)(c - d\alpha)}{(c + d\alpha)(c - d\alpha)} = \frac{ac - bd\alpha^2}{c^2 - d^2\alpha^2} + \frac{bc - ad}{c^2 - d^2\alpha^2}\alpha = e + f\alpha,$$

där

$$e = \frac{ac - bd\alpha^2}{c^2 - d^2\alpha^2} \in R \quad \text{och} \quad f = \frac{bc - ad}{c^2 - d^2\alpha^2} \in R$$

ty $R$ är en kropp. Alltså $x/y \in R[\alpha]$.

Beviset kan te sig avslutat men det finns en punkt som kräver eftertanke. Vi vet att $c+d\alpha \neq 0$ och vi förlänger bråket $x/y$ med $c-d\alpha$. Får vi göra det? Med andra ord, är $c-d\alpha \neq 0$? Antag motsatsen dvs att $c - d\alpha = 0$. Om $d \neq 0$, får vi $\alpha = c/d \in R$ vilket strider mot antagandet om $\alpha$. Om $d = 0$, så ger $c - d\alpha = 0$ att $c = 0$, vilket betyder att $c + d\alpha = 0$ – en motsägelse igen! Alltså är $c - d\alpha \neq 0$ och vårt bevis är fullständigt. □

Låt oss återkomma till allmänna funderingar över talen och deras egenskaper. Våra kunskaper om olika talområden bygger på vår förmåga att hantera talen. I praktiken betyder det att vi följer olika regler när vi utför olika räkneoperationer. Vad är det för regler? Du kan säkert nämna eller skriva ut sådana regler som t ex associativiteten för addition: $a+(b+c) = (a+b)+c$, eller kommutativiteten för multiplikation: $ab = ba$. Hur många sådana regler finns det? Är

*C.F. Gauss (30/4 1777 - 23/2 1855) var en tysk matematiker – en av de mest betydelsefulla i matematikens historia.

alla lika viktiga? När kan man vara säker på att man har alla nödvändiga regler? Sådana frågor har sysselsatt många människor och svaren på dem bygger på matematisk forskning under en ganska lång tidsperiod. Här följer en förteckning över de viktigaste räknelagarna i en talmängd $R$ i vilken de kan vara uppfyllda eller ej – allt beror på hur man väljer $R$ :

### (1.5) Egenskaperna hos addition och multiplikation:

**Addition:**

| | | |
|---|---|---|
| (a) slutenhet: | $\forall_{a,b\in R}$ | $a, b \in R \Rightarrow a + b \in R$, |
| (b) associativitet: | $\forall_{a,b,c\in R}$ | $(a + b) + c = a + (b + c)$, |
| (c) kommutativitet: | $\forall_{a,b\in R}$ | $a + b = b + a$, |
| (d) neutralt element: | $\exists_{0\in R}\forall_{a\in R}$ | $0 + a = a$, |
| (e) motsatt element: | $\forall_{a\in R}\exists_{a'\in R}$ | $a + a' = 0$ ($a'$ betecknas med $-a$). |

**Multiplikation:**

| | | |
|---|---|---|
| (f) slutenhet: | $\forall_{a,b\in R}$ | $a, b \in R \Rightarrow ab \in R$, |
| (g) associativitet: | $\forall_{a,b,c\in R}$ | $(ab)c = a(bc)$, |
| (h) kommutativitet: | $\forall_{a,b\in R}$ | $ab = ba$, |
| (i) neutralt element: | $\exists_{1\in R}\forall_{a\in R}$ | $1a = a$, |
| (j) inverst element: | $\forall_{a\in R\setminus\{0\}}\exists_{a'\in R}$ | $aa' = 1$ ($a'$ betecknas med $a^{-1}$). |

**Addition och multiplikation:**

| | | |
|---|---|---|
| (k) distributivitet: | $\forall_{a,b,c\in R}$ | $a(b + c) = ab + ac$. |

Alla dessa regler gäller då $R$ är en talkropp t ex $\mathbb{Q}$, $\mathbb{R}$ eller $\mathbb{C}$. Om $R = \mathbb{Z}$ så gäller alla räknelagar med undantag av (j) – t ex $2 \in \mathbb{Z}$, men $1/2 \notin \mathbb{Z}$. Egenskapen (j) ger just skillnaden mellan en talkropp och en talring. I en talkropp gäller alla räknelagarna (a) – (k), medan i en talring gäller alla utom (j).

Räknelagarna (a) – (k) är grunden för all manipulation med talen och man måste vara medveten om deras giltighet i det talområde man vill arbeta med. Andra räknelagar som t ex

(i) $a0 = 0$ då $a \in R$,

(ii) $(-1)(-1) = 1$,

(iii) $-(-a) = a$ då $a \in R$,

(iv) $(-a)b = -ab$ då $a, b \in R$,

(v) $(-a)(-b) = ab$ då $a, b \in R$,

kan man bevisa om man vet att $R$ är en ring (se övningar). I själva verket kan man definiera allmänna begrepp *ring* och *kropp* i vilka dessa räknelagar kan härledas:

**(1.6) Definition.** Man säger att en mängd $R$ vars element kan adderas under en operation "+" och multipliceras under en operation "$\cdot$" är en **ring** om dessa operationer har alla egenskaper (1.5) (a) – (k) med undantag av (j). Om alla egenskaper (a) – (k) gäller så säger man att $R$ är en **kropp**. □

Vi möter andra ringar och kroppar än talringar och talkroppar i senare avsnitt om restaritmetiker och polynomringar.

I samband med definitionerna av begreppen ring och kropp har du säkert observerat att man inte nämner subtraktion och division. Förklaringen är att subtraktion och division kan definieras i efterhand med hjälp av addition och multiplikation:

**(1.7) Definition.** (a) Om $R$ är en ring och $a, b \in R$ så säger man att

$$a - b = a + (-b)$$

är **skillnaden** mellan a och $b$.

(b) Om $R$ är en kropp och $a, b \in R$, $b \neq 0$, så säger man att

$$a : b = ab^{-1}$$

är **kvoten** av $a$ genom $b$. Kvoten betecknas också med $\frac{a}{b}$. □

Vårt syfte i detta avsnitt är att förklara hur man definierar talbegreppet. Som vi redan vet finns det oändligt många olika talringar och talkroppar. På vilket sätt intar $\mathbb{Z}, \mathbb{Q}, \mathbb{R}$ och $\mathbb{C}$ en särställning bland dem? Ett kort svar som kräver många förklaringar är följande: $\mathbb{Z}$ är den minsta talringen, $\mathbb{Q}$ är den minsta talkroppen, $\mathbb{R}$ är den största talkroppen som tillåter ordningsrelationen $\leq$ och $\mathbb{C}$ är den största talkroppen överhuvudtaget. Man inser säkert att alla dessa svar förutsätter att man vet vad ett tal är. Svaret på den frågan är inte enkelt och det tog en mycket lång tid i mänsklighetens utveckling innan man kunde komma till ett tillfredsställande svar. Trots det har man sedan en lång tid tillbaka kunnat räkna med alla typer av tal och utveckla vetenskapliga teorier som bygger på beräkningar och som framgångsrikt beskriver världen runt omkring oss. De naturliga talen är med all säkerhet lika gamla som den mänskliga civilisationen, rationella tal (åtminstone positiva) är nästan lika gamla, negativa tal (hela, rationella och reella) användes för ungefär 1000 år sedan, och komplexa tal introducerades under 1500-talet. Därför finns det inte någon större anledning till oro om våra svar inte visar sig bli fullständiga. Vi skall försöka förklara olika aspekter av talbegreppet utan att förutsätta några större förkunskaper. Mera tillfredsställande förklaringar väntar den som läser fortsättningskurser i matematik.

Det finns två möjligheter att introducera talbegreppet. Den ena är att börja med de naturliga talen och försöka steg för steg konstruera andra typer av tal. Den metoden ter sig naturlig och tilltalande men den är mycket arbetsam och, tyvärr, ganska lång om man vill kontrollera alla detaljer. Vi skall berätta om den senare i detta avsnitt.

Den andra möjligheten utgår från att man kan hantera talen om man vet vilka regler som styr deras användning. Det räcker om man kommer överens om dessa regler och följer dem för att kunna använda talen, men man behöver inte bry sig om hur de är konstruerade. En sådan inställning till talen är mycket praktisk, men en matematiker vill gärna veta hur talen konstrueras (och alla andra som använder talen måste tro på möjligheten av dessa konstruktioner). Man kan jämföra den inställningen med inställningen till tekniken – om man har läst en instruktionsbok till en TV-apparat så vet man hur man använder den utan att behöva veta hur den är konstruerad (eller att den finns). En beskrivning av en programvara är troligen ännu bättre som jämförelse – man får en förteckning över kommandon och deras effekt utan att behöva veta hur programvaran är konstruerad eller om den finns tillgänglig.

Vi skall försöka beskriva de egenskaper som karakteriserar de reella talen. Valet av dessa egenskaper är ett resultat av matematisk forskning huvudsakligen under 1800-talet. De reella talen spelar en mycket central roll. Å ena sidan har alla människor en intuitiv uppfattning om dessa tal som kommer från erfarenheten av att räkna och mäta i vardagslivet. Å andra sidan bygger alla vetenskaper, och bland dem matematiken själv, på de reella talens egenskaper.

Som vi redan vet bildar de reella talen en kropp. Men det finns många kroppar så man måste välja egenskaper som utmärker just den. En viktig egenskap är att man kan jämföra de reella talen med hjälp av $\leq$ – de reella talen bildar en ordnad kropp. Låt oss definiera helt allmänt vad detta betyder:

**(1.8) Definition.** Man säger att en kropp $K$ är **ordnad** om den innehåller en delmängd $P$ sådan att:

(a) om $x \in K$ så gäller exakt ett av de tre alternativen: $x \in P$ eller $x = 0$ eller $-x \in P$,

($b$) om $x, y \in P$ så gäller att $x + y \in P$ och $xy \in P$.

Man säger att $P$ är mängden av de positiva elementen i $K$. □

Det är klart att i $K = \mathbb{R}$ kan vi välja $P =$ alla positiva reella tal. Detta betyder att $\mathbb{R}$ är en ordnad kropp. $\mathbb{Q}$ är också ordnad därför att vi kan välja $P =$ alla positiva rationella tal. Vi skall senare visa att $\mathbb{C}$ inte är en ordnad kropp (det är enkelt att visa om man vet att $i^2 = -1$).

Vi skall uppehålla oss en stund vid definitionen (1.8). Man kan definiera:

$$x > y \quad (\text{eller} \quad y < x) \quad \text{om} \quad x - y \in P. \tag{1.9}$$

Man brukar också skriva $x \geq y$ (eller $y \leq x$) om $x > y$ eller $x = y$. $x > 0$ betyder att $x - 0 \in P$ dvs $x \in P$; $x < 0$ betyder att $0 - x \in P$ dvs $-x \in P$.

Om $K$ är en ordnad kropp så kan man definiera de naturliga och de rationella talen i $K$. Först observerar vi att $1 > 0$ ($1 \in K$ är neutralt för multiplikation). Vi vet att $1 \neq 0$ så att $1 \in P$ eller $-1 \in P$. Antag att $-1 \in P$. Då är $1 = (-1)(-1) \in P$ enligt (b) i (1.8). Detta ger att både 1 och $-1$ tillhör $P$ vilket strider mot (a) i (1.8). Därför måste $1 \in P$. De naturliga talen i $K$ får vi som

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

vilka definitionsmässigt betecknas med 1,2,3,4,.... Observera att $1 < 2 < 3 < 4...$ därför att $2 - 1 = 1 > 0$, $3 - 2 = 1 > 0$, $4 - 3 = 1 > 0$ osv. Heltalen i $K$ definieras som: alla naturliga tal $x$, deras motsatta tal $-x$ samt 0 dvs 0, $\pm 1$, $\pm 2$, $\pm 3$, $\pm 4$, .... De rationella talen definieras som alla kvoter $ab^{-1}$ , där $a, b$ är hela och $b \neq 0$ (se (1.7)).

Både $\mathbb{Q}$ och $\mathbb{R}$ är ordnade kroppar så en definition av de reella talen måste bygga på en annan egenskap (utöver det att $\mathbb{R}$ är ordnad). Innan vi formulerar en lämplig egenskap, låt oss återkomma för en stund till definitionen av en ordnad kropp. I en sådan kropp kan man definiera absolutbelopp:

$$|x| = \begin{cases} x & \text{om } x \geq 0, \\ -x & \text{om } x < 0. \end{cases} \tag{1.10}$$

Man kan också säga vad det betyder att en följd $x_1$, $x_2$, $x_3$, ... går mot 0. Man säger så om det för varje naturligt tal $n$ finns ett $N$ sådant att $|x_i| < \frac{1}{n}$ då $i > N$. Nu kan vi formulera en grundläggande egenskap som skiljer $\mathbb{Q}$ från $\mathbb{R}$. Låt $x_1$, $x_2$, ..., $x_i$, ... vara en växande och begränsad följd av rationella tal dvs $x_1 \leq x_2 \leq ... \leq x_i \leq ...$ och det finns ett tal $B$ så att $x_i \leq B$ då $i = 1, 2, ....$ Vad kan man säga om gränsvärdet $\lim_{i \to \infty} x_i$ ? I analyskurser visas att gränsvärdet existerar. Är gränsvärdet ett rationellt tal? Låt oss betrakta ett exempel. Definiera

$$x_n = 1, a_1 a_2 ... a_n, \quad n \geq 1,$$

där $a_i$ är $i$ :te siffran i decimalutvecklingen av $\sqrt{2}$ dvs

$x_1 = 1, 4,$
$x_2 = 1, 41,$
$x_3 = 1, 414,$
$x_4 = 1, 4142,$
. . .

Det är klart att alla $x_n$ är rationella och att följden är växande och begränsad. Ändå är det också klart att $\lim_{n\to\infty} x_n = \sqrt{2}$ dvs följden konvergerar mot ett icke-rationellt tal $\sqrt{2}$ (vi visar om en stund att $\sqrt{2}$ inte är rationellt). Men gränsvärdet är ett reellt tal och det är sant helt allmänt att en växande och begränsad följd av reella tal konvergerar mot ett reellt tal. Man säger att de reella talen bildar en fullständig kropp[†]. Allmänt har man följande begrepp:

**(1.11) Definition.** En ordnad kropp kallas **fullständig** om varje växande och begränsad följd av kroppens element konvergerar mot ett element i kroppen. □

Mera exakt, om $K$ är en ordnad kropp så är den fullständig om för varje följd $x_1 \leq x_2 \leq ... \leq x_n \leq ...$ sådan att $x_n \in K$ och det finns $B \in K$ så att $x_n \leq B$ då $n = 1, 2, ...$ man kan hitta $x \in K$ så att $\lim_{n\to\infty} x_n = x$.

Nu kan vi definiera de reella talen:

**(1.12) Definition.** Med **reella tal** menar man elementen i en ordnad och fullständig kropp $K$. □

Dessa få ord döljer ett ganska sammansatt matematiskt innehåll: $K$ är en kropp dvs uppfyller villkoren (a) – (k) på sidan 5, $K$ är ordnad dvs upfyller (a) och (b) i (1.8), och slutligen är $K$ fullständig dvs uppfyller (1.11). Nu kan man ställa två frågor:

> Finns det en ordnad och fullständig kropp?
> Hur många ordnade och fullständiga kroppar finns det?

Man behöver inte veta svaret på dessa två frågor för att kunna räkna med de reella talen därför att (1.12) är en exakt förteckning över alla grundläggande egenskaper hos dessa tal och det räcker att följa dem och deras logiska konsekvenser. Men svaren på dessa två frågor är mycket viktiga inte bara för en matematiker (en matematiker vill dessutom se själv hur man kommer fram till svaren). De är följande: Det finns ordnade och fullständiga kroppar. Om $K_1$ och $K_2$ är två sådana så finns det en bijektiv funktion $f : K_1 \to K_2$ (dvs enentydig och på hela $K_2$) som uppfyller $f(a+b) = f(a)+f(b)$, $f(ab) = f(a)f(b)$ och om $a > 0$ så är $f(a) > 0$[‡]. Intuitivt säger existensen av $f$ att $K_1$ och $K_2$ skiljer sig bara när det gäller beteckningar dvs om $a \in K_1$ så kan $f(a)$ uppfattas som ett annat namn på $a$. Addition och multiplikation i $K_1$ översätter man med hjälp av $f$ till addition och multiplikation i $K_2$. Likaså positiva element ur $K_1$ övergår med hjälp av $f$ i positiva element i $K_2$. I den meningen är kroppen av de reella talen entydig.

Vi vet redan att om vi har de reella talen så kan vi definiera de naturliga, hela och rationella. På så sätt har vi en möjlighet att tillfredsställa vårt behov av någorlunda ordentlig presentation av talbegreppet. Men även om den för många ändamål är helt tillfredsställande, går vi

[†]Detta bevisas i analyskurser med hjälp av supremumaxiomet som är ekvivalent med den egenskapen.
[‡]En sådan funktion $f$ kallas isomorfism och man säger att $K_1$ och $K_2$ är isomorfa ordnade kroppar.

ett steg längre och försöker beskriva konstruktioner av olika talmängder. Behovet av sådana konstruktioner insåg man under 1800-talet då utvecklingen av matematiken gick så långt att intuitiva föreställningar om talen inte längre kunde accepteras. Man försökte konstruera olika talområden genom att utgå från de naturliga talen och succesivt gå till de hela, rationella, reella och komplexa. Den vägen är ganska lång, arbetsam (man måste kontrollera många detaljer), och det värsta, rätt så tråkig om man bortser från mera allmänna principer som styr dessa konstruktioner och har betydelse i andra sammanhang. Därför behövs möjligen ett varningens ord att inte fördjupa sig i alla detaljer och inte ta vår genomgång på fullt allvar.

**(1.13) De naturliga talen.** De äldsta talen är de naturliga (och de är mest naturliga därför att de är de äldsta). Varifrån kommer de? En stor tysk matematiker L.Kronecker sade någon gång att "Gud skapade de naturliga talen, allt annat är människans skapelse". Det vore för enkelt med detta svar men det är mycket djupsinnigt. Den enda möjligheten att definiera de naturliga talen är den metod som vi använde tidigare för att definiera de reella: Man kan beskriva deras grundläggande egenskaper. Varifrån kommer de egenskaper som betraktas som grundläggande? Svaret är att de kommer från mänsklighetens erfarenhet av experimentell hantering av talen och det faktum att de regler som man har följt under en mycket lång tid ger en bild av verkligheten som överensstämmer med våra observationer. En analys av sådana regler kunde göras enbart av matematiker. Det var R. Dedekind [§] och G. Peano [¶] som föreslog ett urval av sådana grundläggande regler under senare delen av 1800-talet. Den mest kända definitionen kommer från G. Peano och låter så här:

**(1.14) Definition.** Med **naturliga tal** menar man elementen i en mängd $\mathbb{N}$ som satisfierar följande villkor:

(a) det finns ett utvalt element $1 \in \mathbb{N}$;

(b) det finns en injektiv funktion som mot varje element $n \in \mathbb{N}$ ordnar ett element $n^* \in \mathbb{N}$ så att $n^* \neq 1$;

(c) om $X \subseteq \mathbb{N}$ och

($\mathrm{d}_1$) $1 \in X$,

($\mathrm{d}_2$) $\forall_n n \in X \Rightarrow n^* \in X$,

så är $X = \mathbb{N}$. □

Intuitivt betyder $n^*$ talet $n + 1$ ($n^*$ kallas efterföljaren till $n$). Sista villkoret (d) kallas ofta "induktionsaxiomet"(det behandlas närmare i samband med matematisk induktion). Lägg märke till att man inte nämner addition och multiplikation i definitionen. De definieras i

[§]Richard Dedekind (1831-1916) en tysk matematiker.
[¶]Giuseppe Peano (1858-1932) en italiensk matematiker.

efterhand. Peanos definition överenstämmer väl med vår intuition, den är lätt att förstå, den är kort och elegant. Den uppfyller många av de kriterier som man vill uppfylla när man definierar ett matematiskt objekt. Vidare kan man ur den definitionen härleda alla kända egenskaper hos de naturliga talen.

Men hur är det egentligen med existensen och entydigheten av den mängden? När det gäller entydigheten är svaret enkelt: Man kan visa att om $N_1$ och $N_2$ är två mängder som uppfyller villkoren i definitionen (1.14) så är de isomorfa vilket betyder att det finns en bijektiv funktion $f : N_1 \to N_2$ sådan att $f(1) = 1$ samt $f(n^*) = f(n)^*$ (jämför ett liknande påstående om de reella talen på sidan 9). Existensen av de naturliga talen vilar på vår övertygelse om att åtminstone en mängd av de naturliga talen existerar – nämligen den som under mänskligheten historia så troget och framgångsrikt har tjänat till att räkna, resonera och dra korrekta slutsatser om världen runt omkring oss. Med andra ord är existensen av de naturliga talen ett axiom. Här har vi närmat oss matematikens grunder som har mycket gemensamt med vetenskapernas filosofi.

Alla andra talområden kan nu succesivt konstrueras: De hela talen från de naturliga, de rationella från de hela, de reella från de rationella och de komplexa från de reella. När vi sade tidigare att det går att bevisa existensen av de reella talen så menade vi just att det var möjligt att konstruera dessa tal från de naturliga.

Nu skall vi börja vår vandring från de naturliga talen genom rationella och reella till de komplexa. Vi utelämnar många detaljer och begränsar oss till allmänna idéer.

Det finns två huvudorsaker till att talbegreppet utvidgades. Det första var behov i samband med mätningar. Man upptäckte mycket tidigt att det behövdes bråktal för att uttrycka dimensioner (längder och areor) av jordlotter. Men icke-rationella tal dök upp även i samband med mätningar (vi får se det i samband med konstruktionen av de reella talen). Den andra orsaken har en mera abstrakt karaktär. Nya typer av tal behövdes för att kunna lösa ekvationer. Ett typiskt exempel är de komplexa talen. På 1500-talet kände man till formeln:

$$x_{1,2} = -\frac{p}{2} \pm \sqrt{\frac{p^2}{4} - q}$$

för lösningar till andragradsekvationen $x^2 + px + q = 0$. Löser man ekvationen $x^2 - 3x + 2 = 0$ så får man enligt den formeln $x_1 = 1$ och $x_2 = 2$. Tar man i stället $x^2 - 2x + 2 = 0$ så blir $x_1 = 1 + \sqrt{-1}$ och $x_2 = 1 - \sqrt{-1}$ . En del människor skulle kanske säga att ekvationen $x^2 - 2x + 2 = 0$ i så fall saknar lösningar därför att $\sqrt{-1}$ är helt utan mening. Andra skulle acceptera symbolen $\sqrt{-1}$ , tillskriva den egenskapen att $(\sqrt{-1})^2 = -1$ och sätta in $1 + \sqrt{-1}$ i ekvationen $x^2 - 2x + 2 = 0$. Då är

$$(1 + \sqrt{-1})^2 - 2(1 + \sqrt{-1}) + 2 = 1 + 2\sqrt{-1} + (-1) - 2 - 2\sqrt{-1} + 2 = 0$$

dvs $1+\sqrt{-1}$ är en lösning till ekvationen. Så gjorde några italienska matematiker under 1500-talet. Om man anser att $1+\sqrt{-1}$ bör uppfattas som en lösning till ekvationen $x^2-2x+2=0$ så bör man också ha en bra förklaring till varför. Det gäller att motivera användningen av $\sqrt{-1}$. Det tog 300 år innan man kunde ge en tillfredsställande förklaring och rent formellt konstruera de komplexa talen. Men exakt samma situation som med de komplexa talen har man med de hela, rationella och reella. Om man frågar ett barn om $x$ sådant att $2+x=3$ så får man svaret $x=1$. Tar man istället $3+x=2$ riskerar man att bli utskrattad. Ekvationen $2+x=3$ kan lösas i mängden av de naturliga talen, men $3+x=2$ kräver ett nytt talområde – de hela talen (i synnerhet de negativa). På liknande sätt går det att dela 4 i två lika delar (dvs lösa $2x=4$) i heltalen, men det går inte att dela 3 i två lika delar i den mängden (dvs lösa $2x=3$) – det behövs rationella tal för att göra det. Slutligen kan man hitta ett rationellt tal som multiplicerat med sig självt ger 4 (dvs lösa $x^2=4$), men det går inte att hitta ett rationellt tal som multiplicerat med sig självt ger 2 (dvs lösa $x^2=2$) – för att göra det behövs ett nytt talområde. Det naturliga önskemålet att polynomekvationer alltid skall gå att lösa, tvingar oss således att succesivt utvidga talområden. Om det finns en slutstation för denna utvidgningsprocess får vi veta lite senare. Så låt oss börja!

**(1.15) Från de naturliga talen till de hela**. Ekvationen $3+x=5$ definierar $x=2$ som sin lösning. Samma lösning ger $4+x=6$, $5+x=7$ osv. Man kan uppfatta 2 som paret (5,3) eller (6,4) eller (7,5) osv. Paret $(a,b)$ ger lösningen till $b+x=a$ med $a>b$. Paren $(a,b)$ och $(c,d)$ ger samma $x$ om $a-b=c-d$ dvs $a+d=b+c$. Men det finns par $(a,b)$ med $a=b$ och $a<b$. Har de en liknande tolkning ? T ex kan (3,5) uppfattas som lösningen till $5+x=3$. En sådan lösning finns inte bland de naturliga talen men själva tolkningen ger en idé hur man kan definiera heltalen.

Låt oss betrakta alla par $(a,b)$ där $a,b\in\mathbb{N}$. Vi säger att $(a,b)$ och $(c,d)$ tillhör samma klass (eller definierar samma heltal) då och endast då $a+d=b+c$

Alla par som tillhör samma klass som $(a,b)$ betecknas med $[(a,b)]$. En sådan klass kallar vi för ett heltal och kommer överens om följande beteckningar:

$$[(a,b)]=\begin{cases} a-b & \text{om } a>b,\\ 0 & \text{om } a=b,\\ -(b-a) & \text{om } a<b.\end{cases}$$

T ex är $[(1,3)]=-2$ och paren (1,3), (2,4), (3,5) osv tillhör samma klass. Vidare definierar man addition och multiplikation av heltal:

$$[(a,b)]+[(c,d)]=[(a+c,\ b+d)],$$

$$[(a,b)][(c,d)]=[(ac+bd,\ ad+bc)].^{\|}$$

$^{\|}$Tänk på $[(a,b)]$ och $[(c,d)]$ som $a-b$ och $c-d$. Då är $(a-b)(c-d)=(ac+bd)-(ad+bc)=[(ac+bd,ad+bc)]$.

Nu kan man kontrollera att heltalen bildar en ring men att gå igenom alla detaljer är ganska omständligt (se en av övningarna).

**(1.16) Från de hela talen till de rationella**. Konstruktionen är nästan identisk med den förra. Ekvationen $2x = 1$ definierar $1/2$. Samma lösning ger $4x = 2$, $6x = 3$ osv. Vi kan uppfatta $1/2$ som paren (1,2), (2,4), (3,6) osv. $-1/2$ får man som t ex $(-1, 2)$, $(-2, 4)$ osv. Allmänt kan lösningen till $bx = a$ uppfattas som paret $(a, b)$. Observera att $b \neq 0$. Två par $(a, b)$ och $(c, d)$ ger samma rationella tal om $\frac{a}{b} = \frac{c}{d}$. Men vi vill undvika bråk (de skall ju definieras!). Därför skriver vi villkoret på formen $ad = bc$. Nu kan vi starta vår konstruktion.

Betrakta alla par $(a, b)$ sådana att $a, b \in \mathbb{Z}$ och $b \neq 0$. Man säger att $(a, b)$ och $(c, d)$ , $d \neq 0$, tillhör samma klass om $ad = bc$. Alla par som tillhör klassen av $(a, b)$ betecknas med $[(a, b)]$. En sådan klass kallar vi för ett rationellt tal och inför beteckningen

$$[(a, b)] = \frac{a}{b} \text{ (eller } \quad a : b).$$

t ex är $[(1, 3)] = \frac{1}{3}$ och paren (1,3), (2,6), (3,9) tillhör samma klass (definierar samma rationella tal). Nu kan vi definiera addition och multiplikation av rationella tal:

$$\begin{aligned} \frac{a}{b} + \frac{c}{d} &= \frac{ad + bc}{bd}, \\ \frac{a}{b}\frac{c}{d} &= \frac{ac}{bd}, \end{aligned}$$

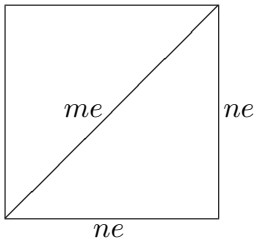
och kontrollera att man verkligen får en kropp (se övningar). Observera att:

$$\begin{aligned} \frac{a}{1} + \frac{c}{1} &= \frac{a + c}{1}, \\ \frac{a}{1}\frac{c}{1} &= \frac{ac}{1}, \end{aligned}$$

dvs talen $\frac{a}{1}$ adderas och multipliceras precis som heltalen $a$. Man kommer överens om att skriva $\frac{a}{1} = a$ så att de vanliga heltalen kan betraktas som en delmängd till de rationella talen.

**(1.17) Från de rationella talen till de reella**. Den biten av vägen är lite annorlunda och utgör ett mycket större steg än de två föregående. Först och främst hittar man lätt ekvationer med rationella koefficienter som saknar rationella lösningar, t ex $x^2 = 2$ (se nedan). Sådana ekvationer kräver en utvidgning av de rationella talen. Men det finns en annan mycket viktig

anledning till att man inser behovet av nya tal. Man upptäckte mycket tidigt att rationella tal inte är tillräckliga för att kunna mäta längder av sträckor. Följande klassiska exempel spelade en mycket viktig roll i matematikens utveckling. Betrakta en kvadrat och anta att man har fixerat en enhet $e$ sådan att kvadratens sida rymmer exakt $n$ enheter och dess diagonal $m$ enheter ($m$ och $n$ är naturliga tal).




Nu vet vi att $(ne)^2 + (ne)^2 = (me)^2$ så att $2n^2 = m^2$ dvs $\sqrt{2} = \frac{m}{n}$. Detta visar att om $e$ finns så är $\sqrt{2}$ ett rationellt tal. Pythagoras ** och hans elever visste mycket väl att det inte var fallet (vi skall visa om en stund att $\sqrt{2}$ inte är rationellt). Sin upptäckt om förhållandet mellan kvadratens sida och dess diagonal betraktade de som något som stred mot naturens ordning och försökte hemlighålla under en tid. Men konsekvensen blev att Euklides †† kort därefter kunde utveckla geometrin och läran om reella tal som mått på sträckor.

Hur visar man att $\sqrt{2}$ inte är rationellt? Vi skall visa det genom att utnyttja entydigheten av primfaktoruppdelningar av de naturliga talen. Antag att $\sqrt{2}$ är rationellt dvs att

$$\sqrt{2} = \frac{m}{n} \ ,$$

där $m, n$ är naturliga tal. Då är $2n^2 = m^2$. Eftersom $m^2$ och $n^2$ är kvadrater av heltal innehåller de ett jämnt antal primfaktorer 2 (möjligen 0 sådana faktorer). Alltså förekommer 2 som primfaktor i $2n^2$ ett udda antal gånger, medan i $m^2$ ett jämnt antal gånger så att $2n^2 \neq m^2$. Detta motsäger likheten $2n^2 = m^2$ och visar att $\sqrt{2}$ inte kan vara rationellt.

Låt oss nu konstruera de reella talen. Vi kan inte längre använda oss av tekniken med par av rationella tal. Men vi kan utnyttja följder av rationella tal. Reella tal (enligt gymnasiekunskaper) är decimaltal av typen $A = a, a_1a_2...a_n...$, där a är heltasdelen och $0, a_1a_2...a_n...$ är decimaldelen av A. Varje sådant tal kan approximeras med rationella tal – följden:

$x_1 = a, a_1$ ,
$x_2 = a, a_1a_2$ ,
$x_3 = a, a_1a_2a_3$ ,
. . .

**Pythagoras (572-500 f Kr)
††Euklides (ca 350 f Kr)

$x_n = a, a_1a_2a_3...a_n$ ,
. . .

består av rationella tal och konvergerar mot A dvs $\lim_{n\to\infty} x_n = A$. T ex är för $A = \pi$:

$x_1 = 3,1$ ,
$x_2 = 3,14$ ,
$x_3 = 3,141$ ,
. . .
$x_8 = 3,14159265$ ,
. . .

Låt nu $A$ vara ett positivt tal. Följden $\{x_1, x_2, ..., x_n, ...\} = \{x_n\}_1^\infty$ består då av rationella tal , den är växande och begränsad (ty $x_n \leq A$ för alla $n$). Vi vet att en sådan följd alltid har ett gränsvärde. Två följder $\{x_n\}$ och $\{x'_n\}$ har samma gränsvärde då och endast då deras skillnad går mot 0 dvs $\lim_{n\to\infty}(x_n - x'_n) = 0$. Positiva reella tal är alltså gränsvärden av växande och begränsade följder av rationella tal och två följder definierar samma reella tal som sitt gränsvärde om deras skillnad går mot 0. Men vi kan inte definiera reella tal som gränsvärden av sådana följder så länge de reella talen inte är konstruerade därför att en sådan definition skulle förutsätta att de reella talen (dvs gränsvärdena) är kända. Ändå identifierar vi varje reellt tal med ett gränsvärde på följande sätt. (Här börjar den formella definitionen.)

Betrakta alla växande och begränsade följder $\{x_1, x_2, ..., x_n, ...\} = \{x_n\}_1^\infty$, där $x_n$ är positiva rationella tal. Man säger att två följder $\{x_n\}_1^\infty$ och $\{x'_n\}_1^\infty$ tillhör samma klass (definierar samma reella tal) om deras skillnad $\{x_n - x'_n\}_1^\infty$ konvergerar mot 0 dvs $\lim_{n\to\infty}(x_n - x'_n) = 0$. Alla följder som tillhör klassen av $\{x_n\}_1^\infty$ betecknas med $[\{x_n\}_1^\infty]$. En sådan klass kallar man för ett positivt reellt tal. Nu kan man definiera addition och multiplikation av de positiva reella talen:

$$[\{x_n\}_1^\infty] + [\{x'_n\}_1^\infty] = [\{x_n + x'_n\}_1^\infty],$$

$$[\{x_n\}_1^\infty][\{x'_n\}_1^\infty] = [\{x_n x'_n\}_1^\infty].$$

För att nu konstruera de negativa reella talen och talet 0 måste man upprepa sama konstruktion som ledde oss från de naturliga talen till de hela: Man betraktar alla par $(a, b)$, där a och $b$ är positiva reella tal, och man identifierar $(a, b)$ med $(c, d)$ om $a + d = b + c$. Kontrollen att man får en kropp, att den är ordnad och fullständig är ganska lång men inte särskilt svår (detaljerna behandlas närmare i fortsättnigskurser i matematik [†]).

[†]Vanligen brukar man i stället för växande och begränsade följder betrakta godtyckliga följder av rationella tal $x_1, x_2, ..., x_n, ...$ sådana att avståndet mellan talen $x_i$ och $x_j$ går mot 0 då i och $j$ växer dvs $|x_i - x_j| \to 0$ då $i, j \to \infty$. Följder av den typen kallas Cauchyföljder.

**(1.18) Från de reella talen till de komplexa.** Vi vet redan att behovet av de komplexa talen upptäcktes i samband med andragradsekvationer med reella koefficienter. En så enkel ekvation som $x^2 = -1$ saknar reella lösningar. Antag att vi har en kropp $K$ som innehåller de reella talen $\mathbb{R}$ och sådan att det finns $\alpha \in K$ som satisfierar ekvationen $x^2 = -1$ dvs $\alpha^2 = -1$. Man kontrollerar utan större svårigheter (se (1.3)) att talen

$$a + b\alpha, \text{ där } \quad a, b \in \mathbb{R} \ ,$$

bildar en kropp. Det finns en mycket lång tradition att $\alpha$ betecknas med $i$ (ibland $j$) [‡]. I den kroppen har vi:

$$\begin{aligned} (a + bi) + (c + di) &= (a + c) + (b + d)i \ , \\ (a + bi)(c + di) &= (ac - bd) + (ad + bc)i. \end{aligned} \tag{1.19}$$

Än så länge har vi inte någon formell konstruktion av de komplexa talen (vi sade ju "Antag att en kropp K..."). Men vi har i alla fall en klar bild av hur en kropp som innehåller lösningen till $x^2 = -1$ måste se ut. Konstruktionen är mycket enkel. Idén är (som flera gånger tidigare) att uppfatta nya tal som par av redan kända: $a + bi$ kan uppfattas som $(a, b)$, där $a, b \in \mathbb{R}$.

**(1.20) Definition.** Med **komplexa tal** menar man alla par $(a, b)$, där $a, b \in \mathbb{R}$, som adderas och multipliceras på följande sätt:

$$(a, b) + (c, d) = (a + c, b + d),$$

$$(a, b)(c, d) = (ac - bd, ad + bc).$$

Mängden av de komplexa talen betecknas med $\mathbb{C}$. □

Beteckningen $(a, b)$ är lite omständlig. Därför observerar man att:

$$(a, 0) + (b, 0) = (a + b, 0),$$

[‡] "$i$" kommer från ordet "imaginär". Det finns ett mycket intressant val av terminologi när det gäller nya typer av tal. De naturliga talen bland de hela kallas positiva, de övriga negativa. Bråktalen bland de reella kallas rationella, de övriga irrationella. Komplexa talen $a + bi$ har realdel $a$ och en imaginärdel $b$. Alltså var allt nytt negativt, irrationellt och imaginärt (samt en lång tid impopulärt).

$$(a,0)(b,0) = (ab,0),$$

dvs paren (a,0) adderas och multipliceras precis som vanliga reella tal a. Man kommer överens om att skriva $(a,0) = a$ så att $\mathbb{R} \subset \mathbb{C}$. Därefter noterar man att $(0,1)(0,1) = (-1,0) = -1$. Man betecknar $(0,1) = i$. Nu har vi $(0,b) = (b,0)(0,1) = bi$ så att

$$(a,b) = (a,0) + (0,b) = a + bi$$

och vi får våra gamla beteckningar (1.19). Det som återstår är kroppstrukturen:

**(1.21) Sats.** *De komplexa talen $a + bi$, där $a, b \in \mathbb{R}$ och $i^2 = -1$, bildar en kropp.*

Satsen visas lätt, men beviset tar lite tid därför att man måste kontrollera alla villkor (a) – (k) på sidan 5.

Innan vi tittar på möjligheten att gå vidare med liknande konstruktioner låt oss summera våra kunskaper. Nu kan vi säga att med ett tal menar man alltid ett komplext tal. I synnerhet kan det vara fråga om ett naturligt, helt, rationellt eller reellt tal. Med en talring (eller talkropp) menas alltid en ring (eller kropp) bestående av tal.

$\mathbb{Z}$ är den minsta talringen därför att om $R$ är en talring så gäller att $1 \in R$ vilket ger att $1+1$, $1+1+1, ... \in R$ dvs $R$ innehåller de naturliga talen. Vidare måste $0 \in R$ och $-x \in R$ om $x \in R$ så att $R$ innehåller $\mathbb{Z}$. $\mathbb{Q}$ är den minsta talkroppen därför att varje kropp $K$ innehåller $\mathbb{Z}$ och därmed också alla tal $\frac{a}{b}$, där $a, b \in \mathbb{Z}$ och $b \neq 0$, dvs $K \supseteq \mathbb{Q}$.

De reella talen bildar den största ordnade talkroppen. Låt oss först konstatuera att $\mathbb{C}$ inte är ordnad. Antag nämligen att man kan välja en mängd $P$ av positiva element i $\mathbb{C}$. Då är $i \in P$ eller $-i \in P$. I varje fall är $(\pm i)^2 = -1 \in P$ vilket är omöjligt ty redan $1 \in P$ (se (1.8)). Man visar (men det är inte helt banalt) att om en talkropp kan ordnas så kan den inte innehålla något komplext tal $a + bi$ med $b \neq 0$ dvs den ligger i $\mathbb{R}$. I den meningen är $\mathbb{R}$ den största ordnade talkroppen.

De komplexa talen bildar den största talkroppen. I vilken mening? Man kan fråga sig som tidigare om det finns polynomekvationer, nu med komplexa koefficienter, som inte kan lösas i det komplexa talområdet. Svaret på den frågan kommer från C.F. Gauss som år 1799 visade följande sats:

**(1.22) Polynomalgebrans fundamentalsats.** *Varje polynomekvation av positiv grad med komplexa koefficienter har en komplex lösning.*

Satsen säger att om $p(X) = a_n X^n + ... + a_1 X + a_0$ , där $a_i \in \mathbb{C}$, $n > 0$ och $a_n \neq 0$ så är $p(z) = 0$

för ett komplext tal $z \in \mathbb{C}$. Man säger också att kroppen av de komplexa talen är algebraiskt sluten. Det finns flera olika bevis för den satsen men alla kräver lite större förkunskaper [§].

Den sista satsen säger att det inte finns något vidare behov att utvidga komplexa talkroppen p g a olösbara polynomekvationer. I den meningen bildar de komplexa talen den största talkroppen. Men en lång tid innan man var medveten om detta, upptäckte man matematiska objekt som kunde användas till att beskriva och utforska naturen och som i många avseenden liknade talen. Du har säkert hört om sådana begrepp som vektor, matris, kvaternion eller tensor. Vektorer och matriser är uppsättningar av tal som också kan adderas och multipliceras på ett lämpligt sätt. De ger en möjlig generalisering av talbegreppet. Kvaternioner, som enklast kan beskrivas med hjälp av matriser, är ett annat exempel på en algebraisk struktur som ligger mycket nära de komplexa talen. Vi skall avsluta detta avsnitt genom att säga några ord om just kvaternioner.

W.R. Hamilton [¶] som gav en formell definition av komplexa tal i form av reella talpar försökte gå vidare med sin idé och betrakta par av komplexa tal. Han ville definiera addition och multiplikation av sådana par och möjligen få en ny kropp. Faktum är att det finns många kroppar som innehåller de komplexa talen, men de måste alltid innehålla element som inte uppfyller någon icke-trivial polynomekvation med komplexa koefficienter (t ex kroppen $\mathbb{C}(X)$ av alla rationella funktioner med komplexa koefficienter dvs alla bråk $\frac{p(X)}{q(X)}$, där $p(X)$ och $q(X)$ är polynom med komplexa koefficienter – variabeln $X$ är inte ett nollställe till något nollskilt polynom med komplexa koefficienter). Därför är det inte längre möjligt att konstruera en kropp större än $\mathbb{C}$ vars element uppfyller polynomekvationer med komplexa koefficienter. Hamilton lyckades dock att konstruera en struktur som har den egenskapen och som uppfyller alla räknelagar för en kropp med bara ett undantag. På Brougham Bridge i Dublin där Hamilton bodde finns idag en tavla med följande text: “Here as he walked by on the 16th of October 1843 Sir William Rowan Hamilton in a flash of genius discovered the fundamental formula for quaternion multiplication $i^2 = j^2 = k^2 = ijk = -1$ and cut it in on a stone of this bridge”. Han publicerade sina resultat år 1853. Konstruktionen av kvaternioner, som spelar en mycket viktig roll i många matematiska och fysikaliska teorier, är följande. Betrakta alla par $(z_1, z_2)$, där $z_1, z_2$ är komplexa tal. Definiera

$$(z_1, z_2) + (z_1', z_2') = (z_1 + z_1', z_2 + z_2'),$$

och

$$(z_1, z_2)(z_1', z_2') = (z_1 z_1' - z_2 \bar{z}_2', z_1 z_2' + \bar{z}_1' z_2),$$

där $\bar{z} = a - bi$ ($z$ konjugat) om $z = a + bi$. Man observerar att

$$(z_1, 0) + (z_1', 0) = (z_1 + z_1', 0),$$

och

$$(z_1, 0)(z_1', 0) = (z_1 z_1', 0).$$

Detta visar att de komplexa talen kan identifieras med paren $(z, 0)$. Därför skriver vi $(z, 0) = z$. Beteckna också $(0, 1) = j$ och $(0, i) = k$. Vi har $j^2 = (0, 1)(0, 1) = (-1, 0) = -1$ och

[§] Beviset ges i kursen “Analytiska funktioner”. Ett nästan rent algebraiskt bevis i “Galoisteori”.

[¶] W.R. Hamilton (1805-1865).

$k^2 = (0,i)(0,i) = (-1,0) = -1$. Dessutom har vi $(0, c+di) = (0,c) + (0,di) = (c,0)(0,1) + (d,0)(0,i) = cj + dk$. Därför kan vi skriva:

$$q = (a+bi, c+di) = (a+bi, 0) + (0, c+di) = a + bi + cj + dk.$$

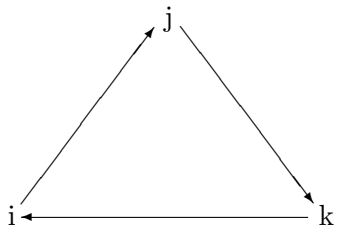
Detta är en typisk kvaternion. Man kan kontrollera direkt att $ijk = -1$ (se övningen om kvaternioner).

Men för att snabbt kunna räkna med kvaternioner är det bäst att kontrollera följande multiplikationsregler:

ij = -ji = k,

jk = -kj = i,

ki = -ik = j.




Vi ser att multiplikation av kvaternioner inte är kommutativ. Låt oss sammanfatta**:**

**(1.23) Sats.** *Alla kvaternioner $a+bi+ci+dk$, där $i^2 = j^2 = k^2 = -1$ och $ij = -ji = k$, bildar en algebraisk struktur* $\mathbf{H}$ *som uppfyller alla villkor i definitionen av en kropp med undantag av multiplikationens kommutativitet. Dessutom uppfyller varje kvaternion en andragradsekvation med reella koefficienter.*

För det sista påståendet i satsen se övningen om kvaternioner. Ibland säger man att $H$ är en icke-kommutativ kropp, men termerna **skevkropp** eller **divisionsring** är mera vanliga. Satsen är inte svår att bevisa.

# ÖVNINGAR

1.1. Vilka av följande talmängder är ringar? Vilka av dem är kroppar?

(a) $\{0, 1\}$,

(b) $a + b\sqrt{3}$, där $a, b \in \mathbb{Z}$,

(c) $a + b\sqrt{5}$, där $a, b \in \mathbb{Q}$,

(d) $a + b\sqrt[3]{2}$, där $a, b \in \mathbb{Z}$,

(e) $a + b\sqrt[3]{2} + c\sqrt[3]{4}$, där $a, b, c \in \mathbb{Z}$,

(f) $a + b\sqrt{2} + c\sqrt{3}$, där $a, b, c \in \mathbb{Z}$.

1.2. Visa att i varje ring $R$ gäller följande likheter:

(a) $a0 = 0$ då $a \in R$,

(b) $(-1)(-1) = 1$,

(c) $-(-a) = a$ då $a \in R$,

(d) $(-a)b = -ab$ då $a, b \in R$,

(e) $(-a)(-b) = ab$ då $a, b \in R$.

1.3. (a) Visa att alla tal av typ

$$a + b\sqrt{2} + c\sqrt{3} + d\sqrt{6}, \text{ där } \quad a, b, c, d \in \mathbb{Q},$$

bildar en kropp.

**Ledning**. Visa att $\sqrt{3} \notin \mathbb{Q}[\sqrt{2}]$ och utnyttja sats (1.3).

(b) Är det möjligt att skriva talet

$$\frac{1}{1 + \sqrt{2} + \sqrt{3} + \sqrt{6}}$$

på formen $a + b\sqrt{2} + c\sqrt{3} + d\sqrt{6}$, där $a, b, c, d$ är rationella tal?

Gör det om Du ser en enkel lösning!

(c) Hur kan man generalisera (a)?

1.4. Motivera att binomialsatsen gäller i varje ring.

1.5. (a) Visa att $\mathbb{Q}[\sqrt{2}] \neq \mathbb{Q}[\sqrt{3}]$.

(b) Försök generalisera (a) och ge exempel på oändligt många olika kroppar.

1.6. (a) Bestäm decimalutvecklingen av talen $\frac{3}{11}$ och $\frac{1}{7}$.

(b) Motivera att decimalutvecklingen av ett rationellt tal är periodisk.

**Ledning:** Analysera divisionsalgoritmen då man decimalutvecklar bråktalen.

**Anmärkning.** Man visar ganska enkelt att om ett reellt tal har periodisk decimalutveckling så är det rationellt.

1.7. Låt $a$ och $b$ vara irrationella tal. Vad kan man säga om talen $a^{-1}$ och $ab$ ? Är de också irrationella?

1.8. Förklara varför 0,999... = 1.

I uppgifterna 1.9 – 1.12 nedan är $K$ en ordnad kropp och $a, b, c \in K$.

1.9. Visa att $K$ har följande egenskaper:

(a) $a < b \Rightarrow a + c < b + c$,

(b) $a < b$ och $c > 0 \Rightarrow ac < bc$,

(c) hur förändras (b) då man ersätter $a < b$ med $a \leq b$?

1.10. Visa att relationen $a \leq b$ är en partiell ordning i $K$ dvs

(a) $a \leq a$ (reflexivitet),

(b) $a \leq b$ och $b \leq a \Rightarrow a = b$ (antisymmetri),

(c) $a \leq b$ och $b \leq c \Rightarrow a \leq c$ (transitivitet).

1.11. Visa att

(a) $|ab| = |a||b|$,

(b) $|a + b| \leq |a| + |b|$ (triangelolikheten).

1.12. Är följande implikationer sanna eller falska?

(a) $a < b \Rightarrow a^2 < b^2$,

(b) $a < b \Rightarrow a^3 < b^3$?

1.13. (a) De naturliga talen bildar en växande följd $1 < 2 < 3 \ldots$ . Visa att den inte är begränsad.

(b) Visa "Arkimedes princip": Om $a, b$ är två positiva reella tal så finns det ett naturligt tal $n$ så att $na > b$.

(c) Låt $a, b$ vara två reella tal och låt $a < b$. Visa att det finns ett rationellt tal $\frac{m}{n}$ sådant att $a < \frac{m}{n} < b$.

**Ledning:** Välj $n$ så att $n(a - b) > 1$. Välj därefter minsta $m$ så att $m > nb$.

1.14. (a) Visa att $\sqrt{3}$ är icke-rationellt genom att jämföra antalet primfaktorer 3 till vänster och till höger i likheten $3n^2 = m^2$.

(b) Visa på liknande sätt att $\sqrt{p}$ är icke-rationellt då $p$ är ett godtyckligt primtal.

(c) Har Du några förslag till hur man kan generalisera (b)?

1.15. (a) Visa att talet ${}^2\log 5$ är icke-rationellt.

(b) Kan Du föreslå några andra tal, i stället för 5 i (a), för vilka påståendet gäller?

1.16. Betrakta alla par $(a, b)$, där $a, b \in \mathbb{N}$ och visa att relationen

$$(a, b)R(c, d) \Longleftrightarrow a + d = b + c$$

är en ekvivalensrelation. Motivera därefter att det finns en bijektion mellan ekvivalensklasserna och heltalen.

1.17. (a) När har ett rationellt tal $\frac{a}{b}$ en invers? Skriv inversen på formen $[(c, d)]$.

(b) Kontrollera att om

$$[(a, b)] = [(a', b')] \quad \text{och} \quad [(c, d)] = [(c', d')]$$

är två rationella tal ($ab' = a'b$ och $cd' = c'd$) så gäller

$$\frac{a}{b} + \frac{c}{d} = \frac{a'}{b'} + \frac{c'}{d'} \quad \text{och} \quad \frac{a}{b}\frac{c}{d} = \frac{a'}{b'}\frac{c'}{d'}$$

(dvs summan och produkten av två rationella tal beror inte på hur dessa tal representeras i form av bråk).

1.18. Skriv följande kvaternioner på formen $a + bi + cj + dk$ :

(a) $(1 + i)(1 + j)$,

(b) $(i + j + k)^2$,

(c) $(1 + 2i + 3j + 4k)(1 - 2i - 3j - 4k)$,

(d) $ijk$.

1.19. (a) Visa att $q = 1 + i + j + k$ och $\bar{q} = 1 - i - j - k$ satisfierar ekvationen $x^2 - 2x + 4 = 0$.

(b) Visa att $q = a+bi+cj+dk$ satisfierar en kvadratisk ekvation med reella koefficienter.

# AVSNITT 2

# RESTARITMETIKER

I detta avsnitt får vi se ringar och kroppar av en annorlunda karaktär. De är nära besläktade med heltalen och har en mycket stor betydelse inom talteorin och dess tillämpningar i datalogi och datateknik.

När man adderar eller multiplicerar två tal som t ex

$$\begin{array}{rr} & 128 \\ + & 39 \\ \hline & \ldots 7 \end{array} \qquad\qquad \begin{array}{rr} & 128 \\ \times & 43 \\ \hline & \ldots 4 \end{array}$$

så bestämmer man först den sista siffran. De operationer som leder till resultatet kallas addition och multiplikation modulo 10. Man adderar $8 + 9$ på vanligt sätt, men sista siffran är resten av $8 + 9$ vid division med 10. På liknande sätt har vi $3 \cdot 8 = 24$, men som sista siffran får vi 4 dvs resten av 24 vid division med 10. Om talen är givna i binära systemet (bas 2) som t ex

$$\begin{array}{rr} & 1011 \\ + & 101 \\ \hline & \ldots 0 \end{array} \qquad\qquad \begin{array}{rr} & 1011 \\ \times & 111 \\ \hline & \ldots 1 \end{array}$$

så räknar man modulo 2 dvs först som vanligt, men därefter tar man resten vid division med 2. Operationerna modulo 10 eller 2 eller modulo ett godtyckligt annat naturligt tal har stor betydelse.

I restaritmetiker arbetar man med rester av heltal vid division med ett fixerat naturligt tal $n$. Vi skall förutsätta att $n > 1$, ty annars har vi bara resten 0. Om $a$ är ett heltal så är

$$a = nq + r,$$

där $q$ är kvoten och $r$ är resten. Resten $r$ kan alltid väljas så att $0 \leq r < n$ dvs det finns $n$ stycken rester : $0, 1, ..., n-1$. Mängden av dessa betecknas ofta med $\mathbb{Z}_n$ (eller $\mathbb{Z}/(n)$). Vi skall skriva $r = [a]_n$ för att uttrycka det faktum att $r$ är resten vid division av $a$ med $n$. Följande egenskaper hos rester kommer att utnyttjas många gånger:

**(2.1) Lemma.** $[a]_n = [b]_n$ *då och endast då* $n|a-b$ [*]. *Med andra ord ger* $a$ *och* $b$ *samma rest vid division med* $n$ *då och endast då* $n$ *är en delare till deras skillnad* $a-b$.

**Bevis.** Om $[a]_n = [b]_n$ så är $a = nq_1 + r$ och $b = nq_2 + r$, vilket ger $a - b = n(q_1 - q_2)$ dvs $n|a-b$.

Omvänt, låt $n|a-b$ dvs $a - b = nq$. Om $a = nq_1 + r_1$ och $b = nq_2 + r_2$ så är

$$a - b = n(q_1 - q_2) + r_1 - r_2$$

dvs

$$r_1 - r_2 = (a-b) - n(q_1 - q_2) = n[q - (q_1 - q_2)].$$

Detta betyder att $n|r_1 - r_2$. Men $0 \leq r_1, r_2 < n$ så att $r_1 - r_2$ är delbart med $n$ endast om $r_1 - r_2 = 0$ dvs $[a]_n = [b]_n$. □

**(2.2) Exempel.** (a) $[3]_5 = [-2]_5$ ty $5|3-(-2) = 5$.
(b) $[n-1]_n = [-1]_n$ ty $n|(n-1)-(-1) = n$. □

**(2.3) Anmärkning.** C.F. Gauss introducerade en mycket viktig beteckning för att uttrycka likheten $[a]_n = [b]_n$ (dvs $n|a-b$). Han skrev:

$$a \equiv b \pmod{n}$$

vilket utläses "$a$ är kongruent med $b$ modulo $n$". Relationen " $\equiv$ " kallas kongruens (här modulo $n$). Vi kommer att använda den beteckningen ganska ofta. □

Kan man helt allmänt addera och multiplicera rester (precis som de sista siffrorna vid addition och multiplikation av heltal)? Det är helt klart att det går men en formell definition är nödvändig. Vi skall skriva $\oplus$ och $\odot$ för att ha en distinktion mellan addition av vanliga heltal och rester. Men den distinktionen är inte nödvändig (man kan skriva " $+$ " och " $\cdot$ " om man så vill).

**(2.4) Definition.** $[a]_n \oplus [b]_n = [a+b]_n$ och $[a]_n \odot [b]_n = [ab]_n$. □

[*]Man skriver $a|b$ och säger att "$a$ är en delare till $b$" om $b = aq$ för något heltal $q$. Man säger också att $b$ är en multipel av a. Om a inte är en delare till $b$ skriver man $a \not| b$.

Definitionen säger att summan av resterna $[a]_n$ och $[b]_n$ får man genom att addera talen $a$ och $b$ på vanligt sätt och därefter ta resten vid division av $a+b$ med $n$. Samma sak gäller för produkten. Här finns det dock en liten detalj som kräver en stunds eftertanke. Om man har två helt godtyckliga heltal $a$ och $b$ som slutar, låt oss säga, på 3 och 8 dvs $[a]_{10}=3$ och $[b]_{10}=8$ så får man alltid samma slutsiffra för $a+b$ och $ab$ dvs $[a+b]_{10}=1$ och $[ab]_{10}=4$. Gäller samma sak helt allmänt då man ersätter 10 med någon annan modul t ex 3 eller 4? Med andra ord är höger led i definitionen (2.4) alltid samma oberoende av $a$ och $b$ till vänster? Frågan kan också formuleras så här: är definitionen (2.4) korrekt? Låt oss kontrollera att den är helt korrekt! Låt:

$$[a]_n=[a']_n \text{ och } [b]_n=[b']_n. \tag{2.5}$$

Vi vill visa att

$$[a+b]_n=[a'+b']_n \quad \text{och} \quad [ab]_n=[a'b']_n. \tag{2.6}$$

Med beteckningen " $\equiv$ " betyder det att

$$a\equiv a' \pmod n \text{ och } b\equiv b' \pmod n$$

ger

$$a+b\equiv a'+b' \pmod n \quad \text{och} \quad ab\equiv a'b' \pmod n$$

dvs kongruenser, precis som likheter, kan adderas och multipliceras ledvis.

**Bevis.** $[a]_n=[a']_n$ och $[b]_n=[b']_n$ betyder att $a-a'=nq_1$ och $b-b'=nq_2$. Alltså är

$$(a+b)-(a'+b')=n(q_1+q_2)\ ,$$

dvs

$$[a+b]_n=[a'+b']_n\,.$$

Vidare är

$$ab-a'b'=(a-a')b+a'(b-b')=n(q_1b+q_2a')$$

dvs

$$[ab]_n = [a'b']_n.$$

□

Nu kan vi konstatera följande:

**(2.7) Sats.** *Alla rester vid division med n bildar en ring $\mathbb{Z}_n$ med avseende på addition och multiplikation av rester:*

$$[a]_n \oplus [b]_n = [a+b]_n$$

*och*

$$[a]_n \odot [b]_n = [ab]_n.$$

**Bevis.** Vi vet redan att summan och produkten av rester är rester (detta ger villkoren (a) och (f) i definitionen av begreppet ring – se (1.5) och (1.6)). Associativiteten:

$$([a]_n \oplus [b]_n) \oplus [c]_n = [a]_n \oplus ([b]_n \oplus [c]_n)$$

får vi enkelt ty

$$VL = ([a]_n \oplus [b]_n) \oplus [c]_n = [a+b]_n \oplus [c]_n = [(a+b)+c]_n,$$

och

$$HL = [a]_n \oplus ([b]_n \oplus [c]_n) = [a]_n \oplus [b+c]_n = [a+(b+c)]_n,$$

så att $VL = HL$. Lika enkelt är det med kommutativiteten:

$$[a]_n \oplus [b]_n = [a+b]_n = [b+a]_n = [b]_n \oplus [a]_n.$$

Vi har

$$[a]_n \oplus [0]_n = [a+0]_n = [a]_n$$

dvs $[0]_n$ är neutral för addition. Likheten

$$[a]_n \oplus [-a]_n = [0]_n$$

säger att $[-a]_n$ är motsatt till $[a]_n$. De övriga villkoren i definitionen av begreppet ring (se (1.6)) lämnar vi som övning. □

Låt oss som exempel skriva ut additions och multiplikationstabellerna för $\mathbb{Z}_3$:

| $\oplus$ | $[0]_3$ | $[1]_3$ | $[2]_3$ |
|---|---|---|---|
| $[0]_3$ | $[0]_3$ | $[1]_3$ | $[2]_3$ |
| $[1]_3$ | $[1]_3$ | $[2]_3$ | $[0]_3$ |
| $[2]_3$ | $[2]_3$ | $[0]_3$ | $[1]_3$ |

| $\odot$ | $[0]_3$ | $[1]_3$ | $[2]_3$ |
|---|---|---|---|
| $[0]_3$ | $[0]_3$ | $[0]_3$ | $[0]_3$ |
| $[1]_3$ | $[0]_3$ | $[1]_3$ | $[2]_3$ |
| $[2]_3$ | $[0]_3$ | $[2]_3$ | $[1]_3$ |

Ofta kommer vi att utelämna $[\ ]_n$ när det är klart vilka rester vi menar. T ex är tabellerna för $\mathbb{Z}_4$ följande:

| $\oplus$ | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| 0 | 0 | 1 | 2 | 3 |
| 1 | 1 | 2 | 3 | 0 |
| 2 | 2 | 3 | 0 | 1 |
| 3 | 3 | 0 | 1 | 2 |

| $\odot$ | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 0 |
| 1 | 0 | 1 | 2 | 3 |
| 2 | 0 | 2 | 0 | 2 |
| 3 | 0 | 3 | 2 | 1 |

I praktiska tillämpningar (utanför matematiken) är $\mathbb{Z}_2$ en av de viktigaste ringarna: Den har följande räknelagar:

| $\oplus$ | 0 | 1 |
|---|---|---|
| 0 | 0 | 1 |
| 1 | 1 | 0 |

| $\odot$ | 0 | 1 |
|---|---|---|
| 0 | 0 | 0 |
| 1 | 0 | 1 |

En viktig fråga är om det kan inträffa att $\mathbb{Z}_n$ är en kropp. Låt oss repetera att $\mathbb{Z}_n$ är en kropp om villkoret (j) i definitionen av begreppet kropp (se (1.6)) gäller dvs om till varje $r \in \mathbb{Z}_n$, $r \neq 0$, existerar en invers $r'$ så att $r \odot r' = 1$. Man inser lätt att $\mathbb{Z}_2$, $\mathbb{Z}_3$ och $\mathbb{Z}_5$ är kroppar. För $\mathbb{Z}_2$ är det klart ($1 \odot 1 = 1$). I $\mathbb{Z}_3$ har vi $1 \odot 1 = 1$ och $2 \odot 2 = 1$ så att både 1 och 2 har invers. I $\mathbb{Z}_5$ är det också klart ty $1 \odot 1 = 1$, $2 \odot 3 = 1$ och $4 \odot 4 = 1$ så att 1,2,3 och 4 har invers. $\mathbb{Z}_4$ är inte en kropp därför att 2 saknar invers (man kan inte hitta $r \in \mathbb{Z}_4$ så att $2 \odot r = 1$). När är $\mathbb{Z}_n$ en kropp? Svaret är, ganska överraskande, att $\mathbb{Z}_n$ är en kropp då och endast då $n$ är ett primtal. Vi skall bevisa det om en stund som ett resultat av en mera allmän observation.

I en godtycklig ring $R$ kan det finnas flera element utöver 1 som har invers. Om $R$ är en kropp så har alla element $\neq 0$ invers. Bland heltalen $\mathbb{Z}$ finns det bara två som har heltalig invers – det är 1 och $-1$. Allmänt har man följande begrepp:

**(2.8) Definition.** Ett element $a$ i en ring $R$ kallas en **enhet** om $a$ har invers dvs om det finns $a' \in R$ så att $aa' = 1$. □

Vi skall hitta alla rester som har invers i $\mathbb{Z}_n$. Tag t ex $\mathbb{Z}_4$. Här är $1 \odot 1 = 1$ och $3 \odot 3 = 1$ så att 1 och 3 har invers (men inte 2). I $\mathbb{Z}_7$ har alla rester $\neq 0$ inverser ty 7 är ett primtal och således är $\mathbb{Z}_7$ en kropp: $1 \odot 1 = 1$, $2 \odot 4 = 1$, $3 \odot 5 = 1$, $6 \odot 6 = 1$.

**(2.9) Sats.** *$r \in \mathbb{Z}_n$ har invers då och endast då $r$ och $n$ saknar gemensamma delare $\neq 1$ dvs $SGD(r, n) = 1$.*

Vårt bevis av satsen utnyttjar en mycket viktig egenskap som Du kommer att möta många gånger: Låt $a, b$ vara två heltal. Då finns det heltal $x, y$ sådana att

$$ax + by = SGD(a, b)^{\dagger}. \tag{2.10}$$

**Bevis.** Om $SGD(r, n) = 1$ så finns det två heltal $x, y$ sådana att

$$rx + ny = 1$$

Alltså är $[rx + ny]_n = [1]_n$. Men $[ny]_n = [0]_n$ så att $[rx]_n = [r]_n \odot [x]_n = [1]_n$ dvs $[x]_n$ är inversen till $[r]_n = r$.

Omvänt. Låt $[r]_n \odot [r']_n = [1]_n$ dvs $[rr']_n = [1]_n$. Enligt (2.1) får vi $n|rr' - 1$ dvs $rr' - 1 = nq$ så att $rr' - nq = 1$. Den likheten säger att $SGD(r, n) = 1$ ty en gemensam delare $d > 0$ till $r$ och $n$ är en delare till 1 dvs $d = 1$. □

Nu får vi omedelbart:

**(2.11) Följdsats.** *$\mathbb{Z}_n$ är en kropp då och endast då $n$ är ett primtal.*

**Bevis.** Om $n = p$ är ett primtal så har varje rest $r \neq 0$ invers därför att resterna $1, 2, ..., p-1$ i $\mathbb{Z}_p$ saknar gemensamma delare med $p$ dvs $SGD(r, p) = 1$ då $r = 1, 2, ..., p-1$. Om däremot $n$ är sammansatt dvs $n = kl$, där $1 < k < n$ och $1 < l < n$ så är $SGD(k, n) = k > 1$, vilket innebär att resten $k$ saknar invers enligt (2.9). □

Nu skall vi gå igenom några mycket berömda satser i talteorin som enkelt kan bevisas med hjälp av restaritmetiker. På senare år visade det sig att dessa satser har mycket väsentliga tillämpningar i samband med datorberäkningar och datorsäkerhet. Men talteori (fast lite mer avancerad) har också kommit in i teoretisk fysik i samband med strängteorin.

[†] Denna likhet är en mycket enkel konsekvens av Euklides algoritm. Se avsnittet om "Delbarhet och primtal".

Vi skall börja med en sats som visades redan år 1682 av G.W. Leibniz [‡], men som kallas Wilsons sats. John Wilson levde senare än Leibniz och lämnade matematiken för juridik.

**(2.12) Wilson's sats.** *Om $p$ är ett primtal så är $p|(p-1)!+1$.*

Innan vi bevisar satsen låt oss betrakta ett exempel. Tag $p = 13$. Satsen säger att $13|12!+1$. Modulo 13 har vi

$$1 \odot 1 = 1,\ 2 \odot 7 = 1,\ 3 \odot 9 = 1,\ 4 \odot 10 = 1,\ 5 \odot 8 = 1,\ 6 \odot 11 = 1,\ 12 \odot 12 = 1.$$

Alltså är (modulo 13):

$$1 \odot 2 \odot 3 \odot 4 \odot 5 \odot 6 \odot 7 \odot 8 \odot 9 \odot 10 \odot 11 \odot 12 =$$
$$= 1 \odot (2 \odot 7) \odot (3 \odot 9) \odot (4 \odot 10) \odot (5 \odot 8) \odot (6 \odot 11) \odot 12 = 12 = -1$$

dvs $13|12!+1$.

**Bevis.** Betrakta kroppen $\mathbb{Z}_p$. Vi skall beräkna $[(p-1)!]_p = [1 \cdot 2 \cdot ... \cdot (p-1)]_p$ och visa att $[(p-1)!]_p = [-1]_p$ vilket just är satsens innehåll.

Varje faktor $r$ i produkten $1 \odot 2 \odot ... \odot (p-1)$ har sin invers $s$ modulo $p$ dvs $r \odot s = 1$. Om $r \neq s$ så kan man utelämna både $r$ och $s$. Men det kan inträffa att $r = s$ dvs $r \odot r = 1$. När? Vi har $[r^2]_p = [1]_p$ då och endast då $p|r^2 - 1 = (r-1)(r+1)$ dvs $p|r-1$ eller $p|r+1$. Men $0 \leq r \leq p-1$ så att $r = 1$ eller $r = p-1$. Alltså finns det två faktorer i produkten $1 \odot 2 \odot ... \odot (p-1)$ som är kvar: 1 och $p-1$ dvs

$$1 \odot 2 \odot ... \odot (p-1) = 1 \odot (p-1)\ .$$

Men $p - 1 \equiv -1 \pmod{p}$ så att $[(p-1)!]_p = [-1]_p$ , vilket visar satsen. □

**(2.13) Anmärkning.** Wilsons sats karakteriserar primtalen i den meningen att om $n|(n-1)!+1$ så är $n$ ett primtal (vi lämnar detta påstående som en bra och enkel övning – se övning 5). Man kan testa med hjälp av datorer om $n$ är ett primtal genom att dividera $(n-1)!+1$ med $n$. Men den metoden är inte särskilt bra därför att $(n-1)!$ växer mycket snabbt med $n$. □

Nu vill vi visa en av de mest berömda satserna inom talteorin – Fermats [§] lilla sats (om den stora får du höra under föreläsningarna). Vi behöver dock en enkel observation som har en mycket allmän karaktär:

[‡]Gottfrid Wilhelm Leibniz (1/7 1646 – 14/11 1716) var en framstående tysk matematiker som skapade differential och integralkalkylen (oberoende av I.Newton).

[§]Pierre de Fermat (20/8 1601 – 12/1 1663).

**(2.14) Proposition.** *Låt $R$ vara en ring.*

*(a) Produkten av två enheter $a$ och $b$ i $R$ också är en enhet.*

*(b) Om $a$ är en enhet i $R$ och $ax = ay$, där $x, y \in R$, så är $x = y$.*

*(c) Om $a$ är en enhet i $R$ och $x_1, x_2, ..., x_n$ är olika element i $R$ så är också $ax_1, ax_2, ..., ax_n$ olika.*

**Bevis.** (a) Om $aa' = 1$ och $bb' = 1$ så $(ab)(a'b') = 1$ dvs $ab$ är en enhet.

(b) Man kan multiplicera $ax = ay$ med $a^{-1}$ vilket ger $x = y$.

(c) Om $x_i \neq x_j$ så är $ax_i \neq ax_j$ ty $ax_i = ax_j$ ger enligt (b) att $x_i = x_j$. □

Nu kan vi visa Fermats lilla sats:

**(2.15) Fermats lilla sats.** *Om $p$ är ett primtal och $a$ är ett heltal så är $p|a^p - a$, med andra ord, $a^p \equiv a \pmod{p}$.*

Tag ett exempel först. Om $p = 5$ och $a = 3$ får vi $5|3^5 - 3 = 240$.

**Bevis.** Om $p|a$ så är påståendet klart. Låt oss anta då att $p \nmid a$ dvs $r = [a]_p \neq 0$. Betrakta resterna $1, 2, ..., p-1 \in Z_p$ och låt oss multiplicera alla dessa rester med $r \neq 0$. Då får vi $(p-1)$ olika enheter i $Z_p$ (se (2.14) (a) och $(c)$) :

$$1 \odot r,\ 2 \odot r, ..., (p-1) \odot r$$

Alltså återfår vi resterna $1, 2, ..., p-1$ (eventuellt i någon annan ordning). I varje fall är

$$1 \odot r \odot 2 \odot r \odot ... \ \odot (p-1) \odot r = 1 \odot 2 \odot ... \ \odot (p-1).$$

Nu kan vi stryka $1, 2, ..., p-1$ till vänster och till höger (se (2.14) (b)) och vi får

$$r^{p-1} = 1$$

dvs

$$[a^{p-1}]_p = [1]_p,$$

vilket betyder att $p|a^{p-1} - 1$. Men i så fall är också $p|a(a^{p-1} - 1) = a^p - a$. □

Fermats lilla sats har en generalisering som visades 100 år senare av L. Euler [¶]. (Eulers sats utgör grunden för konstruktionen av de mest använda krypteringssystemen inom datorsäkerhetstekniken — så kallade RSA-krypton. Se övningarna). Innan vi visar Eulers sats måste vi säga några ord om Eulers funktion $\varphi$.

Hur många rester i $\mathbb{Z}_n$ har invers? Antalet sådana rester betecknas med $\varphi(n)$. Funktionen $\varphi(n)$ kallas Eulers funktion. Enligt villkoret i (2.9) har vi:

(2.16) $$\varphi(n) = \text{antalet } r \text{ sådana att } 0 \leq r < n \text{ och } SGD(r, n) = 1.$$

Det är lätt att beräkna: $\varphi(1) = 1$, $\varphi(2) = 1$, $\varphi(3) = 2$, $\varphi(4) = 2$, $\varphi(5) = 4$, $\varphi(6) = 2$, $\varphi(7) = 6$, $\varphi(8) = 4$, $\varphi(9) = 6$, $\varphi(10) = 4$ osv. Vi återkommer till Eulers funktion i samband med övningarna. Nu kan vi formulera och bevisa Eulers sats:

**(2.17) Eulers sats.** *Låt $a$ och $n$ vara heltal sådana att $SGD(a, n) = 1$. Då är*

$$n | a^{\varphi(n)} - 1,$$

*dvs* $a^{\varphi(n)} \equiv 1 \pmod{n}$.

Först ett exempel. Om $n = 10$ och $a = 3$ så är $10|3^4 - 1 = 80$ (ty $\varphi(10) = 4$).

**Bevis.** Betrakta restklassringen $\mathbb{Z}_n$. Enligt förutsättningen är $r = [a]_n \neq 0$ en enhet i $\mathbb{Z}_n$ (ty $SGD(a, n) = 1$). Låt $r_1, r_2, ..., r_{\varphi(n)}$ vara alla enheter i $\mathbb{Z}_n$ , och låt oss multiplicera alla dem med $r$. Då får vi $\varphi(n)$ olika produkter som alla är enheter (se (2.14) (a) och (c)):

$$r \odot r_1,\ r \odot r_2, \ldots,\ r \odot r_{\varphi(n)}.$$

Alltså får vi alla enheter i $\mathbb{Z}_n$ igen (möjligen i en annan ordning). I varje fall är

$$r \odot r_1 \odot r \odot r_2 \odot \ldots. \odot r \odot r_{\varphi(n)} = r_1 \odot r_2 \odot \ldots. \odot r_{\varphi(n)}.$$

Nu kan vi stryka $r_1, r_2, ..., r_{\varphi(n)}$ till vänster och till höger (se (2.14)(b)) och vi får

$$r^{\varphi(n)} = 1$$

[¶]Leonard Euler (15/4 1707 - 18/9 1783), schweizisk matematiker, den störste matematikern under 1700-talet och en av de mest betydelsefulla i matematikens historia.

dvs

$$[a^{\varphi(n)}]_n = [1]_n$$

vilket betyder att $n|a^{\varphi(n)} - 1$. □

Vi skall avsluta detta avsnitt med ännu en berömd sats som är ca 2000 år gammal. Satsen heter Kinesiska restsatsen och säger följande:

**(2.18) Kinesiska restsatsen.** *Om $n_1, n_2, ..., n_k$ är parvis relativt prima heltal (dvs den största gemensamma delaren till $n_i$ och $n_j$ är 1 då $i \neq j$) och $r_1, r_2, ..., r_k$ är godtyckliga heltal så existerar ett heltal $x$ sådant att*

$$x \equiv r_1 \pmod{n_1}, \;\; x \equiv r_2 \pmod{n_2}, \;\; ..., \;\; x \equiv r_k \pmod{n_k}.$$

*Dessutom finns det bara ett sådant $x$ modulo $n_1 n_2 \cdots n_k$ (dvs ett $x$ med $0 \leq x < n_1 n_2 \cdots n_k$).*

Betrakta ett exempel. Om vi vill hitta $x$ så att $x$ lämnar resten 2 vid division med 3, resten 3 vid division med 4 och resten 4 vid division med 5 så betyder det att $x$ skall uppfylla

$$x \equiv 2 \pmod{3}, \;\; x \equiv 3 \pmod{4}, \;\; x \equiv 4 \pmod{5}. \tag{2.19}$$

Här är $x = 59$ den enda lösningen modulo $60 = 3 \cdot 4 \cdot 5$. Vårt bevis ger också information om hur man hittar $x$ (se exempel (2.22)).

**Bevis.** Låt $n = n_1 n_2 ... n_k$. Betrakta $\mathbb{Z}_{n_i}$. Enligt förutsättningen har vi $SGD(n_i, \frac{n}{n_i}) = 1$. Därför är $\frac{n}{n_i}$ en enhet modulo $n_i$ dvs det finns $x_i \in \mathbb{Z}_n$ så att

$$[\frac{n}{n_i} x_i]_{n_i} = [1]_{n_i},$$

eller med andra ord,

$$\frac{n}{n_i} x_i \equiv 1 \pmod{n_i}.$$

Nu påstår vi att

$$x = \frac{n}{n_1} x_1 r_1 + \frac{n}{n_2} x_2 r_2 + \ldots + \frac{n}{n_k} x_k r_k \tag{2.20}$$

är den sökta lösningen. För att kontrollera det, observera först att

$$[\frac{n}{n_i}x_i]_{n_j} = 0 \text{ då } i \neq j,$$

ty $n_j|\frac{n}{n_i}$. Därför har vi:

$$[x]_{n_i} = [\frac{n}{n_1}x_1r_1]_{n_i} + [\frac{n}{n_i}x_2r_2]_{n_i} + \ldots + [\frac{n}{n_k}x_kr_k]_{n_i} = [\frac{n}{n_i}x_ir_i]_{n_i} = [r_i]_{n_i}$$

dvs $x \equiv r_i \pmod{n_i}$

Om $x$ och $x'$ är två lösningar dvs $[x]_{n_i} = [x']_{n_i}$ då $i = 1, 2, ..., k$ så är $n_i|x - x'$. Men talen $n_1, n_2, ..., n_k$ är relativt prima så att $n = n_1n_2...n_k|x - x'$[‖] dvs $[x]_n = [x']_n$. □

Hur hittar man $x$ rent praktiskt? Det är klart att man behöver $x_i$ dvs man måste lösa

$$\frac{n}{n_i}x_i \equiv 1 \pmod{n_i}. \tag{2.21}$$

Detta betyder att man vill finna tal $x_i$ sådana att $\frac{n}{n_i}x_i - 1 = n_iq$ dvs

$$\frac{n}{n_i}x_i - n_iq = 1.$$

Här känner vi igen (2.10) med $a = \frac{n}{n_i}$, $b = n_i$, $x = x_i$ och $y = -q$. $x_i$ hittar man mycket enkelt med hjälp av Euklides algoritm.

**(2.22) Exempel.** Vi återkommer till (2.19) där $n_1 = 3, n_2 = 4, n_3 = 5$ och $r_1 = 2, r_2 = 3, r_3 = 4$. Alltså är $n = n_1n_2n_3 = 60$ och man måste lösa kongruenserna (2.21) dvs

$$20x_1 \equiv 1 \pmod{3}, \quad 15x_2 \equiv 1 \pmod{4}, \quad 12x_3 \equiv 1 \pmod{5}.$$

Man hittar mycket lätt (utan Euklides algoritm) att $x_1 = 2, x_2 = 3, x_3 = 3$. Alltså är

$$x = \frac{n}{n_1}x_1r_1 + \frac{n}{n_2}x_2r_2 + \frac{n}{n_3}x_3r_3 = 359$$

så att den enda lösningen modulo 60 är 59, ty $359 \equiv 59 \pmod{60}$. □

[‖] Om $a|c$ och $b|c$ samt $SGD(a, b) = 1$ så $ab|c$.

# ÖVNINGAR

2.1. Bestäm sista siffran av talen

(a) $2^{1991}$, (b) $13^{20}$, (c)$7^{7^{7^7}}$.

2.2. Bestäm resten vid division av

(a) $3^{100}$ med 7, (b) $2^{1000}$ med 3,5,11,13, (c) $99^{99}$ med 13.

**Ledning**. Visa först att $99^2 \equiv -1 \pmod{13}$

2.3. (a) Fermat påstod att talen $F_n = 2^{2^n} + 1$, $n = 0, 1, 2, ...$ är primtal. Det är verkligen sant då $n = 0, 1, 2, 3, 4$. Visa det! (en miniräknare kan vara till hjälp).

(b) Ett hundra år senare visade L.Euler att $641|F_5 = 2^{2^5} + 1 = 2^{32} + 1 = 4294967297$. Visa det genom att räkna i $\mathbb{Z}_{641}$ och utnyttja följande likheter: $641 = 5 \cdot 2^7 + 1 = 5^4 + 2^4$.

2.4. (a) För 2500 år sedan påstod kinesiska matematiker att om ett heltal $n > 1$ är en delare till $2^n - 2$ så måste $n$ vara ett primtal. Detta påstående är sant då $n < 341$ men $341|2^{341} - 2$ trots att 341 inte är ett primtal. Visa det!

**Ledning.** $341 = 11 \cdot 31$ och $2^{10} - 1 = 1023 = 3 \cdot 11 \cdot 31$.

**Anmärkning.** P.Fermat kände till den kinesiska hypotesen och han visste att hans tal $F_n = 2^{2^n} + 1$ hade egenskapen

$$F_n | 2^{F_n} - 2.$$

Det var grunden för hans påstående att $F_n$ var primtal.

(b) Visa att $F_n|2^{F_n} - 2$.

2.5. Visa att omvändningen till Wilsons sats gäller, dvs om $n|(n-1)!+1$ så är $n$ ett primtal.

2.6. (a) Visa att $101|1^3 + 2^3 + ... + 100^3$.

**Ledning.** Räkna i $\mathbb{Z}_{101}$.

(b) Visa att $m|1^k + 2^k + ... + (m-1)^k$ då $k$ och $m$ är positiva udda heltal.

2.7. (a) Beräkna inverser $a^{-1}$ till alla $a \in \mathbb{Z}_7$, $a \neq 0$. Beräkna också $\sum a^{-1}$, $a \in \mathbb{Z}_7$, $a \neq 0$.

(b) Låt $p$ vara ett udda primtal. Visa att om

$$1 + \frac{1}{2} + ... + \frac{1}{p-1} = \frac{a}{b},$$

där $a, b$ är heltal så är $p|a$.

**Ledning**. Utnyttja att $\mathbb{Z}_p$ är en kropp.

2.8. Visa att om $x^2 + y^2 = z^2$, där $x, y, z$ är heltal så finns det bland dessa tal ett som är delbart med 3, ett med 4 och ett med 5.

2.9. Visa att $30|n^5 - n$ då $n$ är ett heltal.

2.10. Låt $p, q$ vara två olika primtal och $n = pq$. Visa att:

(a) $\varphi(n) = (p-1)(q-1)$,

(b) $a^{\varphi(n)+1} \equiv a \pmod{n}$.

**Anmärkning.** Man kan visa helt allmänt att $\varphi(ab) = \varphi(a)\varphi(b)$ då $SGD(a,b) = 1$. Beviset är inte svårt. Påståendet i (b) gäller allmänt då $n$ är en produkt av olika primtal. Man får en generalisering av Fermats lilla sats – om $n = p$ så är $\varphi(n) = p-1$ och $\varphi(n) + 1 = p$.

2.11. **RSA-krypteringssystem** **.

(a) Välj två olika primtal $p, q$ och beräkna $n = pq$ ($p, q$ är vanligen mycket stora, säg, av storleksordningen $10^{100}$).

(b) Beräkna $\varphi(n) = (p-1)(q-1)$ och välj $e$ så att $SGD(e, \varphi(n)) = 1$. Beräkna även $d$, så att $ed \equiv 1 \pmod{\varphi(n)}$.

(c) Publicera $n$, $e$ och en ordbok för översättning av meddelanden till exempel:

$$A = 10,\ B = 11, ...,\ Z = 35$$

(då $n > 35$)

(d) Den som vill sända meddelanden till Dig krypterar med hjälp av den kända funktionen

$$E(r) = r^e,\ r \in \mathbb{Z}_n$$

Du är den ende (förhoppningsvis) som kan dekryptera med hjälp av funktionen

$$D(r) = r^d$$

$d$ är hemligt och

$$D \circ E(r) = D(r^e) = r^{ed} = r$$

Visa den sista likheten!

**Ledning.** $ed = 1 + \varphi(n)m$ för ett heltal $m \geq 1$. Utnyttja 10 (b)!

**Anmärkning.** RSA-systemet tillhör sk öppennyckelkrypton dvs kryperingsfunktionen $E$ är allmänt känd. Vad gör den som vill dekryptera? Funktionen $D$ är inversen till $E$ och för att hitta den behöver man $d$. $d$ är lösningen till $ed = 1$ i $\mathbb{Z}_{\varphi(n)}$ och för att hitta $d$ behöver man $p$ och $q$ som inte är kända. Men $n$ är känt så att man måste kunna faktorisera $n$. Här ligger styrkan hos RSA-systemet. Faktoriseringsalgoritmer tar mycket lång tid. De bästa kända algoritmerna för primfaktoruppdelning av $n$ kräver ca $n^{\frac{1}{5}}$ räkneoperationer. Om $p$ och $q$ är ca $10^{100}$ så är $n = 10^{200}$. Om en räkneoperation tar ca 1 $\mu s$ så krävs det $10^{40}\mu s = 3 \cdot 10^{26}$ år för att genomföra beräkningarna för $n$ ($10^6$ datorer var och en kapabel att utföra en räkneoperation på 1 $\mu s$ skulle behöva $3 \cdot 10^{20}$ år för att klara dessa beräkningar!).

(e) Låt $n = 17 \cdot 23 = 391$. Välj krypteringsnyckeln $e = 3$ och kryptera NEJ (med "ordbokensom i (c)). Beräkna $d$ och dekryptera 121 268 358.

**Konstruktionen av systemet publicerades av R.L.Rivest, A.Shamir och L.Adleman 1978.

2.12. Bestäm det minsta positiva heltalet $n$ som lämnar resterna 1,2,3,4,5 vid division med respektive 2,3,4,5,6.

2.13. Bestäm alla $n$ sådana att $4|n$, $9|n+1$, $25|n+2$.

2.14. Låt $x_0$ vara den minsta positiva lösningen till

$$x \equiv r_1 \pmod{n_1},\ x \equiv r_2 \pmod{n_2},\ ...,\ x \equiv r_k \pmod{n_k},$$

där $n_i$ är positiva relativt prima heltal. Visa att varje annan lösning är $x_0 + nq$ där $n = n_1 n_2 \dots n_k$ och $q \in \mathbb{Z}$.

# AVSNITT 3

# POLYNOMRINGAR

Varje ring $R$ ger upphov till polynom med koefficienter i $R$ dvs alla uttryck

$$f(X) = a_0 + a_1X + a_2X^2 + \ldots + a_nX^n,$$

där $a_i \in R$. $a_i$ kallas **koefficienter** till $f(X)$ och $n$ kallas dess **grad** om $a_n \neq 0$. $a_n$ kallas ofta **högsta koefficienten** av $f(X)$. Polynom kan adderas och multipliceras på välkänt sätt: Om

$$f(X) = a_0 + a_1X + a_2X^2 + \ldots \quad \text{och} \quad g(X) = b_0 + b_1X + b_2X^2 + \ldots^*$$

så är

$$f(X) + g(X) = (a_0 + b_0) + (a_1 + b_1)X + (a_2 + b_2)X^2 + ...$$

och

$$f(X)g(X) = a_0b_0 + (a_0b_1 + a_1b_0)X + (a_0b_2 + a_1b_1 + a_2b_0)X^2 + ...$$

dvs koefficienten för $X^k$ i summan $f(X) + g(X)$ är $a_k + b_k$ och för produkten $f(X)g(X)$ är $a_0b_k + a_1b_{k-1} + ... + a_kb_0$.

Med dessa operationer bildar alla polynom med koefficienter i $R$ en ny ring som betecknas med $R[X]$. Man kan således betrakta ringar $\mathbb{Z}[X]$, $\mathbb{Q}[X]$, $\mathbb{R}[X]$, $\mathbb{C}[X]$ med koefficienter i

*Man kan alltid förutsätta att $f$ och $g$ har lika många termer genom att "förlänga"ett av polynomen med ett antal termer med koefficienter 0.

olika talringar och talkroppar, men även $\mathbb{Z}_2[X]$, $\mathbb{Z}_3[X]$ och allmänt $\mathbb{Z}_n[X]$ dvs polynom med koefficienter i ringar av rester. Alla dessa ringar spelar en mycket viktig roll i hela matematiken och har stor betydelse för olika typer av tillämpningar (inte minst gäller det ringarna $\mathbb{Z}_n[X]$).

Vi skall i någon mån formalisera definitionen av begreppen polynom och polynomring i en anmärkning som avslutar detta avsnitt. Där förklarar vi också hur man kan tolka beteckningen X. Vårt sätt att skriva polynom efter växande potenser av $X$ är inte alls nödvändigt, men det underlättar definitionen av addition och multiplikation av polynom. När det gäller formella ting finns det dock några saker som vi vill säga redan nu.

Ett polynom med alla koefficienter lika med 0 kallas **nollpolynomet** och vi definierar dess grad som $-1$. Alla polynom av graden 0 samt nollpolynomet kallas ofta för **konstanta polynom** (dvs $f(X) = a_0$, $a_0 \in K$). Vi skriver $f(X)$, $g(X)$, men man kan skriva kortare $f, g$. I synnerhet betyder $f \neq 0$ att $f$ inte är nollpolynomet. Om $f \neq 0$ och $g \neq 0$ så är

$$\operatorname{grad}(fg) = \operatorname{grad} f + \operatorname{grad} g.$$

Man kan beräkna $f(X)$ för $X = a \in R$. Då får vi polynomets $f$ värde i punkten $a$ dvs $f(a) = a_0 + a_1 a + ... + a_n a^n$.

Vårt största intresse kommer att koncentreras kring polynomringarna $K[X]$, där $K$ är en kropp. Det finns en intressant aspekt av sådana ringar som har långtgående konsekvenser för hela matematiken: Det finns många likheter mellan heltalsringen $\mathbb{Z}$ och polynomringarna $K[X]$. Den första är divisionsalgoritmen:

**(3.1) Divisionsalgoritmen.** *Om $f, g \in K[X]$ och $g \neq 0$ så finns det polynom $q$, $r \in K[X]$ sådana att*

$$f = gq + r, \quad \textit{där } \operatorname{grad} r < \operatorname{grad} g \textit{ eller } r = 0.$$

*Polynomen $q$ och $r$, som kallas kvoten och resten vid division av $f$ med $g$, är entydigt definierade av $f$ och $g$.*

**Bevis.** Vi bevisar satsen med hjälp av induktion efter graden av $f(X)$. Om graden av $f(X)$ är $-1$ (dvs $f(X)$ är nollpolynomet) så är $f(X) = g(X) \cdot 0$ dvs $q(X) = 0$ och $r(X) = 0$. Nu antar vi att satsen gäller för alla polynom $f(X)$ vars grad är $< n$, där $n \geq 0$. Låt $f(X) = a_n X^n + \ldots + a_0$, $g(X) = b_m X^m + \ldots + b_0$ där $a_n \neq 0$, och $b_m \neq 0$. Om $n < m$ så har vi $f(X) = g(X) \cdot 0 + f(X)$ dvs $q(X) = 0$ och $r(X) = f(X)$. Antag att $n \geq m$. Låt

$$f_1(X) = f(X) - \frac{a_n}{b_m} g(X) X^{n-m}.$$

Då är grad $f_1(X) <$ grad $f(X)$ så att

$$f_1(X) = g(X)q_1(X) + r(X), \text{ grad } r(X) < \text{grad } g(X)$$

enligt induktionsantagandet. Men då är

$$f(X) = f_1(X) + \frac{a_n}{b_m}g(X)X^{n-m} = g(X)(q_1(X) + \frac{a_n}{b_m}X^{n-m}) + r(X)$$

dvs

$$f(X) = g(X)q(X) + r(X), \quad \text{grad}\, r(X) < \text{grad}\, g(X)$$

där $q(X) = q_1(X) + \frac{a_n}{b_m}X^{n-m}$.

Entydigheten av $q$ och $r$ bevisas på följande sätt. Antag att även $f = gq_1 + r_1$, där $q_1, r_1 \in K[X]$ och grad $r_1 <$ grad $g$ eller $r_1 = 0$. Då har vi

$$(*) \qquad gq + r = gq_1 + r_1,$$

dvs

$$r - r_1 = g(q_1 - q),$$

vilket betyder att $g$ dividerar $r - r_1$. Men om $r - r_1 \neq 0$ så är grad $(r - r_1) <$ grad $g$, vilket medför att $g$ inte kan vara delare till $r - r_1$. Alltså är $r - r_1 = 0$ dvs $r = r_1$. Likheten $(*)$ ovan ger $gq = gq_1$ så att $q = q_1$ ty $g \neq 0$. □

**Exempel.** Låt $f(X) = 2X^3 + 3X^2 + X + 1$, $g(X) = X^2 + 2$ i $\mathbb{Z}_5[X]$. Vi vill beräkna kvoten och resten vid division av $f(X)$ med $g(X)$. Först divideras den högsta termen $2X^3$ i $f(X)$ med den högsta termen $X^2$ i $g(X)$. Då får man kvoten $2X$ och räknar ut "den första resten" $f_1(X) = f(X) - 2Xg(X)$. Därefter upprepar man proceduren med $f_1(X)$ i stället för $f(X)$.

$$
\begin{array}{rrcl}
 & 2X+3 & = & q(X) \\
\underline{X^2+2}\,| & 2X^3+3X^2+X+1 & & \\
 & -2X^3 \quad -4X & & \\
\hline
 & 3X^2-3X+1 & & \\
 & -3X^2 \quad -6 & & \\
\hline
 & -3X & = & r(X)
\end{array}
$$

(tänk på det att $5 = 0$ i $\mathbb{Z}_5$). □

Delbarhetsbegreppet för polynom liknar samma begrepp för heltalen.

**(3.2) Definition.** Man säger att $g \in K[X]$ är en **delare** till $f \in K[X]$ om $f = gq$, där $q \in K[X]$. Man skriver då $g|f$. □

När man har divisionsalgoritmen kan man utföra Euklides algoritm för att räkna ut största gemensamma delaren ($SGD$) till två polynom $f$ och $g$ (precis som man räknar ut $SGD$ till två heltal i avsnittet om delbarhet). Största gemensamma delaren till två polynom definieras på följande sätt:

**(3.3) Definition.** Om $f, g \in K[X]$ så säger man att $d \in K[X]$ är en **största gemensamma delare** till $f$ och $g$ ($SGD(f, g)$) om

(a) $d|f$ och $d|g$,

(b) $d'|f$ och $d'|g$, där $d' \in K[X]$ implicerar att $d'|d$.

Om $f = g = 0$ definierar man $SGD(0, 0) = 0$. □

**(3.4) Anmärkning.** $SDG(f, g)$ är inte entydig. Om $f \neq 0$ eller $g \neq 0$ och $d_1, d_2$ är två polynom som uppfyller villkoren i (3.3) så är $d_1|d_2$ och $d_2|d_1$. Alltså är $d_2 = d_1 q$, där $q$ har grad 0 ty $d_1$ och $d_2$ har samma grad ($\operatorname{grad} d_1 \geq \operatorname{grad} d_2$ och $\operatorname{grad} d_2 \geq \operatorname{grad} d_1$). Detta betyder att $d_1$ och $d_2$ är lika så när som på en konstant. Genom ett lämpligt val av den konstanten kan vi välja en största gemensamma delare med högsta koefficienten 1. Man kallar en sådan **den största gemensamma delaren**. Två polynom vars största gemensamma delare är 1 kallas **relativt prima**. □

Precis som för heltalen gäller följande sats:

**(3.5) Sats.** *Om $d = SGD(f, g)$, där $f, g \in K[X]$ så existerar $s, t \in K[X]$ så att*

$$d = fs + gt.$$

Man kan visa satsen på liknande sätt som motsvarande sats för heltalen (se avsnittet "Delbarhet och primtal").

Med hjälp av (3.5) visar man som för heltalen följande egenskap som vi snart utnyttjar:

**(3.6) Sats.** *Om $f|h, g|h$ och $SGD(f, g) = 1$, där $f, g, h \in K[X]$ så $fg|h$.*

**Bevis.** Låt $h = fq_f$, $h = gq_g$ och $1 = fs + gt$. Då är $h = hfs + hgt = fgq_g s + fgq_f t = fg(q_g s + q_f t)$ dvs $fg|h$. □

En mycket vanlig uppgift i samband med polynom är att lösa polynomekvationer $f(X) = 0$.

**(3.7) Definition.** Man säger att $a \in K$ är ett **nollställe** till $f \in K[X]$ eller en **rot** till ekvationen $f(X) = 0$ om $f(a) = 0$. □

Ett samband mellan nollställen och delbarhet förklarar vår nästa sats som är mycket enkel att bevisa och samtidigt mycket användbar:

**(3.8) Faktorsatsen.** (a) *Resten vid division av $f \in K[X]$ med $X - a$, $a \in K$, är lika med $f(a)$;*

(b) *$a \in K$ är ett nollställe till $f \in K[X]$ då och endast då $X - a|f(X)$.*

**Bevis.** (a) Enligt divisionsalgoritmen är

$$f(X) = (X - a)q(X) + r,$$

där $\operatorname{grad} r < 1$ eller $r = 0$ dvs $r$ är en konstant. Alltså är $f(a) = r$.

(b) $f(a) = 0 \Leftrightarrow r = f(a) = 0$. □

**(3.9) Definition.** Man säger att $a \in K$ har **multipliciteten** $m$ som ett nollställe till $f \in K[X]$ om $(X - a)^m|f(X)$ och $(x - a)^{m+1} \nmid f(X)$. □

**(3.10) Sats.** *Summan av multipliciteterna av alla nollställen till $f \in K[X]$ är högst lika med grad $f$.*

**Bevis.** Låt $a_1, \ldots, a_r$ vara olika nollställen till $f$ och låt $m_1, \ldots, m_r$ vara deras respektive multipliciteter. Detta betyder att

$$(X - a_1)^{m_1} | f(X), \ldots, (X - a_r)^{m_r} | f(X).$$

Men polynomen $(X - a_i)^{m_i}$ är parvis relativt prima så att

$$(X - a_1)^{m_1} \ldots (X - a_r)^{m_r} | f(X)$$

dvs grad $f \geq m_1 + \ldots + m_r$. □

Ett mycket viktigt begrepp som vi vill diskutera nu är irreducibla polynom som är motsvarigheten till primtalen i heltalsringen:

**(3.11) Definition.** Man säger att ett polynom $f \in R[X]$ är **reducibelt** i $R[X]$ om $f = gh$, där $g, h \in R[X]$ och $1 \leq \operatorname{grad} g < \operatorname{grad} f$, $1 \leq \operatorname{grad} h < \operatorname{grad} f$. Ett polynom av grad minst 1 som inte är reducibelt kallas **irreducibelt**. □

Man säger att en delare $g$ till $f$ sådan att $1 \leq \operatorname{grad} g < \operatorname{grad} f$ är äkta (eller "icketrivial"). Därför är $f$ reducibelt om det har en äkta delare, och irreducibelt om dess grad är minst 1 och det saknar äkta delare.

**(3.12) Exempel.** (a) Varje polynom av grad 1 är irreducibelt.

(b) Ett polynom $f \in K[X]$ av grad 2 eller 3 är reducibelt i $K[X]$ då och endast då $f$ har ett nollställe i $K$ dvs det finns $x_0 \in K$ så att $f(x_0) = 0$. I själva verket, om $f(x_0) = 0$ så är $f(X) = (X - x_0) f_1(X)$ där $f_1(X) \in K[X]$ och grad $f_1(X) \geq 1$ dvs $f(X)$ är reducibelt. Omvänt om $f(X) = g(X)h(X)$ är en faktoruppdelning av $f(X)$ i två icke-konstanta faktorer så måste någon av dessa ha grad 1. Låt $g(X) = b_0 + b_1 X \in K[X]$. Då är $x_0 = -b_0/b_1$ ett nollställe till $f(X)$. Till exempel är $f(X) = X^2 + 1 \in \mathbb{Q}[X]$ irreducibelt i $\mathbb{Q}[X]$ ty det saknar nollställen i $\mathbb{Q}[X]$ ($\pm i \notin \mathbb{Q}$). Det är irreducibelt även i $\mathbb{R}[X]$, men i $\mathbb{C}[X]$ är $X^2 + 1 = (X + i)(X - i)$ så att $X^2 + 1$ är reducibelt i den sista polynomringen.

(c) $f(X) = X^2 + X + 1$ är irreducibelt i $\mathbb{Z}_2[X]$ ty $f(0) = 0^2 + 0 + 1 + 1$ och $f(1) = 1^2 + 1 + 1 = 1$ så att polynomet saknar nollställen i $\mathbb{Z}_2$. Vi har $X^2 + 1 = (X + 1)^2$ i $\mathbb{Z}_2[X]$ så att $X^2 + 1$ är reducibelt i $\mathbb{Z}_2[X]$.

(d) Polynomet $f(X) = X^4 + 4$ saknar rationella (även reella) nollställen. Men man får inte påstå att $f$ är irreducibelt i $\mathbb{Q}[X]$. Detta är ett polynom av grad 4 så att (b) inte är användbar här! I själva verket har vi

$$X^4 + 4 = X^4 + 4X^2 + 4 - 4X^2 = (X^2 + 2)^2 - (2X)^2 = (X^2 + 2X + 2)(X^2 - 2X + 2)$$

så att $X^4 + 4$ är reducibelt i $\mathbb{Q}[X]$. □

Nu skall vi gå igenom några exempel på irreducibla polynom i olika ringar.

**(3.13) Polynomringen** $\mathbb{C}[X]$. I samband med vår diskussion av komplexa tal nämnde vi (polynom)algebrans fundamentalsats (se(4.4)) som säger att varje ickekonstant polynom med komplexa koefficienter har ett komplext nollställe. Detta betyder att om $f \in \mathbb{C}[X]$ och grad $f \geq 1$ så är $f(z_1) = 0$ för ett komplext tal $z_1$. Enligt faktorsatsen har vi

$$f(X) = (X - z_1)f_1(X).$$

Här är $\operatorname{grad} f_1 = \operatorname{grad} f - 1$. Om $\operatorname{grad} f_1 \geq 1$ så har $f_1$ ett komplext nollställe $z_2$. Nu ger faktorsatsen att $f_1(X) = (X - z_2)f_2(X)$ så att

$$f(X) = (X - z_1)(X - z_2)f_2(X).$$

Vi kan fortsätta faktoruppdelningen av $f(X)$ tills vi får

$$f(X) = (X - z_1)(X - z_2)\cdots(X - z_n)f_n(X),$$

där $f_n(X)$ är ett konstant polynom. Detta resonemang leder till följande resultat:

**(3.14) Sats.** *Varje icke-konstant polynom $f \in \mathbb{C}[X]$ är en produkt av förstagradspolynom. Alltså är varje polynom $f \in \mathbb{C}[X]$ av grad $\geq 2$ reducibelt och alla irreducibla polynom i $\mathbb{C}[X]$ är förstagradspolynomen.*

**(3.15) Polynomringen** $\mathbb{R}[X]$. Situationen med irreducibla polynom i den ringen är lite mera komplicerat än i $\mathbb{C}[X]$. Men man kan fortfarande ganska lätt beskriva alla irreducibla polynom. Först noterar vi följande hjälpresultat:

**(3.16) Lemma.** *$f \in \mathbb{R}[X]$ och $z$ är ett komplext tal så är*

$$\overline{f(z)} = f(\bar{z}).$$

*I synnerhet, om $z$ är ett nollställe till $f(X)$ dvs $f(z) = 0$, så är också $\bar{z}$ ett nollställe till $f(X)$ dvs $f(\bar{z}) = 0$.*

**Bevis.** Låt

$$f(X) = a_n X^n + a_{n-1} X^{n-1} + ... + a_1 X + a_0,$$

där $a_i \in \mathbb{R}$. Då är [†]

$$\overline{f(z)} = \overline{a_n z^n + a_{n-1} z^{n-1} + \dots a_1 z + a_0} = \bar{a}_n \bar{z}^n + \bar{a}_{n-1} \bar{z}^{n-1} + ... + \bar{a}_1 \bar{z} + \bar{a}_0 = f(z),$$

ty $\bar{a}_i = a_i$ därför att $a_i$ är reella. □

Nu är det mycket lätt att bevisa följande sats om faktorgruppdelningar i $\mathbb{R}[X]$ :

**(3.17) Sats.** *Varje icke-konstant reellt polynom är en produkt av irreducibla faktorer av grader 1 eller 2. Alltså är varje reellt polynom av grad $\geq 3$ reducibelt och alla irreducibla polynom i $\mathbb{R}[X]$ är förstagradspolynom och andragradpolynom utan reella nollställen.*

**Bevis.** Låt $f(X) \in \mathbb{R}[X]$. Vi visar satsen genom induktion efter graden av $f(X)$. Om $f(X)$ har graden 1 så är påståendet klart ($f(X)$ är irreducibelt). Antag att påståendet gäller för alla polynom av graden $< n$. Vi vill visa att att det också gäller för alla polynom av graden $n$. Låt $f(X)$ vara ett sådant polynom. Om $f(X)$ har ett reellt nollställe $a$ så är:

$$f(X) = (X - a) f_1(X)$$

enligt faktorsatsen, och grad $f_1 = n - 1$. Enligt induktionsantagandet är $f_1$ en produkt av första och andragradspolynom utan reella nollställen så att detsamma gäller för $f(X)$.

Om $f(X)$ saknar reella nollställen så är $f(z) = 0$ för ett icke-reellt tal $z$. Enligt faktorsatsen är

$$f(X) = (X - z) f_1(X).$$

[†]Om $z = a + bi$ så $\bar{z} = a - bi$. Vi har

$$\overline{z_1 + z_2} = \bar{z}_1 + \bar{z}_2 \quad \text{och} \quad \overline{z_1 z_2} = \bar{z}_1 \bar{z}_2$$

samt $\bar{z} = z$ då och endast då $z$ är reellt.

Enligt lemma (3.16) är $f(\bar{z}) = 0$ så att $(\bar{z} - z)f_1(z) = 0$, vilket ger $f_1(\bar{z}) = 0$ ty $\bar{z} - z \neq 0$ (om $\bar{z} - z = 0$ så är $z$ reellt!). Genom att tillämpa faktorsatsen på $f_1(X)$ får vi nu $f_1(X) = (X - \bar{z})f_2(X)$ så att

$$f(X) = (X - z)(X - \bar{z})f_2(X).$$

Vi har

$$(X - z)(X - \bar{z}) = X^2 - pX + q,$$

där $p = z + \bar{z}$ och $q = z\bar{z}$ är reella tal. Därför är också $f_2(X)$ ett reellt polynom (kvoten av $f(X)$ genom $X^2 - pX + q$) och $\operatorname{grad} f_2 = n - 2$. Polynomet $X^2 - pX + q$ är ett andragradspolynom utan reella nollställen. Om $f_2(X)$ är ett konstant polynom så är påståendet klart. Annars säger induktionsantagandet att $f_2(X)$ är en produkt av förstagradspolynom eller andragradspolynom utan reella nollställen så att samma påstående gäller för $f(X)$.

Sista meningen i satsen är en direkt konsekvens av dess första del. □

**(3.18) Polynomringen** $\mathbb{Q}[X]$. Situationen här är mycket mera sammansatt än i $\mathbb{C}[X]$ och $\mathbb{R}[X]$. Det finns inte någon känd beskrivning av alla irreducibla polynom, men man vet att för varje $n \geq 1$ finns oändligt många irreducibla polynom av graden $n$. T ex är alla polynom $X^n - p$, där $n \geq 1$ och $p$ är ett godtyckligt primtal, irreducibla. Detta påstående följer direkt av följande mycket kända resultat:

**(3.19) Eisensteins** [‡] **kriterium.** *Om* $f(X) = a_nX^n + a_nX^{n-1} + ... + a_1X + a_0$, *där* $a_i \in \mathbb{Z}$, *och det finns ett primtal* $p$ *sådant att*

$$p \nmid a_n, p|a_{n-1}, ..., p|a_1, p|a_0 \text{ och } p^2 \nmid a_0,$$

*så är* $f(X)$ *irreducibelt i* $\mathbb{Q}[X]$.

Se övning 5 för ett bevis av Eisensteins kriterium.

För polynom med rationella koefficienter har man en mycket enkel och mycket användbar sats som gör det möjligt att i vissa fall hitta rationella nollställen. Rent allmänt är det ganska svårt att hitta nollställen till en given polynomekvation.

[‡]Ferdinand Eisenstein (16/4 1823 – 11/11 1852) en mycket framstående tysk matematiker.

**(3.20) Sats.** *Om ett rationellt tal $\alpha = \frac{k}{l}$, där $k, l \in \mathbb{Z}$, SGD$(k, l) = 1$, är ett nollställe till polynomet $f(X) = a_n X^n + a_{n-1}X^{n-1} + \cdots + a_1 X + a_0$ med heltaliga koefficienter $a_i$, så är $k$ en delare till den lägsta koefficienten $a_0$, och $l$ är en delare till den högsta koefficienten $a_n$.*

**Bevis.** Enligt förutsättningen har vi:

$$a_n \left(\frac{k}{l}\right)^n + a_{n-1}\left(\frac{k}{l}\right)^{n-1} + \cdots a_1\left(\frac{k}{l}\right) + a_0 = 0,$$

vilket efter multiplikation av bägge leden med $l^n$ ger

$$a_n k^n + a_{n-1}k^{n-1}l + \cdots + a_1 k l^{n-1} + a_0 l^n = 0.$$

Ni noterar vi att $a_0 l^n$ är en multipel av $k$ (flytta alla andra termer till höger!). Alltså är $a_0$ en multipel av $k$ därför att $k$ och $l$ saknar gemensamma faktorer $\neq \pm 1$. På liknande sätt noterar vi att $a_n k^n$ är en multipel av $l$ (flytta alla andra termer till höger som tidigare!). Alltså är $l$ en delare till $a_n$. □

Vi exemplifierar den sista satsen:

**Exempel.** Lös ekvationen $f(X) = X^3 + 2X^2 - 2X - 4 = 0$. Vi försöker hitta rationella nollställen $\alpha = \frac{k}{l}$ med SGD$(k, l) = 1$. Enligt satsen ovan är $k$ en delare till 4, och $l$ är en delare till 1. Alltså $k = \pm 1, \pm 2, \pm 4$ och $l = \pm 1$. Vi kontrollerar $f(\pm 1) \neq 0$, $f(2) = 8, f(-2) = 0$. Alltså har vi hittat ett nollställe $x_1 = -2$. Divisionsalgoritmen ger $f(X) = (X + 2)(X^2 - 2)$ (polynomet $f(X)$ är delbart med $X + 2$ enligt faktorsatsen). Nu kan vi beräkna två andra nollställen $x_2 = \sqrt{2}$ och $x_3 = -\sqrt{2}$ . □

Slutligen ägnar vi några ord åt polynomringarna $\mathbb{Z}_p[X]$. Dessa ringar har en stor praktisk betydelse (särskilt för $p = 2$) i kodningsteori och kryptologi (t ex felkorrigering i datorminnen och säkerhetssystem för dataöverföring). Vi har inte någon möjlighet att fördjupa oss i den problematiken. Men det finns kurser i tillämpad algebra, där man kan bekanta sig med dessa aspekter samt kurser i algebra och talteori, där man studerar rent matematiska tillämpningar på dessa intressanta ringar.

**(3.21) Polynomringarna** $\mathbb{Z}_p[X]$. I dessa ringar finns det också irreducibla polynom av godtyckliga grader (precis som i $\mathbb{Q}[X]$), men det finns exakta formler för deras antal och mycket effektiva algoritmer för att kunna testa irreducibiliteten (beroende på mycket viktiga tekniska tillämpningar finns det färdiga programpaket för dessa ändamål).

Låt oss ägna en stund åt $\mathbb{Z}_2[X]$ som är den enklaste, och för tillämpningarna, den viktigaste, bland ringarna $\mathbb{Z}_p[X]$. Man kan skriva ut alla polynom av given grad $n$ :

**grad 0:** 1

**grad 1:** $X, X+1$

**grad 2:** $X^2$, $X^2+1$, $X^2+X$, $X^2+X+1$

**grad 3:** $X^3$, $X^3+1$, $X^3+X$, $X^3+X+1$, $X^3+X^2$, $X^3+X^2+1$, $X^3+X^2+X$, $X^3+X^2+X+1$

osv. Bland dessa polynom kan man hitta alla irreducibla:

**grad 1:** $X$, $X+1$

**grad 2:** $X^2+X+1$

**grad 3:** $X^3+X+1$, $X^3+X^2+1$.

För att kontrollera att t ex $X^2+X+1$ är irreducibelt finner vi lätt att alla reducibla polynom av grad 2 är $X^2, X(X+1) = X^2+X$ och $(X+1)^2 = X^2+1$ (observera att $2X = 0$ ty $2 \equiv 0 (\text{mod } 2)$!). $X^2+X+1$ finns inte bland dem, vilket betyder att det inte kan faktoriseras i produkt av två polynom av grad 1. På liknande sätt kan man skriva ut alla reducibla polynom av grad 3 och konstatera att $X^3+X+1$ och $X^3+X^2+1$ inte finns bland dem. För andra faktoriseringsmetoder och en tillämpning se övningarna.

Som vi nämnde tidigare påminner ringarna $K[X]$ mycket om heltalen (analogin är starkast då $K = \mathbb{Z}_p$). Irreducibla polynom påminner om primtalen. Vi vet att varje naturligt tal $\neq 1$ har en faktoruppdelning i produkt av primtal t ex

$$10 = 2 \cdot 5 = 5 \cdot 2.$$

Om man betraktar heltalen $\mathbb{Z}$ så har man också

$$10 = (-2) \cdot (-5) = (-5) \cdot (-2),$$

dvs 10 får två "nyafaktoruppdelningar. Talen $\pm p$, där $p$ är ett primtal, har exakt samma egenskap som irreducibla polynom — de saknar äkta delare. Man kan kalla sådana tal för irreducibla (heltal). Om man tillåter $\pm 1$ som faktorer, får man oändligt många faktoruppdelningar som t ex

$$10 = 2 \cdot 5 \cdot 1 = 2 \cdot 5 \cdot 1 \cdot 1 = 2 \cdot 5 \cdot (-1) \cdot (-1)$$

osv. Det är orsaken till att man inte betraktar $\pm 1$ som irreducibla tal (dvs primtal) trots att de saknar äkta delare. För polynom har man en liknande situation. t ex är i $\mathbb{Q}[X]$ :

$$X^2 - 1 = (X-1)(X+1) = 2(X-1)\frac{1}{2}(X+1) = 3(X-1)\frac{1}{3}(X+1)$$

osv. Konstanterna $\neq 0$ kan alltid skrivas in i en faktoruppdelning precis som faktorerna $\pm 1$ i heltalsfallet. I polynomringarna $K[X]$ betraktas därför inte konstanterna $\neq 0$ som irreducibla polynom (eller reducibla polynom). Konstanterna $\neq 0$ är polynom som har invers ($a \cdot \frac{1}{a} = 1$ då $a \neq 0$) (sådana element i ringen kallas för enheter t ex är $\pm 1$ de enda enheterna i $\mathbb{Z}$, konstanterna $\neq 0$ är de enda enheterna i $K[X]$). Följande sats visar på en långtgående likhet mellan primtal och irreducibla polynom:

**(3.22) Sats.** *Varje icke-konstant polynom i $K[X]$ är en produkt av irreducibla polynom. Om $f \in K[X]$ och*

$$f = p_1 p_2 ... p_r = p_1' p_2' ... p_s' \ ,$$

*där $p_1, p_2, ..., p_r$ och $p_1', p_2', ..., p_s'$ är irreducibla så är $r = s$ och man kan numrera faktorerna på ett sådant sätt att $p_i = a_i p_i'$ för en lämplig konstant $a_i \in K$.*

Satsen säger att två olika faktoruppdelningar av samma (icke-konstanta) polynom har lika många irreducibla faktorer, och dessutom kan dessa faktorer paras ihop så att faktorerna i samma par skiljer sig så när som på en konstant (man säger också att sådana faktorer är associerade).

**Bevis.** Existensen av faktoruppdelningen visas med hjälp av induktion efter graden av $f$. Om $\operatorname{grad} f = 1$ så är saken klar. Antag att påståendet gäller för alla polynom av grad $< n$ och låt $\operatorname{grad} f = n > 1$. Om $f$ är irreducibelt så är påståendet klart (det finns bara en irreducibel faktor $p_1 = f$). Om $f$ är reducibelt dvs $f = gh$, där $1 \leq \operatorname{grad} g < \operatorname{grad} f$ och $1 \leq \operatorname{grad} h < \operatorname{grad} f$, så säger induktionsantagandet att både $g$ och $h$ har faktoruppdelningar i produkt av irreducibla polynom så att detsamma gäller för $f$. Vi utelämnar beviset för andra delen av satsen dvs entydigheten. Den visas på precis samma sätt som entydigheten av primfaktoruppdelningar av heltalen med hjälp av en viktig egenskap hos irreducibla polynom som följer nedan. □

**(3.23) Sats.** *Om $p \in K[X]$ är irreducibelt och $p|fg$, där $f, g \in K[X]$ så $p|f$ eller $p|g$.*

**Bevis.** Satsen visas på samma sätt som motsvarande sats för primtal i avsnittet "Delbarhet och primtal". □

Vi skall avsluta detta avsnitt med några ord om definitionen av polynomringarna $R[X]$. Du behöver inte betrakta dessa ord särskilt allvarligt. De är tänkta som en förklaring för den som

känner att det vore bra med en mera stringent definition av begreppet polynom. Men man kan klara sig ganska länge utan den stringensen.

Ett polynom kan nämligen uppfattas som en oändlig följd $(a_0, a_1, a_2, ...)$ där $a_i \in R$. För den följden skall det finnas ett $n$ sådant att $a_i = 0$ då $i > n$. Polynom adderas och multipliceras enligt följande definition:

$$(a_0, a_1, a_2, ...) + (b_0, b_1, b_2, ...) = (a_0 + b_0, a_1 + b_1, a_2 + b_2, ...)$$

och

$$(a_0, a_1, a_2, ...)(a_0, a_1, a_2, ...) = (a_0b_0, a_0b_1 + a_1b_0, a_0b_2 + a_1b_1 + a_2b_0, ...)$$

Vad är $X$ och hur kan man skriva om polynom till den välbekanta formen

$$a_0 + a_1X + a_2X^2 + ... + a_nX^n?$$

Man definierar:

$$X = (0, 1, 0, 0, ...).$$

Då är

$$X^2 = (0, 0, 1, 0, ...),$$
$$X^3 = (0, 0, 0, 1, ...),$$

och allmänt

$$X^n = (0, 0, ..., 0, 1, 0, ...),$$

där $a_n = 1$ och $a_i = 0$ då $i \neq n$. Vidare observerar man att polynomen $(a_0, 0, 0, ...)$ adderas och multipliceras som elementen i $R$ :

$$(a_0, 0, ...) + (b_0, 0, ...) = (a_0 + b_0, 0, ...),$$

och

$$(a_0, 0, ...)(b_0, 0, ...) = (a_0b_0, 0, ...).$$

Därför kommer man överens om att betäckna $(a_0, 0, ...)$ med $a_0$. Man observerar också att

$$(0, 0, ..., 0, a_n, 0, ...) = (a_n, 0, 0, ...)(0, 0, ..., 0, 1, 0, ...) = a_nX^n$$

Om nu $(a_0, a_1, ..., a_n, ...)$ är ett polynom där $a_i = 0$ då $i > n$ så är

$$(a_0, a_1, ..., a_n, ...) = (a_0, 0, 0, ...) + (0, a_1, 0, ...) + (0, 0, 0, ..., a_n, ...) =$$
$$= a_0 + a_1X + ... + a_nX^n$$

och vi får våra välbekanta polynom.

# ÖVNINGAR

3.1. Visa att följande polynom är irreducibla:
(a) $X^2 + 2X + 2$ i $\mathbb{R}[X]$,
(b) $X^3 - 2$ i $\mathbb{Q}[X]$,
(c) $X^2 + 1$ i $\mathbb{Z}_3[X]$,
(d) $X^4 + X + 1$ i $\mathbb{Z}_2[X]$.

3.2. Faktoruppdela polynomet $X^4 + 2X^2 + 9$ i irreducibla faktorer i $\mathbb{Q}[X]$, $\mathbb{R}[X]$ och $\mathbb{C}[X]$.

3.3. Faktoruppdela $X^4 + 1$ i irreducibla faktorer i $\mathbb{Q}[X]$, $\mathbb{R}[X]$ och $\mathbb{C}[X]$.

3.4. Låt $K$ vara en kropp och låt $f \in K[X]$. Visa att
(a) Om grad $f \geq 2$ och $f$ har ett nollställe i $K$ så är $f$ reducibelt i $K[X]$.
(b) Om grad $f = 2$ eller 3 så är $f$ reducibelt i $K[X]$ då och endast då $f$ har nollställen i $K$.
(c) Konstruera ett exempel som visar att (b) inte gäller då grad $f = 4$.
(d) Lös uppgifterna 1 (a)–(c) med hjälp av (b).

3.5. Bevisa Eisensteins kriterium: Om ett polynom

$$f(x) = a_nX^n + a_{n-1}X^{n-1} + ... + a_1X + a_0$$

har hela koefficienter $a_i$ och

$$p \nmid a_n, p|a_{n-1}, ..., p|a_1, p|a_0 \text{ samt } p^2 \nmid a_0$$

för ett primtal $p$ så är $f(X)$ irreducibelt i $\mathbb{Z}[X]$.
**Ledning.** Låt

$$f(X) = (b_kX^k + b_{k-1}X^{k-1} + ... + b_1X + b_0)(c_lX^l + c_{l-1}X^{l-1} + ... + c_1X + c_0)$$

där $1 \leq k < n$ och $1 \leq l < n$. Då är

$$a_0 = b_0c_0$$

och $p|b_0$ eller $p|c_0$ men ej båda (varför?). Låt $p|b_0$ och $p\not|c_0$. Visa succesivt att $p|b_1$, $p|b_2, ..., p|b_k$ genom att studera likheterna:

$$\begin{aligned}
&a_0 = b_0c_0,\\
&a_1 = b_0c_1 + b_1c_0,\\
&a_2 = b_0c_2 + b_1c_1 + b_2c_0,\\
&\dots\\
&a_k = b_0c_k + b_1c_{k-1} + ... + b_{k-1}c_1 + b_kc_0.
\end{aligned}$$

Är $p|b_k$ möjligt? (Observera att $a_n = b_kc_l$!).

**Anmärkning**. Ett resultat känt som Gauss lemma säger att om ett heltaligt polynom är irreducibelt i $\mathbb{Z}[X]$ så är det irreducibelt i $\mathbb{Q}[X]$. Detta resultat ter sig ganska självklart (om man tänker lite på det) men dess bevis är inte helt banalt [§].

3.6. Låt $N = a_na_{n-1}\dots a_1a_0$ beteckna ett naturligt tal med siffrorna $a_i$ (t ex $N = 452 = a_2a_1a_0$ med $a_0 = 2$, $a_1 = 5$, $a_2 = 4$). Betrakta polynomet

$$(*) \qquad f(X) = a_nX^n + a_{n-1}X^{n-1} + \dots + a_1X + a_0.$$

(a) **Delbarhetskriterium vid division med 3 och 9.** Visa att $N$ är delbart med 3 (respektive 9) då och endast då siffersumman i $N$ är delbar med 3 (respektive 9).

**Ledning.** Observera att siffersumman i $N$ är lika med $f(1)$ och dividera $f(X)$ med $X - 1$. Sätt in $X = 10$.

(b) **Delbarhetskriterium vid division med 11.** Visa att $N$ är delbart med 11 då och endast då summan $a_0 - a_1 + a_2 - a_3 + \cdots$ är delbar med 11 (exempel: 1331 är delbart med 11 ty $1 - 3 + 3 - 1 = 0$ är delbar med 11).

**Ledning.** Gör som i (a), men ersätt $X - 1$ med $X + 1$.

3.7. **Derivatan av ett polynom.** Låt $f(X) = a_0 + a_1X + \dots + a_nX^n \in K[X]$. Derivatan av $f(X)$ definieras helt formellt som

$$f'(X) = a_1 + 2a_2X + \dots + na_nX^{n-1}.$$

(a) Visa de vanliga deriveringsreglerna

$$(f + g)' = f' + g', \;\; (fg)' = f'g + fg'.$$

(b) Visa att $a \in K$ är ett multipelt nollställe till $f \in K[X]$ (dvs $a$ har multipliciteten $> 1$) då och endast då $f(a) = f'(a) = 0$.

**Lösning.** "⇒" Låt $f(X) = (X - a)^2q(X)$ (multipliciteten av $a$ är minst 2). Då är $f'(X) = 2(X - a)q(X) + (X - a)^2q'(X)$ så att $f(a) = f'(a) = 0$.

[§]Gauss lemma visas i kursen "Algebraisk talteori".

"⇐" Antag att $f(a) = f'(a) = 0$ och att multipliciteten av $a$ är 1 dvs $f(X) = (X - a)q(X)$ och $q(a) \neq 0$. Då är $f'(X) = q(X) + (X - a)q'(X)$ så att $f'(a) = q(a) \neq 0$ – en motsägelse.

(c) Bestäm reella tal $a$ och $b$ så att polynomet $f(X) = aX^{1998} + bX^{1997} + 1$ är delbart med $(X - 1)^2$.

3.8. % Betrakta följande krets

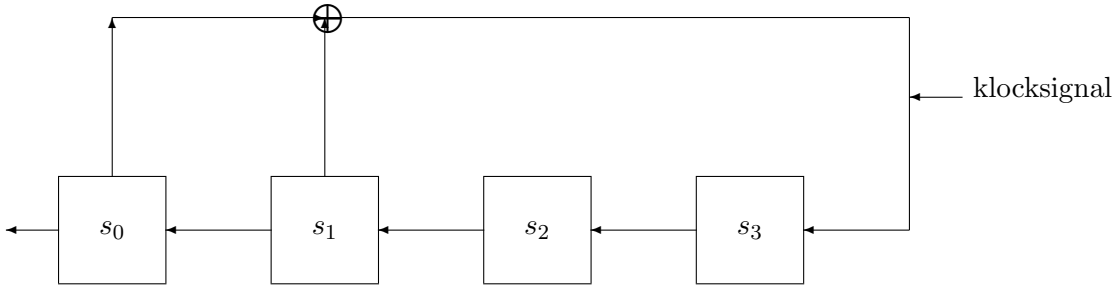



fig 1

Den fungerar så att under verkan av en klocksignal övergår innehållet i vart av ett av registren $s_1, s_2, s_3$ till det föregående registret, $s_0$ emiteras och $s_3$ ersättes med

$$s_4 = s_1 + s_0$$

(dvs efter 1 tidsenhet innehåller registren $s_1, s_2, s_3, s_4$). Allmänt har man

$$s_{n+4} = s_{n+1} + s_n$$

då $n = 0, 1, 2, \ldots$ .

(a) Skriv ut den emiterade sekvensen då $s_0 = 1, s_1 = 0, s_2 = 1, s_3 = 1$ och "+" betyder binär addition. Hur lång är perioden av den sekvensen.

**Anmärkning**. En krets av den typen kallas för ett linjärt återkopplat skiftregister. Sådana kretsar har en stor betydelse i radarkommunikation, kryptering, kodning, avkodning och slumptalsgenerering. Ofta är man intresserad av långa icke-periodiska signalsekvenser. Allmänt betraktar man kretsar:

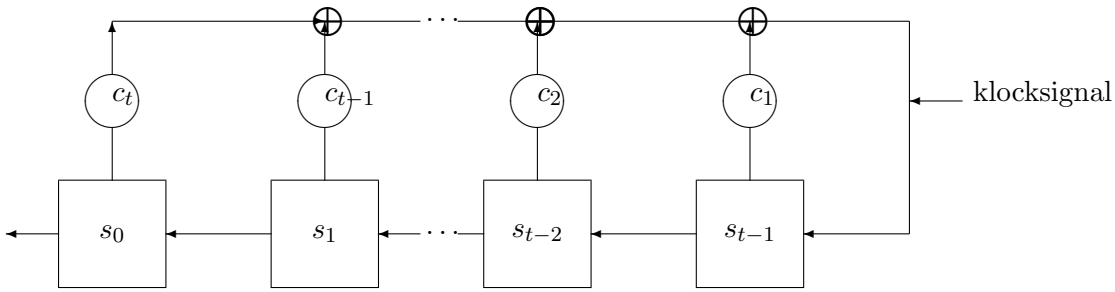



fig 2

där

$$s_{n+t} = c_1 s_{n+t-1} + c_2 s_{n+t-2} + ... + c_{t-1} s_{n+1} + c_t s_n$$

då $n = 0, 1, 2, ...$ . $s_i$ och $c_i$ behöver inte vara 0 eller 1 — de kan tillhöra en godtycklig ring ( men om de är 0 eller 1 och " $\oplus$ " betyder binär addition så förenklas kretsen som i fig 1).

Man säger att

$$p(X) = X^t - c_1 X^{t-1} - c_2 X^{t-2} - ... - c_{t-1} X - c_t$$

är **kopplingspolynomet** till skiftregistret i fig 2. t ex är kopplingspolynomet för kretsen i fig. 1:

$$p(X) = X^4 - X - 1$$

(dvs $p(X) = X^4 + X + 1$ ty $-1 = 1$ i $\mathbb{Z}_2$). Man visar att om $p(X)$ är irreducibelt i $\mathbb{Z}_2[X]$ så är längden av perioden av $s_0, s_1, s_2, ..., s_n, ...$ en delare till $2^l - 1$. Man visar också att det finns irreducibla polynom för vilka man får exakt längden $2^l - 1$ (sådana polynom kallas primitiva).

(b) Visa att polynomet $X^5 + X^2 + 1$ är irreducibelt i $\mathbb{Z}_2[X]$ och konstruera ett linjärt återkopplat skiftregister med detta polynom som kopplingspolynom. Motivera att kretsen genererar en icke-periodisk signalsekvens av längden 31.%