Uppsala universitet
Matematiska institutionen
Christer Kiselman

Övningar
special1
2003-09-17

Specialkurs i matematik NV1, hösten 2003

Den nya kursen med det föreslagna namnet *Specialkurs i matematik NV1* består, kort uttryckt, av tre delar: fördjupning i matematik; introduktion till informationssökning; studier av hur forskning går till inom de matematiska vetenskaperna.

Vad gäller **den andra delen** kommer Ulrika Haak på Ångströmbiblioteket att ge en introduktion till informationssökning 2003-10-01 13:15—17:00. Vidare kommer Ingrid Löfgren att presentera Beurlingbilioteket 2003-10-09 15:15—16:00. Informationssökning är för övrigt en integrerad del av hela kursen, såväl vad avser den första som den tredje delen.

Den tredje delen syftar till att ge kontakt med aktuell forskning inom de matematiska vetenskaperna. Det handlar då om matematik, matematisk logik, matematisk statistik, datoriserad bildbehandling, beräkningsvetenskap, numerisk analys. Det finns andra vetenskaper som använder mycket matematik, till exempel reglerteknik. Dessa kan också tas med här.

Idén är att kursdeltagarna först skall ta reda på vilken forskning som finns i Uppsala inom de nämnda områdena (detta kan ske genom att man letar på den världsvida väven, i årsredogörelser och genomför litteratursökning i olika databaser), och sedan söka upp en forskargrupp eller en enskild forskare för samtal och intervjuer och eventuellt samarbete. (Till forskarna räknas även doktorander.)

Dessa kontakter skall resultera i en skriftlig redogörelse och ett kort föredrag inför kurskamraterna.

Den skriftliga redogörelsen bör åtminstone omfatta följande:

Forskningsområdets vetenskapliga namn
Vetenskaplig beskrivning av forskningsområdet
Forskningsområdets populärvetenskapliga namn
Populärvetenskaplig beskrivning av forskningsområdet
Forskningsprojektets vetenskapliga namn
Vetenskaplig beskrivning av forskningsprojektet
Forskningsprojektets populärvetenskapliga namn
Populärvetenskapligt beskrivning av forskningsprojektet
Varför är forskningsområdet viktigt inomvetenskapligt sett?
Varför är forskningsområdet viktigt sett i ett samhälleligt perspektiv?
Vilka mål har forskningen inom detta område?
Varför är forskningsprojektet viktigt inomvetenskapligt sett?
Varför är forskningsprojektet viktigt sett i ett samhälleligt perspektiv?

Vilka mål har forskningsprojektet?
Vilka arbetar på detta i Uppsala? I världen i övrigt?
Intervjuer med någon eller några forskare

Deltagarna får fritt lägga till andra rubriker efter eget huvud.

Vad så gäller **den första delen** presenteras här litet material i anslutning till den. Det handlar inte om en lärobok utan om underlag för övningarna och uppgifterna.

1. Kardinaltal

Syftet med att införa kardinaltal är att ge en god mening till uttrycken *lika många som*, *högst lika många som* och *minst lika många som*.

1.1. Avbildningar

Vad är en avbildning? Vad betyder skrivsättet: $f\colon X \to Y$? Om $A \subset X$, så definerar vi $f(A) = \{f(x); x \in A\} \subset Y$, kallad *bilden av* A. Om $B \subset Y$, så definierar vi $f^{-1}(B) = \{x \in X; f(x) \in B\}$, kallad *urbilden av* B, eller *den inversa bilden av* B.

Observera att $\operatorname{im} f = f(X)$, kallad *bildmängden* eller *värdemängden*, inte behöver vara lika med hela Y. Man kan också definiera $\ker f = \{(x_1, x_2) \in X^2; f(x_1) = f(x_2)\}$, *kärnan*.

Man säger att en avbildning f är en *injektion* om $f(x_1) = f(x_2)$ medför att $x_1 = x_2$. (Ekvivalent: $\ker f =$ diagonalen i X^2.) Man säger att f är en *surjektion* om $\operatorname{im} f = Y$. Man säger att f är en *bijektion* om f är både en injektion och en surjektion. (Adjektiv: *injektiv, surjektiv, bijektiv.*) Vad betyder ett-ett-avbildning? (Injektion eller bijektion?)

1.2. Definition av kardinaltal

Två mängder X och Y säges ha *samma kardinaltal* om det finns en bijektion $f\colon X \to Y$. Man skriver då $X \approx Y$ och $\operatorname{card} X = \operatorname{card} Y$. Men vad är $\operatorname{card} X$? Vi kan tänka oss att $\operatorname{card} X$ är mängden av alla mängder Y sådana att $Y \approx X$. Kanske leder detta till svårigheter...

Vi ser att relationen $\approx$ mellan mängder är *reflexiv, symmetrisk* och *transitiv*. Vad betyder det? Relationen $=$ mellan kardinaltal har likaså dessa tre egenskaper.

Man säger att X *har kardinaltal högst lika med kardinaltalet för* Y om det finns en injektion $f\colon X \to Y$. Man skriver då $X \preccurlyeq Y$ och $\operatorname{card} X \leqslant \operatorname{card} Y$.

Det är klart att om $X \approx Y$ så gäller $X \preccurlyeq Y$ och $Y \preccurlyeq X$. Man kan visa att relationen $\preccurlyeq$ är reflexiv och transitiv. Man önskar att den också skulle vara asymmetrisk i den meningen att $X \preccurlyeq Y$ och $Y \preccurlyeq X$ skulle medföra att $X \approx Y$. Översatt till kardinaltal betyder detta att $x \leqslant y$ och $y \leqslant x$ medför att $x = y$. Det vore väl naturligt! Är det så? Ja, det är Felix Bernsteins sats från 1898, som vi skall diskutera snart.

1.3. Kardinaltalens aritmetik

Ändliga och oändliga kardinaltal. Uppräkneliga och icke uppräkneliga kardinaltal. Galilei 1638. Cantors diagonalförfarande.

Aritmetik: om X och Y är två mängder och $x = \operatorname{card} X$, $y = \operatorname{card} Y$, så definerar vi *summan* $x + y$ som kardinaltalet hos $X \cup Y$ om X och Y är disjunkta. (Hur gör man annars?) Detta stämmer ju när mängderna är ändliga, och man gör alltså den observationen till en definition i det allmänna fallet.

Analogt definierar vi *produkten* xy som kardinaltalet hos $X \times Y$. Återigen stämmer det i fallet med ändliga mängder.

För potenser observerar vi att y^x är antalet avbildningar av en ändlig mängd X in i en annan ändlig mängd Y, åtminstone om X är icke-tom. Vi gör därför detta till en definition och säger att y^x är kardinaliteten hos mängden av alla avbildningar $X \to Y$ av en mängd $X \neq \emptyset$ med $\operatorname{card} X = x$ in i en mängd Y med $\operatorname{card} Y = y$. Man brukar definiera $y^0 = 1$ om $y \neq 0$ – det är väl naturligt? Det återstår nu endast att definiera 0^0, men vi avstår från att göra det tills vi behöver veta vad det är eller bör vara.

Därmed har vi definierat summor, produkter och potenser av alla kardinaltal utom 0^0.

1.4. Räkneregler för oändliga kardinaltal

Det minsta oändliga kardinaltalet är kardinaltalet för de naturliga talen $\mathbf{N}$. Det betecknas $\aleph_0$ (uttalas alef-noll). Man kan nu visa att $\aleph_0 + \aleph_0 = \aleph_0$ och att $\aleph_0^2 = \aleph_0$. Det är ungefär samma påstående som att $\operatorname{card}\mathbf{Z} = \operatorname{card}\mathbf{N}$ och att $\operatorname{card}\mathbf{Q} = \operatorname{card}\mathbf{N}$.

Man kan mer allmänt visa att om x är ett oändligt kardinaltal, så gäller $x + x = x$, $xx = x^2 = x$, $x^n = x$ för alla heltal $n \in \mathbf{N}^*$. Vidare att $x + y = y$ och $xy = y$ om $x \leqslant y$. Det verkar som om man aldrig skulle få några nya oändliga kardinaltal genom att utföra addition och multiplikation. Detta står i skarp kontrast till de naturliga talen, som ju kan byggas upp genom successiv addition som 0, $0 + 1 = 1$, $1 + 1 = 2$, $2 + 1 = 3$, $3 + 1 = 4$, osv.

Men hur blir det med 2^x? Vi har $2^0 = 1 > 0$, $2^1 = 2 > 1$, $2^2 = 4 > 2$ och allmänt $2^n > n$ för alla $n \in \mathbf{N}$.

Man kan nu visa att man har $x \leqslant 2^x$ men att det aldrig gäller att $2^x = x$. Vi sammanfattar detta genom att skriva $x < 2^x$, där vi alltså nu inför ett nytt tecken: $x < y$ betyder att $x \leqslant y$ men att inte $y = x$. I själva verket skall vi visa att man inte ens har $2^x \leqslant x$.

Sats. *För varje kardinaltal x gäller att $x < 2^x$.*

Bevis. För det första är det lätt att visa att $x \leqslant 2^x$. För varje mängd X kan vi definiera en avbildning $\varphi\colon X \to 2^X$ genom att definiera $\varphi(x)$ som den funktion f som uppfyller $f(x) = 1$ och $f(y) = 0$ då $y \neq x$. Det är klart att olika punkter ger upphov till olika funktioner, så φ är verkligen en injektion. (Vi har här även behandlat det litet konstiga fallet $0 \leqslant 2^0$, dvs. injektionen $\emptyset \to \{1\}$, där resultatet är sant tack vare att vi definierat 2^0 som 1.)

Om det funnes en injektion $\psi\colon 2^X \to X$, så skulle vi kunna definiera en funktion $g \in 2^X$, alltså $g\colon X \to \{0, 1\}$, genom att föreskriva att $g(x) = 1$ om $x \in \operatorname{im}\psi$ med $x = \psi(f)$ och $f(x) = 0$, samt att $g(x) = 0$ annars. Då får vi en punkt $y = \psi(g) \in X$. Vad gäller om denna punkt y? Uppenbarligen tillhör den $\operatorname{im}\psi$, ty $y = \psi(g)$. Och då skall vi enligt definitionen av g ha $g(y) = 1$ om $g(y) = 0$, och $g(y) = 0$ om $g(y) = 1$. Detta är en motsägelse, som visar att det inte kan finnas någon injektion $2^X \to X$; vi kan aldrig ha $2^x \leqslant x$.

Beviset kallas för Cantors diagonalförfarande. Varför?

1.5. Felix Bernsteins[1] sats

Sats. *Om det för två kardinaltal x och y gäller att $x \leqslant y$ och $y \geqslant x$, så gäller $x = y$.*

[1] Felix Bernstein, 1878—1956

Bevis. Översatt till mängder betyder detta att om $X \preccurlyeq Y$ och $Y \preccurlyeq X$, så skall vi visa att $X \approx Y$. Vi skall alltså visa att om vi har två injektioner $f\colon X \to Y$ och $g\colon Y \to X$, så finns det en bijektion $h\colon X \to Y$.

Vi skall inte bara visa existensen av ett sådant h utan faktiskt konstruera avbildningen genom att sy/snickra ihop bitar från f och g. Vilka tygbitar/byggbitar finns det? Jo, vi har ju $\operatorname{im} f \subset Y$ och $\operatorname{im} g \subset X$ och sedan deras bilder under g och f. Sedan kan vi upprepa detta och ta bilderna under f och g och så vidare. Det finns egentligen inga andra bitar, så om det överhuvudtaget går att konstruera h från f och g (och inte bara ge ett abstrakt existensbevis), så måste det gå med just dessa bitar.

Definiera först $A_0 = X \setminus \operatorname{im} g$, och därefter $B_k = f(A_{k-1})$ och $A_k = g(B_k)$, $k \in \mathbf{N}^*$. Om nu $x \in \bigcup A_k$, så definierar vi $h(x) = f(x)$. Om däremot $x \notin \bigcup A_k$, så definierar vi $h(x) = g^{-1}(x)$.

Vi har alltså pusslat ihop restriktioner av de två bijektionerna $f\colon X \to \operatorname{im} f$ och $g^{-1}\colon \operatorname{im} g \to Y$ för att få en avbildning av hela X på hela Y.

Gå igenom de olika fallen och visa att h är injektiv och surjektiv.

1.6. Sierpiński–Mazurkiewicz' paradox

När man vant sig vid oändliga kardinaltal, så är det inte så konstigt att en uppräknelig familj av uppräkneliga mängder har en uppräknelig union. Men det finns värre:

Betrakta mängden av komplexa tal

$$E = \{a_0 + a_1 e^i + a_2 e^{2i} + \cdots + a_m e^{mi}; m \in \mathbf{N}, a_j \in \mathbf{N}, a_m \neq 0\}.$$

Definiera vidare A_k som mängden av alla komplexa tal som har en representation med $a_0 = k$. Då gäller förstås att $E = \bigcup_{k\in\mathbf{N}} A_k$. Vidare är mängderna A_k, $k \in \mathbf{N}$, disjunkta. Detta följer av att e^i är ett transcendent tal, vilket vi accepterar här. Mängden A_0 är kongruent med varje mängd A_k (genom translation). Men A_0 är också kongruent med hela E (genom rotation).

Så vi har funnit en mängd som kan slås sönder i uppräkneligt många disjunkta delmängder, var och en lika mäktig som hela mängden, men inte bara det: varje sådan delmängd är kongruent (isometrisk) med hela mängden. Detta kallas Sierpiński[2]–Mazurkiewicz'[3] paradox (1914). Men det är förstås ingen paradox – om man vant sig vid den. Det verkar svårt att tilldela A_0 ett mått annat än 0 och $+\infty$.

1.7. Övningar

Gå igenom alla frågorna ovan.

Vad betyder ett-ett-avbildning (one-to-one mapping, one-to-one correspondence)? Sök på den världsvida väven och i Beurlingbiblioteket och försök göra statistik över den faktiska betydelsen. Vilka motiveringar finns det för de olika valen?

Gå igenom beviset för att det finns lika många rationella tal som hela tal (i kardinalitetsmening).

Gå igenom beviset för att det finns flera reella tal än rationella tal (i kardinalitetsmening).

Visa att $\aleph_0 + \aleph_0 = \aleph_0$ och att $\aleph_0^2 = \aleph_0$. Visa, eller leta reda på hur man gör för att visa, att $x + x = x$ och $x^2 = x$ för alla oändliga kardinaltal x.

[2]Wacław Sierpiński, 1882—1969.

[3]Stefan Mazurkiewicz, 1888—1945.

Fullborda beviset för Felix Bernsteins sats genom att kolla att den konstruerade avbildningen verkligen blir bijektiv.

Visa allt som påstås ovan i Sierpiński–Mazurkiewicz' paradox.

2. Gaussiska primtal

2.1. Beskrivning av uppgiften

Detta avsnitt kan användas på olika sätt. Ett första syfte är helt enkelt att uppmärksamma att begreppet primtal inte är något absolut utan beror på vad man relaterar det till. Man kan tänka sig en rent grafisk uppgift där det gäller att pricka in de gaussiska primtalen på ett papper för att åskådliggöra hur de fördelar sig i några olika områden. Om man vill gå längre kan man räkna ut tätheten inom några delar av det komplexa planet. En annan möjlighet är att gå igenom teorin för de gaussiska primtalen; då kan följande rader tjäna som ledning, och ett mål kan vara att i detalj visa allt som jag antyder. Ett mer avancerat projekt slutligen är att studera faktorisering i några andra ringar $\mathbf{Z}\big[\sqrt{d}\big]$ för heltal d; här handlar det ju bara om $d = -1$.

2.2. Vanliga primtal och komplexa primtal

Talen 2, 3, 5, 7, 11, ... är primtal, vilket betyder att man inte kan faktorisera dem utan att en faktor måste vara 1 eller -1. Vi kan till exempel skriva

$$19 = 1 \cdot 19 = 19 \cdot 1 = (-1) \cdot (-19) = (-19) \cdot (-1)$$

men inte på något annat sätt med heltal, medan däremot 21 kan faktoriseras som $3 \cdot 7$ eller $(-7) \cdot (-3)$. Man kan också räkna med komplexa tal $z = x + iy$ där x och y är vanliga heltal. Om vi betecknar de vanliga heltalen med $\mathbf{Z}$, så bildar alla tal av formen $z = x + iy$ med $x, y \in \mathbf{Z}$ en mängd $\mathbf{Z} + i\mathbf{Z}$ som också betecknas $\mathbf{Z}[i]$ och kallas *ringen av gaussiska heltal*.[4]

I $\mathbf{Z}[i]$ inträffar nu att några av våra gamla primtal kan faktoriseras på ett nytt sätt. Vi kan t.ex. skriva

$$2 = (1 + i)(1 - i), \qquad 5 = (2 + i)(2 - i) \qquad \text{och} \qquad 101 = (10 + i)(10 - i),$$

vilket visar att 2, 5 och 101, som är primtal i $\mathbf{Z}$, inte är primtal i $\mathbf{Z}[i]$. Begreppet primtal beror alltså på i vilken ring man räknar. Däremot förblir det gamla primtalet 3 ett primtal också i $\mathbf{Z}[i]$, ty det kan faktoriseras som

$$3 = 3 \cdot 1 = 1 \cdot 3 = i(-3i) = (-i)(3i)$$

och på några andra sätt med ± 1 eller $\pm i$ som faktor, men inte utan att en av dessa faktorer med absolutbelopp 1 uppträder. (Visa detta!) Nu kan ju alla gaussiska heltal faktoriseras som

$$z = 1 \cdot z = (-1)(-z) = i(-iz) = (-i)(iz),$$

så faktorer ± 1, $\pm i$ bör vi betrakta som oväsentliga, precis som ± 1 för de vanliga primtalen. Vi kan alltså nu definiera z som ett *gaussiskt primtal* om varje faktorisering $z = ab$ med $a, b \in \mathbf{Z}[i]$ måste ha endera $|a| = 1$ eller $|b| = 1$, dvs. en faktor måste vara

[4] Dessa allmännare heltal är uppkallade efter Carl Friedrich Gauss, 1777—1855.

i^k, $k = 0, 1, 2, 3$. Vi ser att dessa fyra element i $\mathbf{Z}[i]$ är de enda som har invers i $\mathbf{Z}[i]$, precis som ± 1 är de enda heltal som har invers i $\mathbf{Z}$.

Vi kan nu säga att ett tal i $\mathbf{Z}$ som inte är ett primtal inte heller kan vara primtal i $\mathbf{Z}[i]$, ty en faktorisering $z = ab$ med $a, b \in \mathbf{Z}$ gäller ju också i $\mathbf{Z}[i]$. Och som vi sett kan det inträffa att ett primtal i $\mathbf{Z}$ förblir primtal i den större ringen $\mathbf{Z}[i]$ (exempelvis talet 3) men också att det blir sammansatt i $\mathbf{Z}[i]$ (exempelvis $5 = (2 + i)(2 - i)$). Och dessutom finns det nya primtal i $\mathbf{Z}[i]$ som inte ligger i $\mathbf{Z}$ (exempelvis $1 + i$; visa att det är ett gaussiskt primtal!).

2.3. Unik faktorisering av gaussiska heltal

Att varje gaussiskt primtal kan uttryckas som en produkt av primtal är lätt att se. Men är faktoriseringen unik, och hur bevisar man det i så fall? (När man säger unik här bortser man från variationer i ordningsföljden och faktorer som har invers i $\mathbf{Z}[i]$, dvs. i^k, $k = 0, 1, 2, 3$. Vi kanske kan säga att faktoriseringen är väsentligen unik.) Svaret är att Euklides' algoritm fungerar för dem lika väl som för de vanliga heltalen $\mathbf{Z}$.

Sats (divisionsalgoritmen för de vanliga heltalen). *Om $a, b \in \mathbf{Z}$ med $b \neq 0$ så finns heltal q (kvoten) och r (resten) sådana att $a = qb + r$ och $|r| \leqslant \frac{1}{2}|b| < |b|$.*

Bevis. Det finns ett heltal q som ligger så nära a/b att $|q - a/b| \leqslant \frac{1}{2}$. Då gäller att $|r| = |a - qb| = |(a/b - q)b| = |a/b - q||b| \leqslant \frac{1}{2}|b|$.

Sats (divisionsalgoritmen för de gaussiska heltalen). *Om $a, b \in \mathbf{Z}[i]$ med $b \neq 0$ så finns gaussiska heltal q (kvoten) och r (resten) sådana att $a = qb + r$ och $|r| \leqslant \frac{1}{\sqrt{2}}|b| < |b|$.*

Bevis. Det finns ett gaussiskt heltal q som ligger så nära a/b att $|q - a/b| \leqslant \frac{1}{\sqrt{2}}$. Då gäller att $|r| = |a - qb| = |a/b - q||b| \leqslant \frac{1}{\sqrt{2}}|b|$.

I båda fallen gäller $|r| < |b|$, vilket räcker i fortsättningen. Huruvida den exakta faktorn är $\frac{1}{2}$ eller $\frac{1}{\sqrt{2}}$ är inte viktigt.

Euklides' algoritm blir nu likalydande i båda fallen. Följden av absolutbelopp av de successiva resterna bildar en strikt avtagande följd så länge resten är skild från noll. I det första fallet är resterna heltal; i det andra fallet är deras kvadrater heltal. Slutsatsen blir densamma: efter ändligt många steg blir resten noll. Beviset för att faktoriseringen av gaussiska heltal i primtal är unik blir nu ordagrant detsamma som i fallet med vanliga heltal.

2.4. Testa om ett gaussiskt heltal är prima

Skriv ett datorprogram som undersöker om ett givet tal $z \in \mathbf{Z}[i]$ är ett gaussiskt primtal eller ej. Om du har ett program som kan dividera komplexa tal direkt så är det bara att prova om z/c ligger i $\mathbf{Z}[i]$ för olika heltal $c \in \mathbf{Z}[i]$. Det räcker att testa med alla c som uppfyller $1 < |c| \leqslant \sqrt{|z|}$, dvs. ändligt många. Varför?

Om programmet inte kan dividera komplexa tal så får vi låta det undersöka real- och imaginärdelarna för sig. Vi skriver kvoten z/c så här:

$$\frac{z}{c} = \frac{x + iy}{a + ib} = \frac{(a - ib)(x + iy)}{a^2 + b^2} = \frac{ax + by + i(ay - bx)}{a^2 + b^2}.$$

Tydligen är

$$\operatorname{Re} \frac{z}{c} = \frac{ax + by}{a^2 + b^2} \qquad \text{och} \qquad \operatorname{Im} \frac{z}{c} = \frac{ay - bx}{a^2 + b^2},$$

och z/c är ett gaussiskt heltal precis när både $\mathrm{Re}(z/c)$ och $\mathrm{Im}(z/c)$ är vanliga heltal. Som sagt, det räcker att undersöka detta för $c = a + ib$ med $1 < |c|^2 = a^2 + b^2 \leqslant |z|$. Det räcker till och med att kontrollera sådana c som ligger i den första kvadranten, dvs. sådana som uppfyller $a \geqslant 0$ och $b \geqslant 0$. Varför? Dessa anmärkningar kan spara tid.

Gör ett program som trycker ut alla gaussiska primtal upp till en viss gräns. Pricka sedan in dem i ett diagram – eller låt datorn göra det. Under sökandet behöver man bara undersöka en åttondel av planet, t.ex. $z = x + iy$ med $0 \leqslant y \leqslant x$, ty talen $\pm z$, $\pm\bar{z}$, $\pm iz$, $\pm i\bar{z}$ är primtal samtidigt, och ett av dem ligger i oktanten $0 \leqslant y \leqslant x$. Vilka tal är det som är primtal i $\mathbf{Z}$ men upphör att vara det i $\mathbf{Z}[i]$? Går det att säga något om hur de gaussiska primtalen fördelar sig i det komplexa talplanet? Kan du se om de ligger tätare kring origo än långt från origo? (Troligen måste man ha ett ganska stort diagram för att kunna se det.)

2.5. Samband mellan de olika typerna av primtal

Om $z = x + iy$ är ett gaussiskt heltal så beskaffat att $|z|^2 = x^2 + y^2$ är ett primtal i $\mathbf{Z}$, så är z ett primtal i $\mathbf{Z}[i]$. Visa det! Med detta kriterium kan vi till exempel se att $10 + 3i$ är prima, ty dess absolutbelopp i kvadrat är $|10 + 3i|^2 = 10^2 + 3^2 = 109$ som är ett primtal i $\mathbf{Z}$. Omvänt kan vi fråga oss om $|z|^2$ är ett primtal i $\mathbf{Z}$ om z är ett primtal i $\mathbf{Z}[i]$. Svaret är nej, ty 3 är ett gaussiskt primtal medan $|3|^2 = 9$ inte är prima. Men om vi tar ett primtal $z = x + iy$ med imaginärdel $y \neq 0$, är då $|z|^2 = x^2 + y^2$ ett vanligt primtal? Försök visa det!

Vi kan dela in de vanliga positiva primtalen i tre klasser efter vilken rest de ger vid division med fyra: 5, 13, 17, ..., som ger resten 1 vid division med fyra; 3, 7, 11, 19, ... som ger resten 3; och så det återstående primtalet som är 2, det enda jämna primtalet. Det visar sig nu att inget primtal i den första klassen, alltså de som har formen $4k + 1$ för något heltal k, är primtal i $\mathbf{Z}[i]$. De kan alla faktoriseras som $p = (a + ib)(a - ib)$. Man kan nämligen visa att ett sådant primtal p kan skrivas $p = a^2 + b^2$ för några tal a och b; ett bevis för detta finns i till exempel LeVeque [1956, volym I, kapitel 7]. Talen $a + ib$ och $a - ib$ måste vara gaussiska primtal. (Vad är nämligen kvadraten på deras absolutbelopp?)

De positiva primtalen av typen $p = 4k + 3$ däremot är primtal även i den större ringen $\mathbf{Z}[i]$. Försök visa detta! Kanske kan följande vara till hjälp: om p kunde faktoriseras i $\mathbf{Z}[i]$, $p = (a + ib)(c + id)$, så skulle

$$p^2 = |p|^2 = |a + ib|^2|c + id|^2 = (a^2 + b^2)(c^2 + d^2).$$

Och om $|a + ib| > 1$ och $|c + id| > 1$ så måste $p = a^2 + b^2$. Vilka rester kan nu ett tal av formen $a^2 + b^2$ ge vid division med 4?

Talet 2, slutligen, som är primtal i $\mathbf{Z}$, är som vi redan sett inte primtal i $\mathbf{Z}[i]$.

2.6. Fördelning av primtalen

Om fördelningen av de gaussiska primtalen kan vi säga något intressant när vi vet något om fördelningen av de positiva primtalen. Det resultat som uttalar sig om denna kallas primtalssatsen och bevisades år 1896 av Jacques Hadamard och Charles de La Vallée-Poussin. Primtalssatsen säger att antalet primtal p med $2 \leqslant p \leqslant x$, betecknat $\pi(x)$,

uppfyller en asymptotisk relation

$$\pi(x) \sim \frac{x}{\log x},$$

där tecknet $\sim$ betyder att kvoten mellan de bägge leden går mot 1 då $x \to +\infty$, dvs.

$$\pi(x) = \frac{x}{\log x}\,(1 + g(x))$$

där $g(x) \to 0$ då $x \to +\infty$. Läs något mer om primtalssatsen i någon av de böcker som nämns i referenslistan.

Vi delar in π i tre delar, svarande mot resterna vid division med 4,

$$\pi = \pi_1 + \pi_2 + \pi_3,$$

där π_1 räknar primtalen av typ $4k+1$, π_2 det enda jämna primtalet (alltså $\pi_2(x) = 1$ om $x \geqslant 2$, $\pi_2(x) = 0$ annars), och π_3 räknar primtalen av typ $4k+3$.

Till primtalen av typ $4k+1$ hör åtta gaussiska primtal, nämligen $\pm a \pm ib$ och $\pm b \pm ia$, alla med absolutbelopp $\sqrt{p}$. Det blir verkligen åtta olika tal, ty $a \neq 0$, $b \neq 0$ och $a \neq b$. Antalet gaussiska primtal z med $|z| \leqslant r$ som vi får fram genom att ta $p = 4k+1$ blir alltså $8\pi_1(r^2)$.

Till primtalet 2 hör de fyra gaussiska primtalen $\pm 1 \pm i$. Antalet gaussiska primtal z av denna typ är alltså $4\pi_2(r^2)$.

Till primtalen $p = 4k+3$ hör fyra gaussiska primtal $z = i^m p$, $m = 0, 1, 2, 3$. De har alla samma absolutbelopp som p, varav följer att de som uppfyller $|z| \leqslant r$ är $4\pi_3(r)$ till antalet.

När vi räknar ihop alla gaussiska primtal z i cirkelskivan $|z| \leqslant r$ blir deras antal således

$$\gamma(r) = 8\pi_1(r^2) + 4\pi_2(r^2) + 4\pi_3(r).$$

Denna formel gäller exakt, men för att få reda på hur antalet växer med r måste vi veta något om hur stora π_1 och π_3 är jämfört med varandra. Det är nu känt att det är ungefär lika vanligt att ett primtal är av typen $4k+1$ som av typen $4k+3$, dvs.

$$\pi_1(x) \sim \frac{x}{2\log x} \qquad \text{och} \qquad \pi_3(x) \sim \frac{x}{2\log x}.$$

Eftersom $\pi_2(x) \sim 1$ så får vi

$$\gamma(r) \sim 8\frac{r^2}{2\log r^2} + 4 + 4\frac{r}{2\log r} = \frac{2r^2}{\log r} + 4 + \frac{2r}{\log r} \sim \frac{2r^2}{\log r}.$$

Vi ser att de gaussiska primtal som ligger utanför de två axlarna och har absolutbelopp $> \sqrt{2}$ (dvs. de som räknas av den första termen) överväger.

Medeltätheten av de gaussiska primtalen i cirkelskivan $|z| \leqslant r$ blir $\gamma(r)$ dividerat med skivans area:

$$\frac{\gamma(r)}{\pi r^2} \sim \frac{2}{\pi \log r}.$$

Vi kan nu jämföra denna med medeltätheten i intervallet $[-x, x]$ av de vanliga primtalen, som är

$$\frac{2\pi(x)}{2x} \sim \frac{1}{\log x}.$$

2.7. Litteraturhänvisningar

Carleson, Lennart
1968 *Matematik för vår tid.* Stockholm: Prisma.

Crandall, Richard
2001 *Prime Numbers: A Computational Perspective.* New York: Springer. xv, 545 ss.

Hardy, G. H. & Wright, E. M.
1979 *An Introduction to the Theory of Numbers.* 5-e upplagan, Oxford.

Huxley, M. N.
1972 *The Distribution of Prime Nubmers: Large Sieves and Zero-Density Theorems.* x, 128 ss.

Ingham, A. E.
1990 *The Distribution of Prime Numbers.* Cambridge: Cambridge University Press.

Jameson, G. J. O.
2003 *The Prime Number Theorem.* Cambridge: Cambridge University Press. x, 252 ss.

LeVeque, W. J.
1956 *Topics in Number Theory, I & II.* Addison-Wesley.

Newman, D. J.
1980 Simple analytic proof of the prime number theorem. *Amer. Math. Monthly* **87**, 693—696.

Riesel, Hans
1968 *En bok om primtal.* Odense.

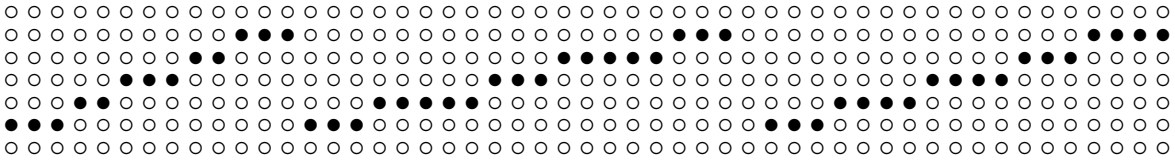
1985 *Prime Numbers and Computer Methods for Factorization.* Boston, Stuttgart: Birkhäuser.

3. Datorskärmens geometri

3.1. Inledning

Punkter, räta linjer och plan har människan studerat i över två tusen år, och vissa kurvor, som ellipser och hyperblar, har varit föremål för vår nyfikenhet nästan lika länge. Allmännare kurvor, som lemniskator och kardioider, har studerats i flera hundra år. Studiet av sådana kurvor grundar sig på att vi kan rita dem på papper och få idéer från handritade bilder. Men med datorerna har vi fått ett nytt sätt att rita. På en datorskärm ser vi bilder, och bilderna består av små bildelement eller pixlar, som ögat sätter ihop till geometriska objekt. En rät linje blir då inte det som Euklides avsåg med en rät linje, utan en ändlig mängd av prickar på skärmen, som ögat ändå uppfattar som ett sammanhängande linjestycke. En kurva är likaså en ändlig mängd av bildelement. (Se figur 3.1.)

Finns det en geometri för dessa bilder på datorskärmen? Svaret är ja. Vi behöver inte nöja oss med att uppfatta bilderna som mer eller mindre noggranna approximationer av ideala räta linjer eller kurvor, utan kan behandla dessa ändliga punktmängder med



Figur 3.1. Tre pixelmängder. Vilka av dem representerar en sträcka?

samma exakthet som Euklides hade i sin geometri. Detta är den digitala geometrin. Den är ung jämfört med Euklides'.

Azriel Rosenfeld gav år 1974 en definition av begreppet digital rät linje. Erik Melin fann år 2003 en annan digitalisering av räta linjer, som respekterar Khalimskys topologi (vi skall snart införa denna). Vi kan också tala om kurvor i det digitala planet. Vi kan ta vilket begrepp som helst i den euklidiska geometrin och försöka översätta det till den digitala geometrin, och se om ett visst resultat i den euklidiska geometrin blir sant i den digitala.

Speciellt skall vi i denna artikel titta på Jordans kurvsats. Denna sats handlar om kurvor i det euklidiska planet och säger att en sluten enkel kurva delar planet i två delar: en inre och en yttre komponent. Beviset är svårt. Om man nu har en kurva i det digitala planet – den består alltså av ändligt många punkter – kan den då dela in planet i två delar? Svaret är ja. Det resultatet bevisades av Efim Khalimsky (E. D. Halimskij) 1970.

Vi skall här beskriva några begrepp inom den digitala geometrin och illustrera dem genom att diskutera Khalimskys digitala version av Jordans kurvsats. Men för att nå dit måste man ompröva en del invanda föreställningar.

Det är nu för tiden lätt att motivera den digitala geometrin med dess tillämpningar inom datorgrafik och bildanalys. Men det kan vara värt att notera att Khalimsky införde sin topologi redan 1969, synbarligen utan några sådan tillämpningar i åtanke.

3.2. Att räkna med kartesiska koordinater

Den klassiska geometrin handlar om punkter, räta linjer och plan, men också om cirklar, klot och andra figurer. Sådana geometriska objekt har studerats i tusentals år, och vi alla är mer eller mindre bekanta med dem. Grekerna skapade i antiken en axiomatisk teori, vilket innebär att man bevisade egenskaper utgående från ett antal grundläggande antaganden, axiom, som inte bevisades. Man kunde räkna ut areor och volymer, men annars räknade man inte så mycket.

En revolution i beräkningsavseende kom med Cartesius (René Descartes, 1596—1650). Han representerade en punkt i planet med ett par av tal (x_1, x_2) (talen kallas *kartesiska koordinater*) och en linje i planet med mängden av alla par av tal (x_1, x_2) som uppfyller en ekvation $a_1x_1+a_2x_2+a_3 = 0$, där inte både a_1 och a_2 är noll. Ett plan i det tredimensionella rummet representeras av mängden av alla tripplar av tal (x_1, x_2, x_3) som uppfyller en ekvation $a_1x_1 + a_2x_2 + a_3x_3 + a_4 = 0$, där inte alla tre talen a_1, a_2, a_3 är noll. Om man har två räta linjer i planet med ekvationerna $a_1x_1 + a_2x_2 + a_3 = 0$ och $a'_1x_1 + a'_2x_2 + a'_3 = 0$, så skär de varandra i en viss mängd, och denna ges av alla lösningar (x_1, x_2) till båda ekvationerna. Kanske finns det ingen lösning (linjerna är skilda och parallella) eller oändligt många lösningar (linjerna sammanfaller) eller så finns det precis en lösning (linjerna skär varandra i en punkt). Man kan alltså

genom räkningar avgöra vad som gäller och uttrycka svaret med hjälp av de sex talen $a_1, a_2, a_3, a_1', a_2'$ och a_3'.

3.3. Att representera figurer i en dator

Nästa revolution i beräkningsavseende kom med datorerna. Man kan nu lagra och bearbeta större informationsmängder än med papper och penna. Speciellt kan man lagra information om geometriska figurer. Vi vet sedan Cartesius att två tal räcker för att beskriva en punkt i planet, tre tal en punkt i rummet. Hur många tal behövs det för att beskriva en cirkel i planet? Jo, tre: centrum ges av två koordinater och radien av ett tal. Hur många tal behövs det för att beskriva en ellipsoid i det tredimensionella rummet? Centrum ges av tre koordinater; för de tre halvaxlarna behövs ytterligare tre tal. Så behöver man ange riktningen hos den största axeln; till detta behövs två vinklar. För att sedan bestämma riktningen hos den näst största axeln behövs ytterligare en vinkel. Alltså kan ellipsoiden beskrivas fullständigt av nio tal. (Om man från början vet att några axlar är lika, så blir det färre.) Ett reellt tal kan behöva oändligt många decimaler för att beskrivas fullständigt, så vi måste nöja oss med att lagra ändligt många av dem. Att lagra nio tal med en viss rimlig precision i en dator kräver inte mycket minnesutrymme.

Men hur blir det om man vill beskriva en godtycklig mängd? Hur många tal behövs för det? Det finns så fruktansvärt många mängder i planet att vi inser att vi måste approximera på något sätt; vi kan ju bara lagra och behandla information som består av ändligt många tecken. Det innebär att vi får dela in planet i små bitar, och så tala om huruvida en viss bit ingår i mängden eller ej. Detta är ju egentligen inte något som har kommit med datorerna, ty ett fotografi i tidningen består av ett raster, vilket man ser om man tittar på det med en lupp. Rastret är så fint att man på litet större avstånd inte kan se det, och ögat uppfattar bilden på ett bra sätt och störs inte av rastret. Också våra ögon har en begränsad kapacitet att ta in information, och det utnyttjar man alltså när man trycker fotografier. På samma sätt består en datorskärm av ändligt många bildelement, och en rät linje är i själva verket en ändlig mängd av punkter; ögat fogar ihop punkterna till en linje om de ligger tillräckligt tätt.

Men hur blir det om vi vill lagra information för att beskriva en godtycklig mängd i planet? Om vi som ett exempel tar en skärm som har 1 024 gånger 768 pixlar, dvs. är indelad i 1 024 små intervall på längden och 768 intervall på höjden, så innebär det att det finns $1\,024 \times 768 = 786\,432$ pixlar. Hur beskriver man nu en mängd? För varje pixel måste man tala om huruvida den ingår i mängden eller ej. Om vi skriver en etta när pixeln är med i mängden och en nolla när den inte är med, behöver vi alltså skriva 786 432 nollor eller ettor för att beskriva en godtycklig delmängd av skärmen. Och det finns $2^{786\,432} \approx 10^{236\,740}$ olika mängder. (Detta tal kan jämföras med universums massa, som några astronomer skattar till 10^{53} kg, vilket är 6×10^{79} protonmassor eller 10^{83} elektronmassor. Man brukar ju tala om stora tal som astronomiska, men det är en metafor som inte bara har bleknat: den är helt missvisande.[5])

Om vi delar in hela planet i kvadratiska eller rektangulära pixlar så kan vi numrera dem med par av heltal. Vi låter helt enkelt (x_1, x_2) vara koordinaterna för centrum

[5]Den som tror att denna slutsats beror på att universum har en ganska låg densitet ombedes beräkna massan hos ett fiktivt universum med radie lika med 14×10^9 ljusår $\approx 1,3 \times 10^{26}$ m och densitet lika med den hos en neutronstjärna, säg 10^{17} kg/m^3. Slutsatsen är densamma.

i en pixel i någon lämplig skala, vald så att x_1 och x_2 blir heltal. Därför kan vi helt enkelt tala om (x_1, x_2) som en pixel, fast det egentligen är pixelns adress.

3.4. Jordans kurvsats

En cirkel i planet delar in detta i två delar: ett inre område, där avståndet till centrum är mindre än radien, och ett yttre område, där avståndet till centrum är större än radien. På samma sätt är vi vana vid att andra, mer oregelbundna kurvor, delar in planet i ett inre område och ett yttre område. Vi kan alltså deformera cirkeln utan att denna egenskap att dela in planet i två delar går förlorad. Camille Jordan (1838—1922) visade 1893 en sats med just denna innebörd. Om man tar en kurva som ligger i ett plan och som är lik en cirkel i en speciell mening, så består dess komplement av exakt två delar. Det är som ett staket som stänger in fåren och stänger ute vargen. Om kurvan ser ut som en åtta så delas planet in i tre delar, så sådana kurvor får inte accepteras om vi skall få den önskade uppdelningen av planet. (Se figur 3.2.)

Figur 3.2. En kurva som inte är sluten; en som inte är enkel; en enkel och sluten kurva.

Jordans kurvsats gäller alltså för kurvor som är tillräckligt lika cirklar. Detta begrepp ”tillräckligt lika” måste förstås preciseras, liksom vi bör förklara vad som menas med att kurvans komplement består av två bitar. För att göra det behöver vi topologi – och det ämnet, som kan beskrivas som studiet av öppna mängder, skall vi ta upp i nästa avsnitt.

Beviset för Jordans kurvsats är svårt. För en cirkel är det ju lätt att avgöra om man ligger innanför eller utanför – man mäter bara avståndet till centrum. Men tänk på en kurva som Helge von Kochs snöflingekurva eller någon annan, oregelbunden fraktal. Om man ligger i närheten av den så ser man en mycket taggig kurva och det är omöjligt att avgöra genom att bara titta på punkter i närheten huruvida man ligger utanför eller inuti kurvan. Detsamma gäller för en labyrint (Se figur 3.3.)

3.5. Topologi

Jordans kurvsats gäller inte bara för cirklar utan även för kurvor som liknar cirklar. Vi behöver förklara vad som menas med att en kurva är tillräckligt lik en cirkel. Det matematiska ordet är *homeomorf.* Vi kräver alltså att kurvan skall vara homeomorf med en cirkel. Vad betyder nu det? Kalla kurvan för K och en cirkel för C. Vi kräver

Figur 3.3. När är vi inne i kurvan?

först att det skall finnas en bijektion av C på K. Det betyder att det skall finns en avbildning f från C in i K (detta skriver man kort $f\colon C \to K$) som dessutom är sådan att varje punkt i K är en bildpunkt under f och två olika punkter i C avbildas på två olika punkter på K. Därmed har vi uteslutit åttan, ty för den kurvan gäller att två olika punkter på cirkeln avbildas på samma punkt. Men vi har inte uteslutit en kurva som är ett linjestycke, ty den uppstår genom att man klipper upp cirkeln; vi har en bijektion mellan cirkeln och linjestycket, till exempel genom att vi låter $g(p)$ vara vinkeln som radien genom punkten p bildar med x-axeln, vald så att $0 \leqslant g(p) < 2\pi$ (vi använder alltså polära koordinater). Vi låter så f vara avbildningen $f(p) = (g(p), 0)$. Då har vi en bijektion av cirkeln på en sträcka i planet. Och för en sträcka gäller ju inte satsen: dess komplement består bara av en enda bit. Tydligen fordras det ytterligare någon egenskap hos avbildningen. Denna egenskap är kontinuitet, som vi nu skall försöka förklara. Det vi skall kräva är att f är en bijektion och att både f och dess invers är kontinuerliga. Då är f en *homemorfism.*

Kontinuitet hos en avbildning f innebär att små ändringar hos ursprungspunkten (argumentet) ger upphov till blott små ändringar hos bildpunkten. Några språng får inte förekomma. Alltså: om punkten q ligger nära punkten p så skall bilden $f(q)$ ligga nära bilden $f(p)$. Detta är ett intuitivt talesätt, som måste göras precist och hållbart, dvs. så att det tål olika generaliseringar.

För att definiera kontinuitet behöver vi topologier. Topologi är, kort uttryckt, studiet av öppna mängder. Man säger att man definierat en topologi på en mängd X om man har angivit en familj av delmängder av X med vissa egenskaper. Mängderna som tillhör denna familj kallas *öppna*, men det är blott ett namn – man skulle kunna kalla dem glada eller gula eller gulliga. De öppna mängderna skall uppfylla två axiom (se nedan).

På den reella axeln $\mathbf{R}$ har vi en topologi: man säger att ett intervall är *öppet* om det är av formen

$$]a, b[= \{x \in \mathbf{R}; a < x < b\}.$$

Observera att $]a, a[$, den tomma mängden $\emptyset$, är ett öppet intervall. Och man säger att

en mängd är *öppen* om den är en union av öppna intervall. Det innebär att en mängd av reella tal är öppen om och endast om det för varje punkt b i mängden finns ett litet intervall $[a, c]$ med $a < b < c$ som också ligger i mängden. Det är inte svårt att verifiera att dessa öppna mängder uppfyller de två axiomen. Man kan hitta en oändlig familj av öppna mängder vars snitt inte är öppet.

I $\mathbf{R}^n$, rummet av alla n-tipplar av reella tal, kan man definiera ett *öppet klot* som en mängd av formen

$$B(c, r) = \{x \in \mathbf{R}^n; d(c, x) < r\}.$$

Det är mängden av alla punkter vilkas avstånd till centrum c är strikt mindre än radien r (vi låter här $d(c, x)$ beteckna avståndet mellan c och x; precis vilket avstånd vi tar spelar ingen roll). Sedan definierar vi en *öppen mängd* som en mängd A sådan att den för varje punkt $c \in A$ innehåller hela klotet $B(c, r)$ för någon tillräckligt liten radie r.

Man kan lätt visa att om vi har en godtycklig familj av öppna mängder, så är unionen av dem alla öppen: om $c \in \bigcup U_j = V$, så ligger c i U_j för något index j, och då finns det ett klot $B(c, r)$ som är innehållet i U_j, alltså även i V. Och om vi har en ändlig familj av öppna mängder, så är dess snitt öppet: om $c \in \bigcap U_j = W$, så finns det klot $B(c, r_j)$ innehållna i U_j för varje index j; därför ligger klotet $B(c, r)$ med radie lika med det minsta av talen r_j i W.

En funktion $f\colon \mathbf{R}^n \to \mathbf{R}^m$ kallas *kontinuerlig* om det för varje öppen mängd B i $\mathbf{R}^m$ gäller att mängden A av alla punkter i $\mathbf{R}^n$ som avbildas in i B är öppen. Vi kan nu lätt se att en sådan funktion inte kan göra några språng. Om en funktion är kontinuerlig och avbildar en punkt a på en punkt $b = f(a)$, så tar vi en öppen mängd som innehåller b, till exempel ett klot $B(b, s)$ med en liten radie. Då är mängden av alla punkter x som avbildas in i $B(b, s)$ öppen, vilket betyder att den måste innehålla något klot $B(a, r)$. Bilden av $B(a, r)$ ligger alltså inne i $B(b, s)$, så f kan inte göra något stort språng i $B(a, r)$ (inte större än $2s$). Och eftersom s kan väljas hur litet som helst, så måste eventuella språng hos f bli allt mindre och mindre i de klot $B(a, r)$ som vi får fram genom resonemanget. Därmed har vi visat att kontinuerliga funktioner inte kan ha språng i någon punkt. Denna iakttagelse är så viktig att man har tagit den som utgångspunkt för nya definitioner, nämligen där det inte finns något avstånd.

I stället för att definiera begreppet öppen mängd med hjälp av avstånd som vi nyss gjort tar vi nu i stället en axiomatisk väg. De egenskaper som vi visade tar vi som axiom. Låt alltså X vara en godtycklig mängd, och låt oss ha en familj $\mathcal{U}$ av delmängder av X som vi kallar öppna.

Axiom 1. *Om U_j, $j \in J$, är öppna, där J är en godtycklig indexmängd, så är unionen $\bigcup_{j\in J} U_j$ öppen.*

Detta innebär att om vi har mängder $U_j \in \mathcal{U}$, så tillhör mängden V av alla punkter som ligger i något U_j också $\mathcal{U}$: $V \in \mathcal{U}$.

Axiom 2. *Om U_j, $j \in J$, är öppna, där J är en ändlig indexmängd, så är snittet $\bigcap_{j\in J} U_j$ öppet.*

Detta innebär att om vi har ändligt många mängder $U_j \in \mathcal{U}$, så tillhör mängden W av alla punkter som ligger i alla U_j också i $\mathcal{U}$: $W \in \mathcal{U}$.

Eftersom det är tillåtet att låta J vara tom, så följer av axiom 1 att den tomma mängden är öppen, och av axiom 2 att hela rummet X är öppet, ty $\bigcup_{j\in\emptyset} U_j = \emptyset$ och

$\bigcap_{j\in\emptyset} U_j = X$. (Den som tycker att detta verkar konstigt kan helt enkelt lägga till som ett axiom 3 att $\emptyset$ och X är öppna.)

Om vi har en familj $\mathcal{U}$ av delmängder av X som uppfyller axiomen, så säger vi att vi har en *topologi* på X, och att X med denna topologi är ett *topologiskt rum.* (Något avstånd behöver inte finnas.) Och om vi har två topologiska rum X och Y och en avbildning f av X in i Y, så säger vi att den är *kontinuerlig* om det är så att urbilden $f^{-1}(B) = \{x \in X; f(x) \in B\}$ är öppen för varje öppen mängd B i Y – definitionen är alltså ordagrant densamma som i det fall då vi hade definierat öppna mängder med hjälp av avstånd.

I ett topologiskt rum kan vi nu definiera begreppet sammanhängande mängd. En mängd A kallas *sammanhängande* om den innehåller minst en punkt och om den inte kan delas upp i en viss mening, nämligen så att om vi har två öppna mängder U och V sådana att $U \cup V$ innehåller A och $A \cap U$ och $A \cap V$ inte har någon gemensam punkt, så gäller antingen $A \subset U$ eller $A \subset V$. En *komponent* är en maximal sammanhängande mängd, dvs. en sammanhängande mängd som inte är äkta delmängd av någon sammanhängande mängd. Vi har nu alla begrepp som behövs för att förstå Jordans kurvsats, både den klassiska och den digitala.

Låt oss nu titta på hur den klassiska satsen lyder exakt.

Sats (Jordans kurvsats). *Antag att $f\colon C \to f(C) \subset \mathbf{R}^2$ är en homeomorfism av en cirkel C in i $\mathbf{R}^2$. Dess bild $K = f(C)$ delar planet $\mathbf{R}^2$ i precis två delar: $\mathbf{R}^2 \setminus K$ har precis två komponenter.*

De två orden *homeomorfism* och *komponent* har nu fått sin förklaring. Med en *Jordankurva* menar man just en homeomorf bild av en cirkel.

3.6. Khalimskys topologi

Efim Khalimsky hittade på en topologi för de hela talen $\mathbf{Z}$ som defineras så här: en mängd A av hela tal kallas *öppen* om den för varje jämnt tal $2n \in A$ också innehåller dess två udda grannar $2n-1$ och $2n+1$. Mängden av alla jämna tal är alltså inte öppen, men mängden av alla udda tal är det. Vidare är $\{1\}$ en öppen mängd, medan $\{0\}$ inte är öppen. Den minsta öppna mängd som innehåller $\{0\}$ är $\{-1, 0, 1\}$. Man kan lätt verifiera att de öppna mängderna som vi definierat dem här uppfyller axiomen 1 och 2 i avsnitt 3.5. Det gäller till och med något starkare än axiom 2: snittet av en oändlig familj av öppna mängder är också öppet.

Figur 3.4. Khalimskylinjen.

De jämna och udda talen spelar tydligen helt olika roller i Khalimskys topologi. Vad innebär detta för kontinuiteten hos en avbildning $f\colon \mathbf{Z} \to \mathbf{Z}$? Om vi tar en godtycklig öppen delmängd B av $\mathbf{Z}$, så skall tydligen mängden A av alla punkter n som avbildas in i B vara öppen. Detta innebär att om ett jämnt tal $2n$ tillhör A, dvs. $f(2n) \in B$, så

skall även $f(2n \pm 1) \in B$. Om $f(2n)$ är udda, så kan vi ta $B = \{f(2n)\}$ – det är ju en öppen mängd. Och då måste $f(2n \pm 1) = f(2n)$, dvs. funktionen måste vara konstant i mängden $\{2n-1, 2n, 2n+1\}$. Om däremot $f(2n)$ är ett jämnt tal, så kan vi ta $B = \{f(2n)-1, f(2n), f(2n)+1\}$ och då måste $f(2n \pm 1)$ avbildas in i den mängden, vilket innebär att $f(2n \pm 1)$ kan avvika högst en enhet från $f(2n)$. En kontinuerlig funktion kan alltså aldrig ändra sig snabbare än ett steg för varje steg som argumentet tar, vilket medför att vi alltid har $|f(x) - f(y)| \leqslant |x - y|$ för alla $x, y \in \mathbf{Z}$, men med den extra inskränkningen att den måste vara konstant i tre punkter $2n-1, 2n, 2n+1$ om den råkar ta ett udda värde i en jämn punkt $2n$. Man kan också uttrycka det så att funktionen kan ta olika värden i x och $x+1$, men bara om antingen både x och $f(x)$ är jämna eller om båda är udda. Detta innebär exempelvis att funktionen $f(x) = x + 2$ är kontinuerlig, men inte funktionen $f(x) = x + 1$.

Vi bildar nu planet $\mathbf{Z}^2$, som består av alla par av heltal, och vi kan ge även det en topologi. En delmängd A av $\mathbf{Z}^2$ kallas *öppen* om den för varje pixel $(x_1, x_2) \in A$ också innehåller pixlarna $(x_1 \pm 1, x_2)$ om x_1 är jämnt och $(x_1, x_2 \pm 1)$ om x_2 är jämnt. Det följer av detta att om A är öppen och $(x_1, x_2) \in A$ med både x_1 och x_2 jämna, så måste A innehålla även de åtta punkterna $(x_1 \pm 1, x_2)$, $(x_1, x_2 \pm 1)$, $(x_1 \pm 1, x_2 \pm 1)$. Därmed har vi en topologi i planet $\mathbf{Z}^2$. Vi kallar även den för *Khalimskys topologi.* Det finns nu flera slag av pixlar (x_1, x_2): de där både x_1 och x_2 är jämna; de där båda är udda; och de där en av koordinaterna är udda och den andra är jämn.

Vi kan nu tala om en kontinuerlig avbildning $f\colon \mathbf{Z} \to \mathbf{Z}^2$, där både $\mathbf{Z}$ och $\mathbf{Z}^2$ ges Khalimskys topologi. Men vi behöver också en motsvarighet till cirkeln som vi använde i Jordans sats. En cirkel får man av $\mathbf{Z}$ genom att identifiera något jämnt tal m med noll. Det betyder att man identifierar talet $j + km$ med j för alla $k \in \mathbf{Z}$. Man säger att man räknar *modulo* m. Vi är ju vana att räkna klockslag modulo 12 eller 24. En resa som startar klockan 22 och tar 9 timmar är ju slut klockan 7; additionen lyder $22 + 9 = 7$ (modulo 24). Låt oss med $\mathbf{Z}_m$ beteckna heltalen $\mathbf{Z}$ när man räknar modulo m. Dessa tal kan representeras av talen $0, 1, 2, ..., m-1$. Om man lägger 1 till $m-1$ så får man 0, dvs. man har gått urtavlan runt. (Att räkna modulo ett udda tal är inte bra i detta sammanhang, eftersom de udda och jämna talen spelar olika roller, och om man räknar modulo ett udda tal, så kan distinktionen inte upprätthållas; m identifieras ju med det jämna talet 0.) Låt oss säga att $\mathbf{Z}_m$ är en *Khalimskycirkel.*

Figur 3.5. Khalimskyplanet.

En *digital Jordankurva* är en kontinuerlig avbildning av en Khalimskycirkel $\mathbf{Z}_m$ in i

Khalimskyplanet $\mathbf{Z}^2$ sådan att dess invers existerar och är kontinuerlig. Vi har alltså en homeomorfism $f\colon \mathbf{Z}_m \to f(\mathbf{Z}_m) \subset \mathbf{Z}^2$. Definitionen är densamma som för de klassiska Jordankurvorna. Kurvan består av m punkter i planet $\mathbf{Z}^2$.

Man kan visa att en Jordankurva kan svänga i en punkt endast om punktens bägge koordinater är jämna eller udda, och den kan svänga 45 eller 90 grader, aldrig 135 grader. Om kurvan svänger i en punkt (x_1, x_2) med x_1 udda och x_2 jämn, så kan f inte vara kontinuerlig. Om kurvan svänger 135 grader, så är f^{-1} inte kontinuerlig. (Se figur 3.6.)

Figur 3.6. Några kurvor.

3.7. Jordans kurvsats i det digitala planet

Vi formulerar nu Khalimskys sats.

Sats (Khalimskys sats för digitala Jordankurvor). *Antag att $f\colon \mathbf{Z}_m \to f(\mathbf{Z}_m) \subset \mathbf{Z}^2$ är en homeomorfism av en Khalimskycirkel $\mathbf{Z}_m$ in i det digitala planet $\mathbf{Z}^2$ försett med Khalimskys topologi. Dess bild $K = f(\mathbf{Z}_m)$ delar planet $\mathbf{Z}^2$ i precis två delar: $\mathbf{Z}^2 \setminus K$ har precis två komponenter.*

Satsen ser precis likadan ut som den klassiska Jordans kurvsats i avsnitt 3.5; det är topologierna som är olika.

Jordans klassiska sats är som sagt svår att bevisa. Beviset för den digitala satsen är mycket lättare. Det beror på att en digital Jordankurva bara kan innehålla ändligt många punkter och har en längd som är av formen $p+q\sqrt{2}$, där p och q är icke-negativa heltal. Om vi kortar av kurvan genom att ta en genväg, så måste längden minska till ett annat tal $p_1 + q_1\sqrt{2}$ (p_1, q_1 icke-negativa heltal) och detta kan bara ske ändligt många gånger. Beviset består av att man visar att alla Jordankurvor utom de allra kortaste verkligen kan kortas på ett sådant sätt att den nya kortare kurvan också är en Jordankurva. Efter ändligt många steg har man fått en kurva som är så kort att den inte kan kortas mera. (Se figur 3.7.) Men då är den av en mycket enkel typ, så enkel och kort att man direkt kan verifiera att slutsatsen gäller. Beviset är alltså ett

induktionsbevis som går över kurvans längd. Något liknande är förstås inte möjligt i det klassiska fallet.

Figur 3.7. De minsta Jordankurvorna.

Den digitala kurvsatsen visar att man med ett digitalt staket kan stänga in och stänga ut lika väl som med en vanlig kurva. Prickarna på datorskärmen kan, trots att de bara är ändligt många, bilda ett staket med samma egenskaper som de klassiska Jordankurvorna. Knepet är att byta ut topologin. Då kan man i stället behålla den invanda formuleringen av satsen.

3.8. Litteraturhänvisningar

Halimskiĭ, E. D.
1970 Applications of connected ordered topological spaces in topology. Conference of Math. Departments of Povolsia.

Khalimsky, Efim; Kopperman, Ralph; Meyer Paul R.
1990 Computer graphics and connected topologies on finite ordered sets. *Topology and its Applications* **36**, 1—17.

Kiselman, Christer O.
2000 Digital Jordan curve theorems. *Discrete Geometry for Digital Imagery,* 9th International Conference, DGCI 2000, Uppsala Sweden, December 13—15, 2000. (Eds. Gunilla Borgefors, Ingela Nyström, Gabriella Sanniti di Baja.) Lecture Notes in Computer Sciences **1953**, pp. 46—55. Springer.
2001 *Digital Geometry and Mathematical Morphology.* Föreläsningsanteckningar. Uppsala universitet, Matematiska institutionen.
Tillgänglig på `www.math.uu.se/~kiselman`

Melin, Erik
2003 *Connectedness and continuity in digital spaces with the Khalimsky topology.* Uppsala universitet, Matematiska institutionen, Project Report 2003:9, 43 ss.
Tillgänglig på `www.math.uu.se/~melin`

Rosenfeld, Azriel
1974 Digital straight line segments. *IEEE Transactions on Computers,* **C-23**, No. 12, 1264—1269.

3.9. Övningar

3.1. Vilka av de tre pixelmängderna i figur 3.1 kan anses representera en sträcka (ett segment av en euklidisk rät linje)? Enligt vilka kriterier?

3.2. Ange ett nödvändigt och tillräckligt villkor på heltalen a och b för att intervallet

$$[a, b] \cap \mathbf{Z} = \{x \in \mathbf{Z}; a \leqslant x \leqslant b\}$$

skall vara öppet för Khalimskys topologi.

3.3. Vilka av följande delmängder av $\mathbf{Z}^2$ är öppna för Khalimskytopologin? Ange för varje mängd som inte är öppen den minsta öppna mängd som innehåller den givna mängden.

(a) $\{x \in \mathbf{Z}^2; |x_1|, |x_2| \leqslant 1\}$;

(b) $\{(1, 1), (2, 1), (3, 1)\}$;

(c) $\{(3, 1)\}$;

(d) $\{(1, 0), (2, 0), (3, 0)\}$;

(e) $\{(1, 0)\}$.

3.4. Visa att $\mathbf{Z}$ är sammanhängande för Khalimskytopologin (definition i avsnitt 3.5). Visa att $\mathbf{Z} \setminus \{a\}$ inte är sammanhängande om $a \in \mathbf{Z}$. Visa motsvarande egenskaper för $\mathbf{R}$, dvs. att $\mathbf{R}$ är sammanhängande medan $\mathbf{R} \setminus \{a\}$ inte är det.

3.5. Visa att funktionen $f(x) = x + 3$, $x \in \mathbf{Z}$, inte är kontinuerlig. Visa att funktionen $g(x) = x + 4$, $x \in \mathbf{Z}$, är kontinuerlig.

3.6. Definiera två funktioner $f, g \colon \mathbf{Z} \to \mathbf{Z}$ genom att sätta $f(x) = 0$, $x \leqslant 0$, $f(x) = 1$, $x \geqslant 1$ och $g(x) = 0$, $x \leqslant 1$, $g(x) = 1$, $x \geqslant 2$. Visa att f men inte g är kontinuerlig.

3.7. Definiera en avbildning $f \colon \mathbf{Z} \to \mathbf{Z}^2$ genom att sätta $f(x) = (x, x+1)$, $x \in \mathbf{Z}$. Visa att f inte är kontinuerlig. Visa att f^{-1} som avbildning från $\operatorname{im} f$ till $\mathbf{Z}$ är kontinuerlig. (Tag först reda på vad detta påstående måste betyda. Vilka är de öppna mängderna i $\operatorname{im} f$?)

3.8. Vilka av de fyra kurvorna i figur 3.6 är digitala Jordankurvor (definition i avsnitt 3.6)?

3.9. Var finns Efim Khalimsky nu?

3.10. Låt $f, g \colon \mathbf{Z} \to \mathbf{Z}$ vara två kontinuerliga funktioner och antag att $f(x) < g(x)$ för alla $x \in \mathbf{Z}$. Visa att $f(x + 2) \leqslant g(x)$ för alla $x \in \mathbf{Z}$.

3.11. Betrakta avbildningar $f \colon X \to Y$ där $X = Y = \{0, 1, 2\}$. (a) Hur många sådana avbildningar finns det? (b) Hur många av dessa är bijektioner? (c) Hur många avbildningar som uppfyller $|f(x) - f(x')| \leqslant |x - x'|$ för alla $x, x' \in X$ finns det? (d) Hur många kontinuerliga avbildningar finns det? Samma frågor om $X = \{0, 1, 2\}$ och $Y = \{1, 2, 3\}$. Samma frågor om $X = Y = \{0, 1, 2, 3\}$.

4. Tessellationer av planet

Professor Gunilla Borgefors har författat en uppsats med titeln *Tessellationer – konsten att dela upp planet i regelbundna mönster* (17 sidor). Hon berättade om tessellationer 2003-09-25 15:15.

5. Grupper

5.1. Cykliska grupper

Är det sant att $23 + 8 = 7$? Eller att $11 + 8 = 7$? Om man lägger sig klockan 23 och sover 8 timmar så vaknar man klockan 7. Eller om man startar en resa klockan 11 och reser i 8 timmar, så kommer man fram klockan 7. Vi kan alla räkna med klockslagen: matematiskt kallas det att räkna *modulo 24* respektive *modulo 12*. Man skriver detta så: $23+8 \equiv 7 \pmod{24}$ och $11+8 \equiv 7 \pmod{12}$ (utläses "elva plus åtta är kongruent med sju modulo tolv"). Utrycket *modulo 24* betyder att likhet gäller efter addition av någon multipel av 24.

Man har en mängd $\mathbf{Z}_{24}$ som består av talen $\{0, 1, 2, 3, ..., 23\}$ och där man sedan identifierar 24 med 0, 25 med 1 och så vidare. På samma sätt har vi till exempel $\mathbf{Z}_4$, som är heltalen modulo 4, och består av $\{0, 1, 2, 3\}$, där man sedan identifierar $3+1 = 4$ med 0, 5 med 1 osv. Att räkna modulo 12 är alltså det man gör när man lär sig att läsa av en klocka (med visare; en digital klocka går direkt på siffrorna och är därför inte lika instruktiv i detta sammanhang).

Vi kan lätt göra en additionstabell för $\mathbf{Z}_4$:

$\mathbf{Z}_4$	0	1	2	3
0	0	1	2	3
1	1	2	3	0
2	2	3	0	1
3	3	0	1	2

Vi kan notera att 0 är ett neutralt element: $x + 0 = x$ för alla $x \in \mathbf{Z}_4$. Vidare finns till varje x ett element y så att $x + y = 0$, dvs. y är en invers till x. Att det är så kan man läsa av från additionstabellen. Slutligen är det så att $x + y = y + x$; det kan likaledes läsas av från tabellen (spegla den i diagonalen).

På liknande sätt kan man göra tabeller för $\mathbf{Z}_m$, $m = 1, 2, 3, 4, 5, ...$, i varje fall om m inte är alltför stort. För de minsta talen, $m = 1, 2, 3$, blir det:

$\mathbf{Z}_1$	0
0	0

$\mathbf{Z}_2$	0	1
0	0	1
1	1	0

$\mathbf{Z}_3$	0	1	2
0	0	1	2
1	1	2	0
2	2	0	1

Vi kallar dessa $\mathbf{Z}_m$ för *cykliska grupper* – definition senare.

5.2. Symmetriska grupper

Vi tittar nu på ett annat exempel, mängden av alla bijektioner av mängden $\{1, 2, 3\}$ av tre element. Dessa bijektioner kan sammansättas: om f och g är två bijektioner (permutationer) så är deras sammansättning $g \circ f$ (f följd av g) också en bijektion. Den definieras av att $(g \circ f)(j) = g(f(j))$ för $j = 1, 2, 3$. Vi skriver nu bara gf för $g \circ f$ och uppfattar operationen som en multiplikation. Är den associativ, dvs. gäller $h(gf) = (hg)f$? Ja, ty både $(h \circ (g \circ f))(j))$ och $((h \circ g) \circ f)(j)$ blir $h(g(f(j)))$ om man använder definitionen av sammansättningen.

Vi kan nu definiera a som permutationen $1 \mapsto 2, 2 \mapsto 1, 3 \mapsto 3$, alltså den som kastar om 1 och 2 men lämnar 3 fix. Vidare kan vi definiera b som permutationen

$1 \mapsto 1, 2 \mapsto 3, 3 \mapsto 2$, alltså den som kastar om 2 och 3 men lämnar 1 fix. Vad blir nu ab? Jo, den blir $1 \mapsto 1 \mapsto 2, 2 \mapsto 3 \mapsto 3, 3 \mapsto 2 \mapsto 1$. Vidare blir ba avbildningen $1 \mapsto 2 \mapsto 3, 2 \mapsto 1 \mapsto 1, 3 \mapsto 3 \mapsto 2$. Vi ser att $ab \neq ba$, ty vi har ju till exempel att $(ab)(1) = 2$ medan $(ba)(1) = 3$. Vi har nu ett nytt fenomen: sammansättningen är inte kommutativ. Vi vet att det finns $3! = 6$ olika permutationer, så denna mängd, som kallas S_3, har sex element. Det finns tydligen ett neutralt element, den permutation som inte förändrar något alls, låt oss kalla den e, alltså $e(k) = k$ för $k = 1, 2, 3$. Vi vet alltså att $ex = xe$ för alla permutationer x. Vi kan nu göra en tabell, som vi denna gång kallar för multiplikationstabell, för S_3. Vi har redan hittat fem element, nämligen e, a, b, ab och ba, och vi har sett att de alla är olika. Det måste alltså finnas ett sjätte element. Vi kan ju bilda aba till exempel, och vi ser att det är avbildningen $1 \mapsto 3, 2 \mapsto 2, 3 \mapsto 1$ som är skild från de fem andra. Vidare ser vi att $aa = e$ och $bb = e$. Så vi kan börja att fylla i tabellen så här:

S_3	e	a	b	ab	ba	aba
e	e	a	b	ab	ba	aba
a	a	e	ab	b	aba	ba
b	b	ba	e		a	
ab	ab	aba	a			
ba	ba	b				
aba	aba					e

Man ser att den första och andra raden är lika. Det beror förstås på att $ex = x$ för alla element x. Vidare är den första och andra kolonnen lika eftersom $xe = x$ för alla x. (När man blir mera van kan man alltså stryka den första raden och den första kolonnen och bara behålla matrisen med de 6×6 produkterna.)

Vi försöker nu fylla i tabellen mera. Vi ser till exempel att $(ab)(ba) = a(bb)a = aea = aa = e$; på samma sätt $(ba)(ab) = b(aa)b = bb = e$. Med några andra liknande lätta räkningar får vi tabellen:

S_3	e	a	b	ab	ba	aba
e	e	a	b	ab	ba	aba
a	a	e	ab	b	aba	ba
b	b	ba	e		a	
ab	ab	aba	a		e	
ba	ba	b		e		a
aba	aba	ab		a		e

Vi ser nu att det fattas två element i vissa rader och kolonner. Exempelvis fattas ba och aba i kolonnen för b. Är alltså $(ba)b$ lika med ba eller aba? Om $(ba)b$ vore lika med ba så skulle $(ab)(ba)b = b$ vara lika med $(ab)ba = e$, vilket ju inte är sant. Den andra möjligheten, alltså att $(ba)b = aba$ är alltså den enda möjliga. Vi har därmed också räknat ut $b(ab)$ och kan fylla i två platser till:

S_3	e	a	b	ab	ba	aba
e	e	a	b	ab	ba	aba
a	a	e	ab	b	aba	ba
b	b	ba	e	aba	a	
ab	ab	aba	a		e	
ba	ba	b	aba	e		a
aba	aba	ab		a		e

Nu ser vi att det till exempel i raden för b fattas bara ett enda element, som måste vara ab, eftersom det är det enda som inte står i denna rad. Det som nu fattas i den sista kolonnen måste vara b. Så kan man nu fortsätta. Vi använder oss alltså av egenskapen att varje rad och varje kolonn måste innehålla alla element i S_3. Genom att resonera så kan vi nu fylla i hela tabellen, multiplikationstabellen för S_3:

S_3	e	a	b	ab	ba	aba
e	e	a	b	ab	ba	aba
a	a	e	ab	b	aba	ba
b	b	ba	e	aba	a	ab
ab	ab	aba	a	ba	e	b
ba	ba	b	aba	e	ab	a
aba	aba	ab	ba	a	b	e

Naturligtvis kan man räkna ut alla permutationer explicit, men vi har varit litet lata och räknat så litet som möjligt.

Man kallar S_3 för en grupp, den *symmetriska gruppen av permutationer av tre element*[6] och vi har sett att den är icke-kommutativ. När är $xy \neq yx$? Vi ser att vi alltid måste ha $xy = yx$ om $x = e$ eller $y = e$ eller $x = y$. Sådan par kommuterar alltid. Det återstår då tio två-mängder som inte kommuterar automatiskt, nämligen $\{a, b\}$, $\{a, ab\}$, $\{a, ba\}$, $\{a, aba\}$, $\{b, ab\}$, $\{b, ba\}$, $\{b, aba\}$, $\{ab, ba\}$, $\{ab, aba\}$, $\{ba, aba\}$. Av dessa tio två-mängder kommuterar endast en, $(ab)(ba) = (ba)(ab) = e$. Så statistiken säger att av de tio oordnade par som inte kommuterar automatiskt, så kommuterar endast ett, dvs. graden av icke-kommutativitet är 90%.

Man kan på samma sätt studera den symmetriska gruppen av permutationer av n element. Exempelvis har S_4 24 element. Man ser att tabellen för S_1 är densamma som för $\mathbf{Z}_1$ och den för S_2 är densamma som för $\mathbf{Z}_2$. Man behöver bara byta ut addition mot sammansättning. Vi säger att S_n är *isomorf* med $\mathbf{Z}_n$ för $n = 1, 2$. Dessa grupper är kommutativa; alla andra S_n från och med S_3 är icke-kommutativa.

5.3. Definition av en grupp

Vi har sett några exempel på grupper. Det är dags att defininiera dem explicit.

Definition. *En* grupp *är en icke-tom mängd G med en operation $G \times G \to G$ som är associativ (dvs. $x(yz) = (xy)z$ för alla $x, y, z \in G$); som har ett neutralt element e (dvs. $ex = xe = x$ för alla $x \in G$); och där varje element har en invers (dvs. för alla $x \in G$ finns ett $y \in G$ sådant att $xy = yx = e$).*

[6]Det kan förefalla konstigt att kalla denna grupp för symmetrisk eftersom den inte är särskilt symmetrisk. Förklaringen ligger i att permutationen lämnar de symmetriska polynomen av tre variabler invarianta. Den hör alltså ihop med dessa polynom.

(Man kan visa att det räcker att kräva att $ex = x$ och att det finns en vänsterinvers y till x (dvs. $yx = e$). Att $xe = x$ och att $xy = e$ följer då.)

En grupp kan vara *kommutativ* (synonym *abelsk*), nämligen om $xy = yx$ för alla x och y, eller icke-kommutativ (synonym icke-abelsk) om det finns x och y med $xy \neq yx$.

En grupp kallas *cyklisk* om det finns ett element a sådant att mängden av alla element $..., a^{-1}a^{-1}a^{-1}a^{-1}, a^{-1}a^{-1}a^{-1}, a^{-1}a^{-1}, a^{-1}, e, a, aa, aaa, aaaa, ...$ är lika med hela gruppen. Man säger att a *genererar* gruppen. Grupperna $\mathbf{Z}$ och $\mathbf{Z}_m$ är cykliska.

De minsta grupperna är $\mathbf{Z}_1 = \{0\}$, $\mathbf{Z}_2 = \{0, 1\}$, där $1 + 1 = 0$, $\mathbf{Z}_3 = \{0, 1, 2\}$. Dessa är de enda med högst tre element. Det finns två grupper med fyra element, nämligen $\mathbf{Z}_4$, som vi redan studerat, och en grupp V_4 som studeras i en övning. Det finns endast en grupp med fem element, $\mathbf{Z}_5$. Det finns två med sex element, $\mathbf{Z}_6$ och S_3; den sistnämnda är den minsta icke-abelska gruppen. Antalet element i en grupp kallas med en klassisk term för gruppens *ordning*. Det finns endast en grupp av ordning sju ($\mathbf{Z}_7$); men fem av ordning åtta, varav tre abelska och två icke-abelska. Av ordningarna 48 och 64 finns det flera: 52 grupper, varav 5 abelska och 47 icke-abelska, har ordning 48; medan det av ordning 64 finns 267 grupper, 11 abelska och 256 icke-abelska.

5.4. Gruppelement som ord

Vi kan uppfatta S_3 som mängden av alla ord som kan skrivas med två bokstäver a och b, alltså alla ord $a, b, ab, ba, aba, baaaab, aaabbababababa$ och så vidare. Vi måste också tillåta det tomma ordet. Det är inte så lämpligt att skriva det tomma ordet som ett tomrum, ty tomrummet behövs för ett annat syfte: det är redan upptaget som en viktig beteckning för ordmellanrum. Alltså hittar vi på någon annan beteckning för det, till exempel e.

I gruppen finns dessutom förenklingsregler, till exempel att $aa = e$ och $bb = e$. Genom att använda dessa regler kan man förenkla varje ord som innehåller upprepningar. Exempelvis kan $abaabbbababababababbbbb$ förenklas till $bababababa$. Vi behöver alltså endast studera ord som innehåller a och b alternerande. Men det räcker inte. Vi har också en regel $bab = aba$, som vi kan avläsa från multiplikationstabellen. Man ser nu lätt att varje ord med mer än tre bokstäver kan förenklas till ett ord med högst tre bokstäver. Exempelvis blir $abab = a(bab) = a(aba) = ba$ och $baba = (bab)a = (aba)a = ab$. Vi ser att alla ord bildade av a och b kan förenklas till ett av de sex orden e, a, b, ab, ba, aba.

Detta betyder att vi kan betrakta gruppen S_3 som ett språk där varje ord bildas av två bokstäver a, b och där vi har en synonymordbok med tre regler: $aa = e$, $bb = e$, $bab = aba$.

På samma sätt kan varje grupp uppfattas som ett skriftspråk med ett antal bokstäver tillsammans med en synonymordbok med ett antal regler som talar om vilka ord som betyder samma sak. För S_3 ser vi att det räcker med två bokstäver och tre regler.

Ibland är detta synsätt inte så fruktbart, till exempel för $\mathbf{Z}$. Där består alfabetet av alla heltal och synonymordboken av alla regler $ab = c$ med $a + b = c$. Därmed kan varje ord förenklas till ett ord som består av högst en bokstav. Det blir många bokstäver, många ord och många regler, men varje ord är mycket kort.

Låt oss titta igen på S_3 och först göra en multiplikationstabell för orden e, a, b, ab, ba, aba, bab utan att använda synonymordboken:

S_3	e	a	b	ab	ba	aba	bab
e	e	a	b	ab	ba	aba	bab
a	a	aa	ab	aab	aba	$aaba$	$abab$
b	b	ba	bb	bab	bba	$baba$	$bbab$
ab	ab	aba	abb	$abab$	$abba$	$ababa$	$abbab$
ba	ba	baa	bab	$baab$	$baba$	$baaba$	$babab$
aba	aba	$abaa$	$abab$	$abaab$	$ababa$	$abaaba$	$ababab$
bab	bab	$baba$	$babb$	$babab$	$babba$	$bababa$	$babbab$

Låt oss sedan förenkla denna genom att använda reglerna $aa = e$ och $bb = e$. Då blir det:

S_3	e	a	b	ab	ba	aba	bab
e	e	a	b	ab	ba	aba	bab
a	a	e	ab	b	aba	ba	$abab$
b	b	ba	e	bab	a	$baba$	ab
ab	ab	aba	a	$abab$	e	$ababa$	b
ba	ba	b	bab	e	$baba$	a	$babab$
aba	aba	ab	$abab$	a	$ababa$	e	$ababab$
bab	bab	$baba$	ba	$babab$	b	$bababa$	e

Till slut inför vi regeln $bab = aba$ och kan förenkla ytterligare:

S_3	e	a	b	ab	ba	aba	aba
e	e	a	b	ab	ba	aba	aba
a	a	e	ab	b	aba	ba	ba
b	b	ba	e	aba	a	ab	ab
ab	ab	aba	a	ba	e	b	b
ba	ba	b	aba	e	ab	a	a
aba	aba	ab	ba	a	b	e	e
aba	aba	ab	ba	a	b	e	e

Vi ser nu att den sista raden och den sista kolonnen är överflödiga. Om vi tar bort dessa, så får vi den tidigare tabellen för gruppen.

Det vi sagt här om ord och förenklingar av ord gäller också halvgrupper. En halvgrupp är en generalisering av begreppet grupp, där man inte kräver att det skall finnas inverser. Den formella definitionen lyder som följer.

Definition. *En* halvgrupp *är en icke-tom mängd G med en operation $G \times G \to G$ som är associativ (dvs. $x(yz) = (xy)z$ för alla $x, y, z \in G$).*

Eventuellt kan det finnas ett neutralt element, och vissa element kan ha inverser.

Vi skall nu titta på en halvgrupp som också kan betraktas som en uppsättning ord, men med svårare synonymregler.

5.5. Conways trassel

Vi tittar på en halvgrupp som består av alla ord som kan skrivas med två bokstäver S och V och med vissa, nu mer komplicerade, förenklingsregler. Vi skall definiera S och

Figur 5.1. Ett icke tilltrasslat rep-par.

V som operationer på tilltrasslingar av två rep. Utgångsläget är ett par av rep som inte alls är tilltrasslat (se figur 5.1).

Två ord skall betraktas som synonyma om deras resultat på ett icke-tilltrasslat rep-par är desamma.

Operationen S (snurra!) skall vara att rotera ett trassel 90° i negativ led (se figur 5.2).

Figur 5.2. Operationen S (snurra!).

Operationen V (vrid!) skall vara att de båda ändarna till höger byter plats, så att den hitersta passerar över den bortersta (se figur 5.3).

Figur 5.3. Operationen V (vrid!).

Halvgruppen består alltså av alla ord som $VSVSVVVVSVSVVSVVSSVVSS$, plus det tomma ordet e. Man ser lätt att $SS = e$, ty om man utför två snurrningar betraktar vi trasslet som uppstår som lika med det vi startade med. Att T-et i figur 5.2 kommer upp-och-ned bryr vi oss alltså inte om. Två S efter varandra kan alltså strykas; med andra ord har S en invers, $S^{-1} = S$. Men man får oändligt många ord V, VV, VVV, $VVVV$,... ty dessa vridningar gör att repen blir mer och mer tvinnade om varandra.

Vilka regler utom $SS = e$ finns det? Låt oss titta på ordet VSV. Om vi successivt utför operationerna V, S, V verkande på ett icke tilltrasslat rep-par så får vi trasslet i figur 5.4, dvs. vi återgår till det icke tilltrasslade paret. (Detta innebär dock inte att V är en invers till SV, ty vi har bara visat att VSV verkande på ett icke tilltrasslat

rep-par ger ett icke-tilltrasslat rep-par. Det finns andra trassel som inte återförs till samma av VSV.)

Figur 5.4. Operationerna V, SV, VSV verkande på ett rep-par.

Vi gör nu samma sak med ordet $VVSVSVV$ (se figur 5.5).

Figur 5.5. Operationerna V, VV, SVV, $VSVV$, $SVSVV$, $VSVSVV$, $VVSVSVV$. Man ser att den sista tilltrasslingen kan lösas upp. Motsvarande följd av rationella tal blir $0 \mapsto 1 \mapsto 2 \mapsto -\frac{1}{2} \mapsto \frac{1}{2} \mapsto -2 \mapsto -1 \mapsto 0$.

Vi skall också skapa en halvgrupp av operationer på de rationella talen $\mathbf{Q}$ med ett extra element, som vi kallar oändligheten och skriver ∞. Vi definierar två operationer s och v genom

$$s(x) = -1/x, \qquad v(x) = x + 1, \qquad x \in \mathbf{Q} \cup \{\infty\}.$$

Speciellt sätter vi $s(0) = -1/0 = \infty$, $s(\infty) = -1/\infty = 0$ och $v(\infty) = \infty + 1 = \infty$. Då får vi en halvgrupp av rationella funktioner genom upprepade sammansättning av dessa två funktioner. En sådan funktion kan till exempel vara vsv; den är alltså $(vsv)(x) = (-1/(x+1)) + 1 = x/(x+1)$. Vi observerar att den tar värdet 0 i origo: $(vsv)(0) = 0$. (Men den är inte identiskt noll.)

Nu gäller en sats om detta, som talar om hur synonymordboken ser ut när det gäller verkan på ett icke tilltrasslat rep-par.

Sats 5.1. (John H. Conway). *En sammansättning av operationerna S och V på ett icke tilltrasslat par av rep ger ett icke tilltrasslat par om och endast om motsvarande sammansättning av funktionerna s och v är noll i origo.*

Denna sats formulerades av Conway (1970) utan bevis. Ett bevis som förutsätter avancerade kunskaper i knutteori publicerades av Burde & Zieschang (1986). Ett någorlunda elementärt bevis publicerades av Goldman & Kauffman (1997).

Sats 5.2. *Givet ett godtyckligt element $x \in \mathbf{Q} \cup \{\infty\}$ så finns det en ändlig sammansättning f av funktionerna s och v sådan att $f(0) = x$.*

Alla rationella tal, liksom ∞, kan alltså nås från 0 med hjälp av funktionerna s och v. Det innebär att alla rationella tal och ∞ kan förekomma som värden för ett trassel.

Sats 5.3. *Givet ett godtyckligt element $x \in \mathbf{Q} \cup \{\infty\}$ så finns det en ändlig sammansättning g av funktionerna s och v sådan att $g(x) = 0$.*

I betraktande av sats 5.1 innebär detta resultat att varje trassel som bildats med operatorerna S och V kan lösas upp av en sammansättning av samma operatorer.

5.6. Litteraturhänvisningar

Burde, G.,; Zieschang, H.
1986 *Knots.* Berlin: de Gruyter.

Conway, John H.
1970 An enumeration of knots and links, and some of their algebraic properties. *I: Computational Problems in Abstract Algebra* (D. Welsh, red.).

Goldman, Jay R.; Kauffman, Louis H.
1997 Rational tangles. *Advances in Applied Mathematics* **18**, 300—332.

5.7. Övningar

5.1. Lös ekvationerna $x^2 = e$ och $y^3 = e$ i S_3.

5.2. Låt V_4 vara mängden $\{e, a, b, c\}$ med reglerna $ea = a$, $eb = b$, $ec = c$, $a^2 = b^2 = c^2 = e$. Visa att det finns ett och endast ett sätt att komplettera dessa regler så att V_4 blir en grupp. Är den kommutativ? Visa att den inte är isomorf med $\mathbf{Z}_4$. (*Ledning:* Hur många lösningar finns det till ekvationen $x^2 = e$?) Det finns således minst två grupper med fyra element (i själva verket exakt två).

5.3. Hur många grupper av ordning sex finns det? Vi har sett att det finns minst två, $\mathbf{Z}_6$ och S_3. Visa att det inte finns flera genom att visa att det inte finns några andra multiplikationstabeller. (*Ledning:* Dela upp i fall efter antalet lösningar till ekvationen $x^m = e$ för $m = 2, 3, 6$.)

5.4. Bevisa sats 5.3.

5.5. Bevisa sats 5.2. *Ledning:* Euklides' algoritm.

5.6. Gör praktiska övningar som visar att VSV, $VVSVSVV$ och $VVVSVSVVSVV$ återger det icke tilltrasslade paret.

5.7. Gör ännu mer övningar genom att starta från noll och så tillämpa s och v på de rationella tal som uppstår samtidigt som S och V tillämpas på rep-paret. Genom att räkna på de tal som uppstår och försöka komma tillbaka till noll kan man trassla ut ett svårt tilltrasslat rep-par. I själva verket kan man lära sig en algoritm för att lösa upp trassel.