

Välj specialarbete i matematik

Idéer till specialarbeten i matematik för gymnasister

Samlade som en del av FRN-projektet
Information om högskolan i gymnasiet

Projektledare: Dan Laksov

$$\frac{\pi}{2} = \frac{2}{1} \cdot \frac{2}{3} \cdot \frac{4}{3} \cdot \frac{4}{5} \cdot \frac{6}{5} \cdot \frac{6}{7} \cdots \frac{2n}{2n-1} \cdot \frac{2n}{2n+1} \cdot \frac{\int_0^{\pi/2} \sin^{2n} x \, dx}{\int_0^{\pi/2} \sin^{2n+1} x \, dx}$$

$$e = 2 + \frac{1}{1} + \frac{1}{2} + \frac{1}{1} + \frac{1}{1} + \frac{1}{4} + \frac{1}{1} + \frac{1}{1} + \frac{1}{6} + \frac{1}{1} + \frac{1}{1} + \frac{1}{8} + \dots$$

$$(\text{Antalet primtal } p \text{ med } p \leq x) \sim \frac{x}{\log x}$$

$$\frac{a_1 + a_2 + \dots + a_{2^m}}{2^m} \geq (a_1 a_2 \dots a_{2^m})^{\frac{1}{2^m}} \geq \frac{2^m}{\frac{1}{a_1} + \frac{1}{a_2} + \dots + \frac{1}{a_{2^m}}}$$

Förord till nätversionen.

Tack vare ett generöst bidrag från *Forskningsrådsnämnden* kunde boken *Välj specialarbete i matematik* delas ut gratis till biblioteken vid alla gymnasier i Sverige med NV-program. Restupplagan skickades gratis till alla intresserade gymnasielärare. Upplagan tog därför snabbt slut.

Det har nu i många år varit en ständig efterfrågan efter en ny upplaga. En tämligen otillfredsställande utgåva utan figurer har funnits tillgänglig på nätet, men många har efterfrågat en nyutgåva.

Det var Anders Thorup i Köpenhamn som föreslog att en nätutgåva, som var så lik originalet som möjligt, borde framställas. Han tog på sig den svåra och arbetskrävande uppgiften att överföra de ursprungliga filerna till ett format som passar för nätet. Speciellt svårt var det att omarbeta de många figurerna. Thorup har ritat dessa "för hand" i ett program han själv har skrivit, vilket gör att figurerna och programmet ligger i själva filen. Slutresultatet har blivit mycket tillfredsställande och så identiskt originalet som man kan önska.

Vi vill rikta ett varmt tack till Anders Thorup och *Stiftelsen för Vetenskaplig Forskning och Utbildning i Matematik* som har bidragit till framställningen av nätutgåvan.

För att få nätversionen så nära originalet som möjligt har vi tagit med introduktionerna till den ursprungliga utgåvan. Läsaren måste därför ta i beaktande att en del av informationen i nyutgåvan inte längre är aktuell. Till exempel kan inte längre exemplar beställas via Institut Mittag-Leffler, och författarnas adresslista är ej uppdaterad.

Dan Laksov

Stockholm 25/8-2004

Välj specialarbete i matematik

Idéer till specialarbeten i matematik för gymnasister. Samlade som en del av FRN-projektet *Information om högskolan i gymnasiet*.

Projektledare

Dan Laksov

Detta häfte är avsett för matematikintresserade gymnasister som letar efter ämnen för sina specialarbeten. Syftet är att få eleverna intresserade av matematik och ge dem en uppfattning om några av de många fascinerande ämnen och problemställningar de kommer att möta, om de fortsätter att studera matematik vid universitet eller högskola. Tack vare att projektet har fått stöd av Forskningsrådsnämnden (FRN) har det varit möjligt att distribuera denna utgåva kostnadsfritt till alla gymnasier i landet och till en rad privatpersoner. Vi hoppas att samlingen skall nå alla elever på alla gymnasier med N-linje i Sverige och att den skall finnas tillgänglig på alla skolbibliotek och lärarrum på dessa gymnasier. Ytterligare exemplar kan beställas på Institut Mittag-Leffler.

Beställningsadress:

Institut Mittag-Leffler
Auravägen 17
S-182 62 DJURSHOLM

Sättning T_EX, computer modern roman
ISBN: 91-7170-851-0
THD AB
BANDHAGEN 1989

Till dig som väljer specialarbete i matematik.

Avsikten med denna samling är att hjälpa dig som är intresserad av matematik att finna ett tema för ditt specialarbete. Vi hoppas att du, bara genom att läsa igenom titlarna eller de kortfattade inledningarna till uppgifterna, skall finna någon av dem spännande och få lust att arbeta vidare med ämnet. Oavsett vilken av uppgifterna du väljer, är avsikten att häftet skall innehålla så rikligt med material och vägledning, att du skall kunna skriva ett specialarbete efter en arbetsinsats på 80–100 timmar, förutsatt att du behärskar andra årskursen i matematik. De flesta ämnena innehåller en rad olika uppgifter, som kan vara av mycket växlande svårighetsgrad. Dessa uppgifter är avsedda att vara inspirationskällor för självständigt arbete snarare än problem som bör lösas ett efter ett i den följd som de står. Ett specialarbete med utgångspunkt från ett ämne i detta häfte kan bestå av att du:

- * löser några av de givna uppgifterna
- * ger en självständig framställning av delar av ämnet
- * fördjupar dig i en mindre del av ämnet
- * diskuterar relationen mellan ämnet och delar av årskursen
- * använder materialet som bakgrund för egna ämnen
- * utreder begrepp och definitioner som förekommer i ämnet
- * gör datorexperiment som bekräftar resultaten i ämnet.

Detta är bara några av många möjligheter och vi har nämnt dessa för att framhålla att det *långt ifrån är nödvändigt att lösa alla uppgifter i ett ämne för att ha fullgjort specialarbetet*. De flesta ämnena innehåller material för många specialarbeten och lämpar sig utmärkt för grupparbete.

Vill du gå vidare med delar av ett ämne och kanske behöver ytter-

ligare material eller förklaringar, är det meningen att du skall kontakta en högskolelärare och då helst den som har skrivit uppgiften. *Var inte rädd för att ringa eller skriva* (telefonnummer och adress finns längst bak i häftet); läraren är minst lika intresserad av matematik som du är och vill gärna veta vad arbetet med hans ämne har resulterat i.

Valet av ämnen i samlingen har naturligtvis begränsats av kursplanen i matematik på gymnasiet. Av den anledningen är en oproportionerligt stor del av uppgifterna inom områdena kombinatorik och elementär talteori, där man behöver förhållandevis små förkunskaper för att förstå problemställningarna. Urvalet ger således en skev bild av den rika och fascinerande flora av matematiska områden som du möter på en högskola. Vi har emellertid gjort ett allvarligt försök att, innanför den givna ramen, välja ämnen som ger en realistisk föreställning om vilken typ av arbetsmetoder och problemställningar som du kan komma i kontakt med under fortsatta högskolestudier.

Vårt hopp är att du skall finna matematiken lika spännande, viktig och nyttig som författarna gör och att specialarbetet skall uppmuntra dig till vidare matematikstudier vid någon av landets högskolor, när du är klar med gymnasiet.

Välkommen!

Lärarhandledning.

Denna samling av ämnen till *Specialarbeten i matematik* för N-linjen på gymnasiet är främst avsedd som en handledning för eleverna. Tanken är att en elev med intresse för matematik, vid en snabb genomläsning av samlingen skall bli inspirerad av något av de föreslagna ämnena, och, med den vägledning som finns i uppgiften, på egen hand kunna göra färdigt sitt specialarbete. Eleverna kan också arbeta i grupp med ett ämne och skriva var sin version, eller del av uppgiften.

Gymnasielärarens uppgift skall vara av mer uppmuntrande och vägledande karaktär. Den kan bestå i hjälp med att hitta referenserna i uppgiften eller med att köpa nödvändig litteratur till skolbiblioteket, fördelning av arbetsuppgifter för en grupp elever, förmedling av kontakt mellan eleven och högskolan om detta är nödvändigt, o.s.v. Om läraren själv blir fascinerad av ämnet och har tid att delta aktivt, är detta naturligtvis den idealiska situationen. Det är emellertid orimligt att begära att en lärare skall behärska alla de olika ämnen som finns representerade i samlingen, eller att läraren utöver sin vanliga arbetsbörda skall ha tid att grundligt sätta sig in i de problem, som eleverna väljer att arbeta med.

Vi hoppas att matematiklärarna på gymnasiet skall uppleva detta häfte som ett stöd i undervisningen och inte som en börda i form av extra arbete och att det kan hjälpa dem att inspirera matematikintresserade elever till att välja specialarbete i matematik och kanske också fortsätta att studera matematik vid någon högskolan efter avslutad skolgång.

Prosjektlederens forord

Interessen for matematikk vekkes ofte tidlig under gymnasietiden. Det kan derfor være av avgjørende betydning at elevene får bra og inspirerende undervisning og får et innblikk i den mangesidige og spennende verden som matematikken utgjør. En viktig del av denne undervisningen er spesialarbeidet som elevene skal skrive i løpet av det tredje året i gymnaset. Prosjektlederen vil med dette takke de kollegene som har innsett betydningen av spesialarbeidet og som har bidratt til samlingen. Uten deres entusiasme ville prosjektet ikke kunne vært gjennomført.

Samtidig vil jeg takke dem som har hjulpet med framstillingen av heftet. Spesielt vil jeg nevne Karin Lindberg ved Institut Mittag-Leffler, som har stått for en stor del av ord- og tekst- behandlingen med $\text{T}_{\text{E}}\text{X}$ og for redigeringen av materialet. Uten hennes innsats og dyktighet ville arbeidsbyrden med framstillingen blitt overveldende. Birgitta Krasse ved KTH har tilpasset forfatternes illustrasjoner til formatet i heftet og har også produsert illustrasjoner der slike ikke fantes. Samarbeidet med begge har vært en fornøyelse. Roswita Graham har vært til hjelp med formateringen av materialet.

Til slutt vil jeg takke Forskningsrådsnämnden som har gjort dette heftet mulig, og Institut Mittag Leffler og Matematiska Institutionen ved KTH, som på ulike måter har bidratt med nødvendige tjenester som kopiering, posthåndtering og bruk av datautstyr.

Innehåll

1	Två formler för talet π	Leif Abrahamsson
10	En trafikmodell	Leif Arkeryd
17	Om rättvisa val	Leif Arkeryd
23	Vinkeln 60 grader kan inte tredelas med enbart passare och linjal	Jöran Bergh
27	Träd och koder	Anders Björner
41	Iteration av kvadratiska polynom	Lennart Carleson
46	Om $\int_0^1 \frac{dx}{1+x^2}$	Lennart Carleson
52	Konvexa funktioner	Urban Cegrell
56	Om Pythagoras hade varit taxichaufför i Luleå	Andrejs Dunkels
64	Resträkning och ekvationer	Torsten Ekedahl
75	Polyedrar och polygoner	Ralf Fröberg
79	Lotto, ett skicklighetsspel!	Jan Grandell
85	Något om algebraiska kurvor	Björn Gustafsson
90	Något om metriker	Björn Gustafsson
94	Möbiusgruppen och icke-euklidisk geometri	Lars Gårding
105	Något om permutationer	Lars Holst
110	Att dela en hemlighet	Johan Håstad
118	Generering av pseudoslumptal	Johan Håstad
127	Offentlig kryptering	Johan Håstad
137	Felrättande koder	Thomas Höglund
146	Euler-Mac Laurins summationsformel och Bernoulliska polynom	Lars Hörmander
156	Reflektionsprincipen	Dag Jonsson
162	Antikens universum	Sten Kaijser
173	Pythagoreiska trianglar	Sten Kaijser
195	Gaussiska primtal	Christer Kiselman

203	Konvexitet i komplexa planet	Bo Kjellberg
208	Genererande funktioner	Göran Kjellberg
215	Något om differenser	Göran Kjellberg
224	Om Möbiustransformationer	Torbjörn Kolsrud
230	Myntveksling	Dan Laksov
234	Fibonaccis talföljd	Bernt Lindström
238	Permutationer med paritet	Bernt Lindström
246	Matrisavbildningar	Kirsti Mattila
252	Volymer av n -dimensionella klot	Mikael Passare
263	Mönster	Johan Philip
286	Gör Din egen kurvkatalog	Hans Riesel
292	Kvadratrotter och kedjebråk	Hans Riesel
300	Periodiska decimalbråk	Hans Riesel
306	Summan av två heltalskvadrater	Hans Riesel
313	Stokastisk geometri	Lennart Råde
318	Frankering og computer-nettverk	Øystein J. Rødseth
325	Kurvlängd och geometri på en sfärisk yta	Peter Sjögren
333	Kampen om sista stickan	Krister Svanberg
337	Ett belysande exempel	Lasse Svensson
342	Explorativ dataanalys (EDA)	Kerstin Vännman
344	Fraktaler och iteration av funktioner	Hans Wallin
350	Något om medelvärden	Pepe Winkler
356	Adresslista

Två formler för talet π

LEIF ABRAHAMSSON

Uppsala Universitet

Denna uppgift syftar till att härleda två formler för talet π . De två formlernas härledning är oberoende av varandra och kan således var för sig utgöra grunden till ett specialarbete. Möjligen kan inledningen också tjäna som en introduktion till algebraiska och transcendent tal, om någon som läser detta hellre skulle vilja skriva ett specialarbete om sådant.

Inledning. Formeln för arean A av en cirkel med radie r ges som bekant av $A = \pi r^2$. Denna formel säger att om vi vet de exakta värdena av π och r , så kan vi beräkna det exakta värdet av arean. Tyvärr (?) förhåller det sig ju dock så att man i den fysikaliska verkligheten endast har tillgång till mätinstrument som tillåter att t ex radien hos en cirkel endast kan bestämmas med ett visst mått av noggrannhet – aldrig exakt. När cirkelns radie väl mätts upp behövs ett numeriskt värde på talet π för att vi skall få veta vad arean (ungefär) är.

Vad som används när man utför numeriska beräkningar (för hand eller med hjälp av datorer) är decimal- (binär-) bråksutveckling av de reella tal man för tillfället räknar med – talen decimalbråksutvecklas och man tar med så många decimaler som noggrannheten kräver.

$$\text{T ex} \quad \frac{1}{3} = 0,333333 \quad \text{med 6 decimalers noggrannhet}$$

$$\frac{1}{7} = 0,1428571 \quad \text{med 7 decimalers noggrannhet.}$$

Olika reella tal är olika *svåra* att decimalbråksutveckla: heltalen $(\dots, -2, -1, 0, 1, 2, \dots)$ är det ju inga problem med, de rationella talen (alla tal som är på formen p/q där p och q är heltal och $q \neq 0$, speciellt är heltal också rationella tal – tag $q = 1$) är också lätta att handha (bl a är det så att varje rationellt tal har en decimalbråksutveckling där siffrorna efter ett tag återkommer periodiskt). Reella tal som ej har en periodisk decimalbråksutveckling kallas irrationella (exempel på sådana är $\sqrt{2}$, π , e för att nämna några). Ett annat sätt att uttrycka att ett reellt tal x är rationellt är att säga att x är lösning till en ekvation

$$qx - p = 0, \quad p, q \text{ heltal och } q \neq 0.$$

Eller, uttryckt litet annorlunda, rationella tal är lösningar till första grads ekvationer med heltalskoefficienter (talen p och q kallas koefficienterna i ekvationen ovan). Om man nu generaliserar lite och tittar på andrags ekvationer med heltalskoefficienter:

$$px^2 + qx + r = 0, \quad p, q, r \text{ heltal och } p \neq 0,$$

t ex $x^2 - 2 = 0$ som har en lösning $\sqrt{2}$, så får vi med fler reella tal – inte bara de rationella. Allmänt kallar man ett reellt tal x algebraiskt om x är lösning till någon ekvation

$$a_n x^n + a_{n-1} x^{n-1} + \dots + a_1 x + a_0 = 0,$$

$$a_n, a_{n-1}, \dots, a_1, a_0 \text{ heltal och } a_n \neq 0.$$

Så $\sqrt{2}$ är ett algebraiskt tal, $\sqrt{1 + \sqrt{2}}$ är ett annat. Det sistnämnda är lösning till ekvationen

$$(x - \sqrt{1 + \sqrt{2}})(x + \sqrt{1 + \sqrt{2}})(x^2 - 1 + \sqrt{2}) = x^4 - 2x^2 - 1 = 0.$$

UPPGIFT.(a) Hitta en fjärdegradsekvation med heltalskoefficienter och med $x = \sqrt{10 - \sqrt{2}}$ som en lösning.

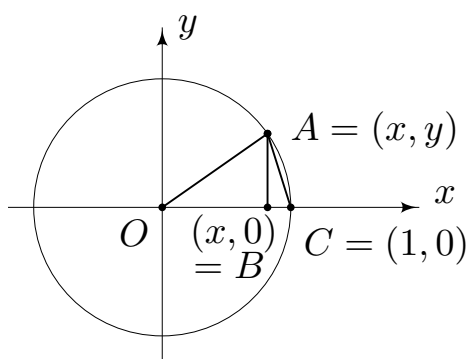
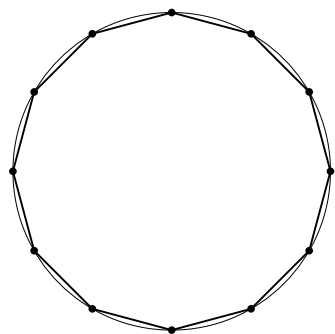
(b) Låt a och b vara två heltal båda större än noll. Försök visa att $\sqrt{a + \sqrt{b}}$ är ett algebraiskt tal, dvs försök hitta en ekvation med heltalskoefficienter som har $x = \sqrt{a + \sqrt{b}}$ som en lösning.

Nu kan man fråga sig: Är alla reella tal algebraiska tal? Dvs är varje reellt tal lösning till någon ekvation med heltals koefficienter? Svaret är nej! Talen π och e är exempel på icke-algebraiska tal (i själva verket utgör de algebraiska talen en förhållandevis liten del av de reella talen – de är i en viss mening inte *fler* till antalet än heltalen!). De icke-algebraiska heltalen kallas transcendent tal och transcendent tals decimalbråksutvecklingar är mer *komplicerade* än algebraiska tals, för ett givet algebraiskt tal finns det ju en ekvation med heltalskoefficienter till vilken det givna talet är en lösning, och det finns *snabba* metoder att t ex med hjälp av en dator finna approximativa lösningar till sådana ekvationer (en sådan metod är *Newton–Raphsons metod*). Detta gör det önskvärt att hitta formler för transcendent tal, t ex talet π som uttrycker π med hjälp av produkter/summor av algebraiska tal. Exempel på två sådana formler presenteras nedan.

En annan historisk notis värd att nämna i sammanhanget är den om cirkelns kvadratur. De gamla grekiska matematikerna formulerade problemet att med hjälp av passare och linjal konstruera en kvadrat med samma area som cirkeln med radie 1 – dvs en kvadrat med area π . Detta förutsätter att man kan konstruera en sträcka med längd $\sqrt{\pi}$ (kvadratens area är ju produkten av sidlängderna). Nu är det dock så att samtliga sträckor som kan konstrueras, med passare och linjal, utgående från en cirkel med radie 1, är algebraiska tal och det dröjde därför fram till 1882 innan grekernas problem fick svaret att det är omöjligt att konstruera en sådan kvadrat i och med att

en matematiker vid namn Lindemann visade att π är transcendent. (Visserligen ryktas det att en amerikansk domstol en gång lagstiftade att talet π är lika med 3, med stöd av en passus i bibeln, men detta får nog betraktas som en *juridisk sanning*.)

Viète's formel. Den formel för π som presenteras här upptäcktes av Francois Viète 1579. Som utgångs-punkt har vi att cirkeln med radie 1 har arean π , och för att få fram ett approximativt värde på π väljer vi att approximera cirkeln med geometriska figurer vars areor lätt kan räknas ut i termer av algebraiska tal. Om vi skriver in en regelbunden månghörning i cirkeln enligt figuren nedan, är det klart att ju fler hörn vi väljer desto bättre ansluter månghörningen till cirkeln och månghörningens area blir en approximation av cirkelns area.



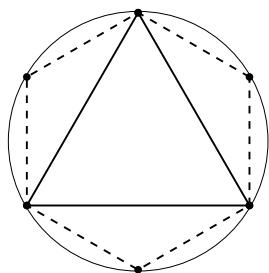
UPPGIFT. Bestäm arean av trianglarna OAC och OAB och visa, med hjälp av att lösa ekvationssystemet

$$\begin{cases} xy = 2(\text{area } OAB) \\ x^2 + y^2 = 1 \end{cases}$$

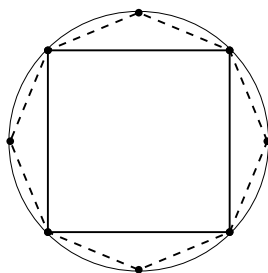
med avseende på y , att

$$(1) \quad \text{area } OAC = \frac{1}{2} \sqrt{\frac{1 - \sqrt{1 - 16(\text{area } OAB)^2}}{2}}.$$

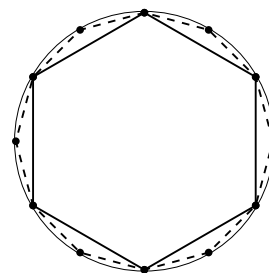
Låt nu P_m vara en regelbunden m -hörning inskriven i cirkeln och låt A_m vara arean hos P_m . Några exempel:



$m = 3$ ($m = 6$)



$m = 4$ ($m = 8$)



$m = 6$ ($m = 12$)

Varje P_m består av m stycken lika stora trianglar och man ser ju lätt att A_m är = (antalet trianglar i P_m) \times (arean hos en av trianglarna).

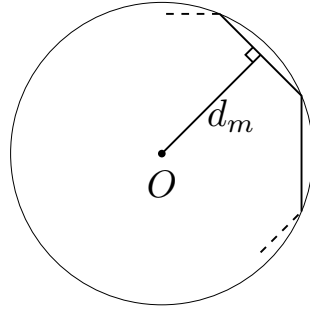
UPPGIFT. Härled, med hjälp av (1), formeln

$$(2) \quad A_{2m} = \frac{m}{2} \sqrt{2 - 2\sqrt{1 - (2A_m/m)^2}}.$$

Genom att använda formel (2) för $m = 2^n$ upprepade gånger, med början för $m = 2^2 = 4$, $A_4 = 2$ kan approximationer av π erhållas.

UPPGIFT. Gör ett datorprogram som, med hjälp av den härledda formeln, beräknar approximationer av π . (Dvs beräkna areorna A_{2^n} för några värden på heltalet n .)

Låt nu d_m beteckna det vinkelräta avståndet från origo till en sida i P_m enligt figur:



Via sambandet

$$\frac{\text{area}(OAB)}{\text{area}(OAC)} = OB$$

kan man härleda att $A_m/A_{2m} = d_m$. (Försök göra detta!) Alltså är t ex $A_4/A_8 = d_4$, $A_8/A_{16} = d_8$ vilket ger att $A_8 = A_4/d_4$ och därmed $A_8/A_{16} = A_4/(A_{16}d_4) = d_8$, dvs $A_4/A_{16} = d_4 \cdot d_8$.

UPPGIFT. Härled formeln

$$(3) \quad \frac{2}{A_{2^n}} = d_4 \cdot d_8 \cdot \dots \cdot d_{2^{n-1}}.$$

Ur detta samband ser vi att $A_{2^n} = 2 \cdot (d_4 \cdot d_8 \cdot \dots \cdot d_{2^{n-1}})^{-1}$ och om vi nu hittar något sätt att beräkna avstånden d_m så har vi alltså en formel som ger godtyckligt bra approximationer av π .

Ur figuren med d_m ovan syns att $d_m = \cos \frac{\pi}{m}$. Vi behöver alltså en värde för $\cos \frac{\pi}{m}$ då $m = 2^n$ för ett heltal $n \geq 2$, för att använda (3).

UPPGIFT. Använd formeln

$$\cos \frac{x}{2} = \sqrt{\frac{1 + \cos x}{2}}$$

för att visa att

$$d_4 = \sqrt{\frac{1}{2}}$$

$$d_8 = \sqrt{\frac{1}{2} + \frac{1}{2}\sqrt{\frac{1}{2}}}$$

$$d_{16} = \sqrt{\frac{1}{2} + \frac{1}{2}\sqrt{\frac{1}{2} + \frac{1}{2}\sqrt{\frac{1}{2}}}}$$

OSV.

Tillsammans ger nu det vi visat Viète's formel:

$$\begin{aligned} \frac{2}{A_{2^n}} &= d_4 \cdot d_8 \cdot \dots \cdot d_{2^{n-1}} = \\ &= \sqrt{\frac{1}{2}} \cdot \sqrt{\frac{1}{2} + \frac{1}{2}\sqrt{\frac{1}{2}}} \cdot \sqrt{\frac{1}{2} + \frac{1}{2}\sqrt{\frac{1}{2} + \frac{1}{2}\sqrt{\frac{1}{2}}}} \cdot \dots \end{aligned}$$

och ur detta kan vi ju lösa ut A_{2^n} och erhålla approximationer av π precis som ovan.

UPPGIFT. Gör ett datorprogram som med hjälp av formeln för $2/A_{2^n}$ ovan approximerar värdet på π . Jämför också dina värden med den bifogade tabellen över π :s decimalbråksutveckling.

Wallis' produkt. Wallis' produkt (från 1665) är, till skillnad från Viète's formel, kanske inte så geometrisk, utan bygger väsentligen på integration av trigonometriska funktioner och lite uppskattningar.

UPPGIFT. Gör en lämplig partiell integration i vänsterledet för att visa att

$$\int_0^{\pi/2} \sin^n x \, dx = \frac{n-1}{n} \int_0^{\pi/2} \sin^{n-2} x \, dx, \quad n \geq 2.$$

Härled ur detta att

$$\int_0^{\pi/2} \sin^{2n+1} x \, dx = \frac{2}{3} \cdot \frac{4}{5} \cdot \frac{6}{7} \cdots \frac{2n}{2n+1}$$

$$\int_0^{\pi/2} \sin^{2n} x \, dx = \frac{\pi}{2} \cdot \frac{1}{2} \cdot \frac{3}{4} \cdot \frac{5}{6} \cdots \frac{2n-1}{2n}.$$

(Ledning: Gör upprepade partiella integrationer!) Härled också ur de båda sista likheterna att

$$\frac{\pi}{2} = \frac{2}{1} \cdot \frac{2}{3} \cdot \frac{4}{3} \cdot \frac{4}{5} \cdot \frac{6}{5} \cdot \frac{6}{7} \cdots \frac{2n}{2n-1} \cdot \frac{2n}{2n+1} \cdot \frac{\int_0^{\pi/2} \sin^{2n} x \, dx}{\int_0^{\pi/2} \sin^{2n+1} x \, dx}.$$

Om vi nu kan visa att kvoten mellan integralerna i den ovanstående formeln är nära 1 då n är stort, så har vi här ytterligare en approximation av talet π , i termer av rationella tal denna gång!

UPPGIFT. Utnyttja olikheterna

$$0 < \sin^{2n+1} x \leq \sin^{2n} x \leq \sin^{2n-1} x \quad \text{för } 0 < x < \pi/2$$

och visa att

$$1 \leq \frac{\int_0^{\pi/2} \sin^{2n} x \, dx}{\int_0^{\pi/2} \sin^{2n+1} x \, dx} \leq \frac{2n+1}{2n}.$$

(Ledning: För den högra olikheten, gör först en partiell integration enligt ovan i nämnaren.)

Eftersom talen $(2n+1)/2n = 1 + 1/2n$ kan fås godtyckligt nära 1 genom att n väljes stort ser vi alltså att produkterna

$$2 \cdot \frac{2}{1} \cdot \frac{2}{3} \cdot \frac{4}{3} \cdot \frac{4}{5} \cdot \frac{6}{5} \cdot \frac{6}{7} \cdots \frac{2n}{2n-1} \cdot \frac{2n}{2n+1}$$

kan fås godtyckligt nära π . Denna produkt kallas Wallis' produkt.

UPPGIFT. Gör ett datorprogram som approximerar π med hjälp av Wallis' produkt. Jämför resultatet med den föregående uppgiften. Tycks någon av metoderna ge en *snabbare* väg till approximationer av π ? Jämför också dina approximationer med den bifogade tabellen över π :s decimalbråksutveckling.

Litteratur

Merparten av materialet till detta specialarbete har jag hittat i boken

Spivak, M., *Calculus*. Publish or Perish Inc., Berkeley, Calif. 1980, en bok som för inte så länge sedan användes i undervisningen av nybörjarstudenter i matematik vid Uppsala Universitet.

Mer om reella tal (bl a π) kan man läsa i
Brun, V., *Alt er tall*. A.s John Griegs Boktrykkeri, Bergen 1964.
Flegg, G., *Numbers – their history and meaning*. Penguin Books Ltd 1984.

Uppslagsverket *Sigma* (som väl torde finnas på de flesta gymnasieskolor) innehåller också en del lättillgängligt material rörande begreppen ovan (om än inte så mycket).

I *Scientific American*, februari 1988, finns en artikel om Ramanujan och π där man kan läsa om andra sätt att beräkna π .

En trafikmodell

LEIF ARKERYD

Göteborgs Universitet

Tänk dig en körfil på en landsväg eller motorväg, modellerad som x -axeln i positiv riktning (fig.1), och med krysset x_j som mittpunkten för bil nummer j på vägen.

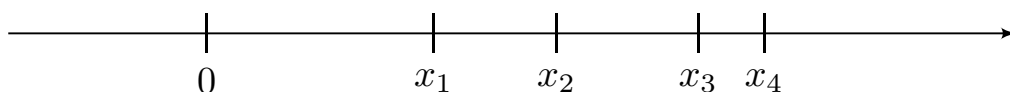


Fig.1

Då blir förstås $\dot{x}_j(t) = dx_j/dt$ hastigheten och $\ddot{x}_j(t) = d^2x_j/dt^2$ accelerationen av bil j . För enkelhets skull antar vi att alla bilar har samma längd L .

Om du sitter på en kulle eller ett berg och tittar på vägen i fjärran, så ser du i stället för de enskilda bilarna en sammanhängande ström (åtminstone i tät trafik), som verkar ha en hastighet $u(x, t)$ i varje punkt x och vid varje tidpunkt t ($u(x, t)$ kallas ett *hastighetsfält*). Du upplever också att bilströmmen har en *täthet* $\varrho(x, t)$ som beror också den på x och t . I lämpliga enheter kan antalet bilar N på sträckan $[0, X]$ anges som $N = \varrho \cdot X$ om tätheten är likformig, och som $N(t) = \int_0^X \varrho(x, t) dx$ i det allmänna fallet (fig.2).

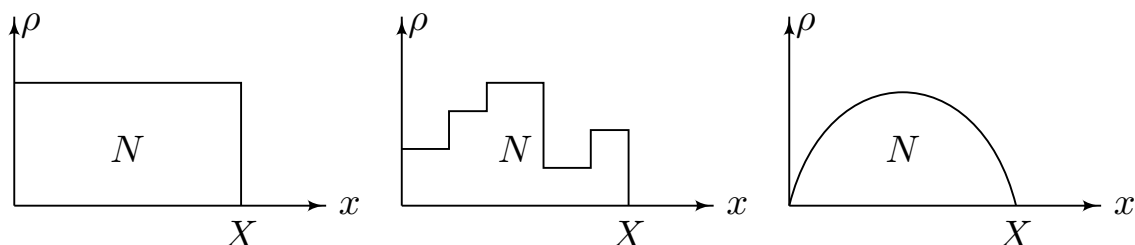


Fig.2

Det är klart att om X är hela vägens längd, så blir medelantalet bilar per längdenhet $\rho = N/X$, och om $X = L$ så blir typiskt $\rho(x, t) = 1/L$ om det finns en bil i punkten x , och $\rho(x, t) = 0$ om det inte finns någon bil där. (För bil i blir dessutom $u(x_i(t), t) = \dot{x}_i(t)$.)

Således för att få en rimlig mening av begreppet biltäthet bör vi observera vägen bortifrån en punkt, där en bekvämt observerbar sträcka är mycket större än L och mycket mindre än vägens längd. Detta definierar begreppet naturlig skala för vägsträckor i den här diskussionen.

ÖVNING 1. Omsätt ovanstående i ett praktiskt experiment på någon kraftigt trafikerad vägsträcka $[0, X]$ i din närhet.

- Fotografera vägsträckan med en halv timmes mellanrum under en halv dag, och upprätta för varje tidpunkt tre diagram som i fig.2.
- Upprita för en lämplig tidpunkt en kurva som beskriver medeltätheten på sträckan $[0, x]$, $\bar{\rho} = N(x)/x$ som funktion av x , $0 \leq x \leq X$. Observera att täthetskurvan fluktuerar kraftigt för små x för att stabiliseras när x blir större.

Om tätheten ρ och hastigheten u är konstanta, d v s oberoende av x och t , så är *bilflödet* q , d v s antalet bilar som passerar en punkt x per tidsenhet, lika med $\rho \cdot u$, d v s $q = \rho \cdot u$. Detta gäller även om q , ρ , u är tidsberoende och rumsberoende (varför?).

Om vägsträckan $[a, b]$ saknar till- och avfarter, så ges ändringen av antalet bilar där på tiden Δt , d v s $N(t + \Delta t) - N(t)$, av antalet bilar som kommer in i $x = a$ minus antalet som går ut i $x = b$, d v s av $\Delta t(q(a, t) - q(b, t))$. Tas nu gränsvärdet av

$$(N(t + \Delta t) - N(t))/\Delta t = q(a, t) - q(b, t),$$

så erhålls

$$\dot{N}(t) = q(a, t) - q(b, t), \text{ eller } \frac{d}{dt} \int_a^b \varrho(x, t) dx = q(a, t) - q(b, t).$$

Detta är en *konserveringslag i integralform (= global form)* för antalet bilar. Om derivatan får flyttas under integraltecknet ger detta

$$(1) \quad \int_a^b \frac{d}{dt} \varrho(x, t) dx = q(a, t) - q(b, t).$$

Fast när en funktion som ϱ här beror på både variabeln t och parametern x , så talar man om den partiella derivatan med avseende på t och skriver detta $\frac{\partial}{\partial t} \varrho(x, t)$. Analogt om t betraktas som en parameter och x som variabel, så skrivs derivatan med avseende på x som $\frac{\partial}{\partial x} \varrho(x, t)$. Högerledet i (1) kan skrivas

$$q(a, t) - q(b, t) = \int_b^a \frac{\partial}{\partial x} \varrho(x, t) dx = - \int_a^b \frac{\partial}{\partial x} \varrho(x, t) dx,$$

och vi får alltså

$$\int_a^b \left(\frac{\partial}{\partial t} \varrho(x, t) + \frac{\partial}{\partial x} \varrho(x, t) \right) dx = 0.$$

Den enda kontinuerliga integrand som ger integralen noll för varje val av a och b är 0-funktionen. (Visa det!) Alltså gäller

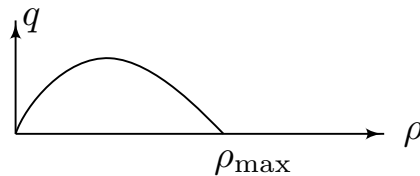
$$(2) \quad \frac{\partial}{\partial t} \varrho(x, t) + \frac{\partial}{\partial x} \varrho(x, t) = 0.$$

Ett samband som (2) mellan en funktion och dess partiella derivator kallas en *partiell differentialekvation*. Då ju $q = u \cdot \rho$, kan den också skrivas

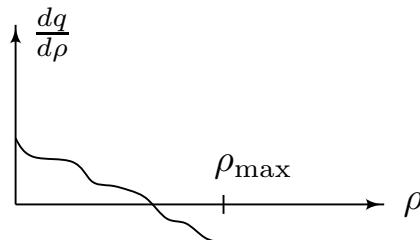
$$(3) \quad \frac{\partial}{\partial t} \rho + \frac{\partial}{\partial x} (u \cdot \rho) = 0.$$

Detta är den förra konserveringslagen för antalet bilar, fast nu i *differentialform* (= *lokal form*). För att kunna lösa ekvationen (3) behövs information om hastighetsfältet u . Det är klart att u beror av trafiktätheten; därför är ett rimligt antagande $u = u(\rho)$. Vi vet erfarenhetsmässigt att ju färre bilar, desto fortare kör bilisterna. Därför antar vi att, när vägen är tom, $u = u_{\max}$ är maximal, t ex en klar sommardag lika med bilarnas maximalhastighet på en tysk motorväg, eller lika med den maximalt tillåtna hastigheten på en svensk landsväg, medan den kanske bara är 20 km/tim en vinterdag med underkyllt regn. Det är också rimligt att hastigheten minskar när tätheten ökar, det vill säga $\frac{\partial u}{\partial \rho} \leq 0$ – ner till noll strax under $\rho = \rho_{\max} = 1/L$.

Eftersom flödet $q = \rho u(\rho)$, så gäller också $q = q(\rho)$. Uppenbarligen är $q = 0$ om $\rho = 0$, eller om $u(\rho) = 0$ som för $\rho = \rho_{\max}$. Däremellan tar vi flödet positivt, och ett plausibelt utseende är



med bl a $\frac{d^2 q}{d\rho^2} < 0$, d v s $\frac{dq}{d\rho}$ minskar då ρ ökar. Detta ger $\frac{dq}{d\rho}$ utseendet



ÖVNING 2. Beräkna q och rita kurvorna för $u(\varrho)$, $q(\varrho)$ och $dq/d\varrho$ om $u(\varrho) = u_{\max}(1 - \varrho/\varrho_{\max})$ där u_{\max} och ϱ_{\max} är konstanter. Vad blir det maximala trafikflödet?

ÖVNING 3. Tänk dig en halvoändlig landsväg $0 \leq x < \infty$ (alltså idealiserad av modellerings-skäl), och antag att $\varrho(x, t)$ är 0 för stora värden på x . Om det inte finns några av- eller påfarter, visa ur (2) att antalet bilar på vägen vid tiden t är

$$N(0) + \int_0^t q(0, s) ds.$$

ÖVNING 4. Om $x(t)$ är rörelsen av en bil i trafikflödet, så är $\dot{x}(t) = u(x(t), t)$, och accelerationen $\ddot{x}(t)$ ges av kedjeregeln

$$(4) \quad \frac{d}{dt}(u(x(t), t)) = \frac{\partial}{\partial t} u(x(t), t) + \frac{\partial}{\partial x} (u(x(t), t)) \dot{x}(t).$$

Antag att accelerationen i detta fall också ges av $-a^2 \varrho^{-1} \partial \varrho / \partial x$, med a en positiv konstant. Använd (3) för att visa att om $u = u(\varrho)$ ej är konstant så följer att $du/d\varrho = -a/\varrho$. Lös denna ordinära differentialekvation under randvillkoret $u(\varrho_{\max}) = 0$. Diskutera varför den ickekonstanta lösningen (som utmärkt beskriver mätdata från flera kraftigt trafikerade tunnlar i USA) inte kan vara särskilt bra som modell för små tätheter.

ÖVNING 5. Förklara utan att använda matematiska formler varför $\int_{a(t)}^{b(t)} \varrho(x, t) dx$ är konstant, dvs oberoende av t , om $a(t)$ och $b(t)$ är läget vid tiden t av två bilar som ligger i trafikflödet.

ÖVNING 6. Antag att $u = \text{konstant} = c$. Inför de nya variablerna $x' = x - ct$ och $t' = t$. Använd kedjeregeln (4) för att visa att konserveringslagen (3) övergår i ekvationen $\partial \varrho / \partial t' = 0$ i de nya koordinaterna.

För fix parameter x' får vi alltså $\rho(x' + ct', t') = \text{konstant}$. För olika x' kan konstanten väljas olika, som en funktion f av x' , dvs $\rho(x' + ct', t') = f(x')$, eller $\rho(x, t) = f(x - ct)$. Således är ρ konstant längs varje *karaktéristisk* $x - ct = \text{konstant}$.

ÖVNING 7. Rita för $t = 1, 2, 3$ kurvan $(x, \rho(x, t))$ om $\rho(x, 0) = (1 + \sin x)/10$ för $0 \leq x \leq \pi$ och $\rho(x, 0) = 0$ annars, samt $u = \text{konstant} = 10$. Du ser alltså en täthetsvåg som rör sig i x -axelns riktning.

Eftersom enligt (4)

$$\frac{d}{dt}(\rho(x(t), t)) = \frac{\partial \rho}{\partial t} + \frac{\partial \rho}{\partial x} \dot{x},$$

så gäller att en observatör i $x(t)$ vid tiden t ser ändringen i täthet som summan av täthetsändringen $\partial \rho / \partial t$ i punkten $x(t)$, och förändringen som beror på att observatören rör sig in i ett område med kanske en annan trafiktäthet. Speciellt om $\dot{x} = c$, så följer observatören med strömmen och $d/dt(\rho(x(t), t)) = 0$, dvs den ser ingen ändring i fallet $u = c$.

För att komma vidare i studiet av den här trafikflödesmodellen, kan du låna boken *Mathematical models* av R. Haberman (Prentice Hall 1977).

ÖVNING 8. Använd den boken för att analysera vad som händer när

- trafiken startar vid grönt ljus (avsnitt 72),
- när trafiken stannar vid rött ljus (avsnitt 78),
- hur chockvågor byggs upp i områden där trafiktätheten ökar (avsnitt 79–82).

På liknande sätt kan man diskutera flöden av gaser uppbyggda av individuella molekyler. Ekvationen blir snarlik, och man får också sådana ekvationer för andra konserverade storheter som energin. De här ekvationerna har varit kända i närmare 200 år, men att lösa dem

utgör fortfarande ett aktivt och komplicerat forskningsområde inom matematiken, där nästan varje verkligt framsteg har fysikaliska och ingenjörsmässiga tillämpningar.

Om rättvisa val

LEIF ARKERYD

Göteborgs universitet

Den *axiomatiska metoden* i matematiken kan sägas bestå i att man studerar konsekvenserna av en uppsättning givna (och förhoppningsvis intressanta) krav (= axiom = postulat). Så kan t ex de naturliga talens egenskaper studeras utgående från Peanos axiom, och geometri i planet har studerats utgående från Euklides' axiom. Också för att matematiskt modellera en *praktisk situation* fastlägger man krav som modellen approximativt eller exakt måste uppfylla. Det här specialarbetet utgår från ett försök att modellera åsikter genom röstning och val, och studerar konsekvenserna av några rimliga krav på ett valförfarande.

Ett exempel. Skolans fritidsklubb har årsmöte och på dagordningen finns bl a följande två punkter.

- i) Val av ny ordförande. Det finns bara ett förslag och medlemmarna skall ta ställning till om den föreslagne skall utses.
- ii) Nya medlemsavgifter. Det finns de tre förslagen A, B, C . Förslag A vill utvidga verksamheten, men innebär en väsentlig ökning av avgifterna. Förslag B vill att man inte gör mer än vad de nuvarande avgifterna räcker till. Förslag C slutligen tycker att avgifterna kan avskaffas, och att man i stället förlitar sig på de bidrag som just har sökts.

Vi antar att diskussionen är slut och skall titta på olika valprocedurer. I i) används förstås en enkel majoritetsomröstning. Men i ii) kan t ex valet göras som en följd av val mellan två olika alternativ. Först genomför årsmötet kanske en enkel majoritetsomröstning mellan förslag A och B , och ställer sedan vinnaren mot C . Men det är

faktiskt inte så okontroversiellt, som det vid första påseendet verkar. Antag att det finns en viss konsistens i hur den enskilde eleven ordnar de tre förslagen, och detta så att om A föredras framför C , och C framför B , så föredras A framför B . Då kan dennes preferenslista skrivas ABC . Detsamma bör förstås gälla för gruppen. Men låt oss betrakta exemplet med 31% av preferenslistorna ABC , 34% med BCA , 35% med CAB och inga andra alternativ föredragna av någon elev.

UPPGIFT 1. Visa att proceduren ovan med första val mellan A och B ger C som vinnare, men att ett första val mellan A och C ger B som vinnare.

Rättvisa procedurer. Vi ser av uppgiften att metoden i det förra exemplet kan manipuleras av mötesordföranden. Så finns det någon "rättvis valprocedur"? För att besvara den frågan börjar vi med att lista några *krav* som ett valförfarande nog bör uppfylla för att kunna kallas rättvist. Sedan skall vi härleda några förvånande konsekvenser av de villkoren. *Målet* är förstås att översätta väljarpreferenser i en grupppreferenslista.

Vi låter x, y, z beteckna alternativa förslag att ta ställning till. Bokstäverna i och j betecknar röstande, och vi antar förstås att det bara finns ändligt många röstande. "Föredras framför" (är en s k relation som nu) skrivs som $>$, "lika bra" skrivs som $=$, och båda tillsammans, dvs \geq , utläses "är minst lika bra som". Vi kräver naturligtvis att om de två förslagen inte är lika bra, så är det ena bättre än det andra, d v s att

a) för alla x, y så gäller precis en av $x > y$, $x = y$, $y > x$;
vidare att

b) för alla x gäller $x = x$;

och slutligen att om ett förslag är bättre än ett andra och detta

bättre än ett tredje, så skall det första vara bättre än det tredje, och analogt för "lika bra", dvs att

c) för alla x, y, z gäller: om $x \geq y$ och $y \geq z$, så $x \geq z$ med $x = z$ precis då $x = y$ och $y = z$.

UPPGIFT 2. Vilka sex rangordningar \geq kan den röstande i göra mellan de tre förslagen x, y, z ? Hur många rangordningar \geq kan de två röstande i, j göra mellan samma förslag?

Nu lägger vi på rättvisekraven.

KRAV 1. *Alla upptänkliga rangordningar är möjliga för de röstande.*

(Rimligt då vi måste förutsätta att de röstande kan ha mycket skilda åsikter, och vi inte vill hindra dem att framföra sin verkliga uppfattning.)

KRAV 2. *Om röstningen $(x \geq y)_i$ i val 1 medför röstningen $(x \geq y)_i$ i val 2, och dessutom i val 1 resultatet är $x \geq y$, så följer $x \geq y$ i val 2.*

(Rimligt att om alla röstande tycker lika i båda valen så skall resultaten bli desamma. Detta utesluter lottning mellan valsedlar, liksom proceduren ovan att jämföra först A och B , och sedan vinnaren med C .)

KRAV 3. *Om $(x \geq y)_i$ för alla i så $x \geq y$, och i detta fall likhet $x = y$ precis då $(x = y)_i$ för alla i .*

(Rimligt att om alla röstande föredrar x framför y , så gör gruppen det.)

UPPGIFT 3. Visa att Krav 1 - 3 har konsekvensen att

KRAV 4. *För alla j har vi att " $(x \geq y)_j$ i val 1 precis då $(x \geq y)_j$ i val 2" medför att " $x \geq y$ i val 1 precis då $x \geq y$ i val 2".*

Det är mycket lätt att konstruera ett röstningsförfarande som uppfyller kraven 1 – 3. Utännamn bara en av de röstande till diktator! Det sista kravet 5 utesluter den möjligheten.

KRAV 5. *Det finns inget i med egenskapen att $x \geq y$ precis då $(x \geq y)_i$.*

UPPGIFT 4. Det naturliga kravet på *en man en röst* behöver inte vara uppfyllt. Visa detta genom att konstruera ett motexempel. Det finns inte heller något krav på att gruppen föredrar x framför y om en enkel majoritet gör det. Visa genom exempel.

UPPGIFT 5. Konstruera ett exempel som uppfyller kraven

a) 1, 2, 5 b) 1, 3, 5 c) 2, 3, 5.

Vad hände med det utelämnade kravet i dina exempel?

Omöjlighetssatsen. Den verkliga överraskningen är, att trots att Krav 1 – 5 är tämligen svaga, så finns det ändå inte något röstningsförfarande som kan uppfylla dem alla samtidigt, om röstningen gäller åtminstone tre förslag x, y, z . Detta resultat är känt som Arrows omöjlighetssats och upptäcktes av den amerikanske nobelpristagaren Kenneth J. Arrow år 1951.

SATS (ARROW). *Det existerar ingen valprocedur med fler än två valalternativ som uppfyller kraven 1 - 5.*

Vi kallar mängden röstande för R , och säger att mängden $R_b \subseteq R$ är *beslutande* för " x mot y " om " $(x \geq y)_i$ för alla i som tillhör R_b medför $x \geq y$, och om i detta fall dessutom $x = y$ medför $(x = y)_i$ för alla i som tillhör R_b ". Uppenbarligen är R beslutande på grund av Krav 1 och Krav 3. Vi skall visa att om Krav 1 - 3 gäller för en valprocedur, så finns det ett beslutande R_b med bara en röstande, och att detta R_b är beslutande för alla par, dvs att det finns en

diktator. Detta i sin tur motsäger Krav 5 och satsen är därmed bevisad. Låt oss genomföra beviset.

BEVIS. Antag att det inte finns något ” x mot y ” för vilket någon enskild röstande är beslutande, och låt R_b vara en minsta beslutande mängd med avseende på alla möjliga ” x mot y ”, och låt den höra till fallet ” \bar{x} mot \bar{y} ”. Vi delar R_b i två disjunkta delmängder R_1 och R_2 med minst en röstande i varje.

Låt z vara ett tredje valalternativ, och betrakta situationen

$$\begin{aligned}
 & (\bar{x} \geq \bar{y} \geq z)_i \quad \text{för alla } i \text{ som tillhör } R_1, \\
 (*) \quad & (z \geq \bar{x} \geq \bar{y})_i \quad \text{för alla } i \text{ som tillhör } R_2, \\
 & (\bar{y} > z > \bar{x})_i \quad \text{för alla } i \text{ som tillhör } R \text{ men inte } R_b.
 \end{aligned}$$

Om nu $\bar{x} \geq z$ så är R_1 beslutande för \bar{x} mot z , och då vore inte R_b minst. Alltså gäller $z > \bar{x}$. Eftersom R_b är beslutande för ” \bar{x} mot \bar{y} ” så ger (*) att $\bar{x} \geq \bar{y}$, och därför $z > \bar{y}$. Det följer härur, väsentligen av Krav 2, att R_2 är beslutande för ” z mot \bar{y} ”. (Likhetsvillkoret ger en del extraarbete.) Detta motsäger att R_b är minst. Vårt antagande är alltså fel och R_b har bara en röstande, \bar{i} .

Vi skall nu också visa att \bar{i} bestämmer alla alternativ ” x mot y ”. Låt z vara ett tredje alternativ, och antag att $(\bar{x} \geq \bar{y} \geq z)_{\bar{i}}$, men att $(\bar{y} > z > \bar{x})_j$ för $j \neq \bar{i}$. Krav 3 ger att $\bar{y} > z$, och \bar{i} är ju beslutande för $\bar{x} \geq \bar{y}$. Alltså är $\bar{x} > z$. Ur Krav 3 följer då att \bar{i} är beslutande för ” \bar{x} mot z ”.

UPPGIFT 6. Visa på samma sätt att för varje $w \neq \bar{x}, z$, så är \bar{i} beslutande för ” w mot z ”.

Således är \bar{i} beslutande för varje par, dvs \bar{i} är diktator. VSB

Vill du veta mer om Arrows omöjlighetssats, kan du läsa Reasonable elections don't exist av John Baylis i tidskriften *the Mathematical Gazette*, 69 (1985), s 95-103.

Låna också band 5 av matematikverket *Sigma* på biblioteket och studera:

UPPGIFT 7. Peanos axiom i artikeln Om den matematiska sanningens natur av C.G. Hempel. Tillämpa axiomen för att visa några räknelagar för de naturliga talen.

UPPGIFT 8. Euklides' axiom i artikeln Den axiomatiska metoden av R.L. Wilder. Tillämpa dem och visa några av de geometriska satserna. Fler finner du i Euklides' *Elementa*.

Vinkeln 60 grader kan inte tredelas med enbart passare och linjal

JÖRAN BERGH

CTH

Att tredela en given vinkel med enbart passare och linjal är ett klassiskt (ca 400 f Kr) geometriskt problem. Detta löstes på artonhundratalet med hjälp av en teori för ekvationers lösbarhet utvecklad av bl a Évariste Galois (1811–1832). Lösningen visar att en sådan tredelning i allmänhet är omöjlig. Dock är den möjlig för vissa vinklar.

- Ange några vinklar som kan tredelas med enbart passare och linjal.

De klassiska problemen med kubens fördubbling och cirkelns kvadratur har visats omöjliga att lösa med bara passare och linjal med samma metoder.

UPPGIFT: Visa att vinkeln 60° inte kan tredelas med bara passare och linjal.

Den klassiska lösningen på problemet är att visa att det är likvärdigt med att finna en lösning till en viss tredjegrads ekvation med enbart kvadratrotutdragningar.

Tredjegrads ekvationen hänger ihop med tredelningen av vinkeln; kvadratrotutdragningarna med konstruktioner som utnyttjar enbart passare och linjal.

Vi tittar först på tredelningen av vinkeln.

Ett samband mellan en vinkel och den tredubbla vinkeln finns i trigonometrin. Vi skall använda formeln

$$\cos 3\alpha = 4 \cos^3 \alpha - 3 \cos \alpha.$$

- Härled denna formel!

Om vi nu låter $3\alpha = 60^\circ$ så får vi

$$\frac{1}{2} = 4 \cos^3 \alpha - 3 \cos \alpha$$

vilket är en tredjegrads ekvation i $\cos \alpha = x$

$$\frac{1}{2} = 4x^3 - 3x.$$

Om man har en lösning x till denna ekvation ger det en möjlighet att konstruera vinkeln $20^\circ = 60^\circ/3$, och omvänt.

Antag nu att vi har bestämt en längdenhet t ex genom att dra en linje med hjälp av linjalen och där avsätta en sträcka med passaren.

- Visa att man kan konstruera (med endast passare och linjal) vinkeln α precis då man kan konstruera sträckan med längd $\cos \alpha$.

Vi ser nu att tredelningen av vinkeln 60° med enbart passare och linjal är likvärdigt med att konstruera en sträcka med längd x som är lösning till ekvationen $\frac{1}{2} = 4x^3 - 3x$.

Vi skall nu undersöka vilka tal (sträckor) som kan konstrueras med enbart passare och linjal, då vi bestämt en längdenhet.

- Visa att alla tal av typen $a + b$, $a - b$, ab och a/b , där a och b är hela tal ($b \neq 0$) går att konstruera. (Ledning: Använd t ex likformighet på en topptriangel.)

Alla tal av typ p/q med p och q heltal, $q \neq 0$, kan alltså konstrueras: med andra ord alla rationella tal.

- Visa att $a \pm \sqrt{b}$ kan konstrueras om a och b är konstruerade men inte \sqrt{b} . (Ledning: Använd t ex likformighet i en rätvinklig triangel med en höjd från räta vinkeln.)

Vi vet nu att alla tal som kan fås ur de rationella talen genom ändligt många kvadratrotutdragningar kan konstrueras med enbart

passare och linjal. Men vi behöver veta vilka som inte går att konstruera för att nå vårt mål: ett nödvändigt villkor för konstruerbarhet.

- Visa att om ett tal är konstruerbart med enbart passare och linjal så måste det vara av formen $a \pm \sqrt{b}$ där a och b ($b \geq 0$) redan konstruerats. (Ledning: Passaren ger cirklar och linjalen ger linjer. Analytisk geometri ger skärningarna som lösningar till vissa ekvationer.)

Konstruerbarhet av ett tal med bara passare och linjal är således likvärdigt med att man har upprepat konstruktionen $a \pm \sqrt{b}$ ett ändligt antal gånger: vid första konstruktionen är a och b rationella.

Vårt sista steg är att visa att ekvationen

$$\frac{1}{2} = 4x^3 - 3x$$

inte har några lösningar av formen $a \pm \sqrt{b}$, där konstruktionen upprepats ändligt många gånger utgående från rationella tal.

- Visa att ekvationen inte har någon rationell rot. (Ledning: Anta att p/q är en rot där p och q är hela tal utan gemensamma faktorer. Detta leder till $8p^3 - 6pq^2 - q^3 = 0$ och det följer att p delar 1, q delar 2.)

Anta nu att ekvationen har en rot av formen $a + \sqrt{b}$ (alternativt $a - \sqrt{b}$) där antalet rotutdragningar vid konstruktionen är minimalt.

- Visa att ekvationen då också har en rot $a - \sqrt{b}$ ($a + \sqrt{b}$). (Ledning: Ekvationen kan skrivas $8x^3 - 6x - 1 = 0$. Dividera högra ledet med lämplig faktor.)

Därmed har $8x^3 - 6x - 1$ faktorn $x^2 - 2ax + a^2 - b$.

- Visa att $a^2 - b \neq 0$ och att ekvationen då har en rot $\frac{1}{8(a^2 - b)}$. (Ledning: Identifiera rötternas produkt som en koefficient i ekvationen.)

Allt är nu klart eftersom roten $\frac{1}{8(a^2-b)}$ kan konstrueras med ett antal rotutdragningar som är strikt mindre än det minimala: $a + \sqrt{b}$ ($a - \sqrt{b}$) hade ett minimalt antal. Vi har alltså fått en motsägelse. Detta innebär att ekvationen $\frac{1}{2} = 4x^3 - 3x$ inte har någon rot på formen $a \pm \sqrt{b}$ (ändligt många rotutdragningar med start i rationella tal) och sådana tal var ju de enda som kunde konstrueras enbart med passare och linjal.

Litteratur

En mycket lättläst bok (jag läste den själv som gymnasist) är Courant, R. & Robbins, H., *What is Mathematics*. Oxford University Press, Oxford 1978, varifrån idén till uppgiften hämtats. Boken borde finnas tillgänglig i skolbiblioteket för elever på NT-linjerna.

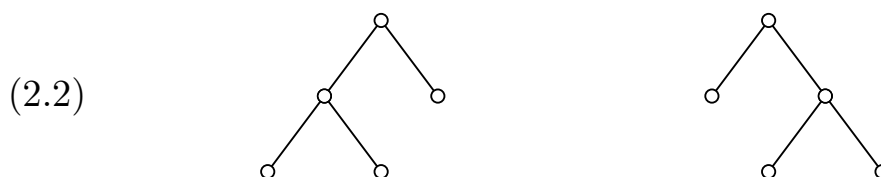
En bok på universitetsnivå är t ex van der Waerden, B.L., *Algebra I*. Springer-Verlag, 1964.

i regel i motsatt riktning mot naturens träd – det har blivit en konvention.) För varje nod v gäller endera av följande två fall:

(1) v har inga barn (d.v.s. inga kanter leder neråt från v), v kallas då ett *löv*.

(2) v har exakt två barn, v kallas då en *inre nod*.

I fallet (2) skiljer man på *vänsterbarn och högerbarn*, så att exempelvis de två träden



anses olika som binära träd (ty i det ena är rotens högerbarn ett löv men ej i det andra).

Binära träd brukar ofta formellt definieras på följande rekursiva sätt:

- (a) en mängd med ett element är ett binärt träd,
- (b) ett binärt träd består av ett element (kallat *rot*) och ett ordnat par av binära träd (kallade *vänsterdelträd* och *högerdelträd*).

ÖVNING 1. Tänk igenom denna formella definition så att du blir övertygad om att den beskriver samma matematiska objekt som de binära träd vi mer intuitivt definierat ovan.

Antalet noder i ett binärt träd måste vara udda. (Varför?) Hur många binära träd med $2k+1$ noder finns det? Svaret för små värden på k ges i följande tabell:

n	1	3	5	7	9	11	13
antal binära träd med n noder	1	1	2	5		42	132

ÖVNING 2. Rita alla binära träd med 9 noder och upptäck på så sätt det värde som saknas i tabellen.

ÖVNING 3. Låt t_k vara antalet binära träd med $2k + 1$ noder. Finn en rekursionsformel för talsviten t_0, t_1, t_2, \dots , d.v.s en formel som uttrycker t_k som funktion av t_0, t_1, \dots, t_{k-1} för alla $k \geq 1$.

ÖVNING 4. Komplettera tabellen till alla udda $n \leq 15$ med hjälp av rekursionsformeln för t_k .

Man kan visa att antalet binära träd med $2k + 1$ noder ges av följande exakta formel

$$(2.3) \quad t_k = \frac{(2k)!}{(k+1)!k!},$$

där $k! = k \cdot (k-1) \cdot (k-2) \cdot \dots \cdot 2 \cdot 1$, och $0! = 1$.

ÖVNING 5. Komplettera tabellen till alla udda $n \leq 25$ med hjälp av (2.3).

Vi betraktar nu något visst binärt träd och låter n beteckna antalet noder, ℓ antalet löv, i antalet inre noder och e antalet kanter. Exempelvis, för träden i (2.2) gäller $n = 5, \ell = 3, i = 2, e = 4$, och för trädet i (2.1) finner vi $n = 15, \ell = 8, i = 7, e = 14$.

ÖVNING 6. Beskriv i formler samband mellan

- (a) n och e ,
- (b) n och i ,
- (c) n, ℓ och i .

Dra ur de tre formler du funnit slutsatsen att om endast *ett* av talen n, ℓ, i och e , är känt så kan övriga bestämmas därur.

3. Viktade binära träd. Låt T vara ett binärt träd. Varje nod v befinner sig på ett visst *djup* $d(v)$ lika med avståndet (antalet

kanter) från roten. Om vi listar lövens djup i växande ordning för träden i figur (2.2) får vi sviten $1, 2, 2$, och för figur (2.1) får vi $2, 2, 3, 3, 3, 4, 5, 5$.

ÖVNING 7.(A) Låt d_1, d_2, \dots, d_ℓ vara lövens djup för ett binärt träd. Visa att då gäller

$$(3.1) \quad \sum_{i=1}^{\ell} \frac{1}{2^{d_i}} = 1.$$

(B) Låt d_1, d_2, \dots, d_ℓ vara en godtycklig svit av ickenegativa heltal för vilka ekvation (3.1) gäller. Visa att det finns ett binärt träd vars löv sitter på djup $d_i, i = 1, 2, \dots, \ell$.

I många datalogiska sammanhang studeras sökstrukturer där information finns lagrad i löven hos ett binärt träd. För att snabbt komma åt denna information är det av intresse att veta vilka träd med l löv som minimerar summan ($L =$ mängden av löv)

$$(3.2) \quad \sum_{v \in L} d(v).$$

ÖVNING 8. Visa att bland alla binära träd med ℓ löv så minimeras summan (3.2) av de träd vars alla löv har djup b eller $b + 1$, där b är heltalet bestämt av $b \leq \log_2 \ell < b + 1$. (Visa också att det alltid finns sådana binära träd, de kallas *balanserade*.)

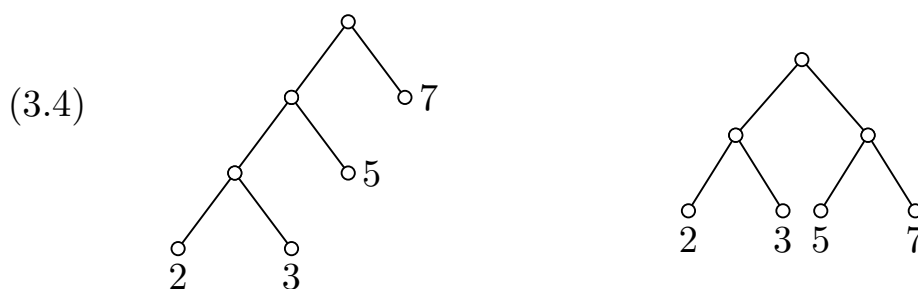
ÖVNING 9. Låt T vara ett godtyckligt binärt träd med n noder. Visa

$$(3.3) \quad \sum_{v \in T} d(v) = 1 - n + 2 \sum_{v \in L} d(v).$$

(Således minimeras även totalsumman av *alla* noders djup av de balanserade träden.)

Vi har sett att de balanserade träden minimerar summan (3.2). De är därför de bästa träden för att lagra information som behöver åtkommas ungefär lika ofta. Men i praktiken är detta sällan fallet. Den information som finns lagrad i ett visst löv kanske efterfrågas mer än all annan information. Det är då intuitivt klart att detta löv bör sitta nära roten och på mindre djup än de andra löven. Detta leder till följande matematiska problemställning.

Först definierar vi begreppet *viktat binärt träd*. Detta är ett binärt träd T där varje löv v är märkt med ett reellt tal $w(v)$. Till exempel:



För varje viktat binärt träd bildar vi summan

$$(3.5) \quad w(T) = \sum_{v \in L} w(v) \cdot d(v),$$

kallad trädets *vikt*. Exempelvis, $w(T) = 2 \cdot 3 + 3 \cdot 3 + 5 \cdot 2 + 7 \cdot 1 = 32$ respektive $w(T) = (2 + 3 + 5 + 7) \cdot 2 = 34$ för träden i figur (3.4).

Problemet är nu att för givna vikter hitta ett binärt träd T_0 som bland alla viktade träd T minimerar $w(T)$. Mer precist, antag givna ℓ tal w_1, w_2, \dots, w_ℓ . Vi vill konstruera ett viktat binärt träd T_0 vars löv är märkta med w_1, w_2, \dots, w_ℓ , och sådant att $w(T_0) \leq w(T)$ för alla andra tänkbara sådana träd T . Vi kallar ett sådant viktat binärt träd T_0 *optimalt*.

Sambandet med det tidigare resonemanget om informationslagring är följande. Antag att ℓ objekt (t.ex. telefonnummer) skall lagras i löven till ett binärt träd. Varje objekt efterfrågas med en viss känd frekvens $w_i, i = 1, 2, \dots, \ell$. När ett objekt skall sökas börjar sökningen alltid vid roten och fortskrider genom successiva höger/vänster-val neråt till det löv där det sökta objektet är lagrat. Antalet steg i sökningen (och därmed proportionellt även åtgången tid och kostnad) är lika med lövets djup. Problemet att lagra de ℓ objekten på effektivaste sätt (d. v. s. så att den totala tiden och kostnaden för sökning minimeras) är då uppenbart ett specialfall av vårt matematiska problem.

Innan vi beskriver problemets lösning är det värt att kommentera specialfallet $w_1 = w_2 = \dots = w_\ell = 1$. I det fallet löste vi problemet redan i övning 8. Märk att de balanserade träden inte är optimala i allmänhet: det högra trädet i figur (3.4) är balanserat men har högre vikt än det vänstra obalanserade trädet (vilket som vi strax skall se är optimalt).

Följande eleganta algoritmiska lösning på problemet gavs av D. Huffman 1952. För $\ell = 2$ och givna vikter w_1 och w_2 är följande träd optimalt (detta är uppenbart):

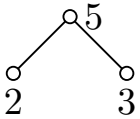
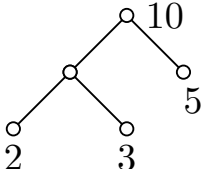
$$(3.6) \quad \begin{array}{c} \circ \\ \swarrow \quad \searrow \\ \circ \quad \quad \circ \\ w_1 \quad \quad w_2 \end{array}$$

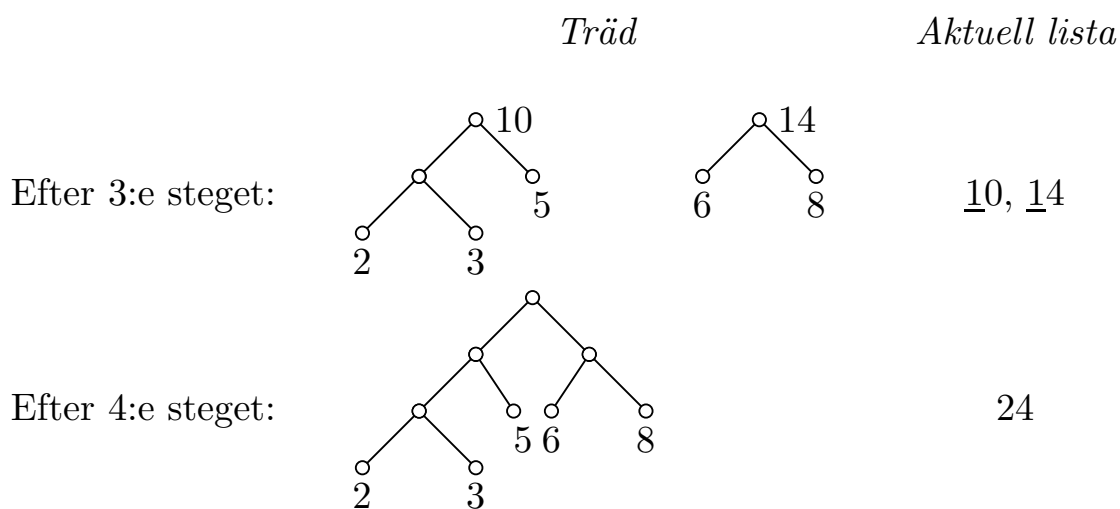
För $\ell \geq 3$ och givna vikter w_1, w_2, \dots, w_ℓ väljer vi ut de två minsta, säg w_1 och w_2 . Rekursivt kan vi anta att vi har en metod att konstruera ett optimalt träd T'_o med $\ell - 1$ löv och vikterna $w_1 + w_2, w_3, \dots, w_\ell$. Ersätt det löv i T'_o som har vikten $w_1 + w_2$ med ett delträd som i (3.6). Det träd T_o som då erhålls kan visas vara optimalt för w_1, w_2, \dots, w_ℓ .

Den rekursiva formuleringen av Huffmans algoritm är lämpad för att bevisa dess korrekthet. Mer om detta strax. För praktiskt bruk är följande omformulering av algoritmen mer användbar. (Tänk igenom att de två versionerna verkligen uttrycker samma sak!)

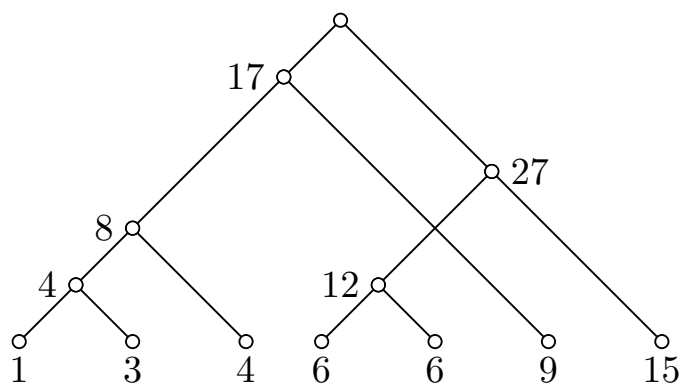
HUFFMANS ALGORITM: Givet tal w_1, w_2, \dots, w_ℓ , som vi uppfattar som märkta noder. Upprepa steget: För de två minsta talen u_1 och u_2 i aktuell lista (d.v.s. $u_1 \leq u_2 \leq u_3 \leq \dots \leq u_k, k \leq \ell$) skapa en *förälder*-nod märkt $u_1 + u_2$ som har u_1 och u_2 som barn. Aktuell lista är till att börja med w_1, w_2, \dots, w_ℓ och senare ersätts i varje steg de två barnen av sin förälder i den aktuella listan. När den aktuella listan har bara ett element stannar algoritmen, och den har då producerat ett optimalt binärt träd.

Vi illustrerar algoritmen med följande exempel. Antag de givna vikterna är 2, 3, 5, 6, 8:

	<i>Träd</i>					<i>Aktuell lista</i>
Utgångsläge:	$\overset{\circ}{2}$	$\overset{\circ}{3}$	$\overset{\circ}{5}$	$\overset{\circ}{6}$	$\overset{\circ}{8}$	<u>2</u> , <u>3</u> , 5, 6, 8
Efter 1:a steget:			$\overset{\circ}{5}$	$\overset{\circ}{6}$	$\overset{\circ}{8}$	<u>5</u> , <u>5</u> , 6, 8
Efter 2:a steget:				$\overset{\circ}{6}$	$\overset{\circ}{8}$	10, <u>6</u> , <u>8</u>



Med denna algoritm bestämmer man förvånansvärt snabbt optimala viktade binära träd även för längre talsviter. Som ett ytterligare exempel utgår vi från talsviten 1, 3, 4, 6, 6, 9, 15:



ÖVNING 10. Bestäm ett optimalt viktat binärt träd för talsviten

(a) 1, 2, 2, 3, 3, 3, 4, 4, 4, 4.

(b) 2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41.

ÖVNING 11. Säg att 7 telefonnummer, som vi kallar A, B, C, D, E, F och G, skall lagras i löven till ett binärt sökträd. Dessa telefonnummer efterfrågas med följande relativa frekvenser: A 9%, B 12%,

C 3%, D 41%, E 6%, F 22%, och G 7%. Hur skulle du välja ett lämpligt träd för detta ändamål?

ÖVNING 12. Bevisa att Huffmans algoritm är korrekt, d. v. s. att det träd som algoritmen producerar verkligen är optimalt.

[*Ledning:* Resonemanget kan föras längs följande linjer. Antag att $w_1 w_2 \leq \dots \leq w_\ell$.

(a) Antag att i ett optimalt träd lövet med vikt w_i sitter på djup d_i , $1 \leq i \leq \ell$. Visa att om $w_i < w_j$ så gäller $d_i \geq d_j$.

(b) Dra ur (a) slutsatsen att det finns ett optimalt träd T_o där w_1 - och w_2 -noderna har maximalt djup och är bröder (har samma förälder).

(c) Låt T_o vara ett optimalt träd som i (b) och låt T_H vara ett träd som producerats enligt Huffmans algoritm. Jämför $w(T_o)$ och $w(T_H)$ och utnyttja att vi rekursivt kan anta att Huffmans algoritm är korrekt för de $\ell - 1$ vikterna $w_1 + w_2, w_3, \dots, w_\ell$.]

4. Prefix-koder. Antag att vi har ett ändligt alfabet A och ett *språk* som består av vissa strängar av symboler ur A . Det kan exempelvis vara ett alfabet till ett naturligt språk som det svenska eller engelska, där symbolsträngarna är naturliga ord. Eller $A = \{0, 1, \dots, 9\}$ och *orden* vissa naturliga tal skrivna i decimalsystemet (d.v.s. i bas 10).

Problemet att koda naturliga språks bokstäver dök upp i telegrafins barndom. Den välbekanta *Morse-koden* ersätter varje bokstav med en kombination av korta och långa tonstötter, vilka vi kan uppfatta som nollor och ettor. Exempelvis:

<i>Bokstav</i>	<i>Morsekod</i>	<i>Binär form</i>
a	· —	01
e	·	0
i	··	00
m	— —	11
o	— — —	111
s	···	000
t	—	1
z	— — ··	1100

Denna kod är konstruerad med hänsyn till att olika bokstäver förekommer med olika frekvenser: vanliga bokstäver som *e* och *t* har kortare kodord än mindre frekventa bokstäver. Morsekoden är dock inte helt binär. För att kunna avkoda måste man veta var ett kodord slutar och nästa börjar. Detta sker genom ett kort uppehåll i överföringen. Denna paus har alltså också innebörd, och Morsekoden är därför *ternär*, d.v.s. den använder 3 symboler: kort, lång och paus. Tag t.ex. den välkända nödanropssignalen *S O S*, som i Morsekod lyder ”···, — — —, ···”. Om pauserna tas bort blir lydelsen ”···— — —···” vilket kan avkodas inte bara som *S O S* utan även som *iamei*, *eetze*, m.m.

Eftersom pauser är så vanligt förekommande vore det en uppenbar effektivitetsvinst att slippa koda dem. Detta kan ske om det är möjligt att ändå otvetydigt avgöra var ett kodord slutar och nästa börjar. Det enklaste sättet att uppnå detta är att välja alla kodord av samma längd. Eftersom det finns 2^k binära ord av längd k , kan ett alfabet med n symboler, där $n \leq 2^k$, alltid kodas med en så kallad *blockkod* av längd k . Exempelvis kan vi för $A = \{a, b, c, d, e\}$ välja följande kod av längd 3:

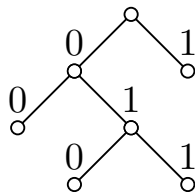
$$(4.1) \quad \begin{array}{l} a \longleftrightarrow 000 \\ b \longleftrightarrow 001 \\ c \longleftrightarrow 010 \\ d \longleftrightarrow 011 \\ e \longleftrightarrow 100 \\ f \longleftrightarrow 101 \end{array}$$

För att avkoda ett budskap som ”010000101101100” delar vi upp det i grupper om 3 och avläser *caffé*.

Nackdelen med blockkoder ur effektivitetssynpunkt är att de inte förmår ta hänsyn till bokstävernans relativa frekvenser: vanliga bokstäver tar onödigt mycket plats och ovanliga tar oskäligt liten plats. Kan Morsekodens frekvensanpassning och blockkodens paus-frihet kombineras? Svaret är *ja*, och har man studerat de binära trädens matematik är det inte svårt att se hur.

En ändlig samling K av binära ord (strängar av nollor och ettor) kallas en *prefix-kod* om det aldrig inträffar att ett ord i K utgör begynnelseavsnitt (prefix) i ett annat ord i K . Exempelvis, $\{00, 01, 1\}$ är en prefix-kod men inte $\{00, 10, 1\}$ eftersom 1 är ett prefix i 10. Det är uppenbart att det exakta villkoret för att en kod skall kunna avkodas utan någon särskild markering av var ett ord slutar och nästa börjar (paus-frihet) är att kodorden bildar en prefix-kod.

Prefix-koder kan på ett enkelt sätt erhållas ur binära träd. För varje löv v följer man stigen från roten till v och betecknar steg till vänster med 0 och steg till höger med 1. Exempelvis avläser vi från trädet



prefix-koden $\{00, 010, 011, 1\}$, och från träden i figur (2.2) får vi koderna $\{00, 01, 1\}$ och $\{0, 10, 11\}$.

ÖVNING 13. (A) Ange den prefix-kod som bestäms av trädet i figur (2.1).

(B) Finn det binära träd som ger upphov till prefixkoden $\{00, 01, 1000, 10010, 10011, 101, 11\}$.

(C) Finns det något binärt träd som ger upphov till koden $\{00, 011, 10, 11\}$? Är detta en prefix-kod?

ÖVNING 14. (A) Visa att en prefix-kod som kommer från ett binärt träd har egenskapen att varje oändlig binär talföljd $a_1a_2a_3\dots$, $a_i \in \{0, 1\}$, entydigt kan uppdelas i kodord.

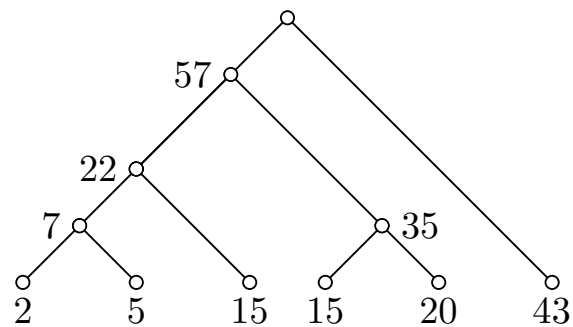
(B) Omvänt, visa att om en prefix-kod har egenskapen i (a) så kommer den från ett binärt träd.

Vi kommer nu till lösningen på det problem som ställdes i inledningen: Hur skall man på effektivaste sätt digitalt koda ett språk vars bokstäver förekommer med vissa kända frekvenser? För att slippa koda pauser mellan kodorden bör vi välja en prefix-kod, och man kan visa (detta är inte svårt) att bland prefix-koderna är de som kommer från binära träd i varje situation minst lika effektiva som de övriga. I en kod som kommer från ett binärt träd svarar kodorden mot trädets löv, och ett kodords längd (antal binära siffror) är lika med lövets djup. För ett alfabet $A = \{x_1, x_2, \dots, x_n\}$ med givna bokstavsfrekvenser $w(x_i)$, $1 \leq i \leq n$, gäller det alltså att finna ett binärt träd vars löv svarar mot A (exakt ett löv för varje bokstav) och så att summan

$$w(T) = \sum_{i=1}^n w(x_i) \cdot d(x_i)$$

är så liten som möjligt. Men detta är ju exakt det problem som Huffmans algoritm löser!

Vi illustrerar metoden med följande exempel. Antag att vi skall koda en text i alfabetet $\{a, b, c, d, e, f\}$ där de individuella bokstäverna förekommer med följande relativa frekvenser: $a = 0,43$; $b = 0,20$; $c = 0,15$; $d = 0,15$; $e = 0,05$ och $f = 0,02$. Vi använder Huffmans algoritm för att finna ett optimalt binärt träd och vikterna 43, 20, 15, 15, 5 och 2 :



Från detta avläser vi den optimala prefix-koden:

$$(4.2) \quad \begin{array}{lll} a & \longleftrightarrow & 1 \\ b & \longleftrightarrow & 011 \\ c & \longleftrightarrow & 010 \\ d & \longleftrightarrow & 001 \\ e & \longleftrightarrow & 0001 \\ f & \longleftrightarrow & 0000 \end{array}$$

Med koden (4.2) kommer den genomsnittliga kodordslängden för varje bokstav att bli

$$0,43 + 3(0,20 + 0,15 + 0,15) + 4(0,05 + 0,02) = 2,21.$$

Detta skall jämföras med kodordslängden 3,00 för blockkoden (4.1), en effektivitetsvinst med 26%.

ÖVNING 15. De svenska bokstäverna förekommer i löpande tidnings-text med de relativa frekvenserna (i procent)

a	9,26	k	3,24	u	1,75
b	1,30	l	5,25	v	2,28
c	1,23	m	3,47	w	0,06
d	4,43	n	8,71	x	0,10
e	9,89	o	4,00	y	0,65
f	2,01	p	1,67	z	0,03
g	3,16	q	0,01	å	1,58
h	1,99	r	8,34	ä	2,02
i	5,67	s	6,46	ö	1,47
j	0,62	t	8,55		

(KÄLLA: Nusvensk frekvensordbok av S. Allén m.fl., Almqvist och Wiksell, Stockholm, 1970, sid 1053.) Konstruera en optimal prefix-kod för svenska språket baserad på dessa data. Bestäm hur mycket effektivare denna kod är än den bästa blockkoden.

(Kommentar: Summan av ovanstående frekvenser är 99,2%. Återstående 0,8% av tidningstexten fördelade sig på 0,58% siffror (symbolerna 0, 1, ..., 9) och 0,22% övriga symboler. Vi bortser i denna övning från dessa, liksom från det faktum att om löpande text (och ej bara enstaka ord) skall kodas så måste alfabetet utökas med en symbol för mellanrum som blir mycket frekvent.)

Litteratur

Knuth, D., *The art of computer programming, Vol. 1: Fundamental Algorithms (2nd Ed.)*. Addison - Wesley, Reading, Ma., 1973.

Iteration av kvadratiska polynom

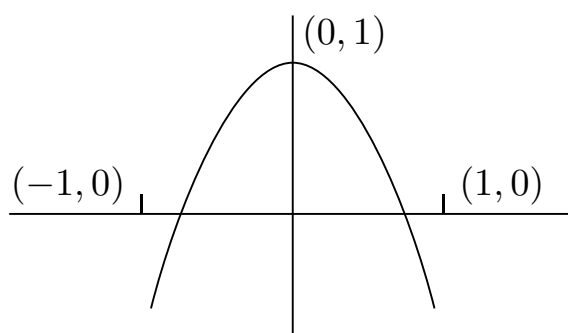
LENNART CARLESON

KTH och Uppsala universitet

Inledning. Man tror gärna att matematiken är en färdig byggnad där forskarna nu för tiden bara putsar av och förfinar de gamla teorierna. Det problem vi skall diskutera visar med all tydlighet att detta är en helt fel föreställning.

Den vanliga beskrivningen av naturfenomen är genom *lineära* ekvationer. Om ekvationen ej är lineär, gör man lineära approximationer, ty dessa är de enda för vilka man har en bra teori. Genom datamaskinerna har man fått möjlighet att med räkningar och bilder studera hur *icke-lineära* ekvationer kan uppträda. Det är en sällsam värld som träder fram, se t.ex. [1], och denna är mycket närmare verkligheten än den vanliga beskrivningen. Vi skall närmare studera ett enkelt sådant problem här.

Om a är ett tal mellan 0 och 2, $0 \leq a \leq 2$, så antar funktionen $y = 1 - ax^2$ bara värden mellan -1 och 1 då x också ligger mellan -1 och 1 .



Detta betyder att vi kan börja i någon punkt x_0 , räkna ut $x_1 = 1 - ax_0^2$, $x_2 = 1 - ax_1^2, \dots$ och få en oändlig följd tal x_0, x_1, \dots alla i $(-1, 1)$. Detta kallas en *iteration* av funktionen $1 - ax^2$.

1. Gör ett dataprogram som utför detta.
2. När vi har en följd x_0, x_1, \dots, x_n : Gör ett dataprogram så att vi kan se hur punkterna fördelas i $(-1, 1)$, ett histogram.

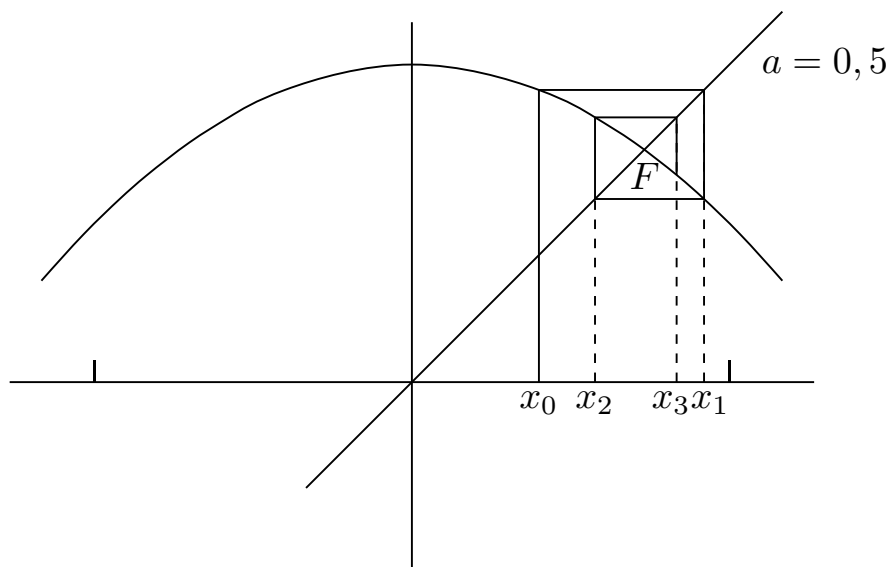
Funktion x^2 kan naturligtvis bytas mot en annan funktion. Vi skall vara intresserade av jämna, monotont växande funktioner $\varphi(x)$ med $\varphi(0) = 0$, $\varphi(1) = 1$.

3. Gör ett program för iteration av en godtycklig sådan funktion $1 - a\varphi(x)$.

Vi återgår till $\varphi(x) = x^2$. Prova lite olika startpunkter x_0 och a -värden.

Man ser att för relativt små a , så hopar sig $\{x_n\}$ i en punkt, sedan i flera punkter och när a är över 1.5 ser det ut som en kontinuerlig fördelning, speciellt när a ligger nära 2.

Man kan åskådliggöra en iteration med följande figur



$F : 1 - ax^2 = x$ kallas fixpunkt.

Man ser att när $a < 0.75$ går spiralen in mot F . Detta beror på att $1 - ax^2$ har en derivata i F som är mindre än 1.

4. Bevisa detta: både påståendet om derivatan och att x_0, x_1, x_2, \dots då för godtycklig startpunkt närmar sig F .

5. För $a = 0.75$ är derivatan $= -1$. Vad händer nu?

När vi experimenterar ser vi att för $0.75 < a < 1.25$ så hoppar punkterna mellan två värden. Vi uttrycker detta så att punkterna x_0, x_1, \dots attraheras till en *cykel* av längd 2.

Matematiskt innebär detta att funktionen

$$f(x) = 1 - a(1 - ax^2)^2$$

har en fixpunkt: $f(x) = x$, där $|f'(x)| < 1$.

6. Bevisa att denna utsaga är sann precis för $a < 1.25$.

Vi ser också att det inte spelar någon roll hur vi väljer x_0 . Följden får alltid samma uppträdande.

När vi sedan låter a växa (ganska långsamt) ser vi att först får vi en cykel av längd 4, sedan av längd 8, etc. Om vi kallar det a -värde där följden slår om från att attraheras till 2^n punkter till 2^{n+1} punkter för a_n så gäller alltså:

$$a_0 = 0.75, \quad a_1 = 1.25;$$

a_n växer men närmar sig inte 2 utan någonstans kring 1.41 finns ett gränsvärde a_∞ .

7. Gör ett dataprogram som bestämmer a_n så noggrant som möjligt.

Detta är inte helt lätt. Skillnaden mellan a_n och a_{n+1} blir snabbt liten och det fordras eftertanke att hitta på ett bra sätt att avgöra om vi närmar oss 2^n eller 2^{n+1} punkter. Programmet bör fungera upp till $n = 8 - 10$.

8. Gör upp en tabell över talen

$$b_n = \frac{a_n - a_{n-1}}{a_{n+1} - a_n}.$$

Vi ser att talen b_n ser ut att närma sig ett visst gränsvärde.

9. Byt nu ut x^2 mot $\varphi(x)$ men se till att $\varphi''(0) \neq 0$, och gör samma beräkningar av a_n och b_n .

10. Välj ytterligare en funktion $\varphi(x)$.

Du har nu upptäckt en naturkonstant av samma typ som π . På just detta sätt upptäcktes denna omkring 1980 av Feigenbaum och kallas Feigenbaums konstant. Tolkningen av experimentet är ganska djup matematik.

11. Gör nu samma experiment med x^4 och funktionen $\varphi(x)$ med $\varphi''(0) = 0$ men $\varphi^{IV}(0) \neq 0$.

Du upptäcker att vi nu får en annan konstant.

När vi sedan fortsätter att göra vårt grundexperiment för $a > a_\infty$ tenderar fördelningen av punkter att bli mer och mer kaotisk. Insprängt bland a -värden med kaotiskt uppförande finns fall med korta attraktiva cyklar.

12. Försök att hitta ett a -värde med attraktiv cykel av längd 3.

Det enda a -värde där man har en jämn fördelning av iterationspunkter är $a = 2$.

13. Gör histogram över fördelningen av punkter för $a = 2$. Vilken fördelning kan det vara?

För att lättare kunna besvara den sista frågan kan vi byta variabler. Om vi gör samma byte för både x och y betyder det bara att vi ändrar skalan. Ett lämpligt byte är

$$x = -\cos u, \quad y = -\cos v, \quad y = 1 - 2x^2.$$

14. Vilket blir sambandet mellan u och v och vilken fördelning får u -värdena? Återgå sedan till 13.

Litteratur

- [1] Mandelbrot, B., *The Fractal Geometry of Nature*. W.H. Freeman förlag 1982.
- [2] Benedicks, M., Periodfördubbling till kaos. *Nordisk matematisk tidskrift (NORMAT)* 31, (1983), s 60–172.

$$\mathbf{Om} \int_0^1 \frac{dx}{1+x^2}$$

LENNART CARLESON

KTH och Uppsala universitet

Vi börjar med att försöka uppskatta ovanstående integral, som vi kallar I , numeriskt. Vi delar in intervallet $(0, 1)$ i $n - 1$ lika delar med delningspunkterna $x_i = i/n$, $i = 0, 1, \dots, n$. Om vi kallar funktionen $f(x)$ och $y_i = f(x_i)$ är den enklaste approximationen

$$(1) \quad \frac{1}{n} [y_0 + y_1 + \dots + y_{n-1}]$$

som vi får genom att ersätta kurvan i (x_i, x_{i+1}) med den räta linjen $y = y_i$.

1. Gör ett dataprogram som utför detta.

Vi ser att det erhållna värdet är för stort eftersom vår approximerande kurva ligger över $f(x)$. Vi kan ersätta (1) med

$$(2) \quad \frac{1}{n} [y_1 + y_2 + \dots + y_n]$$

och gör då motsvarande fel i andra riktningen.

Bättre är att ersätta kurvan med den räta linjen från (x_i, y_i) till (x_{i+1}, y_{i+1}) . Vi har då att beräkna ytan av ett parallelltrapets i varje intervall som har ytan

$$\frac{1}{2n} (y_i + y_{i+1}).$$

Om vi sedan adderar alla dessa får vi den obetydliga förändringen

$$(3) \quad \frac{1}{2n} y_0 + \frac{1}{n} (y_1 + \dots + y_{n-1}) + \frac{1}{2n} y_n.$$

2. Gör dataprogram och notera vad vi får för approximativt värde på I för några olika n .

Ändå bättre borde vara att ersätta kurvan i intervallet (x_i, x_{i+2}) med en parabel $y = ax^2 + bx + c$ som för $x = x_i, x_{i+1}, x_{i+2}$ antar värdena y_i, y_{i+1}, y_{i+2} . Ytan under parabeln är

$$(4) \quad \frac{a}{3}(x_{i+2}^3 - x_i^3) + \frac{b}{2}(x_{i+2}^2 - x_i^2) + c(x_{i+2} - x_i)$$

och villkoren är

$$(5) \quad ax_j^2 + bx_j + c = y_j, \quad j = i, i+1, i+2.$$

Vi måste alltså eliminera a, b, c ur (4) och (5), räkna ut (4) och addera (4) för $i = 0, 2, 4, \dots, n-2$ och vi bör anta att n är ett jämnt tal.

För att förenkla algebran antar vi $i = 0, 1, 2$ och att $x_0 = -\frac{1}{n}$, $x_1 = 0$, $x_2 = \frac{1}{n}$. (4) blir då

$$(6) \quad \frac{2}{3} \frac{a}{n^3} + \frac{2c}{n}.$$

Ur (5) får vi att $c = y_1$ (välj $x_j = x_1 = 0$). a får vi genom att addera (5) för $j = 0, 2$:

$$\frac{2a}{n^2} + 2y_1 = y_0 + y_2.$$

Det slutliga uttrycket för (4) blir

$$\frac{1}{n} \left(\frac{2}{3} y_0 + \frac{2}{3} y_1 + \frac{2}{3} y_2 \right)$$

eller om vi återgår till de ursprungliga beteckningarna

$$(7) \quad \frac{1}{n} \frac{2}{3} (y_i + y_{i+1} + y_{i+2}).$$

Vi skall nu addera (7) för $i = 0, 2, 4, \dots, n-2$ och finner formeln

$$(8) \quad \frac{1}{n} \left\{ \frac{2}{3}y_0 + \frac{2}{3}y_1 + \frac{4}{3}y_2 + \frac{2}{3}y_3 + \frac{4}{3}y_4 + \dots + \frac{4}{3}y_{n-2} + \frac{2}{3}y_{n-1} + \frac{2}{3}y_n \right\}.$$

Kontrollera att om alla $y_i = \text{t.ex. } 1$, formeln ger värdet 1. Observera att n är ett jämnt tal.

3. Gör dataprogram som räknar ut (8).

4. Gör motsvarande kalkyler då vi använder polynom av 3:e graden och 4 konsekutiva punkter.

För vår ursprungliga funktion $f(x) = \frac{1}{1+x^2}$ har vi nu fått 3 olika serier av approximativa värden på I . Om vi noterar vad $4I$ blir ser vi att I måste vara $\pi/4$, och vi ser också hur metoderna förbättrats genom att jämföra samma n för de tre metoderna. För att beräkna I exakt inför vi som vanligt

$$A(x) = \int_0^x \frac{dt}{1+t^2}, \quad A'(x) = \frac{1}{1+x^2}.$$

Vi byter nu variabler genom att sätta $x = \operatorname{tg} u$, $u = 0$ motsvarar $x = 0$ och $u = \frac{\pi}{4}$ motsvarar $x = 1$. Då gäller

$$\begin{aligned} \frac{dA}{du} &= \frac{dA}{dx} \frac{dx}{du} = \frac{1}{1+x^2} \frac{d \operatorname{tg} u}{du} \\ &= \cos^2 u \cdot \frac{1}{\cos^2 u} = 1. \end{aligned}$$

Alltså är $A = u$ eftersom $A(0) = 0$. Vi får alltså mycket riktigt

$$I = \pi/4.$$

Vi kan serieutveckla $\frac{1}{1+x^2}$

$$\frac{1}{1+x^2} = 1 - x^2 + x^4 - x^6 + \dots$$

om $|x| < 1$. Vi har faktiskt följande likhet

$$\frac{1}{1+x^2} = 1 - x^2 + x^4 - \dots \pm x^{2n} \mp \frac{x^{2n+2}}{1+x^2}.$$

Alltså gäller för

$$g(x) = \frac{1}{1+x^2} - (1 - x^2 + x^4 - \dots \pm x^{2n})$$

att

$$|g(x)| \leq x^{2n+2}.$$

Då följer ju också

$$\left| \int_0^1 g(x) dx \right| \leq \int_0^1 |g(x)| dx \leq \int_0^1 x^{2n+2} dx = \frac{1}{2n+3}.$$

Alltså gäller

$$\left| \frac{\pi}{4} - \int_0^1 (1 - x^2 + x^4 - \dots \pm x^{2n}) dx \right| \leq \frac{1}{2n+3}$$

$$\left| \frac{\pi}{4} - \left(1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} \pm \dots \pm \frac{1}{2n+1} \right) \right| \leq \frac{1}{2n+3}.$$

Serien

$$1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} \pm \dots$$

är alltså konvergent och har summan $\pi/4$.

5. Prova att räkna ut summan för några värden på n ; vi ser att detta är en dålig metod att räkna ut π .

Vi kan göra motsvarande studie av $J = \int_0^1 \frac{dx}{1+x}$. Eftersom

$$\frac{d}{dx} \log(1+x) = \frac{1}{1+x}$$

har J värdet $\log 2$ och precis som förut ser vi att

$$1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} \pm \dots = \log 2.$$

Skulle man inte då kunna räkna ut

$$(9) \quad 1 - \frac{1}{4} + \frac{1}{7} - \frac{1}{10} + \frac{1}{13} - \dots$$

genom att studera $K = \int_0^1 \frac{dx}{1+x^3}$? Naturen har tydligen (med Newton som uttolkare) ordnat det så att I och J har ”enkla” uttryck, d.v.s. värden som är relaterade till tal som vi känner i andra sammanhang. Som vi skall se gäller detta även K , även om uttrycken blir mer komplicerade.

Vi skall alltså räkna ut K .

Följande algebraformel kan lätt kontrolleras:

$$\frac{1}{1+x^3} = \frac{1}{3} \frac{1}{1+x} - \frac{1}{3} \frac{x - \frac{1}{2}}{(x - \frac{1}{2})^2 + \frac{3}{4}} + \frac{1}{2} \frac{1}{(x - \frac{1}{2})^2 + \frac{3}{4}}.$$

Vi räknar nu ut integralen från 0 till 1 för de tre termerna i högerledet. Den första ger som tidigare $\frac{1}{3} \log 2$. I den andra observerar vi att

$$f(x) = \frac{x - \frac{1}{2}}{(x - \frac{1}{2})^2 + \frac{3}{4}}$$

antar lika stora värden med motsatt tecken i symmetriska punkter kring $\frac{1}{2}$. Alltså gäller

$$\int_0^{1/2} f(x) dx = - \int_{1/2}^1 f(x) dx.$$

Den andra termen ger alltså bidraget 0. Den tredje, L , är symmetrisk kring $x = \frac{1}{2}$ och alltså är

$$(10) \quad L = \frac{1}{2} \int_0^1 \frac{dx}{(x - \frac{1}{2})^2 + \frac{3}{4}} = \int_0^{1/2} \frac{dx}{x^2 + \frac{3}{4}}.$$

Denna integral räknar vi ut precis som förut. Vi sätter $x = \frac{\sqrt{3}}{2} \operatorname{tg} u$; ($x = 0, u = 0$) och ($x = \frac{1}{2}, u = \frac{\pi}{6}$) motsvarar varandra.

$$\frac{dA}{du} = \frac{dA}{dx} \cdot \frac{dx}{du} = \frac{4}{3} \cos^2 u \cdot \frac{\sqrt{3}}{2} \frac{1}{\cos^2 u} = \frac{2}{\sqrt{3}}$$

så att

$$A = \frac{2}{\sqrt{3}} \cdot u.$$

Alltså gäller för L ur (10), $L = \frac{2}{\sqrt{3}} \cdot \frac{\pi}{6} = \frac{\pi}{3\sqrt{3}}$.

Vi har nu räknat ut allt och funnit att serien (9) har värdet

$$\frac{1}{3} \log 2 + \frac{\pi}{3\sqrt{3}}.$$

6. Gå igenom och redovisa alla detaljer i ovanstående resonemang.

7. Faktum är nu att även serien

$$1 - \frac{1}{5} + \frac{1}{9} - \frac{1}{13} \pm \dots$$

låter sig beräknas. Arbeta med detta baserat på en formel

$$\frac{1}{1+x^4} = \frac{ax+b}{x^2 - \sqrt{2}x + 1} + \frac{cx+d}{x^2 + \sqrt{2}x + 1}$$

där först konstanterna a, b, c, d skall bestämmas.

Litteratur

Hyltén-Cavallius, C.–Sandgren, L., *Matematisk analys I* (Spec. kapitel 10). Studentlitteratur, Lund 1964.

Konvexa funktioner

URBAN CEGRELL

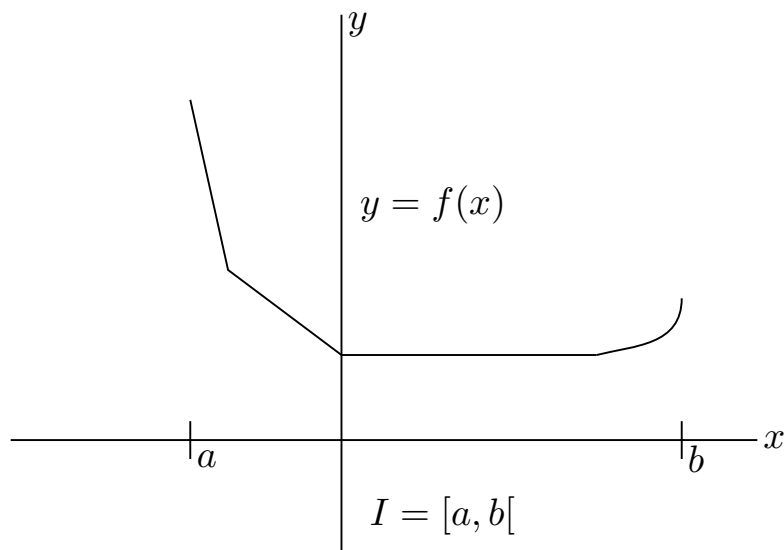
Umeå Universitet

Vi skall titta på reellvärda funktioner som har egenskapen att varje korda till funktionskurvan ligger över funktionskurvan. Sådana funktioner kallas konvexa och vi gör en ordentlig definition.

DEFINITION. Låt I vara ett intervall. Vi säger att f är *konvex* på I om för varje $x \in I$, $y \in I$, $0 \leq \lambda \leq 1$ det gäller

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y)$$

Så här kan det se ut



(Ibland talar man också om konkava funktioner, f är konkav precis då $-f$ är konvex.) Konvexa funktioner dyker upp såväl inom matematiken som i praktiska problem. Avsikten med denna uppgift är att undersöka konvexa funktioner från matematisk synvinkel.

1. Visa att

a) $f(x) = x^2$ är konvex på $I = \mathbf{R} =$ alla reella tal.

b) $f(x) = \begin{cases} x, & x > 0 \\ 0, & x \leq 0 \end{cases}$ är konvex på \mathbf{R} .

2. Visa att om f är konvex på $[a, b]$ och om vi har tal

$$a \leq x_j \leq b, \quad 1 \leq j \leq p$$

och

$$0 \leq \lambda_j \leq 1, \quad 1 \leq j \leq p$$

så att

$$\sum \lambda_j = 1$$

så

$$f\left(\sum_{j=1}^p \lambda_j x_j\right) \leq \sum_{j=1}^p \lambda_j f(x_j).$$

(Ledning: Sant då $p = 2$. Antag sant för p , visa för $p + 1$.)

3. Visa att om f är konvex så är f kontinuerlig. (Om du inte är säker på vad kontinuerlig betyder, se Appendix I i nedanstående bok.)

4. Antag att f och g är konvexa. Visa att $\max(f, g)$ också är konvex och använd detta för att visa att det finns konvexa funktioner som inte är deriverbara överallt. (Jmf.1b.) (Betr. *deriverbar* se kapitel 3 i nedanstående bok.)

5. Visa att om f är konvex på $I = (a, b)$ och om $a < x_0 < b$ så finns en funktion på formen $y = Kx + L$ där K och L är konstanter (dvs en rät linje) så att $f(x) \geq Kx + L$ för alla $x \in I$ och så att $f(x_0) = Kx_0 + L$.

6. Antag att f är två gånger kontinuerligt deriverbar (dvs f'' existerar och är kontinuerlig). Visa att f är konvex om och endast om $f''(x) \geq 0$ för alla $x \in I$.

7. Visa att $f(x) = e^{-\alpha x}$ ($-\infty < \alpha < +\infty$) är konvex på \mathbf{R} .
8. Visa att följande funktioner är konvexa på $\{x \in \mathbf{R} \mid x > 0\}$
- $f(x) = x^p$ ($1 \leq p < +\infty$)
 - $f(x) = -(x)^p$ ($0 < p < 1$)
 - $f(x) = -\log x$.
9. Visa olikheten

$$(x_1 \cdot \dots \cdot x_m)^{1/m} \leq \frac{x_1 + \dots + x_m}{m}$$

då

$$0 \leq x_j, \quad 1 \leq j \leq m$$

(Ledning: Använd 8c.)

10. Antag att ni har en reellvärd funktion $H(x, y)$ definierad på rektangeln

$$a \leq x \leq b \quad (-\infty < a < b < +\infty)$$

$$c \leq y \leq d \quad (-\infty < c < d < +\infty)$$

så att för varje fixt x är $H(x, \cdot)$ konvex och för varje fixt y är $H(\cdot, y)$ konkav. Då kan man visa att

$$\max_{a \leq x \leq b} \min_{c \leq y \leq d} H(x, y) = \min_{c \leq y \leq d} \max_{a \leq x \leq b} H(x, y).$$

Kan du visa en del av detta resultat genom att visa att vänstra ledet aldrig kan vara större än högra ledet.

11. Konvexitet går bra att definiera i rum med större dimension än 1. Vi nöjer oss med att undersöka fallet då definitionsområdet är en rektangel i \mathbf{R}^2 , $Q = \{(\xi, \eta) \in \mathbf{R}^2; a \leq \xi \leq b, c \leq \eta \leq d\}$. Vi säger att en reellvärd funktion f är konvex på Q om $f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y)$ för varje val av x och y i Q och λ , $0 \leq \lambda \leq 1$.

Visa att $(\xi^2 + \eta^2)^{\frac{1}{2}}$ är konvex.

12. Låt $g(\xi, \eta) = \xi\eta$ som ju är en väldefinierad funktion på Q . Det är ju klart (?) att för varje fixt ξ så är $g(\xi, \eta)$ en konvex funktion i η och på samma sätt är för varje fixt η , $g(\cdot, \eta)$ en konvex funktion i ξ . Är g konvex på Q ?

Litteratur

Dunkels, A., Ekblom, H., Grennberg, A., Hedberg, T., Hensvold, E., Kallioniemi, H., Näslund, R., *Derivator, Integraler och sånt ...*. Teknologkårens bokförsäljning, Högskolan i Luleå 1977.

Om Pythagoras hade varit taxichaufför i Luleå

ANDREJS DUNKELS

Högskolan i Luleå

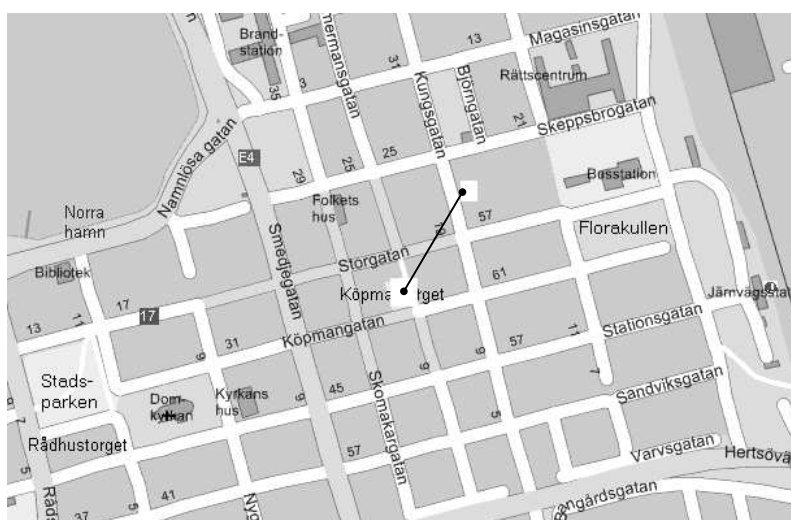


Fig 1.

Om man vill ta sig från P-platsen i hörnet av Köpmangatan och Timmermansgatan till Vinbutikens (se fig 1) så går det inte att snedda fågelvägen. Men man kan promenera – eller åka taxi – Köpmangatan till Kungsgatan och svänga åt vänster där. Om Pythagoras hade varit taxichaufför i Luleå och man frågat honom om avståndet mellan denna P-plats och Vinbutikens, så hade han svarat med summan av biten längs Köpmangatan och biten längs Kungsgatan. Han skulle inte ha sagt att avståndet är kvadratroten ur summan av dessa bitars kvadrater. Många generationer skolbarn skulle ha jublat om Pythagoras hade varit taxichaufför. Tänk att slippa kvadrater och kvadratrötter! Avståndsberäkning utan tandagnisslan och rotutdragning.

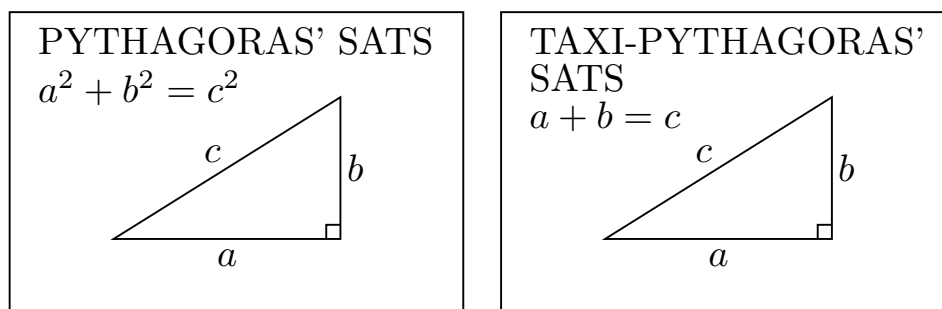


Fig 2.

Låt oss studera några geometriska begrepp och se hur de ter sig om man med avstånd mellan två punkter menar *taxi-avståndet*, dvs den sammanlagda sträckan som man tillryggalägger när man först går horisontellt och sedan vertikalt från den ena punkten till den andra.

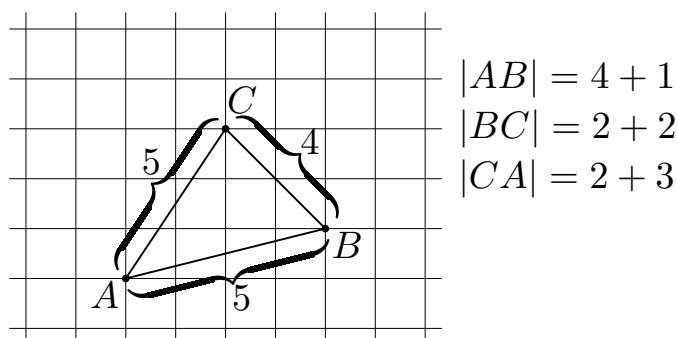


Fig 3.

Hur ser t ex en taxi-cirkel ut? I fig 3 har vi markerat två punkter B och C på taxi-avstånd 5 från A . Och det är lätt att hitta fler. Man går bara först horisontellt därefter vertikalt så att summan blir 5. Mängden av *alla* punkter på taxi-avståndet 5 från A är taxi-cirkeln med centrum i A och taxi-radie 5 (se fig 4).

Hur ser då taxi-mittpunktsnormalen till en sträcka ut? För att förenkla beskrivningen inför vi ett rätvinkligt koordinatsystem med samma skala på båda axlarna. Låt sträckan ha ändpunkterna $(0, 0)$

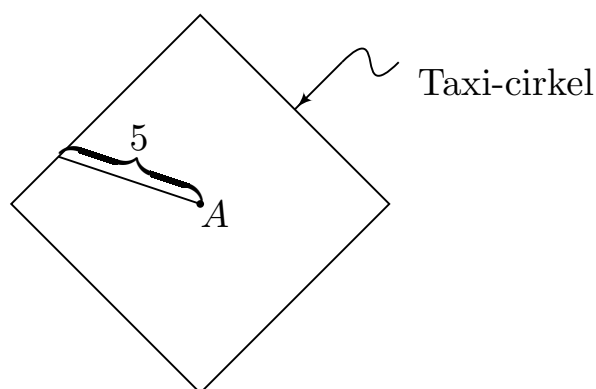


Fig 4.

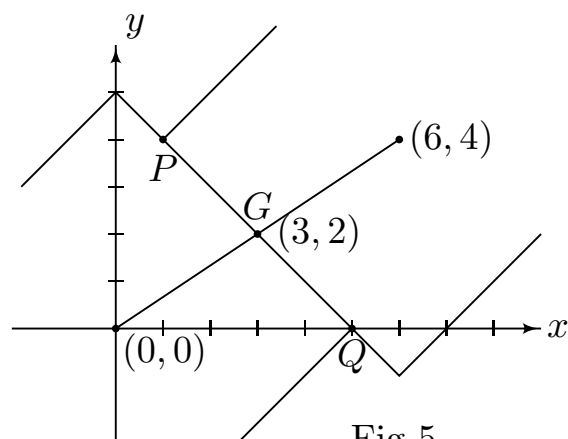


Fig 5.

och $(6, 4)$. Mittpunktsnormalen kan beskrivas utan begreppet vinkel som mängden av alla punkter på lika avstånd från sträckans ändpunkter. I vårt fall finner vi först lätt sträckans mittpunkt $(3, 2)$. Därefter kan vi till exempel rita två taxi-cirklar med radie 5, en med centrum i $(0, 0)$ och en med centrum i $(6, 4)$. Dessa cirkels gemensamma del G ligger då på den sökta taxi-normalen (se fig 5). Innehåller den flera punkter? Om vi rör oss från P (se fig 5) utanför G åt höger, rakt eller snett, så kommer vi att närma oss $(6, 4)$ och avlägsna oss från $(0, 0)$. Inga punkter till höger om P tillhör således vår sökta normal. Om vi rör oss åt vänster från P närmar vi oss $(0, 0)$ och avlägsnar oss från $(6, 4)$, dvs normalen finns inte till vänster om

P heller. Om vi emellertid rör oss rakt upp så ökar taxi-avståndet till $(6, 4)$ med precis lika mycket som till $(0, 0)$. Alla punkter rakt ovanför P ligger alltså på den sökta normalen. Ett liknande resonemang kan föras utgående från Q och vi får taxi-mittpunktsnormalen enligt fig 6.

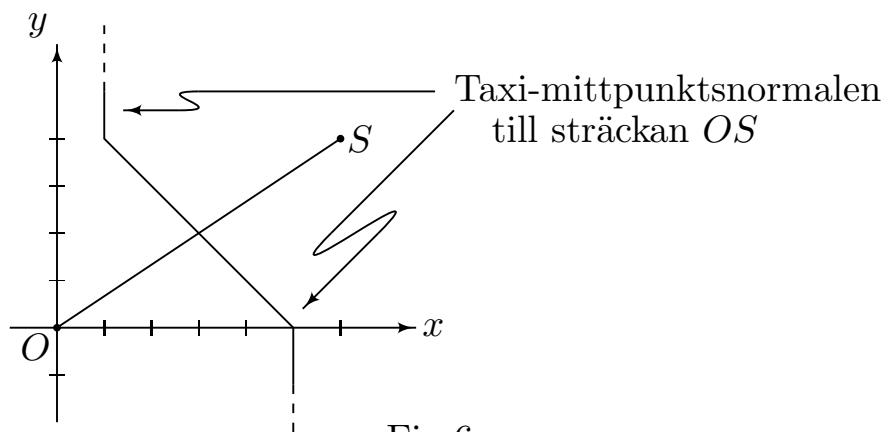


Fig 6.

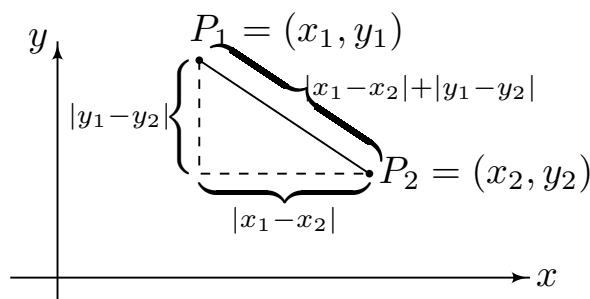


Fig 7.

Lägg märke till att vi också skulle ha kunnat söka alla skärningspunkter mellan cirklar med lika radier och centrum i $(0, 0)$ respektive $(6, 4)$.

Någon kanske föredrar att behandla problemet analytiskt. Hur ser då avståndsformeln ut? Den härleder man lätt med hjälp av Taxi-Pythagoras' sats (se fig 7). Speciellt får vi alltså att $|x| + |y|$ ger avståndet från en godtycklig punkt (x, y) till origo och $|x - 6| + |y - 4|$

ger avståndet från (x, y) till $(6, 4)$. Problemet att bestämma vår taxi-mittpunktsnormal kan således formuleras så här: Lös ekvationen

$$|x| + |y| = |x - 6| + |y - 4|.$$

(I läroböcker förekommer då och då övningar med mystiska ekvationer med absolutbelopp. Alla som undrat var de kommer ifrån har svaret här: Taxi-Pythagoras' värld.) Lösandet av ekvationen överlåter jag åt läsaren – jämför resultatet med fig 6.

Ser alla taxi-mittpunktsnormaler lika ut? Utredningen av detta överlåter jag åt var och en att genomföra. Ett förslag: pröva med änpunkterna $(1, 1)$ och $(4, 4)$ och var beredd på en överraskning.

Anledningen till att jag valt detta ämne är tvåfaldig. För det första är jag ute för att roa. För det andra vill jag ge exempel på ett område där man kan hitta många ämnen till specialarbeten i matematik för gymnasister. Man kan väl inte precis lära ut kreativt tänkande, men man kan stimulera till att öva denna förmåga. Taxi-geometrin tror jag lämpar sig utomordentligt väl för detta. Här finns mycket att undersöka, här finns problem av varierande svårighetsgrad – från direkta undersökningar av geometriska begrepp till djupare studium av euklidisk och icke-euklidisk geometri. Det finns mycket i undersökningen av taxi-mittpunktsnormalen som påminner om matematisk forskning. Dessutom föreställer jag mig att Du skall kunna komma så långt att Du själv kan formulera vettiga problem – har man funnit taxi-cirklar ger man sig på taxi-ellipser. Vad är egentligen en ellips? Hyperbel? Parabel?

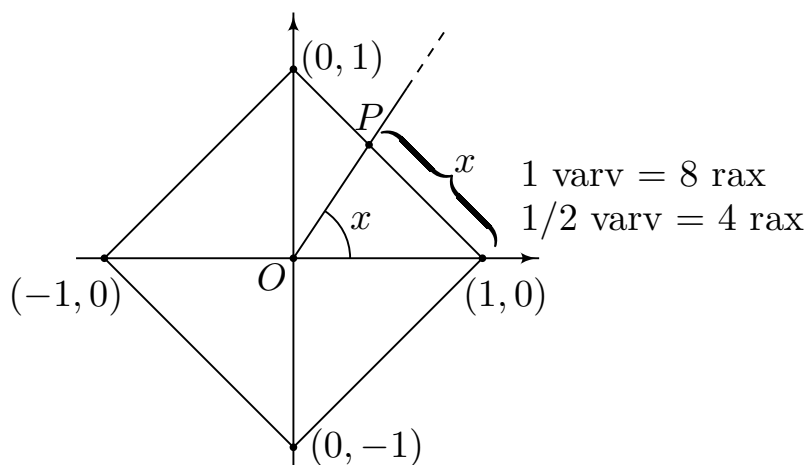


Fig 8.

Taxi-geometrin kan föras vidare till studium av taxi-trigonometri – låt oss kalla den för *trixonometri*. Utgående från taxi-enhetscirkeln gäller det bara att apa efter det vanliga sättet att definiera cosinus och sinus. Men allra först måste man skaffa sig ett lämpligt vinkelmått, det s k *raxianmättet* (se fig 8). Vi inför så de trixonometriska funktionerna (se fig 8) *tosinus* och *tinus* utgående från koordinaterna för punkten P :

$$P = (\text{tos } x, \text{tin } x).$$

Vi får lätt t ex

$$\begin{cases} \text{tos } 0 = 1, \\ \text{tin } 0 = 0, \end{cases} \quad \begin{cases} \text{tos } 1 = 1/2, \\ \text{tin } 1 = 1/2, \end{cases} \quad \begin{cases} \text{tos } 4 = -1, \\ \text{tin } 4 = 0, \end{cases} \quad \begin{cases} \text{tos } 6 = 0, \\ \text{tin } 6 = -1. \end{cases}$$

Man ser också att taxi- π är exakt 4 – avsevärt enklare än i vår vanliga omvärld. Trixonometriska ettan, som ju är en direkt följd av Taxi-Pythagoras' sats, blir

$$|\text{tos } x| + |\text{tin } x| = 1.$$

Detta är bara början. Mycket spännande återstår att pröva. Hur ser till exempel kurvorna $y = \text{tos } x$ och $y = \text{tin } x$ ut? Övriga trixonometriska funktioner? (Problem med benämningen av taxitangensfunktionen uppstår – fritt fram för fantasin!)

Här har jag tagit upp ett fåtal spridda detaljer från ett rikt fält. Som sagt, svaren är inte givna på förhand, de är ibland lätta att komma åt, ibland svåra; ibland förvånande, oftast roande.

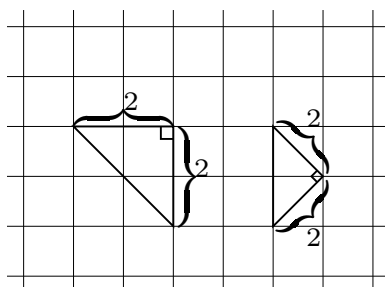


Fig 9.

Låt oss avslutningsvis titta på trianglarna i fig 9. Två sidor och mellanliggande vinkel är lika. Men trianglarna är ju inte kongruenta. I själva verket skiljer sig taxi-geometrin från den traditionella euklidiska geometrin just ifråga om det axiom som motsvarar *första kongruensfallet* och ingenting annat.

Litteratur

Kolmogorov, A.N. och Fomin, S.V., *Elements of the Theory of Functions and Functional Analysis, Vol 1*. Graylock Press, Rochester, N Y 1957. (Speciellt fig 16 s 74.)

Krause, E.F., Taxicab Geometry. *The Mathematics Teacher*, 66:8 (1973), s 695–706.

Straley, H.W., A Metric World. *The Mathematics Teacher*, 66:8 (1973), s 713–721.

Anderson, A.J., What is an Ellipse? *Mathematics Teaching*, 65 (1973), s 19, 39–42.

Georgou, W.J., Square Trigonometry. *The Pentagon. A Mathematics Magazine for Students*, 31:1 (1971), s 3–13.

Detta är en omarbetad version av en artikel med samma namn i *Elementa* 59 (1976), s 16–21. Se även Björn Gustafsson Något om metriker i denna volym.

Resträkning och ekvationer

TORSTEN EKEDAHL

Stockholms Universitet

Beskrivning av uppgiften. Specialarbetet består i att sätta sig in i hur man räknar med rester vid division med primtal, hur man löser ekvationer vid resträkning. Till sist ska man beräkna antalet lösningar till en viss sorts ekvationer och jämföra den statistiska fördelningen av dessa antal med en fördelning som man hoppas att de ska uppfylla.

Resträkning. Vi säger att två tal är *kongruenta* modulo ett tredje om de har samma rest vid division med detta tal. Till exempel så är 12 kongruent med 5 modulo 7 eftersom de båda har resten 5 vid division med 7. Man skriver också detta som $12 \equiv 5 \pmod{7}$. (Referens till detta och det som följer närmast är [Hardy-Wright:Kap. V, spec 5.2, 5.3].) Kongruenser uppfyller de vanliga räknelagarna: $12 \equiv 5 \pmod{7}$ och $(13) \equiv (20) \pmod{7}$ så $12 \cdot 13 \equiv 5 \cdot 20 \pmod{7}$ och $12 + 13 \equiv 5 + 20 \pmod{7}$. Vi har naturligtvis också t ex $2^{12} \equiv (2^3)^4 \equiv 8^4 \equiv 1^4 \pmod{7}$. Man kan också tala om kongruensekvationer: $x \equiv 3 \pmod{7}$ är en lösning till ekvationen $x^2 + 2x - 1 \equiv 0 \pmod{7}$ eftersom $3^2 + 2 \cdot 3 - 1 = 14 \equiv 0 \pmod{7}$. På samma sätt är $(x, y) \equiv (6, 0) \pmod{7}$ en lösning till ekvationen $y^2 \equiv x^3 - 1 \pmod{7}$. Man kan räkna antalet lösningar till sådana ekvationer. Först konstaterar vi att om ett tal (eller ett par av tal) är en lösning till en kongruensekvation beror bara på resterna av talet (talen) vid division med det tal som resterna räknas modulo (i våra exempel 7). Därför är det naturligt att gå genom varje rest precis en gång. Om vi till exempel vill se hur många lösningar ekvationen $x^2 + 2x - 1 \equiv 0 \pmod{7}$ har så ska vi låta x anta värdena $0, 1, \dots, 6$:

$$0^2 + 2 \cdot 0 - 1 = -1 \not\equiv 0 \pmod{7}$$

$$1^2 + 2 \cdot 1 - 1 = 2 \not\equiv 0 \pmod{7}$$

$$2^2 + 2 \cdot 2 - 1 = 7 \equiv 0 \pmod{7}$$

$$3^2 + 2 \cdot 3 - 1 = 14 \equiv 0 \pmod{7}$$

$$4^2 + 2 \cdot 4 - 1 = 23 \not\equiv 0 \pmod{7}$$

$$5^2 + 2 \cdot 5 - 1 = 34 \not\equiv 0 \pmod{7}$$

$$6^2 + 2 \cdot 6 - 1 = 47 \not\equiv 0 \pmod{7}$$

Vi ser alltså att ekvationen $x^2 \cdots x - 1 \equiv 0 \pmod{7}$ har två lösningar. På motsvarande sätt kan vi räkna antalet lösningar till ekvationen $y^2 \equiv x^3 - 1 \pmod{7}$ och då måste vi låta både x och y anta värdena $0, 1, \dots, 6$ dvs i allt måste vi räkna igenom $7 \cdot 7 = 49$ olika fall. Vi kan gör räkningarna på ett enklare sätt än att gå igenom alla dessa 49 fall och verifiera om vänsterledet har samma rest som högerledet vid division med 7. Om vi väljer en rest (t ex 3) för x så kan vi börja med att fråga oss om det överhuvudtaget finns ett y så att ekvationen $y^2 \equiv 3^3 - 1 \pmod{7}$ är uppfylld. Då $3^3 - 1 \equiv 5 \pmod{7}$ så betyder det att vi frågar oss om ekvationen $y^2 \equiv 5 \pmod{7}$ har någon lösning eller inte. Om vi går igenom alla möjligheter för y får vi:

$$0^2 \equiv 0 \pmod{7}$$

$$1^2 \equiv 1 \pmod{7}$$

$$2^2 \equiv 4 \pmod{7}$$

$$3^2 \equiv 2 \pmod{7}$$

$$4^2 \equiv 2 \pmod{7}$$

$$5^2 \equiv 4 \pmod{7}$$

$$6^2 \equiv 1 \pmod{7}$$

Vi ser alltså att ekvationen $y^2 \equiv 5 \pmod{7}$ inte har någon lösning och därför när vi försöker räkna lösningarna till $y^2 \equiv x^3 - 1 \pmod{7}$ så kan vi utesluta alla par (x, y) där $x = 3$. Å andra sidan, om $x = 2$

får vi $2^3 - 1 \equiv 0 \pmod{7}$ och från tabellen ovan ser vi att ekvationen $y^2 \equiv 0 \pmod{7}$ har precis en lösning så att bland paren (x, y) för vilka $x = 2$ får vi 1 lösning till ekvationen $y^2 \equiv x^3 - 1 \pmod{7}$. Vi kan gå igenom alla rester x och vi får då:

$$0^3 - 1 \equiv 6 \pmod{7}$$

$$1^3 - 1 \equiv 0 \pmod{7}$$

$$2^3 - 1 \equiv 0 \pmod{7}$$

$$3^3 - 1 \equiv 5 \pmod{7}$$

$$4^3 - 1 \equiv 0 \pmod{7}$$

$$5^3 - 1 \equiv 5 \pmod{7}$$

$$6^3 - 1 \equiv 5 \pmod{7}$$

Vi ser alltså att vi får sammanlagt 3 lösningar till ekvationen $y^2 \equiv x^3 - 1 \pmod{7}$; 1 för varje x -värde 1, 2 resp 4.

Vi kan byta ut 7 mot ett annat tal; av anledningar som kommer att bli klara om ett ögonblick vill vi låta detta tal vara ett udda primtal p . Om vi vill räkna antalet lösningar till ekvationen $y^2 \equiv x^3 - 1 \pmod{p}$ så kan vi resonera som i det speciella fallet 7. För ett givet värde x_0 på x får vi tre fall:

- I. $y^2 \equiv x_0^3 - 1 \pmod{p}$ har ingen lösning.
- II. $y^2 \equiv x_0^3 - 1 \pmod{p}$ har en lösning och $x_0^3 - 1 \not\equiv 0 \pmod{p}$.
- III. $x_0^3 - 1 \equiv 0 \pmod{p}$.

I det första fallet så finns det inga par (x_0, y) som är lösningar till ekvationen $y^2 \equiv x^3 - 1 \pmod{p}$. I det andra fallet finns det precis två lösningar: Det finns minst två lösningar ty om (x_0, y) är en lösning så är $(x_0, -y)$ en annan och y och $-y$ är olika rester eftersom p är ett *udda* tal. Å andra sidan finns det högst två lösningar till ekvationen $y^2 \equiv t \pmod{p}$. Om $y_0^2 \equiv t \pmod{p}$ och $y_1^2 \equiv t \pmod{p}$ så $y_0^2 \equiv y_1^2 \pmod{p}$ och därför $(y_0 - y_1)(y_0 + y_1) = y_0^- y_1^2 \equiv 0 \pmod{p}$, men om två tal ej är delbara med p så är deras produkt inte heller delbar med

p [Hardy-Wright: 1.3 Thm 3] (här använder vi att p är ett primtal, vi har t ex $2 \not\equiv 0 \pmod{6}$ och $3 \not\equiv 0 \pmod{6}$ men $2 \cdot 3 \equiv 0 \pmod{6}$). Därför har vi antingen $y_0 - y_1 \equiv 0 \pmod{p}$ dvs y_0 och y_1 är samma rester eller $y_0 + y_1 \equiv 0 \pmod{p}$ dvs y_0 och $-y_1$ är samma rester, vilket ger högst två möjligheter för en lösning till $y^2 \equiv t \pmod{p}$. Till sist, i fall III finns bara en lösning (x_0, y) ty vi måste ha $y^2 \equiv 0 \pmod{p}$ vilket igen medför att $y \equiv 0 \pmod{p}$ (eftersom y är ett primtal).

Vi ser alltså att för att räkna antalet lösningar (x, y) till $x^2 \equiv y^3 - 1 \pmod{p}$ så kan vi gå genom alla möjligheter för x ($x = 0, 1, 2, \dots, p-1$), räkna ut resten av $x^3 - 1$ vid division, kontrollera i vilket av fallen I-III vi befinner oss i och, till sist, för varje värde av x lägga 0, 2 resp 1 till antalet lösningar om vi är i fall I, II resp III.

Snabbheten i olika metoder. Det kan tyckas att vi har kommit på en mycket bättre metod än att gå igenom alla par (x, y) om vi följer detta recept och vi som i fallet $p = 7$ börjar med att skriva upp en lista på alla rester som är kvadrater. Låt oss göra en uppskattning av antalet saker vi måste göra för att komma fram till antalet lösningar. I det fall där vi går igenom alla par måste vi för varje par räkna ut y^2 och $x^3 - 1$, räkna ut deras skillnad, ta resten vid division med p och sedan se om denna rest är noll eller inte. Detta innebär 5 multiplikationer, 2 subtraktioner, en division och en jämförelse. Vi vill bara ha en grov uppskattning av hur lång tid det tar att göra våra beräkningar så vi nöjer oss med att konstatera att det tar en viss fix tid att kontrollera om ett par (x, y) uppfyller $y^2 \equiv x^3 - 1 \pmod{p}$ eller ej. I allt tar det alltså en fix tid gånger p^2 , antalet par, för att bestämma antalet lösningar (vi bryr oss inte om att multiplikationer osv tar längre tid ju fler siffror de inblandade talen är, denna tid växer rätt långsamt med p). Eftersom tiden växer

som en multipel av *kvadraten* på p så blir den snabbt stor och denna metod blir snabbt opraktisk. Om vi tittar på den andra metoden så börjar vi med att göra en lista på alla kvadrater vilket tar en fix tid gånger p , sedan räknar vi ut för varje x resten av $x^3 - 1$ och letar sedan i listan för att se om denna rest förekommer. Då listan har längd $\frac{p+1}{2}$ så är söktiden för att se om en rest är en kvadrat eller ej en fix tid gånger p och då vi måste söka för varje x blir den totala söktiden en fix tid gånger p^2 ! Vi kan förbättra tiden om vi istället börjar att göra en lista över *alla* rester vid division och sedan prickar för de som är kvadrater. Att göra detta tar en fix tid gånger p . I steget där vi tidigare sökte igenom listan för att se om resten av $x^3 - 1$ var en kvadrat eller ej kan vi nu gå in i listan och se om resten av $x^3 - 1$ är förprickad eller ej. Detta tar bara en fix tid för varje x så vi ser att den totala tiden för att bestämma antalet lösningar till $y^2 \equiv x^3 - 1 \pmod{p}$ är en fix tid gånger p , vilket är en avsevärd förbättring.

Ett problem med denna senare metod är att vi måste ha en lista av längd p med kvadraterna förprickade vilket tar plats om p är stort. Det finns ett lite mer avancerat sätt att göra det hela på som också tar en fix tid gånger p . Det hela går ut på att hitta ett sätt att avgöra om det för en given rest t finns en lösning till ekvationen $y \equiv t \pmod{p}$ utan att göra upp en lista på alla rester av kvadrater. Närmare bestämt är det så att $t^{\frac{p-1}{2}}$ har rest $p-1$ vid division med p om det ej finns någon lösning, har rest 1 vid division med p om det finns en lösning och t ej har rest 0 vid division med p och har, naturligtvis, rest 0 om t har rest 0 vid division med p (se [Hardy-Wright: 6.5, 6.6 Thm 83]). Detta är det sk Eulers kriterium. Till exempel så $3^3 = 27 \equiv 6 = 7 - 1 \pmod{7}$ och $2^3 = 8 \equiv 1 \pmod{7}$ och vi ser från tabellen ovan att $y \equiv 3 \pmod{7}$ ej har någon lösning medan $y \equiv 2 \pmod{7}$ har det. Till en början kan det tyckas att

detta inte är till någon hjälp då det behövs $\frac{p-1}{2}$ multiplikationer för att beräkna $t^{\frac{p-1}{2}}$. Detta är dock inte sant, det går att göra med ett mycket mindre antal! Ta beräkandet av 3^{11} som ett exempel. Vi börjar med att skriva 11 binärt: $10 = 2^3 + 2 + 1$. Vi beräknar sedan först $3^2 = 9$, sedan $3^{2^2} = 9^2 = 81$ och $3^{2^3} = 81^2 = 6561$. Till sist får vi $3^{11} = 3^{2^3+2+1} = 3^{2^3} \cdot 3^2 \cdot 3^1 = 6561 \cdot 9 \cdot 3 = 177147$. Vi ser på detta sätt att antalet operationer som behövs för att beräkna $t^{\frac{p-1}{2}}$ är proportionellt, inte mot p utan mot längden på den binära utvecklingen av $\frac{p-1}{2}$. Detta är av samma storleksordning som den tid det tar att multiplicera eller addera två tal av storlek ungefär p en tid som vi redan flera gånger har låtsats är konstant.

Vi får alltså följande recept för att beräkna antalet lösningar till ekvationen $y^2 \equiv x^3 - 1 \pmod{p}$, som har fördelen att vara snabb och ej kräva att vi gör långa listor.

Gör följande för $x = 0, 1, \dots, p-1$:

STEG 1: Beräkna resten vid division med p av $x^3 - 1$. Kalla denna rest t .

STEG 2: Beräkna resten av $t^{\frac{p-1}{2}}$ vid division med p om t ej är 0. Kalla denna rest r .

STEG 3: Om t är 0 lägg 1 till antalet lösningar. Om r är 1 lägg 2 till antalet lösningar annars gör inget.

För att få en uppfattning om antalet lösningar är det en bra idé att dra p från antalet lösningar (anledningen till detta blir klar om några rader). Vi kallar det tal vi då får för a_p . Då p är lika med antalet x som vi går igenom så får vi följande recept för att beräkna a_p .

Sätt a_p lika med 0. Gör följande för $x = 0, 1, \dots, p-1$:

STEG 1: Beräkna resten vid division med p av $x^3 - 1$. Kalla denna rest var t .

STEG 2: Beräkna resten av $t^{\frac{p-1}{2}}$ vid division med p om t ej är 0. Kalla denna rest r .

STEG 3: Om t är 0 lägg 0 till a_p . Om r är 1 lägg 1 till a_p annars lägg -1 till a_p .

Förväntade fördelningar. Man kan visa att t är lika med 0 för högst $3x$ (se [Hardy-Wright:VII Thm 107]). Detta fall kommer därför inte att påverka a_p speciellt mycket. I de andra fallen så verkar det rimligt att resten av $x^3 - 1$ för $x = 0, 1, \dots, p-1$ skulle vara en kvadrat ungefär lika ofta som det inte var det (eftersom det finns lika många rester skilda från 0 som är kvadraten av en rest som det finns rester som inte är det). Därför bör det vara så att vi lägger 1 till a_p ungefär lika många gånger som vi lägger till -1. Om detta är sant så bör a_p vara ungefär 0. Mer precist kan vi till och med tänka oss att det är helt slumpmässigt om, för ett givet x , resten $x^3 - 1$ är kvadraten av en rest eller inte. I så fall kan vi naturligtvis inte hoppas på att a_p skulle vara exakt 0. Å andra sidan om det är slumpmässigt så vet vi från sannolikhetsläran att med hög sannolikhet ska a_p högst vara i storleksordningen \sqrt{p} . Ett mycket berömt matematiskt resultat säger att vi alltid har $-2\sqrt{p} \leq a_p \leq 2\sqrt{p}$. Detta bekräftar till en del vår "statistiska modell" att det är slumpmässigt om $x^3 - 1$ har rest en kvadrat eller inte men visar också att situationen är mer komplicerad eftersom om det vore slumpmässigt på samma sätt som slantsingling så skulle vi alltid få åtminstone något a_p som är större än $2\sqrt{p}$ (närmare bestämt en viss proportion av a_p :na skulle vara större än $2\sqrt{p}$ eller för den delen större än ett godtyckligt tal gånger \sqrt{p}). För att få en bättre idé om vad vi kan säga om a_p :na så skalar vi dem med faktorn $2\sqrt{p}$ och sätter $b_p = a_p/2\sqrt{p}$. På så sätt kommer vi alltid att ha att $-1 \leq b_p \leq 1$. Vad kan vi nu säga om dessa tal? Erfarenheten visar att vi inte kan säga något förnuftigt om de enskilda b_p :na utan endast något om deras statistiska fördelning dvs

om vi beräknar b_p för ett antal p så kan vi se hur stor proportion av b_p :na ligger i intervallet mellan a och b för olika a och b med $-1 \leq a \leq b \leq 1$. Ett annat matematiskt resultat, inte lika berömt som det förra, säger att b_p :na är jämnt fördelade bortsett från att hälften av dem är 0 dvs om vi beräknar b_p för tillräckligt många p så kommer andelen b_p som ligger i intervallet mellan a och b att komma godtyckligt nära $\frac{b-a}{4}$ plus $\frac{1}{2}$ om 0 ligger i intervallet (4:an i nämnaren kommer från att vi vill ha proportionen 1 när $a=-1$ och $b=1$). Med andra ord, bortsett från 0, är inget värde på b_p mer sannolikt än något annat.

Vi kan ersätta $x^3 - 1$ med ett annat tredjegradspolynom t ex $x^3 + 2x + 1$. Om vi definierar a_p och b_p på samma sätt men nu för ekvationen $y^2 \equiv x^3 + 2x + 1 \pmod{p}$ så gäller fortfarande att $-2\sqrt{p} \leq a_p \leq 2\sqrt{p}$ och därför $-1 \leq b_p \leq 1$. Denna gång ska b_p :na däremot inte vara jämnt fördelade. Istället ska proportionen av de b_p som finns i intervallet mellan a och b vara nära

$$\frac{2}{\pi} \int_a^b \sqrt{1-x^2} dx$$

(denna integral kan naturligtvis räknas ut men den ser trevligare ut som den är). Att detta gäller för $x^3 + 2x + 1$ är något man inte, till skillnad från fallet $x^3 - 1$, vet utan för tillfället endast hoppas på. Även om man inte kan bevisa det är det något man kan undersöka rimligheten av genom att beräkna b_p för ett antal p och jämföra de proportioner man får med vad man hoppas på.

Situationen för ett godtyckligt tredjegradspolynom förväntas vara densamma som för ett av de polynom vi diskuterat. Detta är inte sant för några sällsynt förekommande polynom: de som har ett nollställe gemensamt med sin derivata som t ex $x^3 + x^2$; $x = 0$ är ett nollställe till både $x^3 + x^2$ och dess derivata $3x^2 + 2x$. Om vi

istället tittar på $x^3 + 2x + 1$ så är dess derivata $3x^2 + 2$ så om de har ett gemensamt nollställe x så är

$$\begin{aligned}x^3 + 2x + 1 &= 0 \\3x^2 + 2 &= 0.\end{aligned}$$

Om vi multiplicerar den första ekvationen med 3 och den andra med x och subtraherar får vi

$$\begin{aligned}4x + 3 &= 0 \\3x^2 + 2 &= 0.\end{aligned}$$

Om vi multiplicerar den första ekvationen med $3x$ och den andra med 4 och sedan subtraherar får vi

$$\begin{aligned}4x + 1 &= 0 \\9x - 8 &= 0.\end{aligned}$$

Om vi multiplicerar den första ekvationen med 9 och den andra med 4 och sedan subtraherar får vi

$$41 = 0.$$

Av detta ser vi att $x^3 + 2x + 1$ inte har något nollställe gemensamt med sin derivata. Om vi istället hade räknat rester vid division med 41 så hade vi på slutet istället fått

$$41 \equiv 0 \pmod{41}$$

vilket faktiskt är sant och man kan kontrollera att $x = 10$ är en gemensam lösning till $x^3 + 2x + 1 \equiv 0 \pmod{41}$ och $3x^2 + 2 \equiv 0 \pmod{41}$. Precis som vi inte är intresserade av polynom som har ett nollställe gemensamt med sin derivata så är vi heller inte intresserade

av vissa ”dåliga” primtal. Dessa är de primtal för vilka det polynom vi är intresserade av har ett nollställe gemensamt med sin derivata vid resträkning vid division med primtalet i fråga. I fallet $x^3 + 2x + 1$ borde vi alltså hoppa över 41 när vi beräknar b_p :na och i alla fall när vi räknar ut proportionen av b_p i ett visst intervall. Då det bara rör sig om ett värde har det dock inte någon större betydelse.

Om vi nu betraktar ett tredjegradspolynom som ej har något nollställe gemensamt med sin derivata så gäller igen att $-1 \leq b_p \leq 1$ för alla udda primtal p . Vidare hoppas man att den statistiska fördelningen antingen ska vara som i fallet $x^3 - 1$ dvs proportionen b_p i intervallet från a till b ska vara ungefär $\frac{b-a}{4}$ plus $\frac{1}{2}$ om 0 ligger i intervallet eller som i fallet $x^3 + 2x + 1$ dvs proportionen b_p i intervallet från a till b ska vara ungefär $\frac{2}{\pi} \int_a^b \sqrt{1-x^2} dx$. Det finns ett sätt att bestämma direkt om det första fallet ska gälla och när det är så så vet man att den statistiska fördelningen är den man hoppas på. Å andra sidan finns det inget exempel på ett polynom som inte faller i den första kategorin för vilket man vet att den statistiska fördelningen är den man önskar. Vad man har gjort är att beräkna b_p för ett antal polynom och ett antal primtal och se om proportionerna är de de borde vara. Det är dessutom det andra fallet som är det vanligaste; om man tar ett polynom på måfå så är chansen mycket liten att den faller i den första kategorin.

Om man tittar på polynom av andra gradtal så händer följande. Om graden är 1 händer inget spännande: a_p är alltid 0. Om graden är 2 är det knappast mer intressant: a_p är alltid 0 eller -1. Om graden är 4 är situationen precis som i fallet grad 3. Om graden är 5 eller större är situationen mer komplicerad t ex om graden är 5 vet man bara att $-4\sqrt{p} \leq a_p \leq 4\sqrt{p}$ och för fördelningen av b_p :na finns det fler än två möjligheter.

Närmare beskrivning av specialarbetet. Man ska till en början förstå kongruensräkning; i den mån ovanstående inte är tillräckligt kan man läsa mer i t ex [Hardy-Wright: Kap V, VI, VII]. Det finns möjlighet att lägga mer eller mindre tid på denna del, t ex är det inte nödvändigt att veta varför Eulers kriterium fungerar för att använda det. Den andra delen av specialarbetet skall sedan ägnas åt att undersöka fördelningen av b_p :na för några tredjegradspolynom och primtal. Det är knappast realistiskt att genomföra detta för hand men den enklaste formen av programmerbar fickräknare är tillräcklig (några exempel måste naturligtvis räknas ut för hand för att kontrollera att man har programmerat rätt). Alla de steg som beskrivits ovan går att programmera, även om en del saker fordrar eftertanke som t ex att man måste se till att heltalsberäkningarna utförs exakt och inte i flyttalsform. Resultaten skall sedan jämföras med den förväntade fördelningen. Detta kan ske i tabell- eller diagramform men man kan också tänka sig någon form av statistisk analys. I det fall det finns intresse för numeriska beräkningar kan man jämföra hastigheten av de olika metoder som beskrivits ovan och göra olika försök att öka genom att ändra i programmen.

Litteratur

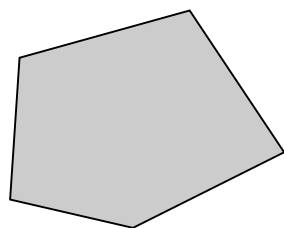
Hardy, G.H. & Wright, E.M., *An Introduction to the theory of numbers*. Fifth edition, Oxford Univ. Press, Oxford 1977.

Polyedrar och polygoner

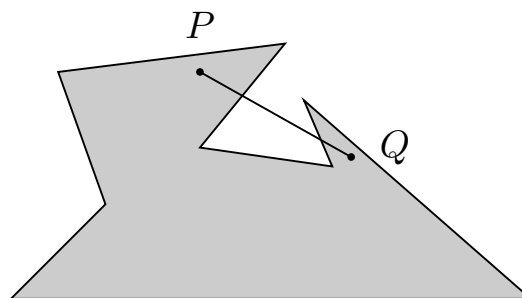
RALF FRÖBERG

Stockholms universitet

En polygon är en plan figur som begränsas av linjestycken. Vi begränsar oss till konvexa, begränsade polygoner. (En polygon F är konvex om den till varje par av punkter P och Q innehåller hela linjestycket mellan P och Q .)



Konveks



Ej konveks

En *kant* till en polygon är ett linjestycke som begränsar polygonen, ett *hörn* är skärningspunkten mellan två kanter. Om K är antalet kanter och H antalet hörn gäller $K - H = 0$. En *regelbunden* polygon är en polygon med "identiska" kanter och hörn. (Alla kanter är lika långa, vinklarna i alla hörn är lika stora.) En polygon med n kanter kallas n -gon. En regelbunden 3-gon är alltså en liksidig triangel, en regelbunden 4-gon är en kvadrat o s v.

ÖVNING 1. Bestäm vinkelsumman i en n -gon.

ÖVNING 2. Bestäm vinkeln vid varje hörn i en regelbunden n -gon.

En polyeder är en 3–dimensionell figur som begränsas av polygoner. Vi begränsar oss till konvexa polygoner. De 2–dimensionella begränsningsytorna kallas *sidor*. Skärningen mellan två närliggande sidor kallas *kant* och skärningspunkten mellan två närliggande kanter kallas *hörn*. En polyeder har minst 4 hörn. En polyeder med 4 hörn har 6 kanter och 4 sidor och kallas *tetraeder*. Om S är antalet sidor, K antalet kanter och H antalet hörn gäller alltså $S - K + H = 2$. Vi ska visa att denna formel (Eulers relation) gäller för alla polyedrar. Vi har visat att den gäller för polyedrar med 4 hörn. Antag nu att en polyeder S har 5 hörn. Välj ut ett hörn P och låt närliggande hörn vara P_1, P_2, \dots, P_n . (Det gäller här att $n = 3$ eller $n = 4$.) Vi ska se vad som händer med formeln $S - K + H$ då vi tar bort hörnet P och alla kanter och sidor som P ligger på och lägger till en ny sida $P_1P_2 \dots P_n$ (om den inte redan finns). Om den nya figuren blir en tetraeder har vi tagit bort ett hörn, tre kanter och tre sidor samt fått en ny sida. Vi ser alltså att $S - K + H$ är lika stor för de två figurerna. I annat fall får vi en 4–gon och vi har tagit bort ett hörn, fyra kanter och fyra sidor.

ÖVNING 3. Visa att formeln $S - K + H = 2$ gäller även i detta fall.

ÖVNING 4. Vi vet nu att $S - K + H = 2$ gäller för alla polyedrar med högst 5 hörn. Visa på samma sätt som ovan att formeln är sann om $H = 6$.

Vi antar nu att Eulers relation är visad för alla polyedrar med N hörn. För en polyeder med $N + 1$ hörn väljer vi ut ett hörn P och antar att P_1, P_2, \dots, P_n är de hörn som är grannar till P . Vi gör som förut en ny figur med N hörn genom att ta bort P och alla kanter och sidor som innehåller P . Vi lägger till sidan $P_1P_2 \dots P_n$ om den inte redan finns. Om den nya figuren är en polyeder med N hörn har vi tagit bort ett hörn, n kanter och n sidor och fått en ny sida, så $S - K + H$ ändras inte i detta fall. Om den nya figuren är en N –gon

har vi tagit bort ett hörn, n kanter och n sidor och inte fått någon ny sida. Vi ser att $S - K + H$ minskar med ett när vi tar bort ett hörn i detta fall. I en N -gon gäller $S - K + H = 1$ eftersom $S = 1$ och $K - H = 0$, så vi får $S - K + H = 2$ för den polyeder vi startade med även i detta fall. Vi har nu visat att formeln är sann för alla polyedrar. (Eftersom den är sann för polyedrar med 6 hörn är den sann för polyedrar med 7 hörn och alltså för polyedrar med 8 hörn o s v.) En polyeder kallas *regelbunden* (eller en platonsk kropp) om alla sidor, kanter och hörn är "identiska". En regelbunden polyeder har alltså regelbundna n -goner som sidor och i varje hörn möts lika många sidor och vinklarna mellan dessa är lika stora.

ÖVNING 5. Visa att en regelbunden polyeder har antingen trianglar, kvadrater eller regelbundna 5-hörningar som sidor. (Använd övning 2.)

Antag nu att r sidor möts i varje hörn och att varje sida är en n -gon.

ÖVNING 6. Visa att $(r - 2)(n - 2) < 4$. (Använd övning 1.)

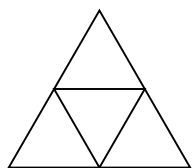
ÖVNING 7. Visa att de enda lösningarna till ovanstående olikhet är $(r, n) = (3, 3), (3, 4), (3, 5), (4, 3)$ och $(5, 3)$.

Det kan bara finnas en figur för varje möjlig lösning eftersom utseendet vid ett hörn bestämmer hela figuren (bortsett från storleken). För varje möjlig lösning kan man konstruera en regelbunden polyeder, de blir i ordning tetraeder, oktaeder, ikosaeder, kub och dodekaeder, Antalet sidor är 4, 8, 20, 6 resp 12.

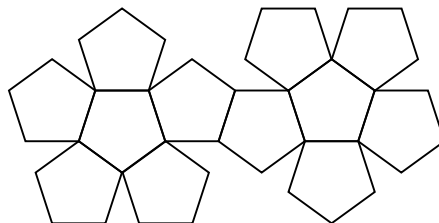
Låt K vara en regelbunden polyeder. Vi bildar en ny polyeder K' genom att som hörn ta mittpunkterna på varje sida i K och som kanter linjestyckena mellan par av närliggande mittpunkter. Den nya polyedern K' kallas dualen till K och blir regelbunden.

ÖVNING 8. Bestäm dualen till var och en av de fem regelbundna polyedrar.

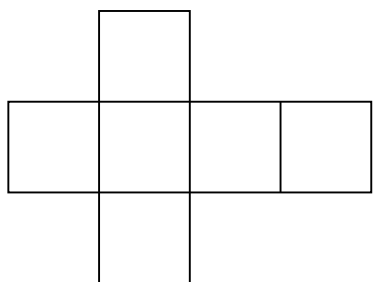
ÖVNING 9. Förstora följande figurer på ett pappark och tillverka de fem regelbundna polyedrar.



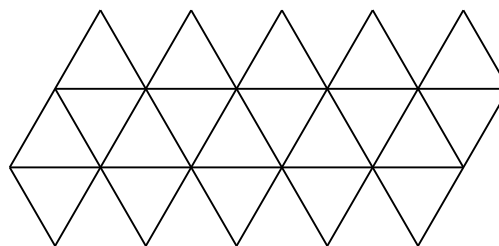
Tetrahedron $\{3, 3\}$



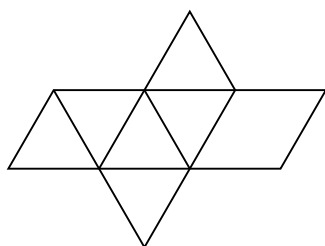
Dodecahedron $\{5, 3\}$



Cube $\{4, 3\}$



Icosahedron $\{3, 5\}$



Octahedron $\{3, 4\}$

Lotto, ett skicklighetsspel!

JAN GRANDELL

KTH

1. Inledning. Du håller nog med om att om man köper en lott så är det bara en fråga om tur om man vinner och hur mycket man vinner. På samma sätt håller du nog med om att om man tippar så beror resultatet inte bara på tur, utan också på hur mycket man vet om fotboll och om olika lag. Ska man spela så ska man sikta högt, så vi bryr oss bara om högsta vinsten och om tretton rätt. Tips är lite komplicerat, eftersom man kan vinna olika mycket beroende på om det är en svår omgång eller inte. Det kan vara dumt att helt följa tidningarnas tips, eftersom om deras tips slår in så har nog många tippat precis som de, och vinsten blir inte så hög.

Hur är det nu med Lotto? I Lotto tippar man sju nummer bland talen 1 till 35. Precis som förut siktar vi högt och bryr oss bara om 7 rätt. Tänk Dig nu att du har spelat på Lotto och sedan får du veta att du har en rad med 7 rätt. När du jublat färdigt så blir du nog lite nervös, och börjar fundera på hur mycket du har vunnit. Det bästa är naturligtvis om du är ensam vinnare, för då är din lycka gjord.

2. Beskrivning av uppgiften. Din uppgift är att utreda följande:

Chansen att få sju rätt på Lotto beror bara på tur, men hur mycket man vinner om man har sju rätt beror även på skicklighet. (Tyvärr kommer du inte att få lära Dig hur du skall tjäna pengar på Lotto, men du får lära Dig hur du ska undvika att förlora onödigt mycket.)

Jag föreslår att du försöker Dig på följande uppgifter.

1. Räkna ut hur stor sannolikheten är för att få 7 rätt på en Lottorad. Försök att motivera varför raderna 1, 2, 3, 4, 5, 6, 7 och 3, 15, 18, 27, 28, 31, 35 har samma sannolikhet att ge 7 rätt. I tidningarna kan man läsa att ”nu har nummer 18 inte kommit på 10 veckor, så nu måste nummer 18 snart komma”. Vad tror du om sånt?
2. (Din huvuduppgift.) Kalla den sannolikhet som du nyss räknat ut för p och antag att N Lottorader har lämnats in. Försök visa att

$$\text{Sannolikheten för att } k \text{ rader har 7 rätt är } \frac{(Np)^k}{k!} e^{-Np},$$

där $k! = k \cdot (k-1) \cdot \dots \cdot 2 \cdot 1$, gäller med (nästan) perfekt noggrannhet om de tippade raderna får 7 rätt oberoende av varandra. Om detta gäller så säger man att antalet rader med 7 rätt är Poissonfördelat.

För att klara den här uppgiften måste du läsa ganska mycket. Ska jag försöka hjälpa dig lite på traven så tycker jag att du ska göra så här.

- (a) Antag att olika rader får 7 rätt oberoende av varandra.
- (b) Övertyga Dig om att antalet rader med 7 rätt är binomialfördelat. Vad man menar med att något är binomialfördelat kan du läsa om i Bloms bok.
- (c) Utnyttja att binomialfördelningen ibland kan approximeras med Poissonfördelningen.

3. I tidningarna kan du få reda på hur många som har fått 7 rätt vecka för vecka. Om vi antar att ungefär lika många Lottorader lämnas in varje vecka, dvs att N är lika varje vecka, så borde ”fördelningen” för 7 rätt stämma med Poissonfördelningen. Om du ritar ett stolpdiagram så kommer du att se att det inte stämmer. Troligen finns det en eller annan vecka där antalet rader med 7 rätt är alldeles ”för många”. När du gör detta så bör du välja Np som medelantalet som har haft 7 rätt under de veckor som du tittat på.

Du bör åtminstone titta på resultaten från ett halvt år. Sådana uppgifter kan du hitta i tidningarna.

4. Fundera på varför Poissonfördelningen inte stämmer.

En möjlighet är att det var fel att anta att ungefär lika många Lottorader lämnas in varje vecka. Lite fel är det naturligtvis, men det är nog inte så fel att det förklarar varför det inte stämmer.

En annan möjlighet är att det var fel att anta att raderna får 7 rätt oberoende av varandra. Det är nog här felet ligger. Nu är begreppet ”oberoende” väldigt svårt, så bekymra Dig inte så mycket om du tycker att stycket som kommer nu verkar krångligt. När du läst det som du måste för att göra specialarbetet tror jag att stycket inte verkar så konstigt längre.

Nu kan man tycka, att eftersom en person ju inte vet hur andra tippas, så borde de inlämnade raderna vara tippade oberoende av varandra. Så är det nog också, men detta är faktiskt inte samma sak som att olika rader får 7 rätt oberoende av varandra. Detta att många tippas system strider mot oberoendet, men det ”drar åt fel håll”. Förklaringen måste i praktiken vara att vissa tippade rader är vanligare än andra. Matematiskt betyder oberoende att om vi har två händelser A och B så är A och B oberoende om $P\{A \text{ och } B\} = P\{A\}P\{B\}$. Om A och B är händelserna att två godtyckligt valda tippade rader har 7 rätt, så blir den betingade sannolikheten för A , när vi vet att B har inträffat, större än sannolikheten för A . Förklaringen är att om B inträffar så var det nog en ”vanlig” rad som fick 7 rätt, och därför är chansen att en annan rad också skall ha 7 rätt större. För att förstå det här kan du tänka dig att bara två rader tippas, t.ex. raderna 1, 2, 3, 4, 5, 6, 7 och 3, 15, 18, 27, 28, 31, 35. Låtsas nu att alla bara tippas en rad och att var och en väljer raden 1, 2, 3, 4, 5, 6, 7 med sannolikheten 0,5 och raden 3, 15, 18, 27, 28, 31, 35 med sannolikheten 0,5. Då tippas alla oberoende

av varandra, men de vinner inte oberoende av varandra. Antingen vinner ju ingen, eller också vinner ju väldigt många. I detta fall är den betingade sannolikheten för A , när vi vet att B har inträffat 0,5. Du tycker kanske att det här exemplet är väldigt fånigt, men det är ofta bra att tänka på "extrema" fall för att förstå vad som händer.

Slutsatsen är att olika rader är olika vanliga, och man kan därför, ungefär som på tips, tala om tala om lätta och svåra omgångar. Skälet kan vara att en del tippar snygga mönster eller har "favorital". Andra tippar "enkla" rader. Om raden 1, 2, 3, 4, 5, 6, 7 ger 7 rätt lär vinsten bara bli några hundra. Det lär också vara så att låga nummer förekommer mer än höga nummer. Det kan bero på att den som tippar börjar i nummerordning och att talen "tar slut". Många försöker nog slumpa ut talen, men det visar sig att de flesta tror att slumpen sprider ut talen jämnare än vad den gör. För att du skall få en känsla för hur slumpen uppträder, så har jag "slumpat" 100 rader.

Fråga gärna några som brukar spela Lotto, hur de väljer sina tal.

5. Fundera på hur man "bör" tippa i Lotto.

Det gäller att tippa en så "ovanlig" rad som möjligt. Tänk på att om ett par till i Sverige har samma rad, så räcker det att ett för att raden ska vara "vanlig" och därför inte bra.

Mitt förslag är att man lägger 35 lappar i en skål och drar 7 stycken. Visserligen har man ingen garanti för att bli ensam om raden, men man undviker nog "vanliga" rader.

Slumpade Lottorader.

1 5 8 13 22 30 35	2 13 14 20 25 29 33	13 17 18 21 23 30 35
9 10 15 18 22 29 34	1 5 11 21 25 31 35	8 13 14 17 18 28 29
5 7 16 17 27 34 35	10 16 21 26 33 34 35	1 3 7 14 19 26 32
6 10 18 19 20 28 30	1 12 19 27 28 31 32	6 7 15 19 22 29 31
3 4 18 20 24 27 30	6 7 8 9 13 14 22	10 11 15 16 23 25 32
3 7 14 16 20 21 25	2 4 5 21 24 33 34	3 9 17 26 30 32 33
2 9 12 13 16 22 35	4 19 22 23 27 29 33	4 8 10 16 18 28 33
4 15 16 25 28 29 35	12 16 17 28 29 30 35	1 9 15 20 30 31 32
6 12 19 23 30 31 33	6 7 10 15 20 22 25	2 11 15 24 33 34 35
11 14 15 19 23 26 32	1 3 12 15 19 23 33	11 16 17 21 23 26 31
1 3 5 8 20 24 32	2 3 13 15 21 22 28	1 2 9 18 24 33 35
14 15 16 20 23 27 33	2 12 16 23 29 33 35	13 14 21 23 24 27 28
2 11 14 16 21 34 35	7 10 11 15 21 28 33	2 3 11 20 22 25 29
4 6 8 11 14 20 24	2 5 6 13 16 19 23	13 14 17 19 20 24 25
6 10 13 15 17 20 34	9 12 18 21 23 27 33	1 7 13 14 18 21 22
5 6 22 24 31 32 34	10 11 14 15 16 20 26	1 11 15 21 22 26 31
4 6 13 15 20 25 29	1 9 12 23 31 34 35	3 8 10 13 15 26 28
7 8 13 25 26 33 35	8 11 14 16 18 23 29	10 11 13 17 22 32 34
3 5 15 16 18 22 23	1 3 6 13 17 24 35	3 5 6 8 23 32 33
2 3 4 5 16 22 35	5 14 18 21 22 26 31	3 6 8 17 27 28 35
4 15 17 24 25 29 32	7 15 16 21 25 34 35	15 17 18 25 31 32 34
3 12 14 16 24 26 28	6 13 25 26 31 33 34	9 19 23 25 28 33 34
1 3 4 10 15 17 35	6 12 15 17 19 32 35	1 9 22 26 29 30 33
1 12 14 15 20 29 30	3 4 6 8 13 20 34	4 6 11 22 23 29 35
11 13 14 15 20 22 32	6 13 20 22 23 24 29	6 7 17 27 32 34 35
2 5 6 10 17 18 31	4 8 17 26 27 30 35	7 16 18 20 29 30 32
9 10 12 20 24 25 29	13 16 17 20 21 30 35	6 11 15 17 22 26 28
4 8 11 15 25 26 27	13 18 21 28 30 31 35	5 18 19 22 23 24 32
1 4 6 10 15 20 35	6 7 10 11 16 28 35	7 9 12 14 15 22 30
1 2 3 4 8 11 27	2 5 10 13 17 26 35	2 16 19 22 24 34 35
5 6 12 14 18 21 23	5 6 11 24 27 28 32	2 7 8 10 19 27 33
10 11 19 23 25 26 27	1 2 4 12 14 24 31	4 6 7 9 13 20 25
4 10 11 13 29 30 34	3 5 7 22 26 29 31	1 5 11 14 19 26 30
3 4 10 18 19 25 30		

Varför bör Du inte använda någon av dessa rader om Du spelar på Lotto?

LYCKA TILL MED DITT ARBETE!

Jag hoppas att du – när du har gjort arbetet – tycker att sannolikhetslära verkar roligt. Vill du lära dig mera så skall du läsa ämnet matematisk statistik. Matematisk statistik finns både på universiteten och på de tekniska högskolorna.

Litteratur

En bra och ganska lättläst bok som passar för din uppgift är:
Blom, G., *Sannolikhets teori och statistikteori med tillämpningar*.
Bok C, Studentlitteratur, Lund 1980.

Något om algebraiska kurvor

BJÖRN GUSTAFSSON

K T H

Inledning. De enklaste matematiska funktionerna är de som kan definieras direkt med hjälp av de fyra räknesätten, dvs polynomen, (bara tre räknesätt behövs) och de rationella funktionerna. Det är därför rimligt att säga att till de enklaste kurvorna i planet hör de som kan parametreras med rationella funktioner samt (allmänare) sådana som är nivåkurvor till polynom (rationella funktioner ger inget ytterligare här). Studiet av dessa typer av kurvor och deras högre-dimensionella motsvarigheter kallas algebraisk geometri och är en gren av matematiken som, trots hundratals år på nacken, fortfarande är en av de mest livskraftiga och kanske t o m är extra aktuell idag på grund av helt nya tillämpningar inom modern fysik.

Denna uppgift skall ge några smakprov på algebraisk geometri med tillämpningar.

- a) Kurvan $x^2 + y^2 = 1$ i planet ($= \mathbf{R}^2$) har som bekant en parametrisering $x = \cos t$, $y = \sin t$ (t reell parameter). Visa att det också finns en rationell parametrisering, dvs att det finns två (icke-konstanta) rationella funktioner $q(t)$ och $r(t)$ så att $q(t)^2 + r(t)^2 \equiv 1$. (Rationell funktion = kvot mellan två polynom. Exempel: $\frac{t^3 - 3t + 1}{5t^2 + 2}$.)
- b) Visa mer allmänt att varje irreducibel andragradskurva $p(x, y) = 0$ är rationell, dvs kan parametreras med rationella funktioner. (*Irreducibel* betyder att polynomet $p(x, y)$ inte kan skrivas som produkten av två polynom av lägre gradtal, inte ens om man tillåter dessa att ha komplexa koefficienter. Exempel: $x^2 + y^2 - 1$ är irreducibelt men inte $x^2 + y^2$, ty $x^2 + y^2 = (x + iy)(x - iy)$.) Detta

resultat har som konsekvens att problemet att beräkna integraler såsom

$$\int \frac{dx}{\sqrt{1-x^2}}, \quad \int \frac{2x+1}{x\sqrt{1+x+x^2}} dx$$

(eller, allmänt, integraler av typen $\int r(x, \sqrt{ax^2+bx+c})dx$ där $r(x, y)$ är en rationell funktion i två variabler) kan återföras till problemet att beräkna en integral av en rationell funktion (i en variabel).

Visa detta.

Ledning: Om exempelvis $\sqrt{1-x^2}$ förekommer i integranden, sätt $y = \sqrt{1-x^2}$ så att $x^2 + y^2 = 1$ och gör variabelsubstitution till en rationell parameter i integralen.

- c) Visa att kurvan $x^3 + y^3 = 1$ (eller allmännare $x^n + y^n = 1$, $n \geq 3$) inte är rationell. *Ledning:* Anta $\frac{p(t)^n}{r(t)^n} + \frac{q(t)^n}{r(t)^n} \equiv 1$, där p, q, r är polynom, som vi kan anta inte innehåller någon gemensam faktor. Härled först relationen

$$\frac{qr' - rq'}{p^{n-1}} = \frac{rp' - pr'}{q^{n-1}} = -\frac{pq' - qp'}{r^{n-1}}.$$

Ovanstående uttryck definierar en rationell funktion, som vi kan skriva u/v där u och v är polynom utan gemensam faktor. Härled nu en motsägelse genom att å ena sidan visa att (på grund av att $p^{n-1}, q^{n-1}, r^{n-1}$ måste innehålla v som faktor) u/v i själva verket är ett polynom (dvs $v = \text{konstant}$) och å andra sidan visa att u :s gradtal blir strängt mindre än v :s gradtal då $n \geq 3$.

- d) En punkt (x_0, y_0) på en algebraisk kurva $p(x, y) = 0$ (algebraisk betyder att $p(x, y)$ är ett polynom) kallas singular om (förutom $p(x_0, y_0) = 0$) $\frac{\partial p}{\partial x}(x_0, y_0) = \frac{\partial p}{\partial y}(x_0, y_0) = 0$.

Här är några exempel på algebraiska kurvor som har en singular punkt i origo.

- 1) $y^3 = x^2 + y^2$,
- 2) $y^3 = x^2 - y^2$,

3) $(x^2 + y^2)^2 = x^2 - y^2,$

4) $(x^2 + y^2)^2 = x^2 y.$

(Hitta gärna på fler exempel själv.) Rita upp dessa kurvor och studera dem speciellt i närheten av origo. Tre av dem har egenskapen att (nästan) varje rät linje $y = tx$ (där t är en konstant) genom origo skär kurvan i precis en punkt utöver origo. Visa att detta ger upphov till en rationell parametrisering av kurvan, nämligen med t som parameter. (Den återstående kurvan är också rationell.) Denna metod att parametrisera vissa kurvor ger att varje tredjegradskurva som innehåller en singulär punkt är rationell. (Metoden kan också tillämpas på alla andragradskurvor, vilket ger en ledning till a) och första delen av b).) Kan en (irreducibel) andragradskurva ha singulära punkter?

- e) Förståelsen för algebraiska kurvor ökar avsevärt om man i den tillhörande ekvationen $p(x, y) = 0$ tillåter x och y att vara komplexa tal. Till ett givet polynom $p(x, y)$ får man därvid en komplex *kurva*

$$\{(z, w) : z, w \text{ komplexa tal sådana att } p(z, w) = 0\},$$

som är en två-dimensionell mängd i ett fyr-dimensionellt rum. Man kan nu visa att en (irreducibel) algebraisk kurva $p(x, y) = 0$ är rationell om och endast om motsvarande komplexa kurva (kompletterad med vissa oändlighetspunkter) topologiskt sett är en sfär (vanligtvis mycket tillknycklad och med gott om singulära punkter där den t ex skär sig själv). Med detta (topologiskt en sfär) menas ungefär att kurvan kan deformerats kontinuerligt, utan att det någon gång uppstår brott, till en sfär. Exempel på ytor som topologiskt sett inte är sfärer är torusen, Kleins flaska eller varje icke slutna yta (alltså en yta som har kanter). Kan du inse att den komplexa kurvan $z^2 + w^2 = 1$ topologiskt sett är en sfär

medan t ex $z^3 + w^3 = 1$ är en torus (jämför a) och c) ovan)? (Inte så lätt! Ta inte denna del av uppgiften alltför allvarligt, men läs gärna om ytors topologi i någon bok, t ex Sigma, band IV .)

För den som vill jobba mera:

f) Studera kurvan

$$C_a : p(x, y) = (x^2 + y^2)^2 - 2(x^2 - y^2) - 2r^2(x^2 + y^2) - a = 0$$

för något fixt värde på $r > 1$, t ex $r = \sqrt{2}$, och alla reella värden på a . Hur många olika sammanhängande kurvor och hur många isolerade punkter innehåller C_a för olika värden på a ? För vilka värden på a ändrar C_a struktur? Kan du hitta några värden på a för vilka C_a är rationell? (För de flesta värden på a är den inte det. Vi kan upplysa om att en rationell (irreducibel) kurva inte kan innehålla mer än en sluten kurva, men kan innehålla ett flertal isolerade punkter.) Är C_a reducibel (dvs kan $p(x, y)$ faktoriseras) för några värden på a ?

g) Kurvskaran i f) kan beskriva vissa fysikaliska fenomen. Om vi t ex låter r växa från noll till oändligheten och väljer a enligt

$$\begin{cases} a = -(1 - r^2)^2 & \text{då } 0 \leq r < 1, \\ a = 0 & \text{då } r \geq 1, \end{cases}$$

så beskriver (under lämpliga antaganden) mängderna

$$D(r) = \{(x, y) \in \mathbf{R}^2 : p(x, y) < 0\}$$

tillväxten av två grunda vattenpölar då det kontinuerligt droppar vatten i punkterna $(x, y) = (\pm 1, 0)$. (Tidsparametern blir proportionell mot r^2 vid konstant dropp-hastighet.) Rita några av mängderna $D(r)$ och studera speciellt vad som händer då r passerar värdet $r = 1$.

Litteratur

a) - f) Shafarevich, I.R., *Basic Algebraic Geometry*. Springer-Verlag 1974.

g) ANMÄRKNING: De ”grunda vattenpölar” avser egentligen så kallade Hele Shaw-flöden (med fria ränder).

Allmänt om Hele Shaw-flöden hittar du i

Bear, J., *Dynamics of Fluids in Porous Media*. Elsevier, New York 1972

eller i

Lamb, H., *Hydrodynamics*. Dover, New York 1932.

För det specifika exemplet i g) se

Richardson, S., Some Hele Shaw flows with time-dependent free boundaries. *J. Fluid Mech.* 102 (1981), s 263–278.

Något om metriker

BJÖRN GUSTAFSSON

K T H

Inledning. När man talar om avståndet mellan två punkter, t ex två orter i Sverige, så kan man mena litet olika saker, såsom avståndet fågelvägen, kortaste avståndet längs ett givet vägnät, eller den tid resan skulle ta med ett visst färdssätt (jfr enheter såsom dagsmarscher, ljusår osv). De flesta rimliga avståndsbegrepp uppfyller vissa elementära krav, som sammanfattas i matematikerns axiom för en sk metrik. Denna uppgift går ut på att studera geometriska egenskaper hos några andra avståndsbegrepp än de vi är vana vid. Trots att de grundläggande axiomen är uppfyllda så bjuds man ofta på överraskningar när man tittar närmare på geometrin. Det bör framhållas att de olika deluppgifterna nedan snarast bör ses som förslag eller idéer till vad som kan göras. Koncentrera dig gärna på någon mindre del av uppgiften och *forska* vidare i den efter eget huvud. Se också Andrejs Dunkels uppgift Om Pythagoras hade varit taxichaufför i Luleå i denna bok.

I planet, som vi identifierar med

$$\mathbf{R}^2 = \{x = (x_1, x_2) : x_1, x_2 \text{ reella tal}\},$$

kan avstånd mellan punkter mätas på olika sätt. Det vanliga *euklidiska* avståndet mellan $x = (x_1, x_2)$ och $y = (y_1, y_2)$ är

$$d(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2},$$

som bla har följande egenskaper (gällande för alla x, y, z).

- 1) $d(x, y) \geq 0$.

- 2) $d(x, y) = 0$ om och endast om $x = y$.
- 3) $d(x, y) = d(y, x)$.
- 4) $d(x, y) \leq d(x, z) + d(z, y)$ (triangelolikheten).

Rent allmänt kallas en funktion d som uppfyller 1) - 4) en *metrik* (avståndsfunktion). Några andra kandidater till metriker på \mathbf{R}^2 är

$$d(x, y) = (|x_1 - y_1|^p + |x_2 - y_2|^p)^{1/p}$$

där p är ett positivt tal ($0 < p < \infty$), och

$$d(x, y) = \max\{|x_1 - y_1|, |x_2 - y_2|\},$$

som svarar mot gränsfallet $p = \infty$.

- a) Undersök 1) - 3) för dessa funktioner d ($0 < p \leq \infty$).
- b) Rita upp "enhetsbollen", d v s mängden

$$B = \{x \in \mathbf{R}^2 : d(0, x) \leq 1\}$$

för några olika p -värden, t ex $p = 1/2$, $p = 1$, $p = 2$, $p = 4$, $p = \infty$.

- c) Undersök om triangelolikheten gäller för $p = 1/2$, $p = 1$, $p = \infty$. (För $p = 2$ gäller den som bekant; kan du bevisa det rent algebraiskt?)
- d) Triangelolikhetens gällande eller inte gällande hänger samman med en viss geometrisk egenskap hos enhetsbollen. Vilken? Med ledning av detta, eller på annat sätt, försök bestämma exakt för vilka p -värden som triangelolikheten gäller.
- e) I analogi med *maxi-metriken* ovan (fallet $p = \infty$) har vi bland metrikerna med $0 < p < \infty$ också en som kan kallas *taxi-metriken* av det skälet att avståndet mellan två punkter är den sträcka en taxiförare måste köra om vi tänker oss \mathbf{R}^2 som en stad med ett

tätt kvadratisk rutmönster av gator, varje gata parallell med en av koordinataxlarna. Vilken metrik är det?

- f) Givet en rät linje L i \mathbf{R}^2 och en punkt Q utanför denna så finns det, mätt i metriker d ovan, alltid minst en punkt på L som ligger på kortaste avstånd från Q . För vissa p -värden finns det emellertid ibland (beroende på L 's läge) flera punkter på kortaste avstånd. Vilka p -värden är det?
- g) De klassiska kägelsnitten (ellipsen, parabeln, hyperbeln) kan som bekant definieras med hjälp av enbart den euklidiska ($p = 2$) metriken: en ellips utgörs av de punkter för vilka summan av avstånden till två givna punkter är en given konstant o s v. Om vi i dessa definitioner byter ut det euklidiska avståndet mot någon av våra p -metriker (med $p \neq 2$) så får vi nya klasser av kurvor (p -ellipser osv). Rita några av dessa (t ex i fallen $p = 1$ och $p = \infty$). Är dessa p -ellipser osv verkligen kägelsnitt i den meningen att de uppkommer som skärningen mellan någon lämplig (icke-cirkulär) dubbelkon och plan i olika lägen? (Författaren vet ej svaret.)

För den som är trött på \mathbf{R}^2 följer här några uppgifter (h) - l) på en litet annorlunda metrik.

- h) Låt \mathbf{Q} vara mängden av rationella tal. På \mathbf{Q} definierar vi en avståndsfunktion genom

$$d(x, y) = \begin{cases} 0 & \text{om } x = y \\ 1/2^k & \text{om } x \neq y, \end{cases}$$

där heltalet k bestäms ur representationen

$$x - y = 2^k \cdot \frac{m}{n},$$

där m och n är heltal som ej är delbara med 2 (alltså udda). (Observera att varje rationellt tal $x - y$ skilt från noll kan skrivas på detta sätt och att k blir entydigt bestämt.) Exempel:

$$d(1/8, 11/28) = 1/2^{-3} = 8, \text{ ty } 1/8 - 11/28 = -15/56 = 2^{-3} \cdot (-15)/7.$$

Beräkna $d(0, 10^{-10})$, $d(-169, 6999)$.

- i) Visa att d uppfyller metrikaxiomen 1) - 3).
 j) Visa att d inte bara uppfyller triangelolikheten 4) utan t o m

$$4') \quad d(x, y) \leq \max\{d(x, z), d(z, y)\}.$$

En metrik som uppfyller 4') kallas en icke-arkimedisk metrik.

- k) Några lustiga egenskaper hos icke-arkimediska metriker är: alla trianglar är likbenta, dvs av de tre talen $d(x, y)$, $d(y, z)$, $d(z, x)$ så är alltid minst två stycken lika; varje punkt i en boll är medelpunkt i den, dvs om vi sätter $B(y, r) = \{x \in \mathbf{Q} : d(y, x) < r\}$ så gäller att $x \in B(y, r)$ medför $B(y, r) = B(x, r)$. Kan du bevisa dessa påståenden?
 l) Kan talet 2 i definitionen av d bytas ur mot något annat tal (så att man fortfarande får en metrik)?

Litteratur

- a) - g) Kreyszig, E., *Introductory Functional Analysis with Applications*. Wiley & Sons 1978.
 h) - l) Koblitz, N., *p-adic Numbers, p-adic Analysis and Zeta-Functions*. GTM 58, Springer-Verlag 1977.

Möbiusgruppen och icke-euklidisk geometri

LARS GÅRDING

Lunds Universitet

Meningen med detta förslag till enskilt arbete är att alla uppgifter U redovisas skriftligt med fulla motiveringar och att alla utelämnade figurer ritas. Möbius (1790-1863) var en tysk matematiker. Den modell av den icke-euklidiska geometrin som presenteras här föreslogs av Poincaré (1854-1912). En enkel bok om icke-euklidisk geometri är

Meschkowski, H., *Non-euclidean Geometry*. Academic paperbacks, Academic Press 1965.

Möbiusfunktioner. En *Möbiusfunktion* är helt enkelt en bruten linjär funktion av en komplex variabel z ,

$$(1) \quad z \rightarrow f(z) = (az + b)/(cz + d)$$

där a, b, c, d är komplexa tal sådana att $ad - bc \neq 0$, ett villkor som förhindrar att funktionen f är konstant.

U. Visa det sista påståendet. (Ledning. Visa att $ad - bc = 0$ då $f(z_1) = f(z_2)$ och $z_1 - z_2 \neq 0$.)

U. Visa att varje Möbiusfunktion kan skrivas i formen (1) där $ad - bc = 1$

Det är uppenbart att $f(z) = 1/z$ är en Möbiusfunktion. För att den ska vara definierad också för $z = 0$, inför man en punkt ∞ i oändligheten i det komplexa planet. Det på det sättet utvidgade planet komplexa planet betecknar vi med C^* . Om f ges av (1), sätter vi naturligtvis $f(\infty) = a/c$ och $f(-d/c) = \infty$. Observera att

villkoret $ad - bc \neq 0$ medför att varken a/c eller d/c har formen $0/0$ och alltså är väl definerade och antar värdet ∞ då en nämnare är noll. Belöningen för dessa djärva steg är följande

SATS. *Varje Möbiusfunktion är en bijektion av C^* och dess invers är också en Möbiusfunktion.*

U. Bevisa satsen, dvs att $f(z)$ avbildar C^* på sig själv och att ekvationen $f(z) = w$ har den entydiga lösningen $z = (dw - b)/(-cw + a)$ då w är ett komplext tal eller ∞ .

U. Sammansättningen $f \circ g$ av två funktioner f och g definieras av

$$(f \circ g)(z) = f(g(z)).$$

Visa att $f \circ g$ är en Möbiusfunktion då f och g är det. Visa att det till varje Möbiusfunktion f finns en invers Möbiusfunktion g sådan att $f \circ g(z) = g \circ f(z) = z$ för alla z . (Ledning. Föregående uppgift.) Man brukar beteckna den inversa funktion med f^{-1} och har då $f^{-1} \circ f = f \circ f^{-1} = e$ där e betecknar identiteten $e(z) = z$ för alla z . Av $f \circ g = h$ följer då att $g = f^{-1} \circ h$. (Observera att f^{-1} är den inversa funktionen och inte detsamma som $f(z)^{-1} = 1/f(z)$.)

Anmärkning. En samling bijektioner av en mängd X som innehåller den identiska funktionen och med varje funktion också dess invers kallas en transformationsgrupp. Alla Möbiusfunktioner bildar alltså en transformationsgrupp.

U. En Möbiusfunktion f sådan att $f(\infty) = \infty$ sägs vara *affin*. Visa att de affina funktionerna har formen

$$f(z) = Az + B$$

där $A \neq 0$. Tolka fallet $A = 1$ geometriskt (en translation), fallet $B = 0, |A| = 1$ (en vridning kring origo) och fallet $B = 0, A > 0$ (en

sträckning eller homoteti med centrum i $z = 0$). Illustrera de olika fallen med figurer.

U. Visa att alla affina funktioner bildar en transformationsgrupp och att varje affin funktion avbildar cirklar i cirklar.

U. Funktionen $f(z) = 1/z$ kallas en *inversion*. Visa att den överför cirkeln $|z| = 1$ i sig själv och det inre av cirkeln till det yttre och omvänt. Studera motsvarande avbildning av cirkeln i detalj.

U. Då $f(z) = (az + b)/(cz + d)$ inte är affin, finn A och B sådana att

$$(az + b)/(cz + d) = A/(cz + d) + B.$$

Visa att detta innebär att $f = f_1 \circ f_2 \circ f_3$ där f_3 är affin, f_2 en inversion och f_1 är affin.

Dubbelförhållande.

U. Visa att det finns precis en affin funktion $z \rightarrow w = f(z)$ som överför två givna punkter $z_1 \neq z_2$ i två andra punkter w_1, w_2 . Visa att sambandet kan skrivas som

$$(z - z_1)/(z - z_2) = (w - w_1)/(w - w_2).$$

Båda sidor i denna formel är något som brukar kallas för enkelförhållande. Det finns ett motsvarande *dubbelförhållande* för Möbiusfunktioner, som definieras av

$$(z_1, z_2; z_3, z_4) = \frac{z_1 - z_3}{z_1 - z_4} : \frac{z_2 - z_3}{z_2 - z_4}$$

där högra sidan definierar den vänstra. Det är alltså fråga om kvoten mellan två enkelförhållanden varav namnet.

U. I det komplexa planet är $1/\infty = 0$ och $1/0 = \infty$ men uttrycken $0/0$ och ∞/∞ är odefinierade. Visa att dubbelförhållandet är väl definierat precis då tre av talen z_1, z_2, z_3, z_4 är olika.

Nu kommer en viktig sats om dubbelförhållande och Möbiusfunktioner.

SATS. Låt z_1, z_2, z_3 vara tre skilda punkter.

1) Formeln

$$z \rightarrow (z, z_1; z_2, z_3)$$

definierar en Möbiusfunktion som avbildar punkterna z_1, z_2, z_3 på $0, \infty, 1$ respektive.

2) Om w_1, w_2, w_3 är tre andra skilda punkter så definierar ekvationen

$$(2) \quad (z, z_1; z_2, z_3) = (w, w_1; w_2, w_3)$$

en Möbiusfunktion $z \rightarrow w$ som avbildar punkterna z_1, z_2, z_3 på punkterna w_1, w_2, w_3 .

3) En Möbiusfunktion är entydigt bestämd av sina värden i tre olika punkter.

U. Visa att om en Möbiusfunktion $f(z)$ överför punkterna $0, \infty, 1$ i sig själva så är den identiska funktionen, dvs $f(z) = z$ för alla z .

U. Visa satsen genom att till att börja med kontrollera 1) och visa att 2) följer av 1). Slutligen, visa att om en Möbiusfunktion f överför z_1, z_2, z_3 i w_1, w_2, w_3 och $g(z) = (z, z_1; z_2, z_3)$, $h(w) = (w, w_1, w_2, w_3)$, så är $h \circ f \circ g^{-1}(z) = z$ då $z = 0, \infty, 1$ och alltså, enligt föregående uppgift, $f = h^{-1} \circ g$.

Cirkeleområden. Genom tre separata punkter i planet går som bekant en entydigt bestämd cirkel. (Observera att en rät linje anses vara en cirkel med centrum i ∞ .) En fjärde punkt ligger på samma cirkel precis då summan av motstående vinklar i den fyrhörning som de uppspanner är π eller noll beroende på om fyrhörningen är konvex eller inte. (Rita en cirkel och fyra punkter 1,2,3,4 på den som inte behöver följa varandra i ordning. Drag linjerna 12,23,34,41. Då är vinkeln med spetsen i 1 lika med vinkeln med spetsen i 3 eller också är deras summa lika med π .)

U. Låt de tre punkterna vara z_1, z_2, z_3 och låt z vara en fjärde punkt. Visa att z ligger på den cirkel som går genom de tre punkterna precis då $\arg(z, z_1; z_2, z_3)$ är noll eller π , dvs dubbelförhållandet är reellt. (Observera att vinklarna ska tas med tecken.)

Om vi definerar ett cirkelområde i planet som det inre eller yttre av en cirkel får vi nu följande viktiga sats som tillsammans med den föregående ger oss en nästan fullständig överblick över ämnet Möbiusfunktioner.

SATS. *En Möbiusfunktion avbildar cirklar på cirklar och cirkelområden på cirkelområden. Ett cirkelområde kan avbildas på varje annat cirkelområde genom en Möbiusfunktion.*

U. Visa att $f(z) = r(z - i)/(z + i)$ avbildar övre halvplanet på det inre av cirkeln $|z| = r$. Räkna ut inversen som alltså avbildar det inre av cirkeln på övre halvplanet.

Om C_1 och C_2 är cirkelområden och $f(z)$ och $g(z)$ är Möbiusfunktioner som avbildar C_1 på C_2 så är det klart att $h(z) = f \circ g^{-1}(z)$ avbildar C_1 på sig självt. Omvänt, om $h(z)$ har denna egenskap så avbildar $f \circ h(z)$ C_1 på C_2 . För att veta vilka Möbiusfunktioner som avbildar ett givet cirkelområde på ett annat behöver man alltså bara känna till alla som avbildar ett givet cirkelområde, t. ex. *enhetsskivan* $|z| < 1$, på sig självt. I så fall avbildas också *enhetscirkeln* $|z| = 1$ och det yttre av enhetsskivan på sig själva. Observera att en Möbiusfunktion som avbildar en cirkel på sig själv mycket väl kan avbildas det inre av cirkeln på det yttre och tvärtom (ge ett exempel på detta).

SATS. *Då $|a| < 1$ och $|b| = 1$ avbildar*

$$(3) \quad f(z) = b(z - a)/(1 - \bar{a}z)$$

enhetsskivan på sig själv och varje Möbiusfunktion med denna egen-
skap har formen (3).

U. Bevisa första delen av satsen genom att verifiera att $|f(z)| = 1$ då $|z| = 1$ och att $|f(0)| < 1$. Bevisa andra delen av satsen så här: om $g(z)$ avbildar enhetscirkeln på sig själv så finns ett a med $|a| < 1$ så att $g(a) = 0$. Men då avbildar $h = g \circ f$ enhetscirkeln på sig själv och $h(0) = 0$. Men då är $h(z) = z/(cz + d)$. Visa att detta medför att $c = 0$ och slutför beviset.

U. Att en cirkelskiva avbildas på en cirkelskiva betyder inte att mittpunkt avbildas på mittpunkt. Visa detta genom att låta $b = 1$ och $a = 0.5$ i (3). Skissera bilderna av cirklarna $|z| = 1/3$ och $|z| = 2/3$.

Spegelpunkter. Två punkter u och v sägs vara *spegelpunkter* med avseende på en linje om de är varandras spegelbilder i linjen. De är spegelpunkter med avseende på en cirkel om de ligger på en och samma linje genom cirkelns medelpunkt, på samma sida om medelpunkten och produkten av deras avstånd till cirkelns medelpunkt är lika med radiens kvadrat. (Rita figur, kontrollera att medelpunktens spegelbild ligger i oändligheten.)

U. Visa att om $f(z)$ är en affin funktion och u och v är spegelpunkter med avseende på en cirkel C så är $f(u)$ och $f(v)$ spegelpunkter med avseende på cirkeln $f(C)$

U. Visa genom att Möbiusfunktionen $f(z) = r(z - i)/(z + i)$ avbildar spegelpunkter med avseende på reella axeln på spegelpunkter med avseende på en cirkel med radien r .

U. Det finns en ekvivalent definition av begreppet spegelpunkt: två punkter u och v sägs vara spegelpunkter med avseende på en cirkel

C genom punkterna a, b, c precis då

$$(u, a; b, c) = \overline{(v, a : b, c)} = (\bar{v}, \bar{a} : \bar{b}, \bar{c}),$$

(den den sista likheten är en anmärkning). För att göra situationen tydlig ska vi skriva spegelpunkt(1) för den första och spegelpunkt(2) för den andra definitionen. Visa att om $f(z)$ är en Möbiusfunktion och u och v är spegelpunkter(2) med avseende på en cirkel C så är $f(u)$ och $f(v)$ spegelpunkter(2) med avseende på cirkeln $f(C)$. Visa att spegelpunkt(1) och spegelpunkt(2) betyder samma sak. (Ledning. Visa det senare först då cirkeln är reella axeln och, allmännare, då cirkeln är en rät linje. Använd en föregående uppgift till att visa att spegelpunkt(1) och spegelpunkt(2) är samma sak först för en cirkel med centrum i origo och sedan för vilken cirkel som helst.)

En modell av den icke-euklidiska geometrin. Ett av axiomen för den euklidiska geometrin, framställd i Euklides' *Elementa* cirka 400 år f. K., är parallellaxiomet som säger följande: genom en punkt utanför en rät linje kan en och endast en linje dragas som icke träffar den förra hur långt den än utdrages. Två linjer som på detta sätt inte har någon punkt gemensam sägs vara parallella och axiomet kallas parallellaxiomet. De övriga axiomen, t.ex. *genom två skilda punkter kan en och endast en rät linje dragas, eller två icke parallella linjer ha precis en punkt gemensam*, är betydligt enklare och man försökte därför länge bevisa parallellaxiomet från de övriga. Dessa bemödande tog slut då Lobachevski och Bolai i början av 1800-talet visade att det finns objekt som man kalla punkter och linjer som uppfyller alla axiom utom parallellaxiomet. Teorin för dessa objekt kallas den icke-euklidiska geometrin. Med hjälp av Möbiusgruppen är det lätt att hitta sådana objekt och dessutom konstruera en motsvarighet till den euklidiska geometrins kongruens-transformationer, dvs förflyttningar i planet som lämnar alla avstånd

oförändrade. Det euklidiska planets motsvarighet är då det inre av en cirkel C där de räta linjernas motsvarighet är cirkelbågar som skär C ortogonalt. Meningen med de uppgifter som följer är att läsaren själv ska härleda de viktigaste satserna i den icke-euklidiska geometrin.

U. Låt C och D vara cirklar som skär varandra under rät vinkel. Visa att de punkter P och Q där en rät linje från C 's medelpunkt skär D är spegelpunkter med avseende på C . (Ledning. Vilket bevis som helst får användas).

U. Låt P och Q vara två punkter i det inre av en cirkel C . Visa att det finns precis en cirkel genom P och Q som skär C ortogonalt och konstruera den.

U. Låt Punkter betyda punkter inuti en fast cirkel C (som vi identifierar med det icke-euklidiska planet) och Räta linjer betyda cirkelbågar i C som skär C under rät vinkel. Enligt det föregående går precis en Rät linje genom två skilda Punkter. Visa att två Räta linjer skär varandra i högst en Punkt men att parallellaxiomet inte är uppfyllt för Punkter och Räta linjer.

Avstånd. I det icke-euklidiska planet kan man införa ett avstånd. För enkelhetens skull låter vi det icke-euklidiska planet, här betecknat med E^* , vara enhetsskivan $|z| < 1$. Med en icke-euklidisk förflyttning menar vi en Möbiusfunktion $f(z)$ som avbildar E^* på sig självt. Alla dessa bildar tydligen en transformationsgrupp. Vi kallar den T .

U. Visa att det finns minst en Möbiusfunktion i T som överför en given Rät linje i en annan given Rät linje. Visa också att det finns en sådan funktion som överför en given Rät linje i sig själv och samtidigt en given Punkt på den Räta linjen till en annan given Punkt på den. (Ledning. Det räcker att verifiera den andra delen av uppgiften då

den Rätta linjen är reella axeln mellan 1 och -1 . Varför?)

U. Då z och w är två Punkter på en Rät linje, definiera ett avstånd mellan dem genom formeln

$$d(z, w) = \log |(z, w; Z, W)|$$

där Z och W är ändpunkterna på den cirkelbåge som innehåller z och w och är vinkelrät mot enhetscirkeln. Visa att $d(z, w) = d(f(z), f(w))$ då $f(z)$ ligger i T och att

- 1) $d(z, w) = d(w, z) \geq 0$,

- 2) $d(z, w) = 0$ precis då $z = w$,

- 3) $d(z, w) = d(z, u) + d(u, w)$ då Punkten u ligger mellan z och w och på samma Rätta linje.

(Ledning. Det räcker att visa detta då den Rätta linjen är den del av reella axeln som ligger i E^* . Varför?)

- 4) $d(z, w)$ går mot oändligheten då en av punkterna går mot enhetscirkeln och den andra är fix.

Av 4) följer att man inte kan räkna enhetscirkeln till det icke-euklidiska planet. Det ligger i oändligheten. Ingen som vistas i det icke-euklidiska planet når någonsin dit.

U. Visa att om z, w, u, v är Punkter och $d(z, w) = d(u, v)$ så finns det en funktion f i T sådan att $f(z) = f(u), f(w) = f(v)$.

Icke-euklidiska cirklar och trianglar. En Cirkel i det icke-euklidiska planet E^* definieras som mängden av Punkter w sådana att $d(w, z)$ är en konstant som kallas cirkelns Radie. Punkten z kallas Cirkelns Medelpunkt.

U. Visa att varje Cirkel är en euklidisk cirkel men att dess radie och medelpunkt inte är Cirkelns Radie och Medelpunkt. (Ledning. Visa att Medelpunkt och medelpunkt är samma då medelpunkten ligger

i $z = 0$ och visa sedan att två Cirklar kan överföras i varandra med en Möbiusfunktion i T precis då deras Radier är lika.)

U. Visa (figur!) att Cirklar med samma Radie som går mot oändligheten representeras av cirklar med allt mindre radier som går mot enhetscirkeln.

Det visas i ett appendix att en Möbiusfunktion bevarar vinklar. Det är därför överflödigt att definiera särskilda Vinklar i det icke-euklidiska planet. Som bekant spelar parallellaxiomet en avgörande roll då man visar att summan av innervinklarna i en euklidisk triangel är π . Detta är inte sant för en Triangel, dvs en icke-euklidisk triangel.

U. Visa att summan av innervinklarna i en Triangel är mindre än π . (Ledning. Visa med en figur att detta är sant då $z = 0$ är en inre punkt. Hur gör man då $z = 0$ inte är en inre punkt?)

U. Visa att det finns Trianglar med hörn i oändligheten där summan av innervinklarna är noll.

Anmärkning. Det förefaller då man ritar Trianglar att vinkelsumman blir mindre ju större Triangelns yta är och det är också sant. Det finns ett precist samband mellan dessa storheter (en sats av Gauss) som vi inte kan gå in på här. Om man bedriver icke-euklidisk geometri i en stor cirkel $|z| < R$ och låter R gå mot oändligheten blir geometrin inom ett begränsat område mer och mer lik den euklidiska.

Några numeriska uppgifter. Med ordet avbilda menas här att finna en Möbiusfunktion som avbildar.

U. Avbilda cirkelskivan $|z| < 1$ på cirkelskivorna $|z-1| < 2$, $|z-i| < 3$ och $|z-4| > 5$.

U. Avbilda cirkelskivan $|z| < 1$ på sig själv så att punkten $z = 1/2$ avbildas på sig själv. Finn sedan alla sådana avbildningar.

U. Avbilda cirkelskivan $|z - 4| < 1$ på övre halvplanet.

U. Avbilda övre halvplanet på sig självt så att punkten i övergår i sig själv. Finn sedan alla sådana avbildningar.

Appendix.

Vinklar. Trots att den avbildning $z \rightarrow f(z)$ av det komplexa planet i sig som förmedlas av en Möbiusfunktion $f(z)$ kan möblera om ganska ordentligt, t.ex. så att 0 avbildas på ∞ , finns det någonting som inte ändras och det är vinklar. Låt h och k vara två komplexa tal som inte är noll. Om då t är ett reellt tal som genomlöper ett litet intervall $0 \leq t < s$ där $s > 0$ är liten och fix, så betyder $t \rightarrow z + th$ och $t \rightarrow z + tk$ två små linjestycken L_1 och L_2 utgående från punkten z . (Rita en figur!). Vinkeln mellan dem är som bekant lika med $\arg h\bar{k}$. Om vi antar att $f(z)$ inte är ∞ , så avbildas linjestyckena genom f på två små kurvstycken $f(L_1)$ och $f(L_2)$ utgående från $f(z)$ och givna av $t \rightarrow f(z + th)$ och $t \rightarrow f(z + tk)$. (Rita figur!)

U. Visa att tangentvektorerna T_1 och T_2 till dessa kurvor i punkten $f(z)$ är $h/(cz + d)^2$ och $k/(cz + d)^2$ (om $ad - bc = 1$ och $f(z) = (az + b)/(cz + d)$). (Ledning. Derivera med avseende på t och sätt $t = 0$.)

Om nu två kurvor som möts i ∞ antas ha samma tangentvektorer där som de har under avbildningen $z \rightarrow 1/z$, så ger denna övning en viktig

SATS. Vinklar ändras inte vid avbildningar givna av Möbiusfunktioner.

Något om permutationer

LARS HOLST

KTH, Stockholm

1. Inledning. I många matematiska resonemang måste man räkna antalet *fall* av olika slag. Den del av matematiken som systematiskt studerar dylikt brukar kallas *kombinatorik*. Flera av de grundläggande frågeställningarna har en lång historia, men man kan nog säga att ett mera systematiskt studium påbörjades under 1600-talet, då även sannolikhetsteorin började utvecklas. Många av de klassiska resultaten i sannolikhetsteorin formulerades först rent kombinatoriskt. Avsikten med nedanstående är att ge underlag för ett specialarbete om några grundläggande resonemang och begrepp i enumerativ kombinatorik som anknyter bl.a. till klassisk sannolikhetsteori. För en fyllig och inspirerande framställning med mängder av olika exempel hänvisas till boken av Feller (1968).

2. Permutationer. *På hur många sätt kan 7 barn fördela platserna i ett sjuannalag i fotboll? Hur många olika blandningar finns av en kortlek omfattande 52 olika kort? På hur många olika sätt kan n personer sätta sig på n stolar?* Många matematiskt sett ekvivalenta formuleringar finns av dylikt, t.ex. ange antalet olika sätt som talen $1, 2, \dots, n$ kan skrivas efter varandra på en rad. För exempelvis $n = 3$ finns de 6 möjligheterna 123, 132, 213, 231, 312, 321. Ett sådant sätt kallas en *permutation* av talen $1, 2, \dots, n$. Antalet sådana brukar betecknas $n!$, läses *n -fakultet*. Av praktiska skäl sätter man $0! = 1$. *Vad är $n!$ uttryckt i $1, 2, \dots, n$? Räkna ut detta antal för de två första frågorna.*

3. Binomialkoefficienter. Betrakta m nollor och $n - m$ ettor, som skrivs efter varandra på en rad i någon ordning. Detta kan även

Som bekant gäller att

$$(x + y)^n = (x + y)(x + y)\dots(x + y),$$

där produkten innehåller n faktorer. För att få den allmänna termen $x^k y^{n-k}$ vid hop-multiplikation tar man ur var och en av k faktorer ett x och från alla de andra ett y . *På hur många sätt kan detta göras? Vad blir b_{nk} i följande viktiga samband,*

$$(x + y)^n = \sum_{k=0}^n b_{nk} x^k y^{n-k},$$

det s.k. *binomialteoremet*?

Generalisera ovanstående till att man har r olika sorters objekt, m_1 av sort 1, m_2 av sort 2, etc. och totalt $n = m_1 + \dots + m_r$ objekt. *På hur många sätt kan de ställas i en rad (ger s.k. multinomialkoefficienter)? Vad är koefficienten framför den allmänna termen i utvecklingen av $(x_1 + \dots + x_r)^n$ (det s.k. multinomialteoremet)?* Många ekvivalenta formuleringar finns även av detta. *På hur många olika sätt kan 7 barn bilda ett fotbollslag om de inte skiljer på de 3 kedjespelarna ej heller på de 3 backarna? Hitta på andra formuleringar.*

4. Fixpunkter i permutationer. Betrakta åter personerna 1, 2, ..., n som sätter sig på stolarna 1, 2, ..., n . Låt $f(i)$ beteckna numret på den stol person i sätter sig på; f kan uppfattas som en funktion med talmängden 1, 2, ..., n både som definitions- och värdemängd. Funktionen, som till ett givet stolsnummer tillordnar numret på den person som sitter på den, är den inversa funktionen f^{-1} . *Hur många olika sådana funktioner f finns det?*

Man säger att en permutation med tillhörande funktion f har *fixpunkten* i om $f(i) = i$, dvs om person i sätter sig på stol i . Ett klassiskt kombinatoriskt problem är att bestämma antalet permutationer som saknar fixpunkter. Låt D_n beteckna detta antal; t. ex.

genom att skriva upp alla möjliga fall finner man att $D_1 = 0, D_2 = 1, D_3 = 2, D_4 = 9$. *Visa genom lämpliga kombinatoriska tolkningar av ingående termer att*

$$D_3 = 3! - \binom{3}{1}2! + \binom{3}{2}1! - \binom{3}{3}0! = 2.$$

Observera att $3!$ är totala antalet permutationer av tre element. *Vad blir motsvarande formel för D_4 ? Generalisera detta till godtyckligt n . Detta är ett exempel på den s.k. inklusions-exklusions principen i kombinatoriken.*

För exponentialfunktionen e^x gäller serieutvecklingen

$$e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \dots$$

och därmed

$$e^{-1} = 1 - 1 + \frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!} - + \dots$$

Visa att

$$D_n = \text{heltalsdelen}\left(\frac{n!}{e}\right).$$

En permutation av talen $1, 2, \dots, n$ väljs slumpmässigt, dvs var och en av de $n!$ olika permutationerna har lika chans att bli vald. *Hur kan detta göras rent praktiskt? Sannolikheten att den valda permutationen inte har någon fixpunkt är*

$$P_{n0} = \frac{D_n}{n!}.$$

Varför kan P_{n0} approximeras med e^{-1} för stora n ? Låt P_{nk} beteckna sannolikheten att den slumpmässigt valda permutationen har exakt k fixpunkter. Ange en formel för P_{nk} , och visa att denna sannolikhet kan approximeras med $e^{-1}/k!$ för stora n och "fixt" k . Detta

är ett exempel på s.k. *Poisson-approximation*. *Undersök approximationens noggrannhet numeriskt.*

Det finns många matematiskt ekvivalenta formuleringar av ovanstående t.ex. följande: n par kommer till en fest, varje dam får en helt slumpmässigt vald herre till bordskavaljer. *Vad är sannolikheten att ingen dam har sin medhavda herre som bordskavaljer? Hitta på andra ekvivalenta formuleringar.*

Fixpunktsproblemet ovan kallas ofta *matchnings-* eller *rencontre-problemet*, se boken av Feller (1968). Det löstes omkring år 1710 i begynnelsen av kombinatorikens och sannolikhetsteorins utveckling. För bordsplaceringssituationen kan man fråga sig vad sannolikheten är för att ingen dam sitter brevid sin medhavda herre (som bekant sitter alltid bordskavaljeren till vänster om sin bordsdam). För enkelhets skull låt bordet vara runt. Detta är det s.k. *ménageproblemet*, som är svårare än det förra problemet. En intressant framställning av detta ges i Bogart and Doyle (1986). Man kan lätt tänka sig andra generaliseringar.

Referenser

- [1] Bogart, K.P. and Doyle, P., Non-sexist solution of the ménage problem. *Amer. Math. Monthly* 93(1986), s 514–518.
- [2] Feller, W., *An Introduction to Probability Theory and Its Applications*. Vol. 1. Third Ed. Wiley, New York 1968.

Att dela en hemlighet

JOHAN HÅSTAD

KTH

1. Inledning. Anta att du har en hemlighet som du inte vill att någon annan skall veta. För att vara konkret, låt denna hemlighet vara numret på ditt hemliga bankkonto i Schweiz. Naturligtvis vill du inte att någon känner till detta kontonummer, men ändå vill du att dina tre barn skall kunna komma åt kontonumret när du är död. Du litar inte på dina barn och de litar inte på varandra. (Ett olagligt bankkonto i Schweiz medför att man blir litet försiktig och inte litar på folk.) Du kommer på följande idé. Anta att kontonumret är 31784 (i själva verket är det längre, men vi sparar skrivarbete). Du väljer två slumpartade 5-siffriga tal, låt oss säga 69241 och 77431. Du ger dina tre barn talen 69241, 77431 och 05112, ett till dem var. Dessutom ger du instruktionerna att om de adderar förstasiffrorna i deras tre tal och stryker tiotal-siffran så får de den första siffran i kontonumret, likadant för den andra siffran o.s.v. Du är mycket nöjd med dig själv och har åstadkommit följande:

- 1) Dina tre barn kan tillsammans få fram kontonumret.
- 2) Två av dem kan inte lista ut något om kontonumret förutom att det har fem siffror.

Visa detta påstående.

Med andra ord har du givit bort din hemlighet utan att någon enskild person vet den. Den här uppgiften går ut på att studera algoritmer för att fördela hemligheter och deras egenskaper och svagheter.

Låt oss börja med att påpeka några problem med ovanstående algoritm.

- 1) Om ett av barnen dör eller glömmer sitt tal är kontonumret försvunnet för alltid.
- 2) Anta att den som fått 69241 istället påstår att han fått 23716. När de andra avslöjar sina tal tror de att kontonumret är 95259, medan den som ljugit vet det rätta numret (förutsatt att ingen annan varit lika listig).

2. Bakgrundsfakta. För att beskriva en algoritm som ger ett annat resultat behöver vi litet bakgrundsinformation.

Låt p vara ett primtal. Vi kommer att räkna med talen $0, 1, 2, \dots, p-1$ modulo p . Detta innebär att vi har de vanliga räknesätten $+$, $-$ och \cdot men att vi bara är intresserade av vilken rest svaret ger vid division med p . För att det ska fungera bra att fortsätta att räkna är det viktigt att notera att resultatets rest vid division med p endast beror på operandernas rest vid division med p . (Visa detta.) Vi kommer att skriva $+$, $-$ och \cdot som vanligt medan vi använder likhetstecken med 3 streck och lägger till $(\text{mod } p)$ efteråt för att vi bara är intresserade av vilken rest talen ger vid division med p . Till exempel har vi

$$4 \cdot 6 \equiv 3 \pmod{7}$$

$$16 + 4 \equiv 1 \pmod{19}$$

$$3 - 7 \equiv 7 \pmod{11}$$

Du kan läsa mer om detta sätt att räkna i boken av Hardy och Wright som nämns i litteraturlistan. För att den här definitionen skall fungera behöver p inte vara primtal men när vi nu skall definiera division underlättar det.

Om $b \not\equiv 0 \pmod{p}$ definieras a/b som det tal $c \pmod{p}$ som uppfyller $c \cdot b \equiv a \pmod{p}$. T.ex. har vi $6/11 \equiv 16 \pmod{17}$.

Visa att ett sådant tal finns och är unikt.

(Studera talen $b, 2 \cdot b \dots p \cdot b \pmod{p}$. Visa att de olika \pmod{p} . Ett måste vara a .)

Det finns effektivare sätt att göra division \pmod{p} om p är stort. *Försök finna något.* Ett sätt presenteras under *Bra att veta* i slutet av uppgiften.

3. En ny delningsalgoritm. Med vår nya notation kan vi säga att om den första siffran i kontonumret är s_1 och första siffran i de tre barnens tal är $s_1^{(1)}, s_1^{(2)}$ och $s_1^{(3)}$,* så gäller

$$s_1 \equiv s_1^{(1)} + s_1^{(2)} + s_1^{(3)} \pmod{10}$$

och att $s_1^{(1)}$ och $s_1^{(2)}$ valdes slumpmässigt. (Vi kommer i fortsättningen bara att beskriva vad som händer med första siffran i kontonumret. Vi förutsätter att de andra siffrorna behandlas på samma sätt och eventuella slumpval väljs oberoende.)

Den nya algoritmen kommer att vara liknande och fungerar på följande sätt. Du väljer ett slumpmässigt tal b_1 , $0 \leq b_1 \leq 10$ och nu blir de tre bitarna

$$s_1^{(1)} \equiv s_1 + b_1 \pmod{11}$$

$$s_1^{(2)} \equiv s_1 + 2 \cdot b_1 \pmod{11}$$

$$s_1^{(3)} \equiv s_1 + 3 \cdot b_1 \pmod{11}$$

Dessa bitar är nu inte siffror i ett tal utan tal i intervallet $[0, 10]$ men det spelar inte så stor roll för problemet. Om vi således delar

*Här använder vi litet matematisk notation. Dessa tal är inte potenser, vi använder övre och undre index för att hålla ordning på informationen. Vi låter $s_2^{(3)}$ t.ex. betyda det tredje barnets andra siffra och $s_1^{(2)}$ betyder det andra barnets första siffra o.s.v.

numret 31784 på detta sätt med $b_1 = 2$ $b_2 = 5$ $b_3 = 0$ $b_4 = 7$ $b_5 = 5$ får de tre syskonen:

(5, 6, 7, 4, 9)	Syskon 1
(7, 0, 7, 0, 3)	Syskon 2
(9, 5, 7, 7, 8)	Syskon 3.

Detta sätt att dela hemligheten har följande egenskaper:

- 1) Två syskon tillsammans kan ta reda på kontonumret.
- 2) Inget syskon har någon aning om numret, förutom antalet siffror.

Visa dessa egenskaper.

Diskutera eventuella problem med denna fördelningsalgoritm.

4. Generalisering. En naturlig fråga är nu i vilken grad de givna algoritmerna går att generalisera. Anta att vi vill fördela en hemlighet bland n personer på ett sådant sätt att t personer tillsammans kan ta reda på hemligheten medan $t - 1$ personer inte får någon information även om de samarbetar. Vi har följande grundidé. Låt s vara en siffra i hemligheten. Ta ett polynom av gradtal $t - 1$

$$Q(x) = s + b_1 \cdot x + b_2 \cdot x^2 \dots b_{t-1} x^{t-1}$$

där $b_1, b_2 \dots b_{t-1}$ är slumpmässigt valda heltal under villkoret att $0 \leq b_i \leq 10$. Sedan får den i 'te personen $Q(i)$ (mod 11).

*Visa att denna delningsalgoritm har de önskade egenskaperna om $n < 11$. D.v.s. om t personer tillsammans kan rekonstruera s medan $t - 1$ personer inte får någon information om s . Det finns information som kan vara till hjälp under rubriken *Bra att veta* i slutet av denna beskrivning.*

Hur klarar man större n ?

Om exakt t personer försöker rekonstruera hemligheten har vi ett liknande problem som tidigare, nämligen att en oärlig person kan ge en felaktig bit och förstöra resultatet. För att motverka det kan man försöka följande. Vi byter ut 11 mot ett stort primtal p och sätter

$$Q(x) = s + b_1 \cdot x + b_2 \cdot x^2 \dots b_{t-1}x^{t-1}$$

med $0 \leq b_i \leq p - 1$ slumpmässigt valda. Som förut är den i 'te biten $Q(i) \pmod{p}$.

Om man nu fuskar genom att ändra sin bit slumpmässigt blir s antagligen något slumpmässigt tal mellan 0 och $p - 1$. Eftersom vi vet att $0 \leq s \leq 9$ upptäcker man troligtvis att någon fuskar. (Dock lyckas ju fusket i den mån att den som fuskar kan räkna ut s .)

Visa att denna metod inte är särskilt bra. I själva verket kan en person som vet de andra personernas i byta ut s mot $s + 1$ eller $s + d$ för något d personen väljer själv. Det kan vara bättre att som i 'te bit ge ut två tal. Dels ett slumpvis valt tal a_i och dels värdet av polynomet i denna punkt, $Q(a_i) \pmod{p}$.

5. Förslag till uppgifter. Efter denna inledning kan din egen kreativitet ta vid.

FÖRSLAG.

1. *Försök att hitta andra problem med de presenterade algoritmerna och försök hitta motåtgärder.*

2. *Försök att hitta praktiska användningar av de givna idéerna.* I dagens digitaliserade samhälle finns mycket information som skall hållas hemlig men som en grupp personer bör veta. Utveckla gärna program och sälj till företag som saknar matematisk expertis.

3. *Hitta på andra sätt att dela hemligheter.*

4. *Implementera någon hemlighets-delningsalgoritm. Antingen den allmänna givna eller din egen.*

Om du vill göra något mer avancerat, läs också *Offentlig kryptering* i denna volym och kombinera metoderna.

Bra att veta. Faktorsatsen är sann om vi räknar med polynom modulo p för ett primtal p . Det vill säga om $Q(a) \equiv 0 \pmod{p}$ kan vi skriva

$$Q(x) \equiv (x - a)(P(x)) \pmod{p} \quad \text{för något polynom } P(x).$$

Använd detta för att visa att ett polynom av gradtal t har högst t nollställen.

Med ett givet talpar $(a_i, b_i) \quad i = 1, 2 \dots t$ kan vi hitta ett polynom av gradtal $t - 1$ som uppfyller $Q(a_i) \equiv b_i \pmod{p}$ förutsatt att $a_i \not\equiv a_j \pmod{p}$ för $i \neq j$. Vi kan nämligen ta

$$Q(x) = \sum_{i=1}^t b_i \prod_{j \neq i} \frac{(x - a_j)}{(a_i - a_j)}.$$

Visa att Q är unikt. (Om Q_1 och Q_2 är två polynom som interpolerar (a_i, b_i) så har $Q_1 - Q_2$ t nollställen.)

Slutligen är Euklides' algoritm användbar.

Euklides används till att beräkna största gemensamma delare av två tal och att med givna y_1, y_2 och m beräkna $y_1/y_2 \pmod{p}$.

Låt oss börja med största gemensamma delare. Största gemensamma delare av två tal a och b betecknas med (a, b) och är det största tal som delar både a och b . Idén bakom algoritmen är att om ett tal delar a och b så delar det också $a - k \cdot b$ för alla heltal k . Algoritmer beskrivs nu kanske enklast med ett exempel. Låt oss ta talen 534 och 114. Anta att d delar dessa två tal, då delar det också

$$534 - 4 \cdot 114 = 78$$

och också

$$114 - 78 = 36$$

och också

$$78 - 2 \cdot 36 = 6$$

och också

$$36 - 6 \cdot 6 = 0.$$

Nu slutar algoritmen eftersom vi fick talet 0. Det sista talet 6 kontrolleras lätt vara svaret.

Låt oss beskriva hur man beräknar $y_1/y_2 \pmod{p}$.

Först beräknar man $e \equiv 1/y_2 \pmod{p}$. Sedan blir svaret $y_1 \cdot e \pmod{p}$. Eftersom p är primtal är $(p_1, y_2) = 1$ då $1 \leq y_2 \leq p - 1$. Vi beräknar ändå (p_1, y_2) med Euklides' algoritm. Låt $p = 91$ och $y_2 = 53$, vilket ger

$$91 - 53 = 38$$

$$53 - 38 = 15$$

$$38 - 2 \cdot 15 = 8$$

$$15 - 8 = 7$$

$$8 - 7 = 1.$$

Låt oss använda ekvationerna baklänges

$$\begin{aligned} 1 &= 8 - 7 = 8 - (15 - 8) = 2 \cdot 8 - 15 = 2 \cdot (38 - 2 \cdot 15) - 15 \\ &= 2 \cdot 38 - 5 \cdot 15 = 2 \cdot 38 - 5 \cdot (53 - 38) \\ &= 7 \cdot 38 - 5 \cdot 53 = 7 \cdot 91 - 12 \cdot 53. \end{aligned}$$

Ur detta följer att

$$12 \cdot 53 \equiv -1 \pmod{91}$$

och

$$(-12) \cdot 53 \equiv 1 \pmod{91}$$

och således

$$79 \cdot 53 \equiv 1 \pmod{91}$$

d.v.s.

$$\frac{1}{53} \equiv 79 \pmod{91}.$$

Litteratur

Att dela en hemlighet på det angivna sättet föreslogs först i Shamir, A., How to share a secret. *Communications of ACM* 22 (11) (1979).

Ett ställe där du kan läsa mer om modulo räkning och elementär talteori är

Hardy, G.H., & Wright, E.M., *An introduction to the theory of numbers*. Fifth edition, Oxford Univ. Press, Oxford 1979.

Generering av pseudoslumptal

JOHAN HÅSTAD

Datalogi, KTH

1. Inledning. På de flesta datorer behövs slumptal för många användningsområden (t.ex. för att stimulera förlopp i verkligheten). Datorer är ju deterministiska och därför kan de inte generera slumptal i ordets sanna mening. Istället tar ofta datorn ett slumpmässigt valt frö som t.ex. användaren bestämmer och sedan genererar från detta frö en lång serie av tal som kan användas som slumptal. Eftersom dessa inte är slumpmässiga brukar de kallas pseudoslumptal. En algoritm som tar ett kort slumpmässigt frö och producerar en längre sekvens av pseudoslumptal kallas en *pseudoslumptalsgenerator*. Denna uppgift går ut på att ge litet teori för sådana algoritmer och att studera ett exempel. Vi börjar med exemplet men vi måste först ge litet bakgrundsinformation.

2. Bakgrundsinformation. Låt m vara ett heltal och definiera att $a \equiv b \pmod{m}$ (utläses $a = b$ modulo m) innebära att a och b ger samma rest vid division med m . Vi har t.ex. att $7 \equiv 1 \pmod{3}$, $124 \equiv -6 \pmod{5}$ o.s.v.

Vårt exempel bygger på möjligheten att istället för att göra normal addition och multiplikation med heltal kan vi göra motsvarande operation och bara intressera oss för vilken rest svaret ger vid division med m . Dessutom behöver vi då bara veta vilken rest operanderna ger vid division med m .

Kontrollera att ovanstående påstående är riktigt.

Varje tal ger en av resterna $0, 1 \dots m - 1$ vid division med m . Vi kommer alltså att räkna med dessa tal. Vi har t.ex. följande

ekvationer

$$14 \cdot 5 \equiv 6 \pmod{16}$$

$$11 + 7 \equiv 1 \pmod{17}$$

$$4 \cdot 3 \equiv 0 \pmod{12}.$$

För ytterligare egenskaper av denna modulatoräkning, se boken av Hardy och Wright.

Nu kan vi definiera en flitigt använd och mycket studerad pseudoslumptalsgenerator.

$$\text{Frö} : (m, a, b, x_0).$$

(Ibland är m, a och b fixt och fröt är bara x_0 . Vi kommer dock att använda det stora fröt.) En följd tal $x_1, x_2 \dots$ genereras genom

$$x_i \equiv a \cdot x_{i-1} + b \pmod{m}.$$

Dessa tal används nu som slumptal. Denna generator kommer vi att kalla lineär kongruens generator och förkorta LKG.

Som ett exempel kan vi ta

$$m = 1241 \quad a = 613 \quad b = 113 \quad x_0 = 51$$

vilket ger

$$x_1 \equiv 613 \cdot 51 + 113 \equiv 31376 \equiv 25 \cdot 1241 + 351 \equiv 351 \pmod{1241}$$

$$x_2 = 613 \cdot 351 + 113 \equiv 583 \pmod{1241}$$

och serien blir

$$351, 583, 84, 724, 888, 899, 196, 1125, 983, 807, 886, 914.$$

Vid en ytlig inspektion ser denna serie ganska slumpmässig ut. Frågan är hur slumpmässig den är. Naturligtvis är den inte riktigt slumpmässig men den har några liknande egenskaper. Ungefär hälften av talen är jämna, ungefär hälften är större än $m/2$ o.s.v.

3. Statistiska test. Att räkna antalet jämna tal och se om det är ungefär hälften är ett exempel på ett statistiskt test. Man skulle vilja att pseudoslumptal uppför sig som riktiga slumptal vid de flesta och helst alla statistiska test. Tyvärr är det omöjligt att framställa pseudoslumptal som klarar alla statistiska test, förutsatt att pseudoslumptalserien är längre än det ursprungliga fröet.

Visa detta. (Fyll i detaljerna på nedanstående resonemang.)

LEDTRÅD. Anta att fröets längd skrivet binärt är N och de genererade pseudoslumptalen har sammanlagd binär längd M , där M är betydligt större än N . Kalla den aktuella algoritmen som genererar pseudoslumptal A . Vi kommer nu att definiera följande väldigt speciella statistiska test.

Är denna serie genererad av A på ett frö av längd N ?

På detta statistiska test kommer svaret alltid att bli *ja* om testet görs på en serie genererad av A . Om vi å andra sidan gör testet på riktiga slumptal är sannolikheten att svaret blir *ja* högst $2^N/2^M$ vilket är litet då M är större än N .

Den moderna definitionen av vad vi kan kräva av en pseudoslumptalgenerator utesluter test som är alltför komplicerade. Låt oss ge en informell version av denna definition:

DEFINITION. En algoritm A som genererar pseudoslumptal är *godkänd* om inget statistiskt test som går att utföra i rimlig tid uppför sig väsentligt olika på en serie tal genererade av A på ett slumpvist frö och på en serie riktiga slumptal.

För att göra den definitionen precis måste vi specificera *rimlig tid* och *väsentligt olika*. Det är inte så komplicerat men vi ger inte detaljerna här. Låt oss bara säga att rimlig tid betyder tid som är polynomiell i antalet siffror i fröt. (Om fröt har N siffror kan testet t.ex. ta N^3 tid att utföra.)

För att belysa punkten om rimlig tid låt oss diskutera testet *Är denna serie producerad av algoritmen A på ett frö av längd N ?* Det naiva sättet att utföra detta test är att prova alla frön av längd N , generera pseudotalföljden genom att köra A och se om någon producerad serie överensstämmer med den aktuella serien. Det finns 2^N frön att prova och detta är ju inte polynomiellt och också i praktiken tar det för lång tid även för ganska små N .

Å andra sidan klarar en liten persondator ofta att generera serien snabbt som t.ex. är fallet om man använder den generator vi beskrev ovan.

4. Är LKG godkända? I denna avdelning kommer vi att visa att LKG inte är godkända. Sedan kommer vi att diskutera problem med att göra godkända generatorer.

Låt oss nu beskriva den algoritm som avslöjar att en till synes slumpmässig följd är genererad av en LKG. Lite hjälp hur man implementerar vissa operationer ges i slutet under titeln *Bra att veta*.

Given en talserie

$$x_1, x_2, x_3 \dots x_n$$

kommer vi att försöka att konstruera m , a och b så att $x_i \equiv a \cdot x_{i-1} + b \pmod{m}$. Om inga sådana m , a och b finns kommer det att visa sig under räkningarna. Om du vill kan du sluta läsa här och försöka utan de tips som följer.

Det visar sig vara bra att definiera $y_i = x_{i+1} - x_i$. Skälet är att om x_i uppfyller $x_{i+1} \equiv ax_i + b \pmod{m}$ så uppfyller y_i relationen $y_{i+1} \equiv a \cdot y_i \pmod{m}$.

Visa detta.

Därför räknar vi först ut serien y_1, y_2, \dots, y_{n-1} . Om y_1, y_2 och y_3 är de tre första talen vet vi att $y_2 \equiv ay_1 \pmod{m}$ och $y_3 \equiv a^2y_0 \pmod{m}$. Av detta följer att $y_1 \cdot y_3 \equiv y_2^2 \pmod{m}$. Således delar m talet $m_0 = y_1 \cdot y_3 - y_2^2$ (förutsatt att det inte är 0). Till att börja med kan vi gissa att $m = m_0$ och $a = a_0 = y_2/y_1 \pmod{m_0}$. (Om det finns flera möjligheter för y_2/y_1 , välj ett. Är y_2/y_1 odefinierat har vi gissat för stort m_0 . Hur detta korrigeras beskrivs under *Bra att veta*.)

Betrakta nu generatorn

$$\overline{y_i} \equiv a_0 \overline{y_{i-1}} \pmod{m_0} \text{ med } \overline{y_1} = y_1.$$

Om $\overline{y_i} = y_i$ för alla i har vi nog rätt generator. Annars kan man visa att eftersom $m|m_0$ att $a_0 \overline{y_i} \equiv ay_i \pmod{m}$, och således $\overline{y_{i+1}} = y_{i+1} \pmod{m}$. (Försök göra detta.) Detta innebär att m delar $\overline{y_{i+1}} - y_{i+1}$ om detta inte är 0. Således kan vi uppdatera vår gissning av m till största gemensamma delaren av m_0 och $\overline{y_{i+1}} - y_{i+1}$ (eftersom m delar båda talen). Konstruera en ny generator och börja om, o.s.v. Om vi kan bestämma a och m på detta sätt är det lätt att sen hitta b genom att

$$b \equiv x_2 - ax_1 \pmod{m}.$$

Fundera ut alla detaljer och implementera sedan ovanstående algoritm, eller gärna någon variant du själv hittar på.

Försök analysera hur lång tid din algoritm tar om det riktiga m har högst n siffror. (Hur många gånger kan du vara tvungen att på nytt gissa m ? Hur många operationer måste du göra för att hitta varje gissning?)

Två exempel att köra algoritmen på är följande:

```
18192 45941 42086 43501 31735 2718
41115 31870 28933 918
```

och

```
15584 34075 5151 39785 21388 35991
18143 31570 37764 15032 10300 .
```

Om du har någon intresserad kamrat kan ni byta exempel.

5. Diskussion. Nu när vi har visat att LKG inte är godkända kommer naturligtvis frågan vilka generatorer som är godkända. Några generatorer har konstruerats som antagligen är godkända, men det har ingen lyckats visa. Skälet är följande: För att visa att en generator inte är godkänd, behöver man bara visa att det finns en algoritm som går rimligt fort som skiljer på tal genererade av generatören och tal som är verkliga slumpal. För att visa att en generator är bra behöver man visa att varje statistisk test som går rimligt fort misslyckas med att skilja tal som generatören producerat från riktiga slumpal. I allmänhet att visa att vissa problem kräver lång tid att beräkna är ett mycket viktigt problem som fortfarande är öppet och det återstår mycket forskning inom detta område, som kallas *komplexitetsteori*.

Vad menar jag då med att det finns generatorer som antagligen är godkända. Jo, man har konstruerat generatorer som om de inte är godkända så går något mycket studerat beräkningsproblem att lösa betydligt effektivare än alla har lyckats med. T.ex. har man visat att det finns godkända generatorer om faktorisering av stora tal är svårt. Men det är inget bevis.

Försök konstruera en generator som är effektiv och verkar godkänd. Lättast är att producera tal på något iterativt sätt som vi gjorde med LKG. Använd någon mer komplicerad funktion än att

bara multiplicera med en konstant och lägga till en annan konstant. Använd gärna modulär räkning men försök att göra något annorlunda.

Ta reda på vilken pseudoslumptalsgenerator din dator använder och försök att visa att den inte är godkänd. Ofta använder den modulär räkning. Försök göra något liknande det vi gjorde med LKG. Gissa först m och hitta sedan de övriga konstanterna som ingår.

Bra att veta. I algoritmen behövs två trick. Nämligen att beräkna största gemensamma delare av två tal och att givet y_1, y_2 och m beräkna $y_1/y_2 \pmod{m}$. Båda görs med Euklides' algoritm.

Låt oss börja med största gemensamma delare. Största gemensamma delare av två tal a och b betecknas med (a, b) och är det största tal som delar både a och b . Idén bakom algoritmen är att om ett tal delar a och b så delar det också $a - k \cdot b$ för alla heltal k . Algoritmen beskrivs nu kanske enklast med ett exempel.

Låt oss ta talen 534 och 114. Anta att d delar dessa två tal, då delar det också

$$534 - 4 \cdot 114 = 78$$

och också

$$114 - 78 = 36$$

och också

$$78 - 2 \cdot 36 = 6$$

$$36 - 6 \cdot 6 = 0.$$

Nu slutar algoritmen eftersom vi fick talet 0. Det sista talet 6 kontrolleras lätt vara svaret.

Låt oss beskriva hur man beräknar $y_1/y_2 \pmod{m}$. Först måste vi definiera vad detta betyder. Låt oss säga att c är det tal c så att

$c \cdot y_2 \equiv y_1 \pmod{m}$. Om det finns flera c så välj ett godtyckligt, finns det inget sådant tal är y_1/y_2 odefinierat.

Beräkna nu y_1/y_2 på följande sätt.

Beräkna först (y_1, y_2) . Om detta är d , använd att

$$y_1/y_2 = (y_1/d)/(y_2/d).$$

Vi kan således anta att $(y_1, y_2) = 1$. Om nu $(y_2, m) > 1$ är y_1/y_2 odefinierat (*visa detta*). Om $(y_2, m) = 1$, beräkna $e \equiv 1/y_2 \pmod{m}$ med Euklides algoritm. Sedan är svaret $y_1 \cdot e \pmod{m}$.

Vi ger ett exempel på hur man beräknar $1/53 \pmod{91}$. Utför Euklides algoritm på 53 och 91

$$91 - 53 = 38$$

$$53 - 38 = 15$$

$$38 - 2 \cdot 15 = 8$$

$$15 - 8 = 7$$

$$8 - 7 = 1.$$

Låt oss nu använda ekvationerna baklänges.

$$\begin{aligned} 1 &= 8 - 7 = 8 - (15 - 8) = 2 \cdot 8 - 15 \\ &= 2 \cdot (38 - 2 \cdot 15) - 15 = 2 \cdot 38 - 5 \cdot 15 \\ &= 2 \cdot 38 - 5 \cdot (53 - 38) = 7 \cdot 38 - 5 \cdot 53 = 7 \cdot 91 - 12 \cdot 53. \end{aligned}$$

Ur detta följer att

$$12 \cdot 53 \equiv -1 \pmod{91}$$

och

$$(-12) \cdot 53 \equiv 1 \pmod{91}$$

och således

$$79 \cdot 53 \equiv 1 \pmod{91}$$

d.v.s.

$$\frac{1}{53} \equiv 79 \pmod{91}.$$

Ibland i vår algoritm kan vi råka ut för tal y_1 och y_2 så att $y_1/y_2 \pmod{m_0}$ inte existerar beroende på att vår gissning m_0 av m är för stor. Om $(y_1, y_2) = 1$ beror detta på att $(y_2, m_0) > 1$ och vi byter nu ut m_0 mot $\frac{m_0}{(m_0, y)}$.

Vi kan vara tvungna att göra detta flera gånger men sedan återstår den största faktor av m_0 för vilken y_1/y_2 existerar.

Visa detta.

Litteratur

För en bok som behandlar elementär talteori Hardy, G.H. & Wright, E.M., *An Introduction to the theory of numbers*. Fifth edition, Oxford Univ. Press, Oxford 1979.

En utförlig diskussion av lineär kongruens generatorer finns i Knuth, D.E., *The Art of Computer Programming, Vol. II. Semi-numerical Algorithms*, Addison Wesley 1981.

Den moderna definitionen av slumpstal finns i Blum, M., Micali, S., How to generate cryptographically strong sequences of pseudo-random bits. *SIAM Journal on Computing*, 13, s 850–864.

Algoritmen att LKG inte är godkänd är delvis tagen ur Plumstead, J.B., Inferring a sequence generated by a linear congruence. *Proceedings of 23rd IEEE Symposium on Foundations of computer Science*, s 153–159.

Offentlig kryptering

JOHAN HÅSTAD

KTH

1. Inledning. Denna uppgift går ut på att studera ett offentligt kryptosystem. Med detta menas ett kryptosystem där det är offentligt hur man krypterar, men trots detta kan bara någon som besitter en hemlighet dekryptera. Metoden vi kommer att beskriva kallas RSA-systemet efter Ron Rivest, Adi Shamir och Len Adleman som föreslog systemet. Huvuddelen av uppgiften går ut på att implementera systemet.

RSA-systemet bygger på talteori och för att beskriva och sedermera förstå det behövs litet bakgrund.

2. Litet elementär talteori. Låt m vara ett heltal. Vi kommer att räkna med talen $0, 1, 2 \dots m - 1$ modulo m . Detta innebär att vi har de vanliga räknesätten $+$, $-$ och \cdot , men vi är bara intresserade av vilken rest svaret ger vid division med m . Vi skriver $+$, $-$ och \cdot som vanligt medan vi använder likhetstecken med 3 streck samt lägger till $(\text{mod } m)$ för att markera att vi bara är intresserade av vilken rest talen ger vid division med m . Vi bör här observera att för att veta vilken rest svaret ger vid division med m är det tillräckligt att veta vilken rest operanderna ger vid division med m och inte behöver deras exakta värde.

Visa detta.

Till exempel har vi

$$4 \cdot 6 \equiv 6 \pmod{9}$$

$$13 + 1 \equiv 14 \pmod{18}$$

$$2 - 8 \equiv 11 \pmod{17}$$

Vi kommer också att behöva division med tal modulo m och här får man tänka till litet eftersom kvoten av två heltal inte vanligtvis är ett heltal. Vi diskuterar inte detta problem närmare utan definierar för tillfället $a/b \pmod{m}$ som det tal $c \pmod{m}$, så att $b \cdot c \equiv a \pmod{m}$. De problem som kan uppstå (inget sådant c (t.ex. $7/4 \pmod{14}$)), flera sådana c (t.ex. $10/4 \pmod{14}$)) kommer inte att spela någon större roll för oss, men du kan fundera på vad som bör göras. Hur man effektivt beräknar c beskrivs i slutet under *Bra att veta*.

En av hörnstenarna i RSA-systemet är följande klassiska sats.

FERMATS LILLA SATS. *Om p är ett primtal och $1 \leq a < p$ så är $a^{p-1} \equiv 1 \pmod{p}$.*

EXEMPEL. Låt $p = 17$ och $a = 3$.

$$3^{16} \equiv 9^8 \equiv 81^4 \equiv 13^4 \equiv 169^2 \equiv (-1)^2 \equiv 1 \pmod{17}.$$

Som vi gjorde här är det ofta bekvämt att använda t.ex. (-1) i stället för 16 vid handräkning. Detta ger inget problem ty (-1) och 16 ger samma rest vid division med 17.

Om du vill kan du försöka visa Fermats lilla sats. Enklast är kanske att visa $a^p \equiv a \pmod{p}$ med induktion över a . Använd binomialsatsen och att $\binom{p}{i}$ är delbart med p då $1 \leq i \leq p-1$.

Låt oss fortsätta med litet fler bakgrundsfakta. Låt N vara produkten av två olika primtal, $N = p \cdot q$.

LEMMA 1. *Anta att vi vet vilken rest ett tal ger vid division med p och vid division med q . Då vet vi också vilken rest det ger vid division med $N = p \cdot q$ förutsatt att $p \neq q$ och p och q är primtal.*

För att verifiera detta behöver vi bara visa att om a och b ger samma rest vid division med p och vid division med q , så ger de

samma rest vid division med N . Men detta är ganska uppenbart då $a - b$ är delbart med både p och q och således med $p \cdot q$.

Skriv ut beviset med alla detaljer.

Lemmat är ett specialfall av en sats som kallas den kinesiska restsatsen. Du kan hitta mer information om elementär talteori i boken av Hardy och Wright som nämns i litteraturlistan.

3. RSA-systemet. Låt N vara produkten av två primtal $N = p \cdot q$. Det är viktigt att det bara är mottagaren som vet p och q , det enda som kommer att offentliggöras är N . Låt M vara minsta gemensamma multipel av $p - 1$ och $q - 1$. Eftersom båda talen är jämna vet vi t.ex. att $M \leq \frac{(p-1)(q-1)}{2}$. Låt nu k och k' vara två heltal så att $k \cdot k' \equiv 1 \pmod{M}$. Talen M och k' kommer att hållas hemliga medan k publiceras. Innan vi fortsätter, låt oss ge ett exempel.

Låt $N = 323 = 17 \cdot 19$, $M = 144$, $k = 5$, $k' = 29$. Anta att 210 symboliserar det hemliga meddelandet. Det krypteras som

$$201^k = 201^5 \equiv 201 \cdot (201^2)^2 \equiv 201 \cdot 26^2 \equiv 201 \cdot 30 \equiv 216 \pmod{323}.$$

Observera att detta kan göras av någon som känner N och k .

Detta dekrypteras genom

$$\begin{aligned} 216^{k'} &\equiv 216^{29} \equiv 216^{16} \cdot 216^8 \cdot 216^4 \cdot 216 \\ &\equiv 273 \cdot 220 \cdot 64 \cdot 216 \equiv 305 \cdot 258 \equiv 201 \pmod{323}, \end{aligned}$$

och vi återfår meddelandet. Observera att mottagaren bara behöver k' och således kan glömma p, q och M .

Låt oss nu beskriva systemet formellt och förklara varför man återfår ursprungsvärdet.

Låt m vara det hemliga meddelandet som kan skrivas som ett tal (a blir 01, b blir 02 o.s.v.). Anta att $1 \leq m < N$. (Om meddelandet

är långt blir talet större, men då tänker vi oss det som flera tal som alla är mindre än N .)

OFFENTLIG KRYPTERINGS ALGORITM. I offentlig kryptering beräknas c (chiffertexten) genom $c \equiv m^k \pmod{N}$. N och k är offentliga som tidigare nämnts.

HEMLIG DEKRYPTERINGS ALGORITM. Meddelandet återfås genom $m \equiv c^{k'} \pmod{N}$.

Låt oss visa att dekrypteringen är korrekt. Vi har att $c^{k'} \equiv m^{k \cdot k'} \pmod{N}$ och därmed behöver vi bara följande.

LEMMA 2. *För alla m , $0 \leq m < N$ är det sant att $m^{k \cdot k'} \equiv m \pmod{N}$.*

Genom att använda Lemma 1 behöver vi bara visa att $m^{k \cdot k'} \equiv m \pmod{p}$ och $m^{k \cdot k'} \equiv m \pmod{q}$. Bevisen är identiska och därmed visar vi bara den första likheten. Genom valet av M, k och k' vet vi att

$$k \cdot k' = a \cdot M + 1 = a \cdot b \cdot (p - 1) + 1$$

för heltal a och b . Således är om $m \not\equiv 0 \pmod{p}$

$$m^{k \cdot k'} \equiv (m^{p-1})^{a \cdot b} \cdot m \equiv 1^{a \cdot b} \cdot m \equiv m \pmod{p},$$

där vi har använt Fermats lilla sats. Om $m \equiv 0 \pmod{p}$ är identiteten självklar. Därmed har vi visat Lemma 2 och dekrypteringsalgoritmen är således korrekt.

4. Diskussion. Varför fungerar RSA som ett kryptosystem? Skälet är att det verkar som om det är svårt att finna k' givet N och k . Den bästa kända algoritmen för att göra detta är att faktorisera N och räkna fram M och sedan k' . Faktorisering tar dock lång

tid och även de bästa algoritmer på specialbyggda maskiner klarar av högst ungefär 80-siffriga tal på en månad. Det är möjligt att det finns bättre algoritmer för faktorisering eller att det finns ett sätt att dekryptera RSA utan att faktorisera N . Det pågår mycket forskning inom detta område, men ännu finns inga sådana resultat.

Om det nu är svårt att räkna ut k' , hur går det då till att konstruera systemet? Skälet är förstås att konstruktören känner till p och q och kan räkna ut M och sedan k och k' . Enda svårigheten är att konstruera stora primtal p och q . Detta görs genom att välja ett slumpvis stort tal och testa om det är primtal. Mycket forskning har ägnats åt att testa om stora primtal, men vi kommer här bara att välja ett simpelt test som fungerar för det mesta.

IDÉ. Om $a^{p-1} \equiv 1 \pmod{p}$ för ett slumpvist valt a , $1 \leq a < p$ så är p antagligen primtal.

Denna idé kan göras mer precis, men för det mesta räcker den. Således blir det vårt primtalstest *Prova om $a^{p-1} \equiv 1 \pmod{p}$ för ett antal slumpvist valda a .*

För en utförligare diskussion om primtalstest kan du läsa artiklarna av Pomerance och Wagon som nämns i litteraturförteckningen.

5. Förslag till uppgifter. Nu har du all information som krävs för att implementera RSA-systemet. Gör det med så stora p och q (och därmed N) du kan. Det kan vara viktigt att komma ihåg att datorers heltal är av begränsad storlek. Några tips finns i avdelningen *Bra att veta*.

RSA kan användas till annat än bara enkel kryptering. Till exempel kan man signera meddelanden. Signaturen av ett meddelande m är $m^{k'}$ (mod N).

Kalla signaturen s . Det är lätt att verifiera signaturen då $s^k \equiv m$ (mod N) och k och N ju är offentliga. Att prestera en korrekt sig-

natur av ett meddelande är lika svårt som att dekryptera ett meddelande.

Försök göra något praktiskt med hjälp av RSA-systemet. I dagens hemlighetsfulla värld finns det mycket digital information som skall skyddas eller verifieras som autentisk. Om du lyckas riktigt bra och börjar tjäna pengar på någon produkt, så bör du veta att RSA är patenterat i USA.

6. Slutord. RSA-system ställer många frågor som är obesvarade.

Hur mycket tid måste en algoritm som faktoriserar heltal ta?

Varför är det lättare att visa att tal är primtal än att faktorisera dem?

Finns det andra bra system för offentlig kryptering?

Forskning pågår för att svara på dessa frågor (se referenslistan). Över huvud taget finns det många obesvarade frågor om existens av effektiva algoritmer. Detta område av matematiken brukar kallas komplexitetsteori.

Bra att veta. För att beräkna $a^k \pmod{N}$ för stora tal a, k och N kan det vara bra att organisera räkningarna med litet eftertanke. Vi tar exemplet $N = 1013$, $a = 514$ och $k = 411$. Börja med att skriva k binärt, d.v.s. som en summa av två-potenser

$$k = 256 + 128 + 16 + 8 + 2 + 1.$$

Beräkna sedan $a^{2^i} \pmod{N}$ för $i = 0, 1 \dots 8$. Detta görs lätt genom

att kvadrera föregående tal.

$$\begin{aligned}
 a &\equiv 514 \\
 a^2 &\equiv 514^2 \equiv 816 \pmod{1013} \\
 a^4 &\equiv 816^2 \equiv 315 \pmod{1013} \\
 a^8 &\equiv 315^2 \equiv 964 \pmod{1013} \\
 a^{16} &\equiv 964^2 \equiv 375 \pmod{1013} \\
 a^{32} &\equiv 375^2 \equiv 831 \pmod{1013} \\
 a^{64} &\equiv 831^2 \equiv 708 \pmod{1013} \\
 a^{128} &\equiv 708^2 \equiv 842 \pmod{1013} \\
 a^{256} &\equiv 842^2 \equiv 877 \pmod{1013}.
 \end{aligned}$$

Det följer att

$$\begin{aligned}
 514^{411} &\equiv 514^{256} \cdot 514^{128} \cdot 514^{16} \cdot 514^8 \cdot 514^2 \cdot 514 \\
 &\equiv 877 \cdot 842 \cdot 375 \cdot 964 \cdot 816 \cdot 514 \\
 &\equiv 970 \cdot 872 \cdot 42 \\
 &\equiv 220 \cdot 872 \\
 &\equiv 383 \pmod{1013}.
 \end{aligned}$$

Man kan använda liknande trick men huvudidén är den samma. Det är också bra att kunna beräkna största gemensamma delare av två tal och att givet y_1, y_2 och m beräkna $y_1/y_2 \pmod{N}$. Båda görs med Euklides' algoritm.

Låt oss börja med största gemensamma delare. Största gemensamma delare av två tal a och b betecknas med (a, b) och är det största tal som delar både a och b . Idén bakom algoritmen är att om ett tal delar a och b så delar det också $a - k \cdot b$ för alla heltal k . Algoritmen beskrivs nu kanske enklast med ett exempel.

Låt oss ta talen 534 och 114. Anta att d delar dessa två tal, då delar det också

$$534 - 4 \cdot 114 = 78$$

och också

$$114 - 78 = 36$$

och också

$$78 - 2 \cdot 36 = 6$$

$$36 - 6 \cdot 6 = 0.$$

Nu slutar algoritmen eftersom vi fick talet 0. Det sista talet 6 kontrolleras lätt vara svaret.

Låt oss beskriva hur man beräknar $y_1/y_2 \pmod{N}$. Först måste vi definiera vad detta betyder. Låt oss säga att c är det tal så att $c \cdot y_2 \equiv y_1 \pmod{N}$. Om det finns flera c , så välj ett godtyckligt, finns det inget sådant tal är y_1/y_2 odefinierat.

Beräkna nu y_1/y_2 på följande sätt. Beräkna först (y_1, y_2) . Om detta är d , använd att $y_1/y_2 = (y_1/d)/(y_2/d)$. Vi kan således anta att $(y_1, y_2) = 1$. Om nu $(y_2, N) > 1$ är y_1/y_2 odefinierat. (*Visa detta.*) Om $(y_2, N) = 1$ beräkna $e \equiv 1/y_2 \pmod{N}$ med Euklides algoritm. Sedan är svaret $y_1 \cdot e \pmod{m}$. Vi ger ett exempel på hur man beräknar $1/53 \pmod{91}$.

Utför Euklides algoritm på 53 och 91.

$$91 - 53 = 38$$

$$53 - 38 = 15$$

$$38 - 2 \cdot 15 = 8$$

$$15 - 8 = 7$$

$$8 - 7 = 1.$$

Låt oss nu använda ekvationerna baklänges.

$$\begin{aligned}1 &= 8 - 7 = 8 - (15 - 8) = 2 \cdot 8 - 15 \\ &= 2 \cdot (38 - 2 \cdot 15) - 15 = 2 \cdot 38 - 5 \cdot 15 \\ &= 2 \cdot 38 - 5 \cdot (53 - 38) = 7 \cdot 38 - 5 \cdot 53 \\ &= 7 \cdot 91 - 12 \cdot 53.\end{aligned}$$

Ur detta följer att

$$12 \cdot 53 \equiv -1 \pmod{91}$$

och

$$(-12) \cdot 53 \equiv 1 \pmod{91}$$

och således

$$79 \cdot 53 \equiv 1 \pmod{91}$$

d.v.s.

$$\frac{1}{53} \equiv 79 \pmod{91}.$$

Litteratur

En referens för elementär talteori är Hardy, G.H., & Wright, E.M., *An introduction to the theory of numbers*. Fifth edition, Oxford Univ. Press, Oxford 1979.

Där finns bl.a. Fermats lilla och kinesiska restsatsen.

RSA-systemet presenterades först i Rivest, R., Shamir, A., Adleman, L., A method for obtaining digital signatures and public-key cryptosystems. *Communications of ACM* 21 (1979), s 120–126.

Om du vill läsa mer om primtalstest finns följande artiklar
Wagon, S., Primality testing. *Mathematical intelligenser*, 8:3 (1986),
s 58–61.

Pommerance, C., Recent developments in primality testing. *Mathe-
matical intelligenser*, 3 (1981), s 97–105.

För andra förslag på offentliga krypteringssystem, se
Merkle, R., Hellman, M., Hiding information and signatures in trap-
door knapsacks. *IEEE Transactions on Information Theory*, IT 24
(1978), s 525–553.

McEliece, R.J., A public-key cryptosystem based on algebraic coding
theory. *DSN Progress Report* 42–44, Jan. and Feb. 1978.

De första av dessa har forcerats, se t.ex.
Shamir, A., A polynomial time algorithm for breaking the basic
Merkle–Hellman cryptosystem. *Proceedings of 23rd IEEE Sympo-
sium on Foundations of Computer Science*, 1982, s 145–152.

Felrättande koder

THOMAS HÖGLUND

KTH, Stockholm

Inledning. Felrättande koder används i CD-skivspelare, vid bildöverföring från satellit till jorden och vid kommunikationsradiotrafik – för att ta några exempel.

I samtliga dessa fall är det fråga om att överföra en bitföljd (d.v.s. en följd av nollor och ettor) från en sändare till en mottagare. Problemet är att följderna kan bli störda under överföringen och att det mottagna meddelandet därför kan få dålig kvalitet.

Denna specialuppgift går att genomföra för hand men det är en fördel om du har tillgång till en dator då du kodar och avkodar meddelandena och då du simulerar störningar.

1. Tag ett pennset med 8 av våra vanligaste färger. Identifiera varje färg med en bitföljd av längd 3. T.ex.

rött	orange	gult	grönt	blått	violett	svart	vitt
000	001	010	011	100	101	110	111

Välj en enkel och tydlig färgbild som endast innehåller de färger du valt - en flagga t.ex. Lägg ett rutnät över bilden. Läs av färgen på rutorna i någon viss ordning. Du har nu fått en bitföljd som svarar mot din bild.

En annan möjlighet är att ta några meningar ur en vanlig text och översätta meningen till bitar t.ex. genom identifikationen: A = 00000, B = 00001, ..., Ö = 11011, mellanslag = 11100, punkt = 11101, komma = 11110, övrigt = 11111.

2. För att få en bild av vad störningar kan ställa till med ska du här simulera störningar. Gör ett kast med två tärningar för varje

bit i följd. Ändra biten (nolla till etta, etta till nolla) om båda tärningarna visar sexor, annars låter du biten stå kvar. I medeltal kommer alltså var trettiosjätte bit att ändras. Rita upp den mottagna bilden.

Allmänt om koder. Ett sätt att upptäcka fel är att sända varje bit 2 gånger, 0 kodas till 00 och 1 till 11. Om den mottagna följd är 10 så ser vi att ett fel uppstått men vi kan inte avgöra om den ursprungliga biten var 1 eller 0. Denna kod är alltså felupptäckande men inte felrättande. För att få en felrättande kod kan vi sända varje bit 3 gånger. Om den mottagna följd är 010 så ser vi att (minst) ett fel uppstått och att den ursprungliga biten (troligen) var 0. Risker finns naturligtvis att två fel uppstått men denna risk är liten jämfört med sannolikheten att *ett* fel uppstått - förutsatt att felsannolikheten är liten och att fel uppstår oberoende av varandra. Denna kod rättar 1 fel på 3 bitar. En nackdel är emellertid att meddelandet förlängs med en faktor 3.

Mer allmänt kan vi dela in en bitföljd i block om k bitar, där k är ett positivt heltal, t.ex. $k = 2$. Låt n vara ett heltal som är större än k , t.ex. $n = 5$. Tildela varje bitföljd av längd k en bitföljd av längd n . De senare bitföljderna kallas kodord och vi säger att vi valt en $[n, k]$ kod.

Om vi t.ex. har valt $[5, 2]$ koden:

$$00 \rightarrow 00000, \quad 01 \rightarrow 01011, \quad 10 \rightarrow 10101, \quad 11 \rightarrow 11110,$$

så kodas bitföljden 011011 så här:

$$011011 = 01 \quad 10 \quad 11 \rightarrow 01011 \quad 10101 \quad 11110.$$

Observera att endast ett fåtal av alla n -bitsföljder är kodord.

Avståndet mellan två bitföljder av längd n är antalet av de n positionerna där de två följderna har olika bitar. Så t.ex. är avståndet mellan 10101 och 11110 lika med 3 eftersom de skiljer sig på positionerna 2, 4 och 5.

Då en bitföljd sänts iväg och mottagits ska den avkodas. Om en mottagen n -bitsföljd inte är ett kodord så ersätts den med det närmaste kodordet - om ett sådant finnes. I exemplet ersätts följderna 01000 med 00000 eftersom avståndet mellan dessa är 1 medan avståndet mellan 01000 och varje annat kodord är minst 2. Däremot kan vi inte avkoda följderna 01100 eftersom denna har avståndet 2 till både 00000 och 11110 och avståndet 3 till de två övriga kodorden.

Denna kod rättar alltså ett fel och den är optimal i den meningen att det inte finns någon $[5, 2]$ kod som rättar fler fel. Problemet att för godtyckliga tal n och k finna en optimal $[n, k]$ kod är fortfarande (1988) olöst.

Vi ska nu införa en addition mellan bitföljder. Bitar adderas så här: $0 + 0 = 0$, $0 + 1 = 1 + 0 = 1$, $1 + 1 = 0$. Två lika långa bitföljder adderas genom att addera bitarna positionsvis. T.ex.

$$\begin{aligned} (1, 1, 1, 0, 0) + (0, 1, 0, 1, 0) &= (1 + 0, 1 + 1, 1 + 0, 0 + 1, 0 + 0) \\ &= (1, 0, 1, 1, 0). \end{aligned}$$

Här satte vi för tydlighetens skull parenteser runt bitföljderna och kommatecken mellan bitarna.

Hammingkoden. Denna kod är definierad för vissa n och k nämligen då $n = 2^r - 1$ och $k = n - r$ för $r = 2, 3, 4, \dots$ eller $[n, k] = [3, 1]$, $[7, 4]$, $[15, 11]$, $[31, 26]$, Det är önskvärt att kvoten $\frac{n}{k}$ är nära 1 eftersom det kodade meddelandet förlängs med denna faktor. För Hammingkoden är dessa kvoter 3, 1.75, 1.36, 1.19, Däremot så rättar denna kod bara 1 fel i varje kodord av längd n .

Hammingkoden $[n, k]$ kan konstrueras på följande sätt:

– Skriv upp alla 2^r bitföljder av längd r . Stryk nollföljden och de r följder som innehåller exakt en etta. Kvar blir $2^r - 1 - r = k$ följder av längd r . I fallet $[7, 4]$ (d.v.s. $r = 3$) får vi de $8-1-3=4$ följderna 011 101 110 111.

– Följden $10\dots 0$ av längd k kodas till det kodord av längd n vars k första bitar är just $10\dots 0$ och vars återstående r bitar är den första av de uppskrivna följderna. Kodordet svarande mot $010\dots 0$ fås genom att lägga till den andra av de kvarvarande följderna till $010\dots 0$ o.s.v. tills slutligen $0\dots 01$ kodas genom att lägga till den sista följden.

I exemplet

$1000 \rightarrow 1000011$ $0100 \rightarrow 0100101$ $0010 \rightarrow 0010110$ $0001 \rightarrow 0001111$.

– Koden är linjär: Om x kodas till \bar{x} och y till \bar{y} så kodas $x + y$ till $\bar{x} + \bar{y}$. T.ex. kodas 1010 till 1010101 eftersom $1010 = 1000 + 0010$ och $1000011 + 0010110 = 1010101$. Vidare kodas 1011 till 1011010 eftersom $1011 = 1010 + 0001$ och $1010101 + 0001111 = 1011010$ o.s.v.

3. Låt för varje kodord x , $B(x)$ beteckna mängden av alla bitföljder av längd n som har avståndet högst 1 till x . Övertyga dig om att följande resonemang är riktigt i första hand för $r = 2$ och $r = 3$ men även för ett allmänt r . Avståndet mellan två olika kodord är alltid minst 3. $B(x)$ består av $n + 1 = 2^r$ kodord. Om x och y är två olika kodord så finns ingen bitföljd av längd n som ligger i både $B(x)$ och $B(y)$. Det finns $2^k = 2^{n-r}$ kodord, så de 2^{n-r} mängderna $B(x)$ innehåller därför $2^{n-r} \times 2^r = 2^n$ olika bitföljder av längd n . Därför finns ingen bitföljd av längd n som ligger utanför alla $B(x)$.

Vi har alltså visat att för varje bitföljd av längd n finns exakt ett kodord på avståndet högst 1.

4. Koda bitföljden under punkt 1 med Hammingkoden [7, 4]. Simulera störningar som under punkt 2. Avkoda. Rita bilden. Jämför med resultatet under punkt 2.

5. I de i inledningen nämnda exemplen kommer störningarna oftast i skurar (flera fel i följd). För att simulera en sådan situation kan du göra så här: Kasta två tärningar. Ändra alla de tre första bitarna om du får två sexor, annars låter du bitarna stå kvar. Upprepa för de tre följande bitarna o.s.v. Här kommer alltså störningarna tre och tre men vi har fortfarande att i medeltal är var trettiosjätte bit störd. Eller hur? Avkoda och jämför med tidigare resultat.

Multiplikation av bitföljder. Vi har tidigare definierat addition av bitföljder. För att konstruera koder som klarar av störningar av den typ du simulerade under punkt 5 ska vi nu även införa en multiplikation.

Varje bitföljd $(b_0, b_1, \dots, b_{m-1})$ svarar mot ett polynom $b_0 + b_1x + b_2x^2 + \dots + b_{m-1}x^{m-1}$ och omvänt. Så t.ex. svarar 101 mot $1 + 0 \cdot x + 1 \cdot x^2 = 1 + x^2$ och 010 mot $0 + 1 \cdot x + 0 \cdot x^2 = x$. Följden 000 svarar mot nollpolynomet (talet 0) och 100 mot det polynom som är konstant 1. Det finns två polynom av grad 1: x och $1 + x$. Det finns fyra polynom av grad 2: x^2 , $x + x^2$, $1 + x^2$ och $1 + x + x^2$. De tre första av andragsgradspolynomen går att skriva som en produkt av två polynom av lägre grad, nämligen $x^2 = x \cdot x$, $x + x^2 = x(1 + x)$ och $1 + x^2 = (1 + x)(1 + x)$. (Här utnyttjade vi att $1 + 1 = 0$.) Polynomet $1 + x + x^2$ går däremot inte att skriva som en produkt av två polynom av lägre grad. Sådana polynom kallas irreducibla (jämför primtal). Det finns alltså ett irreducibelt polynom av grad 2: $1 + x + x^2$.

Betrakta nu bitföljder av längd 2. Det finns 4 sådana: 00, 10, 01 och 11 och de svarar mot polynomen 0, 1, x och $1 + x$. När vi multiplicerar dessa ska vi räkna som om $1 + x + x^2 = 0$ d.v.s. som

om $x^2 = 1 + x$. (Addera $1 + x$ till båda sidor och utnyttja att $2 = 0$.) Vi får $x \cdot x = x^2 = 1 + x$, $x(1 + x) = x + x^2 = x + (1 + x) = 1$, $(1 + x)(1 + x) = 1 + 2x + x^2 = 1 + x^2 = 1 + (1 + x) = x$.

För att slippa skriva så mycket skriver vi 0, 1, 2, 3 i stället för 00, 10, 01 respektive 11. D.v.s. i stället för 0, 1, x respektive $1 + x$. Vi har alltså multiplikationstabellen

\times	0	1	2	3	$+$	0	1	2	3
0	0	0	0	0	0	0	1	2	3
1	0	1	2	3	1	1	0	3	2
2	0	2	3	1	2	2	3	0	1
3	0	3	1	2	3	3	2	1	0

Till höger står den additionstabell den addition vi redan infört ger.

Observera att vi kan räkna med dessa element ungefär som med vanliga tal. Vi kan subtrahera: För varje p och q finns exakt ett r så att $r + p = q$ (nämligen $r = p + q$). Vi kan även dividera med p om $p \neq 0$: För varje q finns exakt ett s så att $sp = q$. Om vi skriver q/p i stället för s så har vi t.ex. $3/2=2$ och $1/2=3$. Lägg också märke till att alla element, p , uppfyller $p^4 = p$.

Det finns två irreducibla tredjegradspolynom: $1 + x + x^3$ och $1 + x^2 + x^3$. Välj t.ex. det första. Räkna som om detta är 0 (d.v.s. ersätt x^3 med $1 + x$) när du multiplicerar polynom av grad högst 2. Det medför att x^4 ska ersättas med $x + x^2$ eftersom $x^4 = x \cdot x^3 = x(1 + x) = x + x^2$. Så till exempel är $7 \cdot 5 = 6$ eftersom

$$\begin{aligned} (1 + x + x^2)(1 + x^2) &= 1 + x + 2x^2 + x^3 + x^4 \\ &= 1 + x + 0 + (1 + x) + (x + x^2) = x + x^2. \end{aligned}$$

6. Gör en multiplikationstabell och en additionstabell för de 8 bitföljderna av längd 3. Gör även en tabell över $p, p^2, p^3, p^4, p^5, p^6, p^7$

för alla de 8 elementen p . Även här kan vi subtrahera och dividera. Observera att alla element p , uppfyller $p^8 = p$ och att vissa element är sådana att de 7 potenserna p^i , $i = 1, \dots, 7$ ger alla element utom 0. Sådana element kallas primitiva. Vilka element är primitiva? Om p är primitivt så gäller $1 + p + p^2 + \dots + p^6 = 0$.

Reed-Solomon koden. På motsvarande sätt kan man definiera multiplikation för bitföljder av längd m . Reed-Solomon koden kodar bitföljder av längd mk bitar till bitföljder av längd $m(2^m - 1)$ bitar. Här är m och k är positiva heltal.

Vi ska i fortsättningen hålla oss till fallet $m = 3$, $k = 5$. De bitföljder som ska kodas har alltså längden $3 \times 5 = 15$ bitar och kodorden har längden $3(2^3 - 1) = 3 \times 7 = 21$ bitar (5 element kodas till 7 element). Meddelandet förlängs alltså med en faktor 1.4 att jämföra med 1.75 för Hammingkoden [7,4].

Reed-Solomon koden kan konstrueras så här: Välj ett primitivt element p , t.ex. $p = 2 = (0, 1, 0)$. Dela in 15-bitsföljden i 5 block av längd 3. Vi får då en följd $(a_0, a_1, a_2, a_3, a_4)$ av element. Denna kodas till de 7 elementen (c_0, c_1, \dots, c_6) , där

$$(\star) \quad c_i = a_0 + a_1 p^i + a_2 p^{2i} + a_3 p^{3i} + a_4 p^{4i}$$

för $i = 0, 1, 2, 3, 4, 5, 6$. Här har vi vanlig multiplikation i exponenterna $2i$, $3i$ och $4i$.

Så t.ex. om vi startar med 15-bitsföljden 100010000110111 och låter $p = 2$ så är $(a_0, a_1, a_2, a_3, a_4) = (1, 2, 0, 3, 7)$ och vi får

$$c_i = 1 + 2 \cdot 2^i + 0 \cdot 2^{2i} + 3 \cdot 2^{3i} + 7 \cdot 2^{4i}.$$

Därför är

$$\begin{aligned} c_0 &= 1 + 2 + 0 + 3 + 7 = 7, \\ c_{16} &= 1 + 2 \cdot 2 + 0 \cdot 2^2 + 3 \cdot 2^3 + 7 \cdot 2^4 \\ &= 1 + 4 + 0 + 5 + 4 = 4 \end{aligned}$$

o.s.v.

Addera de 7 identiteterna (\star) och använd dig av att $1 + p + p^2 + \dots + p^6 = 0$. Resultatet blir

$$c_0 + \dots + c_6 = 7a_0 + 0 \cdot a_1 + \dots + 0 \cdot a_4 = a_0.$$

Att $7a_0 = a_0$ kommer sig av att $a_0 + a_0 = 0$. Låt q vara sådant att $qp = 1$ d.v.s. $q = p^6$ ($q = 5$ med vårt val av p). Multiplicera båda sidorna i (\star) med q^i och addera. Resultatet blir

$$c_0 + c_1q + c_2q^2 + \dots + c_6q^6 = a_1.$$

(Kom ihåg att $p^8 = p, p^9 = p^2, \dots, p^{15} = p, \dots$) Multiplicera sen (\star) med q^{2i} och addera. Fortsätt så med q^{3i}, \dots, q^{6i} . Resultatet blir ett avkodningsrecept:

$$a_k = c_0 + c_1q^k + c_2q^{2k} + c_3q^{3k} + c_4q^{4k} + c_5q^{5k} + c_6q^{6k}$$

för $k = 0, 1, 2, 3, 4$. Dessutom får vi två ekvationer som gäller för kodord (men inga andra):

$$(\star\star) \quad c_0 + c_1q^k + c_2q^{2k} + c_3q^{3k} + c_4q^{4k} + c_5q^{5k} + c_6q^{6k} = 0$$

för $k = 5, 6$. Dessa ekvationer kan användas till att ge en algoritm som rättar ett fel men det behöver du inte göra.

7. Denna kod kan rätta fel i ett av de 7 elementen (och därmed 3 fel i de 3 motsvarande bitarna). Du ska nu övertyga dig om att detta påstående är sant genom att fylla i detaljerna i följande resonemang.

– Om $c' = (c'_1, \dots, c'_6)$ och $c'' = (c''_1, \dots, c''_6)$ är två kodord så är $c = c' - c'' = (c'_1 - c''_1, \dots, c'_6 - c''_6)$ också ett kodord.

– Det räcker därför att visa att om $c = (c_1, \dots, c_6)$ är ett kodord sådant att $c_i = 0$ för alla utom möjligen två värden på i så är $c_i = 0$

för alla i .

– Antag att $c_i = 0$ för alla utom möjligen två värden på i ($i = 0$ och $i = 1$ t.ex.). Eftersom de två ekvationerna ($\star\star$) är uppfyllda så måste även c_0 och c_1 vara 0.

8. Koda din bild med denna kod. Simulera störningar (i bitföljden) som under punkt 5. Avkoda och rita upp bilden. Du behöver inte använda avkodningsreceptet. Eftersom du vet vad rätt svar ska vara så behöver du i regel bara kontrollera att den störda följd ($\tilde{c}_0, \dots, \tilde{c}_6$) skiljer sig från det ostörda kodordet (c_0, \dots, c_6) på högst en av de 7 positionerna. Om så inte är fallet kan du nöja dig med att konstatera att fel uppstått.

Litteratur

MacWilliams, F.J. and Sloane, N.J.A., *The theory of error-correcting codes*. North-Holland 1977.

Euler-Mac Laurins summationsformel och Bernoulliska polynom

LARS HÖRMANDER

Lunds Universitet

Datorer gör det möjligt att genomföra räkningar som tidigare varit otänkbara, exempelvis att beräkna summan av en oändlig serie genom att helt enkelt addera ett stort antal termer. I praktiken är detta dock inte särskilt effektivt. Ett syfte med följande uppgift är att visa att matematiska metoder för att reducera räknearbetet inte blivit överflödiga trots de nya tekniska hjälpmedlen.

Uppgiften består av två delar. I den första införs Bernoullital och Bernoullipolynom som hjälpmedel för att jämföra en summa med en integral. I den andra diskuteras Fourierserier med utgångspunkt från Bernoullipolynomen. Detta ger speciellt flera möjligheter att beräkna π .

Om man vill beräkna summan av en serie som

$$S = \sum_{n=1}^{\infty} \frac{1}{n^2}$$

med hög precision, exempelvis 8 siffrors noggrannhet, så verkar det först som om man skulle behöva räkna ut en delsumma

$$S_N = \sum_{n=1}^N \frac{1}{n^2}$$

med ett orimligt stort antal termer. Vi har nämligen

$$\frac{1}{n} - \frac{1}{n+1} = \frac{1}{n(n+1)} < \frac{1}{n^2} < \frac{1}{n(n-1)} = \frac{1}{n-1} - \frac{1}{n},$$

vilket medför att för godtyckliga heltal $N < M$

$$\frac{1}{N+1} - \frac{1}{M+1} < \sum_{N+1}^M \frac{1}{n^2} < \frac{1}{N} - \frac{1}{M}.$$

Summan $R_N = S - S_N$ av termerna efter den N :te ligger alltså mellan $1/(N+1)$ och $1/N$ så man skulle behöva summera 10^8 termer för att få tillräckligt liten rest. Närmare eftertanke visar emellertid att eftersom

$$0 < S - S_N - \frac{1}{N+1} < \frac{1}{N} - \frac{1}{N+1} = \frac{1}{N(N+1)}$$

så räcker det att ta $N = 10^4$, beräkna s_N och addera $1/N$ till resultatet. Ändå krävs 10001 termer.

1. Använd att

$$\frac{1}{n^2} - \frac{1}{2} \left(\frac{1}{n-1} - \frac{1}{n+1} \right) = \frac{1}{n^2(n^2-1)}$$

för att sänka antalet termer under 500! Kan Du hitta ytterligare förbättringar?

Om Du känner till något om numerisk integration så ser Du att man kan uppfatta exemplet ovan så, att $\sum_N^\infty 1/n^2$ approximerats med $\int_N^\infty dx/x^2 = 1/N$, först med rektangeluppskattning, sedan med trapetsuppskattning av felet. Vi skall nu ge en allmän metod som inte kräver att man i varje särskilt fall hittar på ett knep.

För att förenkla börjar vi med att undersöka $\int_0^1 f(x) dx$ för en funktion f som antas ha många kontinuerliga derivator. Om dessa blir allt mindre, som fallet är för $f(x) = 1/(x+n)^2$ då n är stort, så lönar det sig att integrera partiellt:

$$\begin{aligned} \int_0^1 f(x) dx &= [(x-c)f(x)]_0^1 - \int_0^1 (x-c)f'(x) dx \\ &= (1-c)f(1) + cf(0) - \int_0^1 (x-c)f'(x) dx. \end{aligned}$$

Vi får en symmetrisk formel genom att välja $c = \frac{1}{2}$,

$$(1) \quad \int_0^1 f(x) dx = \frac{1}{2}(f(1) + f(0)) - \int_0^1 (x - \frac{1}{2})f'(x) dx,$$

där den första termen i högerledet kallas trapetsapproximationen till integralen; den är integralen av den lineära funktion som är lika med f i 0 och 1. Man kallar

$$(2) \quad B_1(x) = x - \frac{1}{2}$$

för det första Bernoullipolynommet och bestämmer successivt Bernoullipolynomen $B_2(x)$, $B_3(x)$, ... så att

$$(3) \quad B'_n(x) = nB_{n-1}(x), \quad B_n(0) = B_n(1), \quad n > 1.$$

Det andra villkoret kräver att

$$(4) \quad \int_0^1 B_{n-1}(x) dx = 0, \quad n > 1,$$

vilket gäller enligt (2) då $n = 2$. Av (3) får vi $B_2(x) = x^2 - x + c$, och (4) kräver att vi väljer konstanten c så att $\frac{1}{3} - \frac{1}{2} + c = 0$, alltså $c = 1/6$. Det är klart att man på ett och endast ett sätt kan fortsätta att beräkna polynomen $B_n(x)$ genom integration av det föregående polynommet enligt (3) och bestämning av integrationskonstanten enligt (4).

2. Beräkna koefficienterna i polynomen B_3 , B_4 , B_5 , B_6 .

3. Visa att

$$B_n(1 - x) = (-1)^n B_n(x), \quad n \geq 1.$$

4. Visa att det finns en talföljd b_0, b_1, \dots (de Bernoulliska talen) sådana att

$$B_n(x) = \sum_{l=0}^n \binom{n}{l} b_l x^{n-l}, \quad n \geq 1; \quad \binom{n}{l} = \frac{n!}{l!(n-l)!};$$

och att dessa kan beräknas genom rekursionsformeln

$$\sum_{l=0}^{n-1} \binom{n}{l} b_l = 0, \quad n \geq 2.$$

Vi har alltså $b_0 = 1, b_1 = -\frac{1}{2}, b_2 = \frac{1}{6}$. Beräkna b_3, \dots, b_6 , och visa att $b_k = 0$ om k är udda > 1 .

5. Visa att i intervallet $[0, 1]$ är B_{2k+1} för $k > 1$ noll i punkterna $0, \frac{1}{2}, 1$ och inga andra.

6. Visa att

$$B_n(2x) = 2^{n-1} (B_n(x) + B_n(x + \frac{1}{2})),$$

och beräkna $B_n(\frac{1}{2})$.

7. Visa att för $k \geq 1$ gäller

$$|B_{2k}(x)| \leq |b_{2k}|, \quad |B_{2k+1}(x)| \leq (2k+1)|b_{2k}|/4, \quad \text{då } 0 \leq x \leq 1.$$

8. Visa att $B_n(x+1) - B_n(x) = nx^{n-1}$ och använd det för att beräkna summan

$$\sum_{j=0}^N j^n$$

då n och N är positiva heltal.

Genom upprepade partialintegration i (1) får vi nu

$$(5) \quad \int_0^1 f(x) dx = (f(1) + f(0))/2 - \left[\sum_{k=2}^n (-1)^k B_k(x) f^{(k-1)}(x)/k! \right]_0^1 + (-1)^n \int_0^1 B_n(x) f^{(n)}(x)/n! dx.$$

Eftersom $B_k(0) = B_k(1) = b_k$ då $k \geq 2$ och $b_k = 0$ då k är udda, så får vi om vi väljer $n = 2p + 1$ udda

$$(5)' \quad \int_0^1 f(x) dx = \frac{1}{2}(f(1) + f(0)) - \left[\sum_{k=1}^p b_{2k} f^{(2k-1)}(x)/(2k)! \right]_0^1 \\ - \int_0^1 B_{2p+1}(x) f^{(2p+1)}(x)/(2p+1)! dx.$$

Om f är en $2p + 1$ gånger kontinuerligt deriverbar funktion i intervallet $N \leq x \leq M$ där N, M är heltal, så kan vi tillämpa detta på funktionerna $x \mapsto f(x + n)$ för $N \leq n < M$ och addera resultaten. De utintegrerade termerna i $N + 1, \dots, M - 1$ tar då ut varandra tack vare att vi i (5)' har samma koefficienter i 0 och i 1. Det har vi på grund av det andra villkoret (3) som ställdes precis för att uppnå detta. Beteckna med $\bar{B}_\nu(x)$ den funktion med perioden 1 som är lika med B_ν i intervallet $[0, 1]$; den är $\nu - 2$ gånger kontinuerligt deriverbar enligt andra delen av (3). Vi har då bevisat *Euler-Mac Laurins summationsformel*

$$(6) \quad \int_N^M f(x) dx = \frac{1}{2}(f(N) + f(M)) + \sum_{N < n < M} f(n) \\ - \left[\sum_{k=1}^p b_{2k} f^{(2k-1)}(x)/(2k)! \right]_N^M \\ - \int_N^M \bar{B}_{2p+1}(x) f^{(2p+1)}(x)/(2p+1)! dx.$$

9. Använd formeln för att beräkna

$$\sum_1^\infty \frac{1}{n^2} \quad \text{och} \quad \sum_1^\infty \frac{1}{n^4}$$

med 8 siffrors noggrannhet utan att summera ett stort antal termer. (Beräkna summan av de $N - 1$ första termerna och hälften av den

N :te exakt, samt använd (6) för att approximera resten! Experimentera med olika ganska små val av N och p för att se hur stor integralen i högerledet av (6) blir! Hur långt ner kan Du minska antalet termer som måste beräknas?)

10. Använd formeln för att skriva ett program som beräknar Riemanns ζ funktion $\zeta(s) = \sum_1^\infty n^{-s}$ med 8 siffrors noggrannhet då $1 < s \leq 6$.

Vi skall nu diskutera *Fourierserien* för den periodiska funktionen \overline{B}_n . De enklaste funktionerna som är periodiska med perioden 1 är de trigonometriska funktionerna $\sin(2\pi kx)$ och $\cos(2\pi kx)$, eller

$$e^{2\pi ikx} = \cos(2\pi kx) + i \sin(2\pi kx), \quad \text{Eulers formler.}$$

Man säger att en funktion f med perioden 1 har utvecklats i Fourierserie om den framställs som en summa

$$(7) \quad f(x) = \sum_{-\infty}^{\infty} c_k(f) e^{2\pi ikx},$$

vilket också kan skrivas i den mindre lätthanterliga formen

$$f(x) = \sum_0^{\infty} a_k(f) \cos(2\pi kx) + \sum_1^{\infty} b_k(f) \sin(2\pi kx);$$

$$a_0 = c_0, a_k = c_k + c_{-k}, b_k = i(c_k - c_{-k}).$$

I fortsättningen använder vi (7). Om (7) gäller i den starka meningen att

$$\max |f(x) - \sum_{-N}^N c_k(f) e^{2\pi ikx}| \rightarrow 0 \text{ då } N \rightarrow \infty,$$

så kan man multiplicera med $e^{-2\pi i\nu x}$ och integrera varje term för sig. Eftersom

$$\int_0^1 e^{-2\pi i\mu x} dx = \begin{cases} 0, & \text{om } \mu \neq 0 \\ 1, & \text{om } \mu = 0, \end{cases}$$

så får man genast

$$(8) \quad c_\nu(f) = \int_0^1 f(x)e^{-2\pi i\nu x} dx.$$

Man kallar talen som ges av (8) för Fourierkoefficienterna till f . Problemet är om (7) gäller med dessa koefficienter.

Enligt (4) har vi $c_0(\overline{B}_n) = 0$, om $n \neq 0$, och då $\nu \neq 0$ får vi med hjälp av (3)

$$\begin{aligned} c_\nu(\overline{B}_n) &= \int_0^1 B_n(x)e^{-2\pi i\nu x} dx = \frac{1}{2\pi i\nu} \int_0^1 B'_n(x)e^{-2\pi i\nu x} dx \\ &= \frac{n}{2\pi i\nu} \int_0^1 B_{n-1}(x)e^{-2\pi i\nu x} dx = \dots \\ &= \frac{n!}{(2\pi i\nu)^{n-1}} \int_0^1 B_1(x)e^{-2\pi i\nu x} dx = \frac{-n!}{(2\pi i\nu)^n}. \end{aligned}$$

Det gäller nu att visa att

$$\overline{B}_n(x) = \sum_{\nu \neq 0} \frac{-n!}{(2\pi i\nu)^n} e^{2\pi i\nu x}$$

om $n \geq 2$. Då konvergerar i varje fall serien eftersom $\sum_{\nu \neq 0} |\nu|^{-n} < 2 + 2 \int_1^\infty t^{-n} dt < \infty$.

Låt alltså $n \geq 2$, $0 \leq y \leq 1$, och betrakta funktionen

$$f(x) = (\overline{B}_n(x) - \overline{B}_n(y))/(1 - e^{2\pi i(x-y)}), \quad 0 \leq x \leq 1.$$

Med lämplig definition då $x = y$ eller $x = y \pm 1$ så är f kontinuerligt deriverbar för $0 \leq x \leq 1$, så vi har

$$\begin{aligned} c_\nu(f) &= \int_0^1 f(x)e^{-2\pi i\nu x} dx \\ &= [f(x)e^{-2\pi i\nu x}/(-2\pi i\nu)]_0^1 + \int_0^1 f'(x)e^{-2\pi i\nu x}/(2\pi i\nu) dx \rightarrow 0, \end{aligned}$$

då $\nu \rightarrow \infty$. Eftersom

$$\overline{B}_n(x) = \overline{B}_n(y) + f(x) - f(x)e^{2\pi i(x-y)}, \quad 0 \leq x \leq 1,$$

så ger en enkel räkning

$$c_\nu(\overline{B}_n) = \begin{cases} c_\nu(f) - e^{-2\pi iy}c_{\nu-1}(f), & \text{om } \nu \neq 0, \\ \overline{B}_n(y) + c_\nu(f) - e^{-2\pi iy}c_{\nu-1}(f), & \text{om } \nu = 0. \end{cases}$$

Alltså är

$$\begin{aligned} \sum_{-N}^N c_\nu(\overline{B}_n)e^{2\pi i\nu y} &= \overline{B}_n(y) + \sum_{-N}^N (c_\nu(f)e^{2\pi i\nu y} - c_{\nu-1}(f)e^{2\pi i(\nu-1)y}) \\ &= \overline{B}_n(y) + c_N(f)e^{2\pi iNy} - c_{-N-1}(f)e^{2\pi i(\nu-1)y} \\ &\rightarrow \overline{B}_n(y), \quad N \rightarrow \infty, \end{aligned}$$

vilket bevisar att för $n \geq 2$ gäller

$$(9) \quad \overline{B}_n(x) = -n! \sum_{\nu \neq 0} (2\pi i\nu)^{-n} e^{2\pi i\nu x}.$$

Speciellt får vi då $x = 0$

$$(10) \quad b_n = -n! \sum_{\nu \neq 0} (2\pi i\nu)^{-n}.$$

11. Använd (10) och övning 9 för att beräkna π .

12. Visa att om f är en två gånger kontinuerligt deriverbar funktion med perioden 1 så gäller att

$$\frac{1}{2} \int_0^1 f''(x-y)B_2(y) dy = c_0(f) - f(x).$$

Visa med hjälp av det att (7) följer av (9) med $n = 2$.

Eftersom $\sum_{\nu \neq 0} 1/|\nu| \geq 2 \int_1^\infty dt/t = \infty$ så är frågan om konvergensten av Fourierserien för \bar{B}_1 mera komplicerad. Delsummorna är

$$s_N(x) = - \sum_{0 < |\nu| \leq N} e^{2\pi i \nu x} / (2\pi i \nu).$$

Observera att $s_N(x) = -s_N(-x) = -s_N(1-x)$, vilket speciellt ger $s_N(0) = s_N(\frac{1}{2}) = 0$. Vi kan summera derivatan som en geometrisk serie,

$$\begin{aligned} s'_N(x) &= - \sum_{0 < |\nu| \leq N} e^{2\pi i \nu x} = 1 - \sum_{-N}^N e^{2\pi i \nu x} \\ &= 1 - \frac{e^{2\pi i(N+\frac{1}{2})x} - e^{-2\pi i(N+\frac{1}{2})x}}{e^{\pi i x} - e^{-\pi i x}} = 1 - \frac{\sin(2\pi(N+\frac{1}{2})x)}{\sin(\pi x)}. \end{aligned}$$

Eftersom $s_N(0) = 0$ får vi för $0 \leq x \leq \frac{1}{2}$,

$$s_N(x) = x - \int_0^x \frac{\sin(2\pi(N+\frac{1}{2})t)}{\sin(\pi t)} dt.$$

Nu är $f(t) = 1/\sin(\pi t) - 1/(\pi t)$ en kontinuerligt deriverbar funktion i $[0, \frac{1}{2}]$ (visa det som övning) och vi får därför med upprepning av ett bevis ovan att om

$$e_N(x) = \int_0^x \sin(2\pi(N+\frac{1}{2})t) f(t) dt$$

så är $|e_N(x)| \leq C/(N+\frac{1}{2})$. Vidare är

$$s_N(x) + e_N(x) = x - \int_0^x \frac{\sin(2\pi(N+\frac{1}{2})t)}{\pi t} dt = x - G(2\pi x(N+\frac{1}{2}))$$

där

$$(11) \quad G(T) = \int_0^T \frac{\sin t}{\pi t} dt.$$

Sätter vi $x = \frac{1}{2}$ så ser vi eftersom $s_N(\frac{1}{2}) = 0$ att

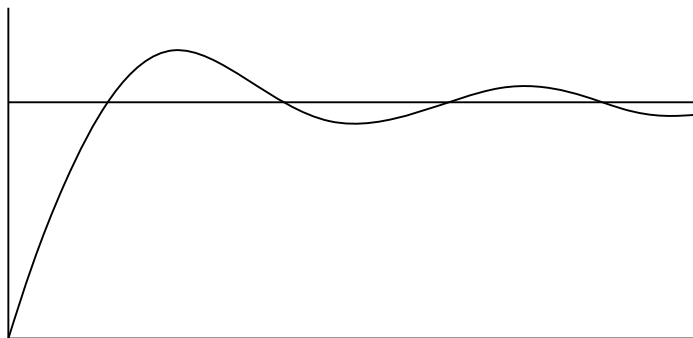
$$G(\pi(N + \frac{1}{2})) \rightarrow \frac{1}{2} \quad \text{då } N \rightarrow \infty.$$

Om $\pi N \leq T \leq \pi(N + 1)$ så är

$$|G(T) - G(\pi(N + \frac{1}{2}))| \leq \frac{1}{\pi N} \int_0^{\frac{\pi}{2}} |\sin t| dt = \frac{1}{\pi N},$$

så vi kan dra slutsatsen att $G(T) \rightarrow \frac{1}{2}$ då $T \rightarrow \infty$ på godtyckligt sätt. Detta medför att $s_N(x) \rightarrow x - \frac{1}{2}$ då $N \rightarrow \infty$ och $0 < x \leq \frac{1}{2}$, alltså för alla icke heltaliga x .

13. Visa att $0 \leq G(y) \leq G(\pi)$ och beräkna $G(\pi)$ numeriskt (eventuellt med hjälp av Euler-Mac Laurins formel eller Simpsons formel)! Värdet är 0,58949 Minimum av s_N ligger alltså nära $-G(\pi)$ då N är stort, fastän $\overline{B}_1 \geq -\frac{1}{2}$. Denna översvängning kallas Gibbs fenomen. Rita gärna upp några delsummor på bildskärmen för att se detta, om Du har tillgång till dator med god grafik!



Grafen för $y = G(x)$ och $y = \frac{1}{2}$, då $0 \leq x \leq 4\pi$.

Reflektionsprincipen

DAG JONSSON

Uppsala Universitet

1. Inledning. Något om permutationer.

EXEMPEL 1. Vi skriver bokstäverna A, B, C i rad. På hur många olika sätt kan de tre bokstäverna ordnas inbördes dvs hur många olika ord bildade av dessa tre bokstäver finns det?

Svar: Det finns 6 ordningar eller *permutationer*: $ABC, ACB, BAC, BCA, CAB, CBA$. Första bokstaven kan väljas på 3 olika sätt. När första bokstaven är vald har vi 2 möjligheter för den andra bokstaven och när de båda första bokstäverna är valda finns det bara en möjlighet kvar för den tredje bokstaven.

UPPGIFT 1. Hur många olika permutationer av bokstäverna A, B, C, D, E finns det?

Allmänt betecknar vi antalet permutationer av n olika bokstäver med $n!$

UPPGIFT 2. Ge ett uttryck för $n!$ i talen $1, 2, \dots, n$.

EXEMPEL 2. *Kommittéproblemet*. Fem personer A, B, C, D, E sitter i en styrelse. Man vill utse en kommitté bestående av en ordförande, en sekreterare och en kassör. Hur många olika sådana kommittéer kan man bilda? Jo, det finns 5 sätt att välja ordförande. När denne är vald har vi 4 sätt att välja sekreterare och när även denne är vald finns det 3 sätt att välja kassör. Det finns alltså $5 \cdot 4 \cdot 3 = 60$ olika sätt. Observera att detta kan skrivas på formen $\frac{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{2 \cdot 1} = \frac{5!}{2!}$.

UPPGIFT 3. Hur många olika kommittéer med ordförande, sekreterare, kassör och klubbmästare kan man bilda med 7 personer?

Allmänt blir antalet ordnade kommittéer av storlek k valda bland n personer

$$n(n-1)\dots(n-k+1) = n!/(n-k)!$$

Motivera denna formel! Antag nu att ordningen mellan kommittémedlemmarna inte spelar någon roll, dvs alla är rätt och slätt medlemmar.

Bland de 60 ordnade kommittéerna ovan hittar man t ex ABC , ACB , BAC , BCA , CAB , CBA . Dessa fall skiljer vi inte åt i det icke ordnade fallet. Här uttrycker ABC enbart att dessa tre bokstäver finns med. På samma sätt har vi 6 ordningar av bokstäverna A, B, D . Därför har vi 6 gånger så många kombinationer i det ordnade som i det icke ordnade fallet. Dividerar vi 60 med 6 får vi alltså 10 icke ordnade fall.

UPPGIFT 4. Hur många olika icke ordnade kommittéer om 4 personer kan man bilda bland 7 personer?

Allmän formel. Vi hade tidigare $\frac{n!}{(n-k)!}$ ordnade fall när k personer skulle väljas ut bland n personer. Om vi i varje kommitté om k personer bortser från ordningen reduceras antalet fall med faktorn $k!$ och vi får $n!/(n-k)!k!$ icke ordnade fall. Detta antal betecknar vi med $\binom{n}{k}$.

2. Reflektionsprincipen

EXEMPEL 3. I ett skolval med 6 röstande har 4 elever röstat på A -partiet och 2 elever på B -partiet. Vid rösträkningen drar man en röst i taget och noterar varje gång den aktuella ställningen. På hur många olika sätt kan rösterna dras så att A -partiet hela tiden befinner sig i ledningen? Efter prövning finner vi 5 olika kombinationer: $AAAABB$, $AAABAB$, $AAABBA$, $AABAAB$, $AABABA$. För varje draget röst har vi dragit fler A -röster än B -röster. Finns

det någon allmän formel som ger antalet kombinationer med denna egenskap?

Låt oss börja med att bestämma *totala* antalet rösträkningskombinationer. Vi har 6 röster varav 4 *A*- och 2 *B*-röster. På hur många olika sätt kan de 4 *A*'na placeras i raden av 6 tecken?

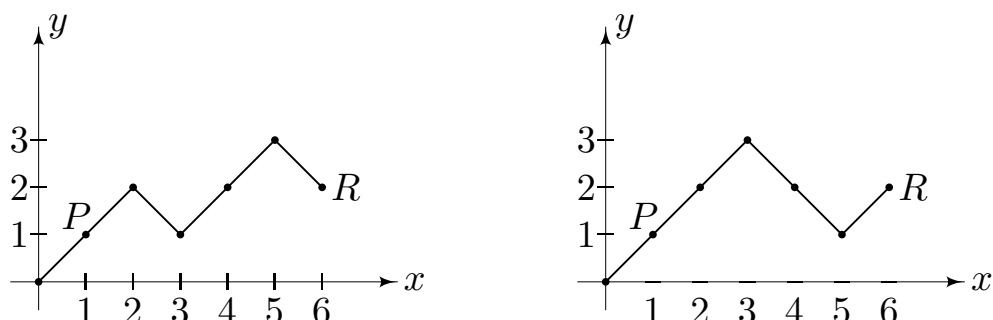
UPPGIFT 5. Visa att detta är en variant av kommittéproblemet. Visa sedan att totala antalet kombinationer blir $\binom{6}{4} = \frac{6!}{4!2!} = 15$. Ange dessa kombinationer (5 av dem är redan givna).

Allmänna fallet. Vi har a st *A*-röster och b st *B*-röster, där $a > b$. Med $n = a + b$, $k = a$ får vi $\binom{a+b}{a}$ olika kombinationer. I hur många av dessa är *A* hela tiden i ledningen?

För att lösa detta problem ska vi använda oss av den så kallade *reflektionsprincipen*.

Inför ett rätvinkligt koordinatsystem med $x =$ antalet dragna röster och $y =$ differensen mellan antalet dragna *A*-röster och antalet dragna *B*-röster i ett givet skede. Vid starten befinner vi oss således i origo. Om den först dragna rösten är en *A*-röst, hamnar vi i punkten (1,1). Denna punkt betecknas i fortsättningen med P . Om i stället en *B*-röst dras hamnar vi i stället i punkten (1,-1), i fortsättningen betecknad med P' . Om de två först dragna rösterna båda är *A*-röster går vi via (1,1) till punkten (2,2) osv. När rösträkningen är klar ska vi tydligen befinna oss i punkten $(a+b, a-b)$, i fortsättningen betecknad med R .

För $a = 4, b = 2$ ska vi alltså förflytta oss från origo till (6,2). De fem ovan nämnda fallen motsvaras av vägar, som hela tiden förutom i origo ligger helt ovanför x -axeln. Två av vägarna är uppritade i nedanstående figur. Rita upp de tre övriga vägarna!

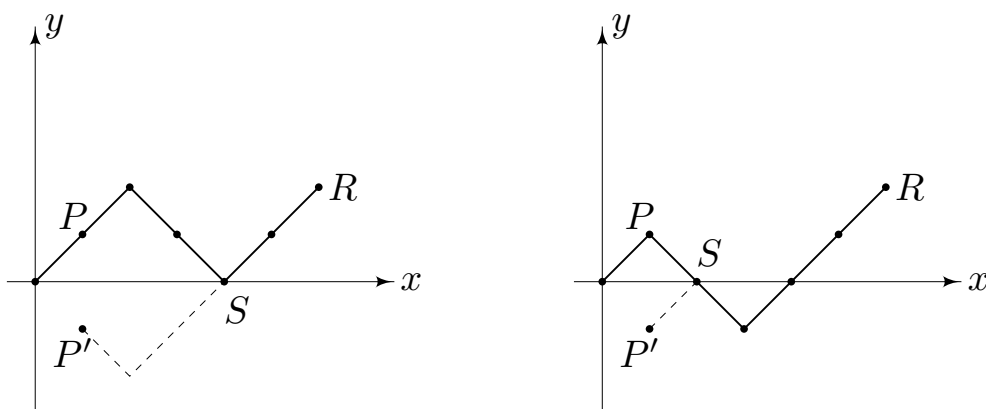


Figuren illustrerar fallen $AABAAB$ och $AAABBA$. Finns det någon allmän formel för antalet vägar av detta slag?

Först noterar vi att alla sådana vägar passerar punkten P . Låt N vara totala antalet vägar (utan krav på att de ska ligga ovanför x -axeln) från P till R .

UPPGIFT 6. Visa att $N = \binom{a+b-1}{a-1}$.

Betrakta vägar från P till R av icke önskat slag, dvs som når x -axeln i minst en punkt. Antag att en sådan väg når x -axeln i punkten $S = (s, 0)$. Vi bildar en ny väg genom att i x -axeln spegla det avsnitt av vägen som ligger mellan P och S medan resten av vägen, dvs den mellan S och R , sammanfaller med den gamla.



Figuren illustrerar fallen $AABBAA$ respektive $ABBAAA$.

Observera att punkten $P' = (1, -1)$ alltid ligger på en sådan (delvis) speglad väg men att den, liksom varje ursprunglig väg, alltid slutar i R .

UPPGIFT 7. Motivera varför antalet vägar mellan P och R och som når x -axeln är lika med totala antalet vägar mellan P' och R . Visa att detta antal är $\binom{a+b-1}{a}$ dvs i exemplet $= \binom{5}{4} = 5$

UPPGIFT 8. Visa att antalet vägar mellan P och R som helt ligger ovanför x -axeln är

$$\frac{a-b}{a+b} \cdot \binom{a+b}{a}.$$

Totala antalet vägar mellan origo och R är som vi tidigare konstaterat $\binom{a+b}{a}$. Antag att vi väljer en väg på måfå bland dessa möjliga vägar. Vad är då sannolikheten att vägen är av önskat slag, med andra ord vad är sannolikheten att vi vid rösträkningen hela tiden har A som ledare?

EXEMPEL 4. Betrakta följande situation. En lärare säljer kompendier för 50 kr stycket till sina 10 elever. Antag att hälften av eleverna betalar med en 50-kronorssedel medan övriga inte har mindre valör än 100 kr. Vad är sannolikheten att läraren klarar av att hela tiden ge växel tillbaka på 100-kronorssedlar om eleverna betalar i slumpmässig ordning?

UPPGIFT 9. Betrakta vägar mellan origo och punkten $(2n, 0)$ på x -axeln (detta svarar mot att antalet elever är $2n$, ett jämnt heltal). Visa att av de $\binom{2n}{n}$ möjliga vägarna mellan nämnda punkter är det exakt $\frac{1}{n+1} \cdot \binom{2n}{n}$ vägar som ligger ovanför eller på x -axeln. Visa vidare att det är exakt $\frac{1}{n} \cdot \binom{2n-2}{n-1}$ vägar som ligger helt ovanför x -axeln (utom i ändpunkterna). Det senare fallet svarar mot att läraren hela tiden har reservväxel, som går åt först vid den sista betalningen.

Litteratur

Feller, W., *An Introduction to Probability Theory and its Applications, Volume 1*. Wiley, New York 1957 (universitetsnivå).

Honsberger, R.A., *Mathematical Gems III*. The Mathematical Association of America, 1985 (klart mera lättillgänglig än den föregående).

Antikens universum

STEN KAIJSER

Uppsala Universitet

Inledning. Detta specialarbete har fyra syften. Det första är att ge en allmän orientering om antikens och medeltidens världsbild, främst för att visa att denna var mycket mer utvecklad än vad vi normalt tror. Det andra är att visa hur antikens matematiker med enkel geometri (t.o.m. nästan utan trigonometri) och primitiva mätinstrument beräknade jordens och solsystemets mått genom att låta eleven utföra beräkningarna, och därigenom själv ta reda på hur stort antikens vetenskapsmän trodde att universum var. Det tredje syftet är att jämföra antikens föreställningar med våra, genom att låta eleven utföra liknande beräkningar med de siffror som dagens vetenskap erbjuder. Det ingår i förutsättningarna för arbetet att eleven endast behöver använda matematik som var känd i Alexandria 200 år före kristi födelse. Det fjärde och sista syftet är att framföra det enkla budskapet: *Den som är beredd att räkna kan alltid ta reda på saker själv, t.o.m. hur många atomer som ryms i universum.*

Den medeltida världsbilden. När vi läser om Columbus resa över Atlanten ges vi ofta intrycket av att han ensam trodde att jorden var rund, medan alla andra trodde att den var platt. Dessutom antyds det att sjömännen varje dag väntade på att nå fram till världens kant där de skulle trilla av och därmed hamna i helvetet.

Det enda som är riktigt i den bild av Columbus som jag antydde ovan, är att *han* trodde att jorden var rund. Men det trodde naturligtvis också Ferdinand och Isabella som utrustade hans skepp,

och det trodde även hans sjömän, ja i själva verket trodde varenda vettig människa att jorden var rund — åtminstone om vi med vettig människa menar ”bildat folk”, sjöfolk och några till. Och redan då var idén att jorden var rund nära tvåtusen år gammal. De som först framförde tanken var *pythagoréerna*, d.v.s. medlemmarna av den rörelse som PYTHAGORAS (han med satsen) grundade. Pythagoréerna ansåg att cirkeln är den fulländade geometriska figuren och att klotet är den fulländade kroppen. Av religiösa skäl ansåg de därför att alla himlakropparna var klot som rörde sig på cirklar över det sfäriska himlavalvet. Pythagoréernas ursprungliga tanke var att solen liksom månen lyste med lånat ljus, och att ljuskällan för alla himlakroppar var ”centralelden” i universums medelpunkt. Att vi aldrig får se denna eld beror på att jorden alltid vänder ”ryggen” mot den. Denna ”rygg” på jorden är därmed också obeboelig på grund av värmen. Så småningom utvecklades Pythagoréernas världsbild därhän att jorden placerades orörlig i världens medelpunkt, medan alla himlakropparna, sol, måne, planeter och fixstjärnor satt fast på sfärer som på sinnrika sätt snurrade runt den orörliga (men fortfarande runda) jorden. Att fixstjärnorna befann sig på en sfär som roterade runt en axel som sträckte sig från polstjärnan i norr till en okänd punkt långt nere i söder var ju inte så svårt att se. Solen var något svårare men dess rörelse är ändå så pass regelbunden att tar man bara hänsyn till ekliptikan så kan även dess rörelse förstås. Värre var det med himlens luffare, planeterna (ja, själva ordet planet kommer faktiskt från ett grekiskt ord för luffare). Den djupt religiöse, och av pythagoréerna inspirerade, PLATON var övertygad om att även planetbanorna kunde beskrivas med hjälp av cirklar och/eller sfärer. Han uppmanade därför sin tids matematiker att skapa ordning bland himlakropparna genom att finna de cirklar som beskrev rörelserna. Den som löste problemet (åtminstone tillfälligt)

var EUDOXUS (som faktiskt var verksam hos Platon vid dennes skola *Akademien* i Aten). Eudoxus visade att man kunde tänka sig sfärer kopplade innanför varandra som alla roterar med varandra med konstanta rotationshastigheter och som tillsammans väl beskriver himlakropparnas rörelser. Denna idé övertogs av ARISTOTELES som dock inte ville ha *tänkta sfärer* utan som istället gjorde dem materiella. Eudoxus' *tänkta axlar* blev också i Aristoteles' modell till verkliga fysiska axlar. Aristoteles' och Eudoxus' modell för himlakropparna stog sig i något hundratal år, ända tills grekerna efter Alexander den stores fälttåg fick del av babyloniernas månghundraåriga observationer.

Efter Alexanders död blev Alexandria centrum i den vetenskapliga världen. Där satt EUKLIDES och skrev sin *Elementa* (den lärobok i geometri som sedan användes i över två tusen år) och där lärde sig ARKIMEDES matematikens grunder. En i Alexandria verksam matematiker som gav två viktiga bidrag till den matematiska beskrivningen av solsystemet var APOLLONIUS (från Perga). Den för framtiden viktigaste insatsen var naturligtvis utarbetandet av den teori för kägelsnitten, d.v.s. ellipser, parabler och hyperbler som blev underlaget för Keplers arbete i början av 1600-talet. Men den idé som fick omedelbar tillämpning var införandet av epicykler vilket kan betraktas som det första försöket att beskriva (nästan) periodiska förlopp med (nästan) Fourierserier. Modellen är att planeten rör sig på en (liten) cirkel, epicykeln, vars centrum rör sig på en cirkel runt jorden. Den stora cirkeln har därvid en omloppstid som svarar mot planetens rörelse (runt solen som vi nu ser det) och den lilla cirkeln har en omloppstid som svarar mot ett jordiskt år.¹ Modellen utvecklades ytterligare genom att medelpunkten för den stora cirkeln

¹ Detta gäller för de yttre planeterna, för de inre har den större cirkeln jordåret som omloppstid medan epicykeln svarar mot planetåret.

inte nödvändigtvis var jorden. Därigenom erhöll man ett system med inte mindre än fyra *frihetsgrader*, nämligen centrum för en *excentrisk stor cirkel*, radien i denna, samt radien i epicykeln. Därigenom kunde de alexandrinska astronomerna HIPPARCHOS och PTOLEMAIOS utarbeta en beskrivning av stjärnhimlen som stod sig i ett och ett halvt årtusende.

Vi som sedan snart fyrahundra år levt i Keplers och Newtons solsystem med solen i centrum föreställer oss ett geocentriskt system (med jorden i centrum), som oerhört primitivt. Vi bör dock komma ihåg att grekernas astronomiska instrument inte tillät några andra observationer än vinkelmätningar — de hade inte ens tillgång till de enklaste kikare. Den moderna tidens utveckling av astronomin beror framför allt på tekniska förbättringar av astronomernas observatorier. Ett faktum är att med endast vinkelmätningar är det matematiskt möjligt att beskriva universum geocentriskt med en *fixstjärnesfär* som universums yttersta gräns.

Det finns dock två himlakroppar som påverkar oss mer än några andra, nämligen solen och månen. Båda dessa befinner sig tillräckligt nära jorden för att uppta en mätbar bråkdel av synfältet. Båda upptar normalt cirka 30 bågminuter, d.v.s. en halv grad. Därigenom får vi för båda ett förhållande mellan avståndet och diametern.

UPPGIFT 1. Beräkna förhållandet mellan månens diameter och avståndet till månen om månens vinkel på himlen är exakt en halv grad.

Att månen och solen upptar ungefär samma vinkel kan man inse om man vet att då månen är som närmast jorden blir en eventuell solförmörkelse total, medan då månen är som längst från jorden blir en eventuell solförmörkelse endast partiell.

Aristarchos från Samos. För den astronomiskt obevandrade finns det inget självklart samband mellan solförmörkelser och månen.

(Skuggsidan av månen syns ju naturligtvis inte alls på dagen när en eventuell solförmörkelse inträffar!) Nu var inte de alexandrinska astronomerna och matematikerna på något sätt astronomiskt obehövande. De höll mycket noga reda på solens och månens inbördes positioner och förstod därför mycket väl dels att månen lyste med lånat ljus, dels månens faser och även sol- och månförmörkelser. Med utgångspunkt från månens faser kan man också räkna ut det relativa förhållandet mellan avstånden till månen och avståndet till solen. Den som först lär ha gjort ett försök att beräkna detta förhållande var ARISTARCHOS från Samos. Aristarchos' idé var att utgå ifrån halvmånen. Resonemanget är ju enkelt — när månen är exakt halv så utgör ju solen, jorden och månen hörn i en rätvinklig triangel med den räta vinkeln i månen. Kan vi därför mäta vinkeln mellan månen och solen så kan vi räkna ut hur mycket längre det är till solen än till månen.

UPPGIFT 2. a) Aristarchos uppmätte vinkeln till *en rät vinkel minus en trettiondel av en rät vinkel*. Hur många gånger större än månen trodde Aristarchos att solen var?

b) I själva verket är solens diameter ungefär 400 gånger större än månens. Vilken vinkel (uttryckt i grader och minuter) borde Aristarchos därför ha uppmätt?

Den fråga som egentligen intresserade Aristarchos mer än avstånden var dock den relativa storleken av solen och månen i förhållande till jorden. Eftersom han redan visste (eller trodde sig åtminstone veta) förhållandet mellan solen och månen räckte det därför att beräkna förhållandet mellan jorden och månen. För att beräkna detta avstånd använde sig Aristarchos av månförmörkelserna. Aristarchos utgångspunkt var det enkla faktum att det finns totala månförmörkelser. Om man förutsätter att avståndet till solen är mycket större än avståndet till månen så att månens avstånd till solen

kan betraktas som konstant så inser man lätt att månen är högst hälften så stor som jorden. Nu är månen i själva verket mindre än så vilket medför att en total månförden tid som en månförmörkelse tar så kunde Aristarchos beräkna hur stor jordskuggan är på månens avstånd. Eftersom jordens diameter är ungefär jordskuggans diameter + en måndiameter kunde han därigenom beräkna förhållandet mellan jordens och månens diametrar. I detta sammanhang bör det påpekas att *månförmörkelser* kan ses antingen ur ett jordiskt perspektiv (d.v.s. som månförmörkelser) eller ur ett månperspektiv (som förmörkelser av solen). Står vi på månen så börjar förmörkelsen (för oss) i det ögonblick som jorden börjar skymma solen, och den blir total när vi kommer in i jordens *kärnskugga*. Står vi på jorden börjar inte månförmörkelsen förrän någon del av månen börjar komma in i kärnskuggan och den är total när hela månen befinner sig i kärnskuggan. I nedanstående uppgift ses allt ur ett jordiskt perspektiv, d.v.s. förmörkelsen varar endast medan någon del av månen befinner sig i kärnskuggan.

UPPGIFT 3. a) Antag att månförmörkelsen är total under halva tiden. Hur stor är då jordskuggan i förhållande till månen?

b) Visa med figur varför jordens diameter är summan av jordskuggans och månens diametrar. (Detta beror på att avståndet till solen är så stort att avståndet är detsamma antingen det uppmäts från jorden eller från månen, och på att solen och månen upptar samma vinkel på himlen.)

c) Hur många gånger större än månen är jorden enligt dessa beräkningar?

d) Hur många gånger större än jorden var i så fall solen (enligt Aristarchos' beräkning)?

e) Vilken vinkel upptar jorden sedd från månen?

f) Ta reda på de rätta förhållandena mellan storlekarna av jorden

och månen och använd dessa för att beräkna under hur stor del av en månförmörkelse som förmörkelsen är total (Du får förutsätta att månen rör sig i en cirkel runt jorden och att månens centrum passerar genom centrum för kärnskuggan). Vad ger detta värde för svar på c) och e)?

g) Vad får vi för svar på fråga d) om vi även antar att förhållandet mellan solens och månens diametrar är 400?

Som framgår av uppgift 3 kunde Aristarchos räkna ut att solen var många gånger större än jorden. En anledning till att Aristarchos inte "vågade" anta att hans vinkel var ännu större var förmodligen att hans resultat var alldeles för fantastiskt redan som det var. Skulle han ha räknat ut att solen var ungefär 100 gånger större än jorden skulle absolut ingen ha trott honom. (Före Aristarchos trodde de djärvaste att solen kanske var större än Peloponesos!)

Trots att Aristarchos grovt underskattade solens storlek räknade han ju ändå ut att den var mycket större än jorden och han drog därav den naturliga slutsatsen: *Jorden kretsar runt solen!*

Som Aristarchos anat var det inte många som trodde honom, och det blev den geocentriska världsbilden som förblev förhärskande under hela antiken och medeltiden. Dock fanns Aristarchos' arbete bevarat och både Hipparchos och Ptolemaios nämner Aristarchos' teori om att jorden kretsar kring solen.

Erathostenes. För att återgå till Columbus tro att jorden var rund, så var detta trots allt den "akademiska världens" övertygelse alltifrån Platons och Aristoteles' Aten på 300-talet f.Kr. Utifrån denna övertygelse hade man också beräknat Jordens storlek. Uppskattningen hade gjorts av ERATHOSTENES i Alexandria. Hans beräkning utgick ifrån två kända data, varvid den väsentligaste var *brunnen i Syene* (nuvarande Assuan). I Syene fanns det en djup brunn där solen en gång om året, nämligen vid Zenit på midsommardagen, kunde

lysa ända ner i botten. Erathostenes antog att Alexandria låg rakt norr om Syene och eftersom han kände till avståndet mellan de två städerna, så mätte han helt enkelt solhöjden i Alexandria vid samma tidpunkt.

UPPGIFT 4. Antag att Alexandria ligger på $31^{\circ} 27'$ och att Syene ligger exakt på Kräftans vändkrets $23^{\circ} 27'$. Antag vidare att Erathostenes' antagande om att Alexandria ligger rakt norr om Syene är riktigt. Om vi förutsätter att Erathostenes' beräkning av jordens omkrets var exakt och att avståndet mellan Alexandria och Syene var 5000 stadier, hur lång var då en stadion?²

UPPGIFT 5. Om man kombinerar uppgift 3 e) (eller f)) med Erathostenes' beräkning av jordens storlek kan man beräkna avståndet från jorden till månen. Gör detta.

Som vi såg ovan så var Aristarchos' och Erathostenes' mätningar och därpå grundade beräkningar tillfredsställande utom vad gäller vinkeln mellan solen och månen vid halvmåne. Detta är också den känsligaste mätningen. Svårigheten är främst att avgöra när månen är exakt halv. Ett sätt att förbättra noggrannheten i mätningar är det som används vid laborationer, nämligen att utföra mätningen flera gånger och sedan bilda ett medelvärde. Ett av Aristarchos' problem var också att han utförde sina beräkningar innan trigonometrin hade utvecklats. Detta innebar dels att han inte kunde beräkna vinkeln vid månen, utom då vinkeln var 90° , dels att han även om han kunnat beräkna vinkeln vid andra faser inte hade kunnat beräkna förhållandet mellan sidorna i triangeln sol, jord och måne ändå.

²Erathostenes' förutsättningar gällde inte exakt. Syene ligger något norr om Kräftans vändkrets, på 24° nordlig bredd och Alexandria ligger något väster om Syene. Dessutom finns det olika uppgifter om längden av en Stadion. Den för Erathostenes gynnsammaste längden av en stadion ger dock ett fel på jordens omkrets som är mindre än 100 km.

UPPGIFT 6. Ange hur du skulle bära dig åt för att beräkna vinkeln vid månen och därmed också förhållandet mellan sidorna då månen inte är exakt halv. (Du skulle t.ex. kunna använda månens *höjd* och *bredd*.)

Arkimedes. Aristarchos' och Erathostenes' beräkningar användes av ARKIMEDES som inleder det berömda arbetet *SANDRÄKNAREN* med att uppskatta hela universums storlek. I detta arbete förutsätter Arkimedes att universum är inneslutet i en stor sfär – fixstjärnesfären – och han gör en beräkning av radien i denna sfär. I sina beräkningar förutsätter Arkimedes att kvoten mellan fixstjärnesfärens radie R och avståndet r mellan jorden och solen är lika stor som kvoten mellan r och jordens radie j , d.v.s. han antog att ekvationen $R/r = r/j$ gällde. Nu vet ju vi att Erathostenes' beräkning (av j) låg nära det rätta värdet, medan Aristarchos grovt underskattade r . Det visste visserligen inte Arkimedes men han garderade sig genom att anta att jordens storlek var 10 gånger större än vad Erathostenes hade beräknat och avståndet till solen var en och en halv gång större än vad Aristarchos trodde.

UPPGIFT 7. a) Vad fick Arkimedes för uppskattning av fixstjärnesfärens radie R ?

b) Ta reda på hur stort astronomerna idag anser universum vara.

Det egentliga syftet med Arkimedes' arbete var dock inte att beskriva universum utan att skriva stora tal. I Sandräknaren fortsätter Arkimedes därför med att namnge och beskriva allt större tal. Han är inte nöjd förrän han beskrivit ett tal som vi numera skulle skriva som en etta med 80.000 biljoner nollor efter sig eller något kortare $10^{8 \cdot 10^{16}}$. Sedan han beskrivit sina stora tal så beräknar han antalet sandkorn som skulle få plats innanför fixstjärnesfären och visar att detta är ett litet pluttetal i jämförelse med de verkligt stora talen som han beskrivit. (Anm. Det bör påpekas att positionssystemet

inte var uppfunnet på Arkimedes' tid och att det största tal som hade ett namn var talet 10.000 som kallades en *myriad*.)

UPPGIFT 8. a) Arkimedes antog att ett sandkorn hade en diameter om ungefär en halv millimeter. Hur många sandkorn fick han plats med?

b) Om en atom antas ha radien 1 Ångström = 10^{-10} meter, hur många atomer får vi plats med i Arkimedes' universum och i vårt?

UPPGIFT 9. Läs i en uppslagsbok eller liknande om någon eller några av de matematiker och astronomer som formade antikens och medeltidens världsuppfattning och ta reda på mer om dem och om vad de gjorde. De intressantaste personerna är Arkimedes, Pythagoras, Aristarchos, Platon, Aristoteles och Eudoxos.

För att avsluta där jag började så vill jag tala om att den verkliga orsaken till att Columbus vågade sig ut på sin resa var att han hade hört om en ny uppskattning av jordens omkrets som bara var en tredjedel av Erathostenes. Amerikas upptäckt beror alltså i själva verket på en felräkning!

Litteratur

1. I *Sigma band 1* finns Arkimedes' uppsats Sandräknaren översatt.
2. Sinnerstad, U., *Från stjärnskådning till rymdforskning*. Doxa, Lund 1985, (Om astronomins historia).
3. Kline, M., *Matematiken i den Västerländska Kulturen*. Prisma, Stockholm 1968.

Morris Kline har skrivit flera läsvärda böcker om matematikens historia och dess roll i västerlandets utveckling. De övriga böckerna finns dock endast på engelska.

4. Hawking, S., *Kosmos, en kort historik*. Prisma 1988.

Pythagoreiska trianglar

STEN KAIJSER

Uppsala Universitet

Kort beskrivning av specialarbetet. Pythagoreiska trianglar har varit kända i minst 4000 år och kanske ännu längre. De utgör därmed ett av de äldsta vittnesbörderna om matematisk aktivitet. Specialarbetet syftar till att visa hur modern algebra kan användas för att förstå gamla problem. Detta specialarbete kan gärna kombineras med specialarbetet *Om gaussiska primtal* av Christer Kiselman, så att någon eller några arbetar med pythagoreiska trianglar och någon/några andra gör specialarbetet om gaussiska primtal. Det bör påpekas att detta arbete kan utföras även av den som inte har tillgång till dator (eller ens fickräknare), men att flera av uppgifterna kan undersökas noggrannare för den som har en dator.

Kort historik. Pythagoras' sats tillhör de äldsta vittnesbörderna om mänsklig matematisk aktivitet. Vi får alla i skolan lära oss om den *Egyptiska triangeln* med sidorna 3, 4 och 5, och vi får ibland höra att egypterna använde denna triangel för att åstadkomma räta vinklar när de byggde sina pyramider. Även om detta förmodligen inte är sant, helt enkelt för att de antagligen föredrog att tillverka vinkelhakar, så är det förvisso sant att egypterna redan för tre och ett halvt årtusende sedan kände till både denna triangel och andra rätvinkliga trianglar vars sidor är heltal. Det som förmodligen också är sant är att antingen PYTHAGORAS själv³, eller någon i hans skola, faktiskt bevisade både satsen och dess omvändning.

³Pythagoras kom till staden Kroton i södra Italien omkring år 530 f.Kr. och var verksam där till sin död ungefär 30 år senare.

PYTHAGORAS' SATS. *Om T är en rätvinklig triangel med kateterna a och b och med hypotenusan c så råder sambandet $a^2 + b^2 = c^2$.*

OMVÄNDNINGEN TILL PYTHAGORAS' SATS. *Om T' är en triangel med sidorna a, b och c sådan att $a^2 + b^2 = c^2$ så är T' rätvinklig med kateterna a och b och med hypotenusan c .*

Det är däremot inte sant att Pythagoras eller pythagoréerna *upptäckte* satsen. Av bevarade kilskrifter framgår att babylonierna i Mesopotamien kände till ett flertal s.k. pythagoreiska taltripplar, d.v.s. taltripplar av *naturliga tal* (a, b, c) sådana att $a^2 + b^2 = c^2$. Sådana tripplar var kända även i Indien vid ungefär samma tid som Pythagoras var verksam i Grekland, och möjligen hade de redan då varit kända i tusentals år.

Vi kommer i fortsättningen att säga att en rätvinklig triangel T vars alla tre sidor har heltalslängd är pythagoreisk (eller en pythagoreisk triangel). Vi kommer likaså att säga att en taltrippel (av naturliga tal) (a, b, c) är pythagoreisk om $a^2 + b^2 = c^2$. Samtidigt bör det påpekas att vi inte skiljer mellan taltripplarna (a, b, c) och (b, a, c) eller mellan en viss triangel och dess spegelvändning.

Som vi också får lära oss ledde Pythagoras sats till upptäckten av irrationella tal — något som fick stor betydelse för den grekiska matematiken. (Ännu större betydelse lär denna upptäckt ha haft för den som gjorde den, eftersom han som straff kastades överbord vid nästa sjöresa — kom sedan inte och påstå att matematisk forskning är riskfri.)

Pythagoreiska taltripplar har varit kända i två och ett halvt årtusende och även generella metoder att konstruera dem har varit kända, åtminstone sedan 300-talet (e.Kr.) av både kinesiska och västländska matematiker. Den som i väst angav en metod var Diofantos. Även om metoder att konstruera dem alltså varit kända dröjde

det länge innan man kunde ge en fullständig teori som direkt kunde beskriva mängden av alla möjliga pythagoreiska trianglar. Den som löste problemet var PIERRE FERMAT (1601 - 1665) som bevisade följande vackra resultat.

SATS 3. Ett primtal p kan skrivas som en summa av två (heltals-) kvadrater om och endast om $p = 4n + 1$.

Vi ska senare se vilken roll denna sats (och dess konsekvenser) spelar för beskrivningen av pythagoreiska trianglar.

Ett viktigt bidrag till förståelsen av pythagoreiska trianglar gavs av GAUSS (1777 - 1855), som med introduktionen av det komplexa planet och de s.k. Gaussiska heltalen skapade nya möjligheter att använda algebraiska metoder för studiet av pythagoreiska trianglar.

Några förberedande observationer. Innan vi övergår till att studera pythagoreiska trianglar med algebraiska metoder ska vi börja med några enkla observationer. Låt oss till att börja med införa ett slags ordning på mängden av dem, så att vi kan tala om att en triangel är mindre än en annan, genom att i första hand gå efter längden av hypotenusan och i andra hand efter längden av den kortaste kateten. Detta innebär t.ex. att (12, 16, 20) är mindre än (7, 24, 25) och att (7, 24, 25) är mindre än (15, 20, 25).

1. Visa att de tre minsta pythagoreiska trianglarna, med avseende på denna ordning, är (3, 4, 5), (6, 8, 10) och (5, 12, 13).

En första naiv fråga som man kan ställa sig när det gäller pythagoreiska trianglar är om alla (naturliga) tal kan vara sida i någon sådan. Svaret på denna fråga får ni genom att lösa följande uppgifter.

2. Låt u vara ett udda tal (≥ 3). Visa att det finns ett tal b så att triangeln $(u, b, b + 1)$ är pythagoreisk. Bestäm också sambandet mellan b och u .

3. Låt j vara ett jämnt tal (≥ 4). Visa att det finns ett tal a så att $(j, a - 1, a + 1)$ är pythagoreisk. Bestäm sambandet mellan a och j .
4. Kan 1 eller 2 vara sidor i en pythagoreisk triangel?

Dessa uppgifter visar att alla naturliga tal, utom 1 och 2, kan förekomma som kateter i en pythagoreisk triangel, så att den naturliga följdfrågan är därför om alla tal också kan vara hypotenusor. För att få en idé om svaret på denna fråga bör ni innan ni läser vidare lösa följande uppgift.

5. Det finns 10 pythagoreiska trianglar med en hypotenusor ≤ 29 . Bestäm dessa.

Ledning: Alla utom en av dessa pythagoreiska trianglar kan erhållas med hjälp av de två föregående problemen. Om ni inte kan hitta den sista nu kommer ni säkert att göra det då ni läst lite längre.

Räkning med tal. När vi som barn började med räkning eller matematik i skolan så fick vi börja med att räkna från 1 till 10. Snart fick vi lära oss addera dessa tal och vi fick räkna längre och längre tills vi så småningom fick en känsla av att det fanns (*nästan?*) hur stora tal som helst. Detta innebär att vi hade en aning om mängden \mathcal{N} av naturliga tal. Innan vi kom så långt hade vi ju naturligtvis börjat med subtraktion och som ett resultat av denna började de negativa talen tränga in i vårt medvetande. Vi fick också lära oss multiplikation och även division. I samband med att vi lärde oss multiplikation upptäckte vi också att vissa tal ständigt dök upp i svaren medan andra aldrig gjorde det, vilket förklarades (i samband med divisionen) av att vissa tal hade många delare medan andra hade få eller ibland inga alls. På så sätt fick vi lära oss om primtalen. Sedan visade det sig att division inte alltid gick jämnt upp, så att vi blev tvungna att lära oss att räkna med bråk. Vi lärde oss att multiplicera och dividera bråk, vilket var lätt så snart vi lärt oss att förkorta bråk. Det var svårare att addera och subtrahera bråken och

för det tvingades vi lära oss begrepp som *minsta gemensam multipel* och *största gemensamma delare*. En viktig egenskap hos de naturliga talen som vi fick lära oss, men som vi aldrig fick se något bevis för var *satsen om entydig primtalsfaktorisering*.

Efter några år i skolan kunde vi därför handskas med om inte mängderna själva så åtminstone elementen i dem för såväl mängden av alla hela tal \mathbf{Z} som mängden av rationella tal (kvoterna, **quotients**) \mathbf{Q} . Något senare lärde vi oss funktioner och började därmed lära oss att arbeta med reella tal, och t.o.m. mängder av reella tal. Eftersom vi också fick lära oss att lösa andragradsekvationer så fick vi åtminstone höra talas om det mystiska talet i , d.v.s. kvadratroten ur -1 , och om de komplexa talen.

6. Ett komplext tal $z = a + bi$ kan skrivas som $z = r(\cos \theta + i \sin \theta)$, varvid $r = |z| = \sqrt{a^2 + b^2}$ och $\tan \theta = b/a$.

$$\begin{aligned} \text{Om } z &= a + bi = r(\cos \theta + i \sin \theta) \\ \text{och } w &= c + di = s(\cos \varphi + i \sin \varphi) \end{aligned}$$

så är

$$zw = (a + bi)(c + di) = (ac - bd) + (ad + bc)i = R(\cos \psi + i \sin \psi).$$

Bevisa att

- (i) $zw = wz$ och att
- (ii) $R = rs$ och $\psi = \theta + \varphi$.

På universitetet får man lära sig mer om komplexa tal, men eftersom de oftare förekommer i samband med analys än med algebra, så ägnas *de hela komplexa talen* ingen större uppmärksamhet. Ändå är dessa, mängden av s.k. *Gaussiska heltal*, en både viktig och intressant matematisk struktur. Denna mängd brukar skrivas som $\mathbf{Z}(i)$ för att

ange att den innehåller dels mängden av heltal \mathbf{Z} , dels talet $i = \sqrt{-1}$. Den grundläggande egenskapen är att $\mathbf{Z}(i)$ är en *Ring* vilket betyder att man kan både addera och subtrahera och dessutom multiplicera två gaussiska heltal med varandra (och resultatet blir på nytt ett gaussiskt heltal). Mängden $\mathbf{Z}(i)$ och operationerna på den definieras på följande sätt:

Låt a, b, c och d vara heltal. Då är $a + bi$ och $c + di$ gaussiska heltal. Vidare definieras summan och produkten som för vanliga komplexa tal, d.v.s. genom att

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

Innan vi fortsätter kan det vara lämpligt att antyda sambandet mellan gaussiska heltal och det problem som vi egentligen håller på med d.v.s. att på något sätt beskriva mängden av alla pythagoreiska trianglar. Vi ska göra detta genom att införa ännu en tolkning av dessa genom att säga att gaussiskt heltal $z = a + bi$ är pythagoreiskt om $|z|$ ($= \sqrt{a^2 + b^2}$) är ett (vanligt) heltal. Vi kommer i fortsättningen helt enkelt att tala om pythagoreiska tal, varvid det är underförstått att talet är ett (pythagoreiskt) gaussiskt heltal. Detta ger oss tre sätt att uppfatta pythagoreiska trianglar, som trianglar, som taltripplar eller som gaussiska heltal. Vi ska snart se att den algebraiska strukturen hos $\mathbf{Z}(i)$ gör det möjligt att ge en enkel och tilltalande beskrivning av de pythagoreiska *talen*. Eftersom vi inte skiljer på de pythagoreiska tripplarna (a, b, c) och (b, a, c) är det värt att notera att talen $a + bi$ och $b + ai$ ger samma pythagoreiska *triangel*. Dessutom är det praktiskt att även tillåta att *realdelen* och/eller *imaginärdelen* av ett pythagoreiskt tal är negativ. (Om $z = a + bi$, så är realdelen $\Re(z) = a$ och imaginärdelen $\Im(z) = b$.) Sammantaget

innebär detta att alla de (vanligen) åtta talen $\pm a \pm bi$ och $\pm b \pm ai$ svarar mot samma pythagoreiska triangel.

De Gaussiska heltalen. Det vi närmast ska ägna oss åt är primtal och primtalsfaktorisering i $\mathbf{Z}(i)$. Vi ska börja med några definitioner.

DEFINITION 1. Om x och z är gaussiska heltal så sägs x vara en *delare* i z om det finns ett gaussiskt heltal y sådant att $xy = z$.

DEFINITION 2. En delare i talet 1 kallas för en *enhet* (i $\mathbf{Z}(i)$).

DEFINITION 3. Två gaussiska heltal x och y sägs vara *associerade* om det finns en enhet ε i $\mathbf{Z}(i)$ så att $x = \varepsilon y$ d.v.s. om kvoten mellan dem är en enhet.

Vi ska använda den vanliga beteckningen $x|z$ för att ange att x är en delare i z . Om $x|z$ och varken x eller $y = z/x$ är enheter så sägs x vara en *äkta* delare.

För att kunna tala om primtalsfaktorisering måste vi naturligtvis först och främst veta vad ett primtal (i $\mathbf{Z}(i)$) är.

DEFINITION 4. Ett gaussiskt heltal p kallas för ett *primtal* (i $\mathbf{Z}(i)$) om det inte har någon äkta delare i $\mathbf{Z}(i)$.

Ett viktigt hjälpmedel för att bevisa satser om naturliga tal är induktionsprincipen, d.v.s. det faktum att en icke-tom mängd av naturliga tal har ett minsta element. Induktionsprincipen kan användas även vid studiet av hela tal eftersom $|n|$ alltid är positivt (eller åtminstone icke-negativt), så att varje mängd av hela tal innehåller något (möjligen t.o.m. 2) tal med minsta belopp. För att på motsvarande sätt kunna använda induktion även vid studiet av $\mathbf{Z}(i)$, så behöver vi en lämplig funktion som till varje gaussiskt heltal tillordnar ett (välvalt) naturligt tal. För detta behöver vi ytterligare ett par nya begrepp.

DEFINITION 5. Om $z = a + bi$ är ett gaussiskt heltal (eller mer allmänt ett komplext tal) kallas talet $z^* = a - bi$ för det *konjugerade talet* till z , eller för *z -konjugat*.

7. Visa att $(z^*)^* = z$ och att $z = xy$ om och endast om $z^* = x^*y^*$.
8. Visa att z är ett (gaussiskt) primtal om och endast om z^* är det.
9. Låt x vara ett gaussiskt heltal och låt n vara ett naturligt tal. Visa att $x|n$ medför att $x^*|n$ och att $n|x$ medför att $n|x^*$.

DEFINITION 6. Om $z = a + bi$ är ett gaussiskt heltal, så kallas talet

$$N(z) = |z|^2 = z^*z = zz^* = a^2 + b^2$$

för *normen* av z .

ANMÄRKNING. Vanligen används ordet norm i modern matematik för något som snarare motsvarar $|z|$ än $|z|^2$ men för gaussiska heltal infördes benämningen redan av Gauss och det har därför förblivit den gängse beteckningen.

10. Visa att $N(z^*) = N(z)$

Eftersom normen av ett gaussiskt heltal alltid är ett naturligt tal så innehåller varje (icke tom) mängd av gaussiska heltal, något element med minimal norm. En annan viktig egenskap framgår av följande uppgift.

11. Visa att normen är *multiplikativ* d.v.s. att $N(zw) = N(z)N(w)$.

Detta innebär att om $x|z$ i $\mathbf{Z}(i)$, så är $N(x)|N(z)$ i \mathbf{Z} .

12. Visa att z är ett primtal i $\mathbf{Z}(i)$ om $N(z)$ är ett primtal i \mathbf{Z} .
13. Visa att z är en enhet i $\mathbf{Z}(i)$ om och endast om $N(z) = 1$ och bestäm alla enheter i $\mathbf{Z}(i)$.
14. Visa att om $z \neq 0$ är ett gaussiskt heltal, så finns det någon enhet ε (i $\mathbf{Z}(i)$), sådan att om $w = \varepsilon z$ så är
 - (i) realdelen av w positiv, och

(ii) antingen är beloppet av imaginärdelen *strikt mindre* än realdelen eller så är den *lika med* realdelen (och därmed positiv). Detta innebär att

$$(\star) \quad -\Re(w) < \Im(w) \leq \Re(w).$$

RITA FIGUR!

(Vi ska säga att ett primtal är skrivet på *normalform* om (\star) gäller.)

Innan vi fortsätter ska vi göra en enkel observation som omedelbart kommer att ge oss ett lätt sätt att konstruera pythagoreiska trianglar.

15. Visa att det nödvändiga och tillräckliga villkoret för att ett gaussiskt heltal z ska vara pythagoreiskt är att dess norm $N(z)$ är en jämn kvadrat.

ANMÄRKNING. Detta innebär att om $w = z^2$ så är $N(w) = N(z^2) = N(z)N(z) = N(z)^2$ så att talet w är pythagoreiskt (om det inte är rent reellt eller rent imaginärt).

16. Av de 10 pythagoreiska trianglar med hypotenusan högst 29, som du (förhoppningsvis) hittat, så kan alla utom två erhållas som kvadrater. Bestäm vilka som inte är det.

Den entydiga primtalsfaktoriseringen i $\mathbf{Z}(i)$. Vi ska börja med att formulera och bevisa den lätta delen av satsen om entydig primtalsfaktorisering, nämligen att en faktorisering existerar.

SATS 4. (*Existensen av primtalsfaktorisering.*) *Varje gaussiskt heltal, som inte är en enhet i $\mathbf{Z}(i)$, kan skrivas som en produkt av primtal.*

BEVIS. Vi ska använda induktion och börjar med att observera att om $N(z) = 2$ så är z ett primtal, helt enkelt därför att om 2 är en produkt av två naturliga tal så måste något av dem vara 1. Vi antar nu att alla gaussiska heltal med en norm som är mindre än n har en

primtalsfaktorisering och att z är ett gaussiskt heltal med normen n . Om z är ett primtal så är det sin egen primtalsfaktorisering och då finns det inget mer att bevisa. Om z inte är ett primtal kan vi skriva $z = xy$ där både x och y är äkta delare. Eftersom $N(z) = N(x)N(y)$ (och båda är större än 1) så är $2 \leq N(x) < N(z)$ och $2 \leq N(y) < N(z)$, så enligt antagandet har båda primtalsfaktoriseringar och produkten av dessa är vår sökta faktorisering av z .

17. Bestäm alla gaussiska heltal med normen 2 och visa att de är associerade.

18. Är 2 ett primtal i $\mathbf{Z}(i)$?

Medan existensen av en primtalsfaktorisering som regel är lätt att bevisa, så brukar entydighet vara ett betydligt svårare problem. Den centrala egenskapen hos primtalen i \mathbf{Z} , och som också måste bevisas i $\mathbf{Z}(i)$ är att om ett primtal delar en produkt så delar det också en av faktorerna.

Utgångspunkten för alla undersökningar av entydigheten av faktoriseringen i $\mathbf{Z}(i)$ är följande

SATS 5. (*Divisionsalgoritmen*) Om a och b är gaussiska heltal, så finns det gaussiska heltal q och r sådana att

$$(i) \quad a = bq + r \text{ och}$$

$$(ii) \quad 0 \leq N(r) \leq \frac{N(b)}{2}.$$

Innan vi bevisar divisionsalgoritmen för *gaussiska heltal* kan det vara lämpligt att ge motsvarande sats för (vanliga) heltal.

SATS 5'. (*Divisionsalgoritmen*) Om a och b är heltal, så finns det heltal q och r sådana att

$$(i) \quad a = bq + r \text{ och}$$

$$(ii) \quad -|b|/2 < r \leq |b|/2.$$

(Denna sats är självklar om vi helt enkelt väljer q så att $|a - bq|$ blir så liten som möjligt. Rita figur!)

BEVIS AV DIVISIONSALGORITMEN FÖR GAUSSISKA HELTAL. Vi antar först att talet b är ett naturligt tal, och för att göra beteckningarna klarare ska vi skriva n istället för b . Detta innebär att vi har ett gaussiskt heltal $a = \alpha + \beta i$ och ett naturligt tal n och enligt divisionsalgoritmen för hela tal finns det hela tal q_1, r_1, q_2 och r_2 sådana att $|r_1| \leq n/2$ och $|r_2| \leq n/2$ och dessutom gäller

$$a = \alpha + \beta i = (q_1 n + r_1) + (q_2 n + r_2) i = (q_1 + q_2 i) n + (r_1 + r_2 i) = qn + r.$$

Eftersom vidare

$$N(r) = r_1^2 + r_2^2 \leq 2 \frac{n^2}{4} = \frac{n^2}{2} = \frac{N(n)}{2}$$

så gäller satsen i detta fall.

Om nu b inte är ett naturligt tal så börjar vi med att multiplicera både a och b med b^* vilker ger talen $A = ab^*$ och $N = bb^*$. Enligt vad vi nyss såg finns q och r' så att

$$A = qN + r' \quad \text{med} \quad N(r') \leq N(N)/2 = N(b)^2/2.$$

Vi kan nu definiera $r = a - bq$ och eftersom

$$r' = (A - qN) = (ab^* - qbb^*) = (a - qb)b^* = rb^*,$$

så är

$$N(r) = \frac{N(r)N(b^*)}{N(b^*)} = \frac{N(r')}{N(b^*)} \leq \frac{N(b)^2}{2N(b)} = \frac{N(b)}{2}.$$

Förutom att den möjliggör induktion är normen användbar också på andra sätt, bl.a. genom att den gör det möjligt att tala om att ett gaussiskt heltal är "större" än ett annat (trots att det egentligen inte finns någon *ordning* i $\mathbf{Z}(i)$). Vi kan därför definiera t.ex. *en största gemensam delare* $\text{sgd}(x, y)$ till två gaussiska heltal x och y som en delare med största möjliga norm. Med hjälp av divisionsalgoritmen kan vi nu bevisa följande viktiga

SATS 6. Låt a och b vara gaussiska heltal, som inte båda är 0, och låt d vara en största gemensam delare. Då finns det gaussiska heltal x och y sådana att

$$d = xa + yb.$$

BEVIS. Bilda mängden $M = M(a, b)$ av alla gaussiska heltal m som kan skrivas på formen $m = xa + yb$ för något val av talen x och y . Det är lätt att se att varje gemensam delare till a och b är en delare till varje tal i M . Speciellt är d en gemensam delare för hela mängden M . Låt nu $d' = x'a + y'b$ vara något tal i M med den minsta möjliga (strikt positiva) normen. Vi vill bevisa att d' är en delare till alla tal i M och väljer ett godtyckligt $m = xa + yb$ i M . Med hjälp av divisionsalgoritmen kan vi skriva $m = qd' + r$, där $0 \leq N(r) < N(d')$. Då är

$$r = m - qd' = (xa + yb) - q(x'a + y'b) = (x - qx')a + (y - qy')b$$

vilket innebär att $r \in M$. Eftersom enligt förutsättningen $N(d')$ är den minsta möjliga normen, så innebär detta att $N(r) = 0$, d.v.s. att $d' | m$. Eftersom både a och b tillhör M så är d' en gemensam delare till dem, och eftersom d har den största möjliga normen av alla delare så är $N(d') \leq N(d)$. Å andra sidan är d en gemensam delare till a och b , vilket innebär att det finns gaussiska heltal s och t så att $a = sd$ och $b = td$. Men då är

$$d' = x'a + y'b = x'sd + y'td = (x's + y't)d = zd.$$

Detta innebär att $N(d') = N(z)N(d)$ och eftersom $N(d') \leq N(d)$ så är $N(z) = 1$. Detta innebär att $d = z^*d' = (z^*x')a + (z^*y')b$.

ANMÄRKNING. Av beviset följer att det finns tal x och y sådana att $d = xa + yb$ men beviset ger ingen metod för att hitta varken dem

eller den största gemensamma delaren. Det finns dock en *konstruktiv* metod som finns beskriven redan i EUKLIDES' ELEMENTA. Enligt denna metod (som brukar kallas *Euklides' Algoritm*) konstrueras $\text{sgd}(a, b)$ för två hela tal a och b genom upprepad användning av divisionsalgoritmen. Om vi antar att $|a| > b$ så får vi en följd $r_1 > r_2 > \dots r_{n-1} > r_n = 0$ genom att

$$\begin{aligned} r_1 &= a - bq_1 && \text{med } |r_1| \leq b/2 \\ r_2 &= b - r_1q_2 && \text{med } |r_2| \leq r_1/2 \\ r_3 &= r_1 - r_2q_3 && \text{med } |r_3| \leq r_2/2 \\ &\vdots \\ 0 &= r_n = r_{n-2} - r_{n-1}q_n. \end{aligned}$$

Det är uppenbart att processen tar slut efter ett ändligt antal steg. Av konstruktionen följer också att $r_1 \in M$ (där M är mängden av alla $xa + by$).

19. Visa (med induktion) att varje $r_i (\neq 0)$ som erhålles i denna följd är av formen $xa + yb$, där x och y är heltal.

20. Visa att den sista resten r_{n-1} är en gemensam delare till a och b . (Euklides' algoritm ger alltså en metod att hitta element av formen $xa + by$ med allt mindre belopp.)

21. Visa att Euklides' algoritm kan användas även i $\mathbf{Z}(i)$, och att den ger en konstruktiv metod för att hitta den största gemensamma delaren till två givna gaussiska heltal.

22. Bestäm (exempelvis med användning av Euklides' algoritm) den största gemensamma delaren till

a) $21 + 20i$ och $5 + 2i$

b) $15 + 10i$ och $13 - 26i$

c) $47 + 4i$ och $26 + 6i$

Därmed är vi färdiga för det viktigaste lemmat i beviset av den entydiga primtalsfaktoriseringen i $\mathbf{Z}(i)$.

LEMMA 1. *Om ett gaussiskt primtal p delar en produkt*

$$z = b_1 b_2 \dots b_n$$

så är p en delare i någon av faktorerna b_k .

BEVIS. Vi ska använda induktion och noterar först att påståendet är trivialt om $n = 1$. Vi antar därför att det gäller för varje produkt med högst $n - 1$ faktorer. Vi sätter $a = b_1 b_2 \dots b_{n-1}$. Om $p|a$ så följer det av induktionsantagandet att p delar någon av faktorerna b_k , $1 \leq k \leq n - 1$ vilket var vad vi ville veta. Vi antar därför att p inte är en delare i a . Eftersom ett primtal inte har några andra delare än enheter och associerade tal (som inte kan vara delare i a) så är 1 en största gemensam delare till a och p . Enligt föregående sats finns det tal x och y sådana att $1 = xa + yp$ vilket innebär att $b_n = 1 \cdot b_n = xab_n + ypb_n$. Enligt förutsättningen är p en delare till båda termerna i högerledet, och därmed också till deras summa, vilket betyder att $p|b_n$.

Äntligen är vi framme vid den stora satsen.

SATS 7. (ARITMETIKENS FUNDAMENTALSATS FÖR GAUSSISKA HELLTAL). *Varje gaussiskt heltal har en primtalsfaktorisering. Denna är entydig bortsett från faktorernas ordning (och förekomst av associerade primtal).*

BEVIS. Eftersom vi redan bevisat existensen behöver vi bara visa entydigheten. Om satsen inte är sann så finns det ett gaussiskt heltal med minimal norm för vilket den inte gäller. Låt x vara ett sådant tal. Vi antar alltså att $x = p_1 p_2 \dots p_n$ och $x = q_1 q_2 \dots q_m$ båda är

primtalsfaktoriseringar av x . Eftersom p_1 är ett primtal och dessutom $p_1|x$ så gäller enligt lemma 1, att $p_1|q_k$ för något k , $1 \leq k \leq m$. Eftersom även q_k är ett primtal så är $q_k = \varepsilon p_1$ där ε är en enhet. Men då är $y = p_2 \dots p_n = (\varepsilon q_1) q_2 \dots q_{k-1} q_{k+1} \dots q_m$ och eftersom $N(y) < N(x)$ så följer det av induktionsantagandet att dessa två primtalsfaktoriseringar av y är lika bortsett från faktorernas ordning, vilket naturligtvis innebär detsamma för faktoriseringarna av x , och därmed har vi fått en motsägelse som bevisar satsen.

23. Visa att varje gaussiskt heltal z har en kanonisk primtalsfaktorisering av formen $z = \varepsilon p_1 p_2 \dots p_n$ där ε är en enhet och alla p_k är skrivna på normalform.

Bestämning av primtalen i $\mathbf{Z}(i)$. För att riktigt kunna utnyttja de gaussiska heltalen för att studera pythagoreiska trianglar, räcker det inte med att känna till satsen om entydig primtalsfaktorisering, vi måste också kunna använda den, varmed menas att vi ska kunna utföra en faktorisering.

Om vi förutsätter att vi kan faktorisera naturliga tal (något som vi med en programmerbar fickräknare kan göra åtminstone för tal upp till 100 000) så vill vi om möjligt utnyttja faktoriseringen av $N(z)$ för att faktorisera z . Vi vill alltså veta om en faktorisering av $N(z)$ som en produkt av naturliga tal på något sätt svarar mot en faktorisering av z som en produkt av gaussiska heltal och hur vi i så fall ska använda den. För att se hur detta ska gå till ska vi börja med att bestämma primtalen i $\mathbf{Z}(i)$ (givet de naturliga primtalen).

Vi börjar med att notera att $j_0 = 1 + i$ och alla med detta tal associerade tal är primtal. Det är naturligt att säga att ett gaussiskt heltal är jämnt om och endast om det är delbart med j_0 .

24. Bevisa att ett gaussiskt heltal z är jämnt om och endast $s = \Re(z) + \Im(z)$ är ett jämnt heltal.

25. Bestäm alla primtal p i $\mathbf{Z}(i)$, sådana att $0 < \Re(p) \leq 5$ och $0 \leq \Im(p) \leq \Re(p)$. (Det finns 7 stycken.)

ANMÄRKNING. Eftersom $N(1+i) = 1+1 = 2$ så är talet $j_0 = 1+i$ i och för sig ett specialfall av vår tidigare observation att z är ett primtal (i $\mathbf{Z}(i)$) om $N(z)$ är ett primtal (i \mathbf{Z}), men det är samtidigt speciellt eftersom $j_0^* = -ij_0$ så att j_0 och j_0^* är associerade.

Utöver talet $1+i$ så är alltså alla gaussiska heltal z , sådana att $N(z)$ är ett *naturligt* primtal, primtal i $\mathbf{Z}(i)$. Frågan är om det finns några andra. Låt därför p vara ett primtal i $\mathbf{Z}(i)$, och antag att $N(p)$ *inte* är ett primtal. Vi kan då skriva $N(p) = kl$ varvid vi antar att k är den minsta primfaktorn i $N(p)$. Detta innebär att $N(k) = k^2 \leq kl = N(p)$. Nu finns det två möjligheter, för antingen är k ett primtal även i $\mathbf{Z}(i)$ eller så är det inte det. Om k är ett primtal så är det en faktor i antingen p eller p^* och därmed i båda, vilket (eftersom p är ett primtal) innebär att $p = \varepsilon k$ där ε är en enhet och att $N(p) = k^2$ (där k är ett primtal i \mathbf{Z}). Om k inte är ett primtal i $\mathbf{Z}(i)$, så innehåller det en primfaktor q , och eftersom q då är en delare i antingen p eller p^* , så är antingen q eller q^* en faktor i p , men eftersom $N(q) < N(k) \leq N(p)$ så ger detta en motsägelse mot antagandet att p är ett gaussiskt primtal. Sammanfattningsvis har vi därmed bevisat följande

SATS 8. *Om p är ett Gaussiskt primtal så är $N(p)$ antingen ett primtal eller en primtalskvadrat i \mathbf{Z} .*

Därmed har vi återfört problemet att bestämma primtalen i $\mathbf{Z}(i)$ till ett problem för primtal i \mathbf{Z} . Det vi gjort är nämligen att uppdelade gaussiska primtalen i två klasser, som exempelvis kan beskrivas av att $p \neq p^*$ (om $N(p)$ är ett primtal) eller $p = p^*$ (annars) och därmed har vi också delat in de vanliga primtalen i två klasser, nämligen dels de som kan faktoriseras i $\mathbf{Z}(i)$, dels de som inte kan det. Frågan är

nu om denna uppdelning kan beskrivas på något annat sätt. Vi observerar därför först att om det naturliga primtalet q innehåller den gaussiska primfaktorn $p = a + bi$ så är $q = N(p) = p^*p = a^2 + b^2$, vilket innebär att q kan skrivas som en summa av två kvadrater. Omvänt är det uppenbart att om $q = a^2 + b^2$ så är $q = (a + bi)(a - bi)$ vilket innebär att q inte är ett gaussiskt primtal. För att bestämma primtalen i $\mathbf{Z}(i)$ så gäller det alltså att ta reda på vilka naturliga primtal som kan skrivas som en summa av två kvadrater. Som ett första steg mot detta noterar vi att om vi bortser från talet 2 är alla naturliga primtal udda, så att om q är en summa av två kvadrater är den ena av dessa jämn och den andra udda. Detta innebär då att q kan skrivas som $(2k)^2 + (2l + 1)^2 = 4k^2 + 4l^2 + 4l + 1 = 4n + 1$. Ett nödvändigt villkor för att q ska kunna faktoriseras i $\mathbf{Z}(i)$ är därför att det är av formen $4n + 1$. Därmed vet vi alltså att alla naturliga primtal av formen $4n + 3$ är primtal också i $\mathbf{Z}(i)$ vilket alltså gäller för talen 3, 7, 11, 19, 23 o.s.v.

Fermats sats. Frågan är nu om alla primtal av formen $4n + 1$ verkligen kan faktoriseras i $\mathbf{Z}(i)$, (d.v.s. skrivas som en summa av två kvadrater) och nu kommer äntligen Fermats sats till användning.

SATS 9. *Varje primtal av formen $4n + 1$ är en summa av två kvadrater.*

Eftersom beviset för denna sats innehåller helt andra begrepp än de som annars ingår i detta specialarbete, så hoppar vi över det.⁴ Ett bevis finns i Le Veque [1956, volym I, kapitel 7]. Däremot finns

⁴Fermats intresse för detta problem tycks till stor del ha berott på ett intresse för det vi håller på med i detta specialarbete, nämligen pythagoreiska trianglar. De algebraiska strukturer, *ändlig grupp* och *ändlig talkropp* som beviset bygger på fanns inte utvecklade på Fermats tid. Därför studerade Fermat vissa specialfall av dem, men de satser han bevisade spelade stor roll för de matematiker, främst LAGRANGE, (1736-1813) och GALOIS, (1811-1832) som senare utvecklade strukturerna.

det all anledning att övertyga sig om att satsen förefaller vara sann genom att utföra följande uppgift.

26. Det finns 11 primtal av formen $4n + 1$ som är mindre än 100. Skriv dem som summor av två kvadrater och faktorisera dem i $\mathbf{Z}(i)$.

27. Skriv ett datorprogram som tar reda på om ett givet naturligt tal kan skrivas som en summa av två kvadrater, och på hur många sätt det kan ske. (Kan du se något mönster?)

Kombinerar vi Fermats sats med sats 7 ovan får vi följande beskrivning av primtalen i $\mathbf{Z}(i)$.

SATS 8. *Primtalen i $\mathbf{Z}(i)$ är dels talen $\{\pm 1 \pm i\}$, dels alla tal av formen $\pm a \pm bi$ sådana att $a^2 + b^2 = p$ där p är ett primtal (i \mathbf{Z}) av formen $4n + 1$, dels alla tal av formen $i^k p$ där p är ett primtal av formen $4n + 3$.*

Beskrivning av mängden av alla pythagoreiska trianglar.

Med hjälp av sats 8 kan vi nu ge en fullständig beskrivning av mängden av alla pythagoreiska trianglar, genom att ge ett kriterium för när $z = a + bi$ är pythagoreiskt. För att göra det så observerar vi först att varje gaussiskt heltal kan skrivas som ett naturligt tal gånger en produkt av *icke-reella primfaktorer*, d.v.s. som $K \cdot (a' + b'i)$ där $(a' + b'i)$ inte innehåller någon *reell* primfaktor. Vi vet att z är pythagoreiskt om $N(z)$ är en jämn kvadrat och eftersom $N(z) = K^2((a')^2 + (b')^2) = K^2P$ så gäller detta om och endast om $P = (a')^2 + (b')^2$ är en jämn kvadrat. Nu är P en kvadrat om och endast om alla dessa primfaktorer förekommer ett jämnt antal gånger, vilket innebär att $a' + b'i$ är en jämn kvadrat. Sammanfattningsvis ger detta

SATS 9. *Ett gaussiskt heltal z är pythagoreiskt om och endast om det är av formen $N \cdot w^2$ där N är ett naturligt tal och w^2 varken är reellt eller rent imaginärt.*

Sats 9 är naturligtvis perfekt som beskrivning av mängden och den ger en utmärkt metod för att konstruera pythagoreiska trianglar. Däremot är det inte så lätt att omedelbart besvara frågan om vilka naturliga tal som kan vara hypotenusor eller hur många olika pythagoreiska trianglar, som har en viss hypotenusor. För att ge ett svar på denna fråga (som är en fråga om naturliga tal) så ska vi dela upp de vanliga primtalen i två klasser \mathcal{P}_1 och \mathcal{P}_2 där den första består av alla primtal av formen $4n + 1$ medan den andra består av 2 och alla primtal av formen $4n + 3$. Vi har då följande

SATS 10. *Ett naturligt tal H kan vara hypotenusor i en pythagoreisk triangel om och endast om det innehåller någon primfaktor av formen $4n + 1$.*

Om $H = K \cdot h$ där $h = p_1^{k_1} p_2^{k_2} \cdots p_n^{k_n}$ är produkten av alla primtal av formen $4n + 1$ så ges antalet N av olika trianglar med hypotenusan h av formeln

$$N = \frac{\mathcal{P} - 1}{2} = \frac{(2k_1 + 1)(2k_2 + 1) \cdots (2k_n + 1) - 1}{2}.$$

BEVIS. Ett gaussiskt heltal z med beloppet H , d.v.s. normen H^2 kan faktoriseras som

$$z = K(a_1 + b_1 i)^{l_1} (a_1 + b_1 i)^{2k_1 - l_1} \cdots (a_n + b_n i)^{l_n} (a_n - b_n i)^{2k_n - l_n},$$

där $a_i^2 + b_i^2 = p_i$, och där $0 \leq l_i \leq 2k_i$. Det finns \mathcal{P} stycken val av talen l_i och alla val utom det där alla $l_i = k_i$ ger ett gaussiskt heltal med imaginärdel $\neq 0$. Om vi sedan påminner oss att z och z^* ger samma triangel, så är det lätt att inse att det totala antalet är $(\mathcal{P} - 1)/2$.

28. Bestäm alla naturliga tal $h < 100$ sådana att det finns mer än en pythagoreisk triangel med hypotenusan h , och bestäm alla de till

dessa h hörande pythagoreiska triangelarna.

Ledning: Det finns fem sådana tal h , tre av dem har två tillhörande trianglar, de övriga har fyra.

29. Bestäm det minsta tal h för vilket det finns fler än 4 pythagoreiska trianglar med hypotenusan h och bestäm de tillhörande triangelarna.

Den inskrivna cirkeln. Om T är en godtycklig triangel så har den som bekant en inskriven cirkel. Medelpunkten för denna kan erhållas som skärningen för triangelns bisektriser. (En bisektris är en linje från ett hörn in i triangeln sådan att vinkeln till båda de bredvidliggande sidorna är lika stora. Rita figur.) Alla tre bisektriserna möts i en punkt, och denna punkt ligger lika långt från alla sidorna. De tre bisektriserna delar därför in T i tre deltrianglar med varsin sida som bas och den inskrivna cirkelns radie som höjd. Eftersom summan av dessa tre trianglars ytor är ytan av den ursprungliga så ger detta att

$$2|T| = r(a + b + c),$$

där $|T|$ är ytan, a , b och c är sidorna och r är den inskrivna cirkelns radie. Det är vanligt att summan $a + b + c$ kallas för $2p$, och då ger ovanstående formel att

$$r = \frac{|T|}{p}.$$

I en rätvinklig triangel (a, b, c) är $2|T| = ab$ och $2p = a + b + c$ så att

$$r = \frac{ab}{a + b + c}.$$

Förlänger vi detta bråk med *konjugatkvantiteten* $a + b - c$ får vi den vackra formeln

$$r = \frac{ab(a + b - c)}{(a + b)^2 - c^2} = \frac{ab(a + b - c)}{2ab} = \frac{a + b - c}{2}.$$

I en pythagoreisk triangel är $a+b-c$ alltid jämnt, så att den inskrivna cirkelns radie är alltid ett heltal. Om $z = a+bi$ är ett gaussiskt heltal i den första kvadranten så ska vi identifiera z med den rätvinkliga triangel som har hörnen i punkterna $0, a, a + bi$.

30. Bestäm den inskrivna cirkelns radie och medelpunkt för alla pythagoreiska trianglar med hypotenusan högst 29.

31. När du bestämde medelpunkterna för cirklarna i de trianglar, som kunde skrivas som jämna kvadrater $A + Bi = (a + bi)^2$ så upptäckte du säkert att medelpunkten var en heltalsmultipel av $(a + bi)$ (exempelvis var $2 + i$ medelpunkt i den inskrivna cirkeln i triangeln $3 + 4i$). Bevisa att detta alltid gäller för sådana trianglar, och bestäm n (som funktion av a och b) sådant att den inskrivna cirkelns medelpunkt i den triangel som ges av $(a + bi)^2$ är $n(a + bi)$. (Härvid förutsättes att såväl $a + bi$ som $(a + bi)^2$ ligger i första kvadranten, d.v.s. att $0 < b < a$.)

Som avslutning kan det vara värt att påpeka att det finns andra intressanta egenskaper hos pythagoreiska taltripplar än att hypotenusan är av en speciell form. De intressantaste egenskaperna finns hos de trianglar vars sidor inte innehåller någon gemensam delare större än 1. Sådana trianglar kallas primitiva och ges alltid av kvadrater i $\mathbf{Z}(i)$.

32. Visa att i en primitiv pythagoreisk triangel är hypotenusan udda, liksom en av kateterna, medan den andra kateten är jämn.

33. Försök att bevisa att i varje pythagoreisk triangel är en av kateterna delbar med 3 och att ytan är delbar med 6.

Litteratur

Carleson, L., *Matematik för vår tid*. Prisma, Stockholm 1968.

Hardy, G.H. & Wright, E.M., *An Introduction to the Theory of Numbers*. Fifth edition, Oxford Univ. Press, Oxford 1979.

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

Ogilvy, C.S. och Andersson, J.T., *Talteori för alla*. Prisma 1968.

Riesel, H., *En bok om primtal*. Studentlitteratur Lund, Odense 1968.

Boken är slutsåld från förlaget, men författaren har några exemplar kvar.

Gaussiska primtal

CHRISTER KISELMAN

Institut Mittag-Leffler & Uppsala universitet

1. Beskrivning av uppgiften. De förslag som presenteras här kan behandlas på flera 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 skriva en uppsats som går 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 eller uppmanar läsaren att visa. Ett mer avancerat projekt slutligen är att studera faktorisering i några andra ringar $\mathbf{Z}[\sqrt{d}]$ för heltal d ; här handlar det ju bara om $d = -1$.

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*. Dessa allmännare heltal är uppkallade efter Carl Friedrich Gauss (1777–1855).

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

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

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 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!).

3. 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 en dator 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 $1 < |c| \leq \sqrt{|z|}$, dvs. ändligt många. Varför?

Om din dator inte kan dividera komplexa tal så får du låta den 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

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

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

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 $0 \leq y \leq x$, ty talen $\pm z$, $\pm \bar{z}$, $\pm iz$, $\pm i\bar{z}$ är primtal samtidigt, och ett av dem ligger i oktanten $0 \leq y \leq x$. 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.)

4. 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 realdel $x \neq 0$ och imaginärdel $y \neq 0$, är då $|z|^2 = x^2 + y^2$ ett vanligt primtal? Försök visa att det är så! (Ledning: använd att $\mathbf{Z}[i]$ har unik faktorisering; en faktorisering av $z\bar{z} = ab$ leder till en faktorisering

$$z = \frac{a}{\bar{z}} \cdot b = a \cdot \frac{b}{\bar{z}}$$

av z , där a/\bar{z} eller b/\bar{z} är ett gaussiskt heltal.)

Vi kan dela in de vanliga positiva primtalen i tre klasser efter vilken rest de ger vid division med fyra: $5, 13, 17, \dots$ som ger resten 1 vid division med fyra; $3, 7, 11, 19, \dots$ 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 $\mathbf{Z}[i]$.

5. 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 $2 \leq p \leq x$, 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 $x \rightarrow +\infty$, dvs.

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

där $g(x) \rightarrow 0$ då $x \rightarrow +\infty$. Läs något mer om primtalssatsen i till exempel Carleson [1968], Hardy & Wright [1979], Newman [1980] eller Riesel [1968; 1985].

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 $x \geq 2$, $\pi_2(x) = 0$ 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 $|z| \leq r$ 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 $|z| \leq r$ är $4\pi_3(r)$ till antalet.

När vi räknar ihop alla gaussiska primtal z i cirkelskivan $|z| \leq r$ 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.

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

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 $|z| \leq r$ 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

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

Vi kan uttrycka dessa resultat så här: sannolikheten att ett vanligt heltal x med stort absolutbelopp skall vara ett primtal i \mathbf{Z} är $\sim 1/\log |x|$, medan sannolikheten att ett gaussiskt heltal z med stort absolutbelopp skall vara ett primtal i $\mathbf{Z}[i]$ är $\sim 2/\pi \log |z|$. (Sannolikheten i ett visst område $\{z; |z - a| \leq r\}$ blir ungefär lika med sannolikheten i hela skivan $\{z; |z| \leq |a|\}$ om $|a|$ är stort jämfört med r .)

Låt datorn räkna ut $\gamma(r)$ och $\pi(x)$ för några värden på r och x och se hur det stämmer.

En bättre approximation än $\pi(x) \sim x/\log x$ är den som Legendre fann 1808:

$$\pi(x) \sim \frac{x}{\log x - 1,08366}.$$

Man kan därför bestämma $a(r)$ så att

$$\gamma(r) = \frac{2r^2}{\log r - a(r)}.$$

Då bör $a(r)$ få ett värde som inte varierar så mycket.

Litteratur

Carleson, L., *Matematik för vår tid*. Prisma, Stockholm 1968.

Hardy, G. H. & Wright, E. M., *An introduction to the theory of numbers*. Fifth edition, Oxford Univ. Press, Oxford 1979.

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

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

Riesel, H., *En bok om primtal*. Studentlitteratur 1968. Odense 1968.

Boken är slutsåld från förlaget, men författaren har några exemplar kvar.

Riesel, H., *Prime Numbers and Computer Methods for Factorization*. Birkhäuser 1985.

Konvexitet i komplexa planet

BO KJELLBERG

KTH

Som vanligt får z beteckna en komplex variabel: $z = x + iy$, alternativt $z = re^{i\theta} = r \cos \theta + i \sin \theta$. Varje z motsvarar en punkt i det komplexa planet. Beloppet av z , dvs avståndet från z till origo, är $|z| = \sqrt{x^2 + y^2} = r$.

Polynom, t ex $Az + B$, $Az^2 + Bz + C$, där A, B, C är komplexa konstanter, är enkla exempel på så kallade *hela funktioner*, definierade och deriverbara i hela planet. Av intresse här är den så kallade maximummodulen $M(r)$, som för en funktion $f(z)$ är maximum av $|f(z)|$ på cirkeln $|z| = r$.

PROBLEM 1. Beräkna $M(r)$, då $f(z) = z - a$, $a > 0$.

LÖSNING. $|f(z)| \leq |z| + a = r + a$, detta största möjliga värde erhålles för $z = -r$, då är ju $f = -r - a$ och sålunda $M(r) = r + a$.

PROBLEM 2. Samma uppgift för $f(z) = (z - 1)(z - ae^{i\frac{2\pi}{3}})$, $a > 0$.

DISKUSSION. Maximum av $|z - 1|$ och $|z - ae^{i\frac{2\pi}{3}}|$ var för sig på $|z| = r$ fås lätt men eftersom maximumvärdena antas i olika punkter så erhåller man inte produktens maximum på detta sätt. Skriver man $z = re^{i\theta}$, räknar ut beloppen och logaritmerar, så fås

$$(1) \quad \ln |f(re^{i\theta})| = \frac{1}{2} \ln(r^2 + 1 - 2r \cos \theta) + \frac{1}{2} \ln(r^2 + a^2 - 2ar \cos(\theta - \frac{2\pi}{3})).$$

Om man vill ha maximum av detta då θ varierar så kan man studera derivatan som bli i maximumpunkten blir noll:

$$(2) \quad \frac{r \sin \theta}{r^2 + 1 - 2r \cos \theta} + \frac{ar \sin(\theta - \frac{2\pi}{3})}{r^2 + a^2 - 2ar \cos(\theta - \frac{2\pi}{3})} = 0.$$

Men den ekvationen är för svår att lösa, vi ger upp!

Vi får pröva ett annat sätt. Tag fram din räknare!

PROBLEM 3. Välj $a = 100$ och sök dig med räknarens hjälp fram till maximumvärdet $\ln M(r)$ av (1) på cirkeln $r = 10$. Vilket värde på θ svarar mot maximum?

Den franske matematikern Hadamard visade för hundra år sedan, att $\ln M(r)$ alltid är en konvex funktion av $\ln r$. Det är samma fantastiska Hadamard som Kiselman omtalar i sin artikel om primtal. Det var en stor upplevelse för mig att 1950 höra ett livfullt föredrag av Hadamard, då 85 år gammal.

Låt oss prova konvexitetssatsen på den enkla funktionen i problem 1!

Där har vi funnit, att $M(r) = r + a$, dvs, $\ln M(r) = \ln(r + a) = \ln(e^t + a)$, om man sätter $\ln r = t$. Två deriveringar med avseende på t ger

$$(3) \quad \frac{d^2 \ln M(r)}{dt^2} = \frac{ae^t}{(e^t + a)^2} = \frac{ar}{(r + a)^2}.$$

Det faktum att denna andra derivata alltid är positiv medför att $\ln M(r)$ är en konvex funktion av $t = \ln r$.

Walter Hayman, en mycket klipsk engelsman, observerade 1966 att man för varje hel funktion kan säga mer än att andraderivatan i (3) är ≥ 0 . Han visade, att det finns en konstant $H_0 > 0$ sådan att det alltid finns r -värden för vilka denna andraderivata är $\geq H_0$ eller åtminstone kommer H_0 hur nära som helst.

PROBLEM 4. Sök maximum av (3) då r varierar!

Jag gissar att du tämligen lätt får fram att uttrycket har ett maximum 0,25 för $r = a$. Det visar att H_0 inte kan överstiga 0,25. Hayman bevisade att $0,18 < H_0 \leq 0,25$ med gissningen $H_0 = 0,25$. Jag tyckte att saken var intressant och kunde konstruera ett exempel,

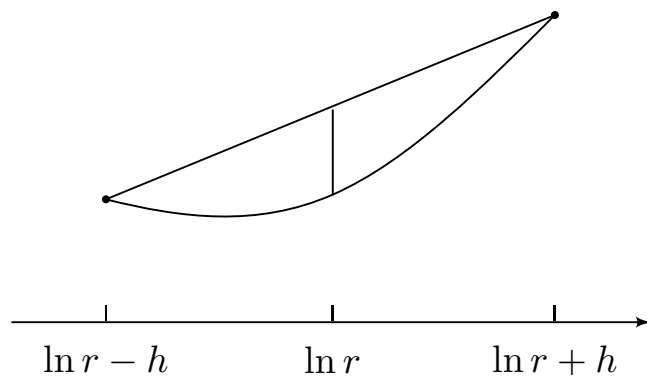
ett polynom som i problem 2, som visade att $H_0 < 0,25$. Därvid valdes a i intervallet $140 < a < 180$. Andraderivatan av $\ln M(r)$ med avseende på $\ln r$ visade sig ha tre lokala maxima, nära $r = 1$, $r = \sqrt{a}$ och $r = a$. Med mycket besvär kunde jag också studera hela mängden av hela funktioner och visa att $0,24 < H_0 < 0,25$.

Det är svårt att göra exakta beräkningar i det här sammanhanget, det visar diskussionen av problem 2. Man måste i varje fall vara van vid att räkna med partiella derivator. Det finns också funktioner, där andraderivatan inte existerar för vissa r -värden, vilket fordrar ett extra resonemang.

Det finns därför anledning att åtminstone vid en första undersökning införa ett enklare mått på konvexitet. Välj ett $k > 1$ och betrakta ett intervall $(\frac{r}{k}, kr)$, logaritmiskt $(\ln r - \ln k, \ln r + \ln k)$. Ju mer $\ln M(r)$ avlägsnar sig från kordan, dvs ju större d i figuren är, desto mer konvex är $\ln M(r)$. Som mått på konvexiteten väljes därför:

$$(4) \quad b(r, h) = \frac{2d}{h^2} = \frac{\ln M(\frac{r}{k}) + \ln M(kr) - 2 \ln M(r)}{h^2}$$

där för korthets skull $h = \ln k$.



Varför tages just faktorn $\frac{2}{h^2}$? Jo, den som är van vid derivator och

integraler finner att $b(r, h)$ är ett medelvärde över intervallet av den andraderivata, som här diskuteras, i varje fall om den är kontinuerlig.

Om man kan få fram resultat om $b(r, h)$ kan man förmodligen senare erhålla fakta om andraderivatan.

För enkelhets skull väljes här ett enhetligt värde $k = 1,01$, dvs $h = \ln 1,01$. Med detta val skrives i fortsättningen kort $b(r)$ i stället för $b(r, \ln 1,01)$.

I analogi med den tidigare definitionen av H_0 kan man definiera en konstant H_0^1 (troligen mycket nära H_0) med egenskapen att varje hel funktion har vissa värden $b(r)$ som är $\geq H_0^1$ eller godtyckligt nära H_0^1 om alla $b(r) < H_0^1$.

PROBLEM 5. Återgå till det enkla polynomet i problem 1 och visa, att $b(r)$ har ett maximum som är $< 0,25$, dvs $H_0^1 < 0,25$.

PROBLEM 6. Återgå till polynomet i problem 3, beräkna också $\ln M(\frac{10}{1,01})$ och $\ln M(1,01 \cdot 10)$ samt till slut $b(10)$!

PROBLEM 7. Betrakta $P_2(z) = (z - 1)(z - ae^{i\alpha})$, $a > 0$. Bevisa att $b(\frac{a}{r}) = b(r)$. Detta resultat minskar räknearbetet om man vill ha $b(r)$ för en följd r -värden.

För några år sedan hade jag kontakt med en ung engelsman, J R. Hilditch. Han blev road av detta och räknade ut en följd av $b(r)$ i ett antal fall. Det ledde honom till gissningen att $0,246 < H_0^1 < 0,247$, vilket kan förmodas leda till att $0,246 < H_0 < 0,247$. Det vore trevligt att kunna bevisa i första hand olikheten för H_0^1 .

Det som är intressant här är ju att hitta funktioner med minsta möjliga konvexitet, dvs så små värden på $b(r)$ som möjligt. Som visas i [2] måste då nollställena $z_1, z_2, \dots, z_n, \dots$ ligga glest, kvoten $|\frac{z_{n+1}}{z_n}|$ måste vara > 25 , säkert större om man gör sig besvär att undersöka saken.

Avlägsna nollställen tycks inte påverka $b(r)$ så mycket, varför studiet av enkla polynom är viktigt.

PROBLEM 8. Försök att välja a och α i problem 7 så att du känner dig övertygad om att $b(r)$ har ett maximum $< 0,247$, vilket innebär att $H_0^1 < 0,247$.

PROBLEM 9. Tag nu i stället ett polynom av grad 3, t ex $P_3(z) = (z - 1)(z - ae^{i\alpha})(z - a^2e^{2i\alpha})$ och försök att genom lämpligt val av a och α pressa ned $b(r)$:s maximum ytterligare!

PROBLEM 10. Möjligen är ett polynom som $P_3(z)$ det extremala, dvs man når ned till det ideala H_0^1 (och H_0 också, gissar jag). Är det så att man genom att tillfoga nollställena kan minska $b(r)$:s maximum, så är ju gissningen fel. Vill någon göra några numeriska experiment och se om de pekar i någon riktning?

Litteratur

[1] Hayman, W.K., Note on Hadamards convexity theorem. *Entire functions and related parts of analysis*. Proc. Sympos. Pure Math., La Jolla, Calif., 1966, s 210-213. Amer. Math. Soc., Providence, R.I., 1968.

[2] Kjellberg, B., The convexity theorem of Hadamard-Hayman. *Proc. Roy. Inst. of Techn.* June 1973, s 87-114.

Kan erhållas genom hänvändelse till förf., Matem. inst., KTH, 100 44 Stockholm.

Genererande funktioner

GÖRAN KJELLBERG

Inledning. Frågar man på hur många olika sätt kan man ... (göra si eller så) ... med n objekt så vill man ha ett svar för varje värde av heltalet n , dvs att lösningen skall bestå av en följd av tal $a_0, a_1, a_2, \dots, a_n, \dots$. Sådana problem kallas kombinatoriska. De förekommer t ex när man vill räkna ut sannolikheter, och kan vara svårlösta.

Genererande funktionen för en talföljd innehåller följdens egenskaper i kompakt form, och dessa funktioner blir därigenom ett verkamt hjälpmedel att finna lösningar till kombinatoriska problem, men inte bara sådana utan också t ex till differensekvationer. De är dessutom högst intressanta i sig själva, eftersom de är ett speciellt fall av det vidare begreppet transform, som spelar stor roll i många av matematikens grenar.

Definition och exempel på genererande funktioner. Den oändliga geometriska serien

$$(1) \quad 1 + x + x^2 + \dots$$

har summan

$$(2) \quad \frac{1}{1-x}$$

när $|x| < 1$.

Man säger också att (1) är potensserieutvecklingen av funktionen (2) i intervallet $|x| < 1$. En potensserie i allmänhet ser ut så här:

$$(3) \quad a_0 + a_1x + a_2x^2 + \dots,$$

där a_i är tal. Man säger att en given funktion $f(x)$ har potensserieutvecklingen (3) omkring noll om vi har att

$$f(x) = a_0 + a_1x + a_2x^2 + \dots$$

för alla x i något intervall $|x| < \varepsilon$ omkring noll.

UPPGIFT 1. Visa att en funktion $f(x)$ endast har *en* potensserieutveckling. Med andra ord, om

$$\begin{aligned} f(x) &= a_0 + a_1x + a_2x^2 + \dots \\ &= b_0 + b_1x + b_2x^2 + \dots \end{aligned}$$

omkring noll, så är $a_i = b_i$ för alla i .

Funktionen (2) och serien (1) bestämmer därför varandra entydigt. Mer generellt kommer funktionen $f(x)$ härövan och serien (3) att bestämma varandra entydigt. Speciellt bestämmer funktionen $f(x)$ koefficienterna a_0, a_1, a_2, \dots och man säger därför också att funktionen $f(x)$ är *genererande* funktion för talföljden a_0, a_1, a_2, \dots . Funktionen (2) är alltså genererande funktion för talföljden $1, 1, 1, \dots$. Dessa tal är koefficienterna i serien (1).

Ett annat exempel på en genererande funktion ger binomialteoremet:

$$(1+x)^n = 1 + \binom{n}{1}x + \binom{n}{2}x^2 + \dots + \binom{n}{n}x^n$$

om n är ett positivt heltal. Det vänstra ledet $f(x) = (1+x)^n$ är genererande funktion för binomialkoefficienterna $1, \binom{n}{1}, \binom{n}{2}, \dots$, där

$$\binom{n}{k} = \frac{n(n-1)(n-2)\dots(n-k+1)}{1 \cdot 2 \cdot 3 \dots k} \quad \text{med} \quad \binom{n}{0} = 1.$$

Som du förmodligen vet, är binomialkoefficienten $\binom{n}{k}$ också lösningen till det kombinatoriska problemet: på hur många sätt kan man välja ut k föremål ur en samling av n föremål?

UPPGIFT 2. Vilka är de genererande funktionerna för talföljderna

$$1, \quad \frac{1}{2!}, \quad \frac{1}{3!}, \quad \frac{1}{4!}, \quad \dots$$

$$1, \quad \frac{1}{3!}, \quad \frac{1}{5!}, \quad \dots$$

$$1, \quad -\frac{1}{2}, \quad \frac{1}{3}, \quad -\frac{1}{4}, \quad \dots ?$$

UPPGIFT 3. Ge exempel på genererande funktioner för andra kända talföljder.

Vad kan genererande funktioner användas till? Manipuleringar med funktionsuttrycket kan leda till samband som gäller för talföljden som uttrycket genererar. Ett välkänt exempel är

$$(1+x)^{n+1} = (1+x)^n(1+x).$$

För potensserierna får vi motsvarande:

$$\begin{aligned} 1 + \binom{n+1}{1}x + \binom{n+1}{2}x^2 + \dots + \binom{n+1}{n+1}x^{n+1} \\ &= (1 + \binom{n}{1}x + \binom{n}{2}x^2 + \dots + \binom{n}{n}x^n)(1+x) \\ &= 1 + (1 + \binom{n}{1})x + (\binom{n}{1} + \binom{n}{2})x^2 \\ &\quad + \dots + (\binom{n}{n-1} + \binom{n}{n})x^n + \binom{n}{n}x^{n+1}. \end{aligned}$$

Av uppgift 1 ovan följer att koefficienterna för varje särskild potens av x måste vara lika. Alltså får man den välkända lagen:

$$\binom{n}{k-1} + \binom{n}{k} = \binom{n+1}{k}, \quad k = 1, 2, \dots, n.$$

Identifiera koefficienterna för var x -potens i de båda leden. Vi får då:

$$\begin{aligned} 1 &= a_0 \\ 0 &= a_1 - a_0 \quad \text{som ger } a_1 = 1 \\ 0 &= a_2 - a_1 - a_0 \quad \text{som ger } a_2 = 2 \\ 0 &= a_3 - a_2 - a_1 \quad \text{som ger } a_3 = 3 \\ 0 &= a_4 - a_3 - a_2 \quad \text{som ger } a_4 = 5 \\ &\vdots \end{aligned}$$

Funktionen $\frac{1}{1-x-x^2}$ är alltså genererande funktion för sekvensen $1, 2, 3, 5, \dots$.

De tal som räknas fram på detta sätt brukar kallas Fibonacci-talen och betecknas med F_n .

UPPGIFT 6. Visa att Fibonacci-talen satisfierar *rekursionsformeln* $F_{n+1} = F_n + F_{n-1}$, med $F_0 = F_1 = 1$.

Leonardo Fibonacci var en framstående matematiker i Italien omkring år 1200. Han introducerade talen F_n som lösningar till *kaninproblemet*: hur många par kaniner finns det efter n månader, om de fortplantar sig enligt följande? Vid tiden 0 finns ett kaninpar. Efter två månader får de sina första barn, två ungar av olika kön, och fortsätter att föda ett par ungar varje månad. De nya kaninerna bildar par med sin tvilling och får också ett par ungar varje månad, men börjar inte med detta förrän de uppnått åldern två månader.

UPPGIFT 7. Visa att Fibonacci-talen kan tolkas på detta sätt.

Genom att omforma uttrycket (4) kan man härleda en explicit formel för det n :te talet F_n på följande sätt:

Om rötterna till ekvationen $1 - x - x^2 = 0$ kallas r_1 och r_2 , så gäller $1 - x - x^2 = (r_1 - x)(r_2 - x)$ och (4) kan delas upp i partialbråk:

$$\frac{1}{1 - x - x^2} = \frac{a}{r_1 - x} + \frac{b}{r_2 - x} = \frac{a}{r_1} \frac{1}{1 - x/r_1} + \frac{b}{r_2} \frac{1}{1 - x/r_2}.$$

De två sista bråken kan utvecklas i geometrisk serie och man får:

$$\frac{1}{1 - x - x^2} = \left(\frac{a}{r_1} + \frac{b}{r_2} \right) + \left(\frac{a}{r_1^2} + \frac{b}{r_2^2} \right) x + \left(\frac{a}{r_1^3} + \frac{b}{r_2^3} \right) x^2 + \dots,$$

varav $F_n = \frac{a}{r_1^{n+1}} + \frac{b}{r_2^{n+1}}$.

UPPGIFT 8. Utför detta, dvs beräkna r_1, r_2, a och b ! (Är det inte underligt!)

UPPGIFT 9. Skriv datorprogram som bestämmer F_n både med hjälp av rekursionsformeln och av formeln ovan. Vilken av formlerna lämpar sig bäst för datorberäkningar?

UPPGIFT 10. Hur ser den genererande funktionen ut för en talföljd G_n som lyder samma rekursionsformel som F_n , men startar med talen $G_0 = a, G_1 = b$, där a och b är godtyckliga reella tal?

Vilken talföljd genereras av funktionen $\frac{1}{1 + x + x^2}$?

UPPGIFT 11. Anta att vi har en potensseriutveckling

$$\frac{1}{a + bx + cx^2} = a_0 + a_1x + a_2x^2 + \dots$$

för funktionen $f(x) = \frac{1}{a + bx + cx^2}$. Visa att talen a_0, a_1, a_2, \dots satisfierar en rekursionsformel

$$a_{n+1} = Aa_n + Ba_{n-1}$$

liknande den för Fibonaccitalen. Uttryck A och B med hjälp av a, b och c .

Finn en explicit formel för a_n med hjälp av rötterna r_1 och r_2 till ekvationen $cx^2 + bx + a = 0$ och a, b och c liknande den du fann ovan för F_n .

UPPGIFT 12. Formulera och lös en uppgift liknande den ovan för funktionen

$$\frac{1}{a_0 + a_1x + a_2x^2 + \cdots + a_nx^n} .$$

UPPGIFT 13. På hur många sätt kan man växla n kronor i enkronor och femkronor? Visa att lösningen ges av den genererande funktionen

$$\frac{1}{1-x} \cdot \frac{1}{1-x^5} !$$

På hur många sätt kan man växla n kronor i enkronor, femmor och tior?

Litteratur

Riordan, J., *An introduction to Combinatorial Analysis*. Wiley 1958.

Något om differenser

GÖRAN KJELLBERG

Differenser har i hundratals år varit ett viktigt hjälpmedel när man framställde eller använde tabeller över funktioner. Efter datorernas tillkomst har denna tillämpning minskat i betydelse, men differenskalkylen utgör ändå en intressant och lättillgänglig teori, rik på anknytningar till andra delar av matematiken, t ex kombinatoriken och differentialkalkylen, och till stor nytta vid numerisk lösning av ordinära och partiella differentialekvationer.

Definition. Antag att vi har en funktion $f(x)$ av en reell variabel x , och ett fixt reellt tal h . Differensen av f är en ny funktion Δf som definieras av

$$\Delta f(x) = f(x + h) - f(x).$$

Man kan bilda differensen av differensen; den kallas 2:a-differensen och betecknas $\Delta^2 f$.

$$\begin{aligned} \Delta^2 f(x) &= \Delta f(x + h) - \Delta f(x) \\ &= f(x + 2h) - f(x + h) - f(x + h) + f(x) \\ &= f(x + 2h) - 2f(x + h) + f(x). \end{aligned}$$

På analogt sätt bildar man 3:e-differensen, 4:e-differensen, osv. För tydlighetens skull säger man också ibland 1:a-differensen av f . Behöver man markera beroendet av h skriver man Δ_h, Δ_h^2 i stället för Δ, Δ^2, \dots . I det följande sätter vi oftast $h = 1$. Det allmänna fallet återföres lätt på detta med en variabeltransformation. (Se uppgift 10 nedan.)

Symbolen Δ ensam representerar en *operator*, dvs något som verkar på en funktion och som resultat ger en annan funktion. En annan välkänd operator är deriveringsoperatoren $D = \frac{d}{dx}$.

UPPGIFT 1. Skriv ner och hyfsa uttrycken för $\Delta^3 f$ och $\Delta^4 f$, analogt med $\Delta^2 f$ ovan. Härled ett uttryck för $\Delta^n f$!

Polynom och differenser. Betrakta $p(x) = x^2 - x + 41$ samt följande tabell där $\Delta = \Delta_1$

x	$p(x)$	$\Delta p(x)$	$\Delta^2 p(x)$
0	41		
		0	
1	41		2
		2	
2	43		2
		4	
3	47		2
		6	
4	53		2
		8	
5	61		2
		10	
6	71		

Aha - 2:a-differensen är konstant. Inte så konstigt förstås; om p är av grad 2 måste Δp ha grad 1 och $\Delta^2 p$ ha grad 0. (Se uppgift 4 nedan.)

Tabellen ovan kallas ett *differensschema*. Där har var term en betydelse som beror på dess läge i förhållande till x :s aktuella värde.

Principen kan åskådliggöras så här:

$x - 2h$	$f(x - 2h)$	$\Delta f(x - 2h)$	$\Delta^2 f(x - 2h)$	$\Delta^3 f(x - 2h)$
$x - h$	$f(x - h)$	$\Delta f(x - h)$	$\Delta^2 f(x - h)$	$\Delta^3 f(x - h)$
x	$f(x)$	$\Delta f(x)$	$\Delta^2 f(x)$	$\Delta^3 f(x)$
$x + h$	$f(x + h)$	$\Delta f(x + h)$	$\Delta^2 f(x + h)$	$\Delta^3 f(x + h)$
$x + 2h$	$f(x + 2h)$	$\Delta f(x + 2h)$	$\Delta^2 f(x + 2h)$	$\Delta^3 f(x + 2h)$

De termer som har samma argument står på en rad som sluttar ned åt höger och summan av en term och den nedanför till höger är lika med termen rakt under den övre termen.

UPPGIFT 2. Visa att Δ är en *linjär operator*, dvs att

$$\Delta(f + g) = \Delta f + \Delta g, \quad \text{och}$$

$$\Delta(af) = a\Delta f, \quad \text{där } a \text{ är en reell konstant.}$$

UPPGIFT 3. Vad är n :e-differensen av x^n ?

UPPGIFT 4. Visa att n :e-differensen av ett n :e-gradspolynom är konstant. Ge gärna flera olika bevis.

Satsen i uppgift 4 gör att vi bara behöver känna till en uppåt-lutande rad för att beräkna differensschemat under denna rad. Betrakta till exempel de sista raderna i differensschemat för $p(x) = x^2 - x + 41$:

x	p	Δp	$\Delta^2 p$
4	53		
		8	
5	61		2
		10	
6	71		

Vad är $p(7)$? Eftersom 2:a-differensen är 2 måste nästa 1:a-differens vara 12 och nästa p -värde 83. Det går alltså att beräkna polynomets fortsatta värden med enbart additioner.

UPPGIFT 5. Gör ett dataprogram som beräknar $p(8), p(9), \dots$ för polynomet $x^2 - x + 41$ ur differensschemat ända till dess det för första gången har beräknat ett p -värde som inte är ett primtal. För vilket x -värde inträffar det? (Obs - att det här polynomet ger en ovanligt lång svit av primtal bör nog betraktas som en kuriositet; i alla fall har det inget samband med differenskalkylen).

UPPGIFT 6. Ett tredjegradspolynom $p(x)$ har värdena $p(2) = -4$, $p(3) = 0$, $p(5) = 7$, $p(6) = -8$. Bestäm $p(4)$!

UPPGIFT 7. Binomialkoefficienten $\binom{n}{k}$ kan ses som en funktion av n , när man håller k fixt. Vad är denna funktions första-differens?

Olika sätt att skriva polynom. Ett n :e-gradspolynom är entydigt bestämt av sina $n + 1$ koefficienter. Av det föregående ses att det också är entydigt bestämt av ett differensschema, som är utbyggt till och med n :e-differensen, dvs antingen av ett funktionsvärde plus n lämpligt valda differenser eller av de $n + 1$ funktionsvärden som är differensschemats bas.

Med kännedom om koefficienterna kan man räkna ut funktionsvärden och differenser, men hur kommer man från differenserna till koefficienterna?

Ett sätt är att använda *faktorialer*, $x^{(k)}$, som definieras:

$$x^{(k)} = x(x-1)(x-2)\dots(x-k+1), \quad x^{(0)} = 1.$$

Dessa funktioner har en viss analogi med potenser.

UPPGIFT 8. Visa att $\Delta x^{(k)} = kx^{(k-1)}$ ($k = 1, 2, \dots$).

Faktorialerna passar alltså ihop med operatoren Δ på samma sätt som potenserna med deriveringsoperatoren D .

Gör ansatsen

$$p(x) = \sum_{k=0}^n b_k x^{(k)}.$$

Sätt $x = 0$: vi får $b_0 = p(0)$. Bilda

$$\Delta p(x) = \sum_{k=1}^n k b_k x^{(k-1)}; \text{ sätt } x = 0; \text{ Vi får } b_1 = \Delta p(0).$$

Bilda

$$\Delta^2 p(x) = \sum_{k=2}^n k(k-1)x^{(k-2)}; \text{ sätt } x = 0; \text{ vi får } 2b_2 = \Delta^2 p(0)$$

o s v.

Man inser att $b_k = \frac{\Delta^k p(0)}{k!}$ för $k = 1, 2, \dots, n$, där $k! = 1 \cdot 2 \cdot k$. Alltså är

$$(*) \quad p(x) = \sum_{k=0}^n \frac{\Delta^k p(0)}{k!} x^{(k)}.$$

Denna formel uttrycker $p(x)$ med hjälp av värdet $p(0)$ samt de differenser som står på en från $p(0)$ nedåt sluttande rad. Den kallas Newtons interpolationsformel, närmare bestämt *Newtons framåtformel*. Det finns nämligen också en Newtons bakåtformel (se uppgift 11) som uttrycker $p(x)$ med hjälp av $p(0)$ och de differenser som står på en från $p(0)$ uppåt sluttande rad.

Man kommer vidare till den vanliga framställningen av $p(x)$ genom att skriva faktorialerna $x^{(n)}$ som ett polynom:

$$x^{(n)} = \sum_{k=1}^n s(n, k) x^k \text{ i } 1, x, \dots, x^n.$$

Här är $s(n, k)$ hela tal. För små värden av n och k är de givna i följande tabell.

Tabell över $s(n, k)$

$n \backslash k$	1	2	3	4	5	...
1	1					
2	-1	1				
3	2	-3	1			
4	-6	11	-6	1		
5	24	-50	35	-10	1	
6					

Dessa tal kallas Stirlings tal av 1:a slaget. (Stirlings tal av 2:a slaget ger x^n som funktioner av $x^{(k)}$.)

UPPGIFT 9. Givet polynomet

$$p(x) = -1 + x^{(1)} + 10x^{(2)} + 20x^{(3)} + 9x^{(4)} + x^{(5)}.$$

Skriv $p(x)$ utvecklat efter potenser, dvs i *vanlig form*.

UPPGIFT 10. Antag att man har värden av ett n :e-gradspolynom $g(y)$ för $y = y_0, y_0 + h, \dots, y_0 + nh$ samt differensschemat

$$\begin{array}{r}
 g(y_0) \\
 \Delta g(y_0) \\
 g(y_0 + h) \\
 \Delta g(y_0 + h) \\
 \dots \dots \dots \Delta^n g(y_0) \\
 \dots \dots \dots \\
 \Delta g(y_0 + (n-1)h) \\
 g(y_0 + nh)
 \end{array}$$

baserat på dessa värden. (I detta schema står Δ för Δ_h).

Uttryck $g(y)$ med hjälp av Newton's interpolationsformel (*) och de givna differenserna! (Tips: sätt $y = y_0 + xh$ och betrakta $p(x) = g(y_0 + xh)$ och differensen $\Delta_h g(y_0) = \Delta p(0)$ osv.)

UPPGIFT 11. Man kan definiera *bakåtdifferenser* $\nabla f(x) = f(x) - f(x - h)$ och få en teori som är en spegelbild av den man får med Δ . Även till ∇ finns en familj av *basfunktioner* som betecknas $x^{[n]}$, och för vilka $\nabla x^{[k]} = kx^{[k-1]}$, $k = 1, 2, \dots$. Ge uttryck för $x^{[k]}$ och använd dem för att ställa upp Newtons bakåtformel.

Differenser och funktioner som inte är polynom.

UPPGIFT 12. Det finns en långtgående parallellitet mellan egenskaperna hos operatorerna Δ och $D = \frac{d}{dx}$. Summation motsvarar integration, och många formler liknar varandra, t ex de för derivatan av en produkt och differensen av en produkt. Operatorn D har en *fixpunkt*, dvs en funktion som inte ändras av D : $Df = f$, nämligen $f = e^x$. Har operatorn Δ också en fixpunkt, en funktion g sådan att $\Delta g = g$, och i så fall vilken?

Många funktioner ger, om de tabuleras tillräckligt tätt, dvs med litet tabellsteg h , differensscheman, som liknar polynomens. Differenserna för positiva funktioner blir mindre utåt höger till och med en viss ordning m . Om de är tillräckligt små, indikerar detta att funktionen kan approximeras bra i detta intervall med ett $(m - 1)$:a-gradspolynom. Detta är bakgrunden till användningen av differenser i samband med funktionstabeller och interpolation.

Differenser som heuristiskt hjälpmedel (*heuristiskt* betär man sig då man gissar och provar sig fram). Om man söker efter en okänd funktion och har möjlighet att räkna ut värden av den kan det vara en hjälp att ställa upp ett differensschema. Om den okända funktionen är ett polynom, så syns det, och polynomet kan också bestämmas. Om differenserna i stället betär sig ungefär som funktionsvärdena själva, så innehåller den okända funktionen förmodligen någon exponentiell term. Ett exempel: Hur många diagonaler finns det i en n -hörning? Vi ställer upp en liten tabell; ritar 5- och 6-hörningar

och räknar:

Antal hörn	Antal diagonaler		
3	0		
		2	
4	2		1
		3	
5	5		1
		4	
6	9		

2:a-differenserna verkar konstanta. Använd Newtons formel, utgående från $n = 3$: Antal diagonaler $= 0 + 2(n - 3) + \frac{(n-3)(n-4)}{2} = \frac{n(n-3)}{2}$.

Återstår att visa att detta gäller för alla n , vilket i detta fall inte är svårt.

UPPGIFT 13. Betrakta en konvex n -hörning. Hur många skärningspunkter mellan diagonaler finns det inom n -hörningen? (Vi antar att aldrig mer än två diagonaler möts i en punkt.) Ledning: antalet $= 0$ för $n = 1, 2, 3$. Rita och räkna antalen för $n = 4, 5$ och 6 . (Denna uppgift är hämtad ur tidskriften Normats (Nordisk matematisk tidskrift) problemspalt. En smartare lösning ges i häfte 4, årgång 1988, problem 149. Men differensschemat kan kanske leda en in på rätt spår.)

Den här tekniken är användbar när man vill veta hur många multiplikationer och/eller additioner som krävs i en algoritm för olika ordning på problemet, t ex lösning av ett linjärt ekvationssystem med n obekanta.

UPPGIFT 14. Hur många positiva heltalslösningar har ekvationen $x + y + z = 2n$, som också uppfyller villkoren $x \leq y + z$; $y \leq z + x$; $z \leq x + y$. (Hämtat ur Polya-Szegö, Aufgaben und Lehrsätze, Abschn. I, kap. 1, problem 31.)

UPPGIFT 15. I hur många delar delas planet av n räta linjer? Vi förutsätter att inga linjer är parallella och att aldrig mer än två linjer möts i en punkt.

Litteratur

Läroböcker i numerisk analys brukar innehålla avsnitt om differenser.

Fröberg, C.-E., *Lärobok i numerisk analys*. Svenska bokförlaget 1962,

har en mycket fyllig framställning.

Om Möbiustransformationer

TORBJÖRN KOLSRUD

KTH

En *Möbiustransformation* är en komplexvärd funktion f av en komplex variabel z på formen

$$f(z) = \frac{az + b}{cz + d}.$$

Här är a, b, c och d komplexa tal. Ofta skriver vi bara

$$z \longrightarrow \frac{az + b}{cz + d}.$$

Observera att om talen a, b, c och d multipliceras med ett godtyckligt komplext tal $k \neq 0$ får vi ändå samma funktion. Det betyder att en Möbiustransformation bara beror av tre parametrar.

Specialfallen $z \rightarrow az$ (en rotation om $|a| = 1$, en *homoteti* om $a > 0$), $z \rightarrow z + b$ (translation) och $z \rightarrow 1/z$ (inversion) kallas nedan för *elementära transformationer*.

1. Beräkna värdet i punkterna $z = \frac{1}{2}$, $z = i$, $z = -2i$ och $z = 1 + i$ av Möbiustransformationerna $z \rightarrow iz$, $z \rightarrow 2z$, $z \rightarrow z + 1 + i$ och $z \rightarrow 1/z$. Rita också så att du ser vad som sker under de elementära transformationerna.

Om f och g är två funktioner kan man bilda deras sammansättning $f \circ g$, definierad genom $f \circ g(z) = f(g(z))$. I allmänhet (se övning 2 nedan) är $f \circ g$ och $g \circ f$ olika funktioner.

2. Bestäm alla sammansättningar av funktionerna $z \rightarrow iz$, $z \rightarrow z + 1 + i$, $z \rightarrow 1/z$. Du ser då att också sammansättningarna är Möbiustransformationer.

3. Visa allmänt att sammansättningen av två Möbiustransformationer, säg

$$z \longrightarrow \frac{a_1 z + b_1}{c_1 z + d_1} = w$$

och

$$w \longrightarrow \frac{a_2 w + b_2}{c_2 w + d_2},$$

är en Möbiustransformation.

Vi är inte intresserade i de triviala fall då f är konstant, dvs då för något komplext tal α , $f(z) = \alpha$ oberoende av hur z väljs. (t ex $a = b = 1$, $c = d = 2$.)

4. Visa att f är konstant precis då $ad - bc = 0$.

Från och med nu antas alltid att $ad - bc \neq 0$.

Under detta villkor kan man lösa ut z ur ekvationen

$$w = \frac{az + b}{cz + d}.$$

Vi får då vad som kallas den *inversa funktionen*. Om denna betecknas med g , gäller att $f(g(w)) = w$ för alla w och $g(f(z)) = z$ för alla z . Om det finns en invers funktion säger vi också att f är *omvändbar*. Varje Möbiustransformation är alltså omvändbar (om $ad - bc \neq 0$).

5. Kontrollera att i fallen $w = iz$, $w = z + 1 + i$ och $w = 1/z$ är också den inversa funktionen en Möbiustransformation.

6. Visa sedan detta i det allmänna fallet, helt enkelt genom att härleda en formel för z som funktion av w .

Om $c = 0$ är det klart att $|w| \rightarrow \infty$ då $|z| \rightarrow \infty$, dvs w växer obegränsat med z .

7. Visa att om $c \neq 0$ gäller att $|w| \rightarrow \infty$ då $z \rightarrow -d/c$. (Det gör du enklast genom att använda uttrycket för den inversa funktionen som härletts i övning 6.) Beräkna också gränsvärdet då $|z| \rightarrow \infty$.

Låt oss införa beteckningen \mathbf{C} för de komplexa talen. Med $\hat{\mathbf{C}}$ betecknar vi \mathbf{C} plus punkten ∞ . Varje Möbiustransformation kan då ses som en omvändbar funktion från $\hat{\mathbf{C}}$ till sig själv. Vi definierar $f(\infty) = \infty$ om $c = 0$. För $c \neq 0$ låter vi $f(\infty) = a/c$ och $f(-d/c) = \infty$. I exemplet $f(z) = 1/z$ är således $f(0) = \infty$ och $f(\infty) = 0$.

8. Skriv Möbiustransformationerna $z \rightarrow (z + 1)/z$ och $z \rightarrow (z + 2i)/(z + 1)$ som sammansättningar av elementära transformationer.

9. Visa allmännare att varje Möbiustransformation kan uttryckas som sammansättning av elementära transformationer.

En av avsikterna med detta arbete är att studera Möbiustransformationers verkan på cirklar och linjer. Låt oss därför behandla ekvationerna för dessa geometriska figurer. Eftersom en cirkel består av alla punkter vars avstånd till en fix punkt z_0 är $r > 0$, är det klart att cirkeln med centrum i z_0 och radie r kan beskrivas genom ekvationen

$$|z - z_0| = r,$$

dvs

$$(z - z_0)(\bar{z} - \bar{z}_0) = r^2.$$

(På reell form blir detta, om $z = x + iy$ och $z_0 = x_0 + iy_0$, $(x - x_0)^2 + (y - y_0)^2 = r^2$.)

Varje linje i planet kan skrivas på formen $Ax + By + C = 0$, där A, B och C är reella tal, och där något av A och B är $\neq 0$.

10. Visa att ekvationen för en linje kan uttryckas med komplexa tal som

$$\alpha z + \bar{\alpha} \bar{z} + \beta = 0,$$

där $\alpha \neq 0$ är ett komplext och β ett reellt tal.

11. Bestäm bilden av linjen $x - y + 1 = 0$ under transformationerna $z \rightarrow z + 1$, $z \rightarrow 1/z$ och $z \rightarrow (z + 1)/z$.

12. Låt allmänare L vara en cirkel eller en linje och låt f vara en Möbiustransformation. Visa med hjälp av resultaten ovan att bilden av L under f , $f(L) = \{w : w = f(z), z \in L\}$, också är en cirkel eller en linje.

Vi har tidigare sett att en Möbiustransformation bara beror på tre parametrar. Det förefaller därför klart att varje Möbiustransformation är bestämd av att det för tre olika punkter z_1, z_2 och z_3 gäller att $z_1 \rightarrow 0, z_2 \rightarrow 1, z_3 \rightarrow \infty$. Vi ska nu visa detta.

13. a) Kontrollera att om ingen av z_1, z_2, z_3 är ∞ , så gäller detta för funktionen

$$f(z) = \frac{z - z_1}{z - z_3} \cdot \frac{z_2 - z_3}{z_2 - z_1},$$

och, om z_1, z_2 eller z_3 är ∞ , för

$$\frac{z_2 - z_3}{z - z_3}, \quad \frac{z - z_1}{z - z_3} \quad \text{resp.} \quad \frac{z - z_1}{z_2 - z_1},$$

att $z_1 \rightarrow 0, z_2 \rightarrow 1, z_3 \rightarrow \infty$.

b) Låt f vara som i a) och anta att också g uppfyller $z_1 \rightarrow 0, z_2 \rightarrow 1, z_3 \rightarrow \infty$. Om h är den inversa funktionen till g , så gäller för $f \circ h$ att $0 \rightarrow 0, 1 \rightarrow 1$ och $\infty \rightarrow \infty$. Vilka Möbiustransformationer uppfyller detta? Vilken slutsats dras om g ?

Funktionen $f(z)$ i 13a) skrivs ofta (z, z_1, z_2, z_3) och kallas för korskvoten. Denna kan användas för att hitta en Möbiustransformation som avbildar tre givna distinkta punkter z_1, z_2, z_3 på tre andra distinkta punkter w_1, w_2, w_3 . Om $f(z) = (z, z_1, z_2, z_3)$ och g är inversen till $w \rightarrow (w, w_1, w_2, w_3)$ gäller för $g \circ f$ att $z_1 \rightarrow 0 \rightarrow w_1, z_2 \rightarrow 1 \rightarrow w_2$ och $z_3 \rightarrow \infty \rightarrow w_3$.

Vitsen med detta är att man, givet två områden, båda begränsade av en cirkel eller en linje, kan finna en Möbiustransformation som överför det ena området i det andra. Anta t ex att det gäller två halvplan, begränsade av linjerna L resp M . Tag två punkter z_1, z_2

på L och anta att en Möbiustransformation f avbildar z_1 på w_1 och z_2 på w_2 där $w_1, w_2 \in M$. Linjerna L och M delar z - resp w -planet i två delar. Av kontinuitetsskäl är det klart att alla punkter i "ena halvan" av z -planet hamnar i samma halva av w -planet. Det räcker därför att välja punkter z_3 och w_3 i de ursprungliga halvplanen och ordna så att också $z_3 \rightarrow w_3$.

14. Finn en Möbiustransformation som avbildar

- a) halvplanet $\{y < -1\}$ på halvplanet $\{x - y > 1\}$,
- b) cirkelskivan $\{|z| < 1\}$ på cirkelskivan $\{|z - 1 + i| < 3\}$,
- c) cirkelskivan $\{|z| < 1\}$ på området $\{|z - 1 + i| > 3\}$ (en "yttre cirkelskiva"),
- d) cirkelskivan $\{|z| < 1\}$ på halvplanet $\{y > 0\}$.

15. a) Låt L vara reella axeln. Kolla att $z \rightarrow (az + b)/(cz + d)$ avbildar L på L precis när a, b, c och d är reella, om vi också antar att $ac - bd$ är reell.

b) Låt H vara övre halvplanet $\{y > 0\}$. I vilket av fallen $ac - bd > 0$, resp < 0 , gäller att H avbildas på sig självt?

Betrakta alla Möbiustransformationer

$$f(z) = \frac{az + b}{cz + d}$$

där a, b, c, d är *heltal* ($0, \pm 1, \pm 2, \dots$) och $ad - bc = 1$. Vi skriver då $f \in \Gamma$.

16. Kontrollera att om f och $g \in \Gamma$ gäller också $f \circ g \in \Gamma$. Visa också att varje $f \in \Gamma$ har en invers funktion som tillhör Γ . (Detta betyder att Γ är en *grupp*.)

Två transformationer i Γ är

$$f_1 : z \rightarrow z + 1$$

med invers funktion

$$g_1 : z \rightarrow z - 1,$$

och

$$f_2 : z \rightarrow -1/z$$

som är sin egen invers. Man kan visa att varje $f \in \Gamma$ kan fås genom sammansättningar av f_1, g_1 och f_2 . Vi säger att de *genererar* Γ .

17. Visa att

$$\text{a) } g(z) = \frac{-1}{z+1} = f_2 \circ f_1(z),$$

$$\text{b) } -(1 + 1/z) = g \circ g(z),$$

$$\text{c) } \frac{z}{z+1} = f_1 \circ f_2 \circ f_3(z).$$

Låt F vara ett område i H . F kallas för ett *fundamentalområde* (för Γ) om varje punkt w i H är bild av en punkt z i F under en lämplig sammansättning av f_1, g_1 och g_2 .

18. Visa att $|z| \geq 1, -\frac{1}{2} < x \leq \frac{1}{2}$ är ett fundamentalområde.

19. Anta nu att vi ersätter Γ med Γ' , där Γ' har funktionerna f_2 (som ovan), $f_3 = f_1 \circ f_1 : z \rightarrow z + 2$ och $g_3 = g_1 \circ g_1 : z \rightarrow z - 2$ som generatorer. Kan du då finna ett fundamentalområde, F' säg, för Γ' ?

20. Vad händer om du har kvar f_3 och g_3 , men ersätter f_2 med $f_4(z) = z/(z + 1)$ och dess invers $g_4(z) = z/(-z + 1)$?

Litteratur

Brinck, I.-Persson, A., *Elementär teori för analytiska funktioner*. Studentlitteratur 1967.

Myntveksling

DAN LAKSOV

KTH, Stockholm

Beskrivelse av oppgaven. Denne oppgaven kan behandles helt eksperimentelt. Den egner seg imidlertid best som et samspill mellom eksperimenter (på dator eller for hånd) og teoretiske betraktninger. Teorien som kan brukes hentes mest fra elementær tallteori og ligger vel innenfor hva Du kjenner til fra læreboken. Metoder som kjedebrøk, rekursjonsformler, kombinatorikk, ... kan også anvendes. En grundig gjennomgang av den første delen av oppgaven (de eksperimentelle delene av a–n nedenfor) burde Du klare om Du er interessert i matematikk og dette burde rekke for en spesialoppgave. Man kommer imidlertid fort frem til forskningsfronten og kan finne en lang rekke tiltalende, men vanskelige problemer i nær tilknytning til dette stoffet. Vi har nevnt noen av dem i slutten av oppgaven.

Den kjente tyske matematikeren F.G. Frobenius (1849–1917) fremhevet ofte at problemstillingene vi tangerer i denne oppgaven er interessante og oppstår naturlig i mange ulike sammenhenger i matematikken og i anvendelser. Han var imidlertid klar over at problemet i full generalitet var meget vanskelig og at man ikke kan vente å finne eksplisitte uttrykk for de tallene man betrakter. På grunn av disse vanskelighetene har dette området aldri blitt sentralt i matematikken, men det har fascinert en lang rekke matematikere.

Problemet. I et land har de bare to myntsorter. Den ene av valøren $a = 5$ (kroner, dollar, pund, lire, ...) og den andre av valøren $b = 9$. Din oppgave er å finne hvilke priser man kan sette på varene i landet for at man skal kunne betale varen med et eksakt antall mynter.

a. Vis at man kan betale følgende beløp

$$5, 9, 10, 14, 15, 18, 19, 20, 23, 24, 25, 27, 28, \\ 29, 30, 32, 33, 34, 35, 36, \dots$$

Fortsett tabellen. Ser Du noe mønster?

- b. Kan Du vise at varene i landet kan ha prisen 32 og alle høyere priser? Hint, det rekkes å vise at man kan betale 5 priser som følger etter hverandre med en differens på en enhet.
- c. Hold myntsorten a fast og prøv med noen andre verdier av myntsorten b f.eks. $b = 1, b = 2, \dots, b = 10$. Bestem i hvert tilfelle den største prisen Du *ikke* kan betale. Ser Du noe mønster?

Dersom det finnes en største pris Du *ikke* kan betale betegner Du denne prisen med $g(a, b)$. Du kan altså betale alle priser $g(a, b) + 1, g(a, b) + 2, g(a, b) + 3, \dots$. F.eks. er $g(5, 9) = 31$.

- d. For hvilke av myntsortene du prøvet i del (c) ovenfor eksisterte $g(5, b)$ og hva er verdien av $g(5, b)$? Ser Du noe mønster? Kan Du gjette når $g(5, b)$ eksisterer for vilkårlige b ? Kan Du gjette en enkel formel for $g(5, b)$ uttrykt ved b ?
- e. Prøv å bevise gjetningene fra punkt d.

Et annet viktig tall er antallet priser Du ikke kan betale. I tilfellet $a = 5, b = 9$ kunne du ikke betale prisene

$$1, 2, 3, 4, 6, 7, 8, 11, 12, 13, 16, 17, 21, 22, 26, 31,$$

det vil si 16 priser.

Når $g(a, b)$ eksisterer betegner vi med $n(a, b)$ antallet priser vi *ikke* kan betale. F.eks. $n(5, 9) = 16$.

- f. Bestem $n(5, b)$ for de myntsortene du eksperimenterte med i del c. Kan Du gjette hva $n(5, b)$ er for vilkårlig b ? Kan Du vise at Din gjetning er sann?

- g. Eksperimenter med å finne $g(a, b)$ for ulike verdier av a og b . Kan Du gjette for hvilke par av tall a, b tallet $g(a, b)$ eksisterer. Kan Du gjette en enkel eksplisitt formel for $g(a, b)$ uttrykt ved a og b ? Bestem også $n(a, b)$ og prøv å gjette en enkel formel for dette tallet uttrykt ved a og b .
- h. Kan Du bevise gjetningene Du gjorde i del g?

Mer hjelp med problemene a til h kan Du finne i [1] i referenselisten.

Dersom landet har 3 myntsorter a, b og c kan man stille samme spørsmål, men disse er mye vanskeligere å svare på. Vi betegner med $g(a, b, c)$ det største tallet som *ikke* kan veksles, når dette finnes, og med $n(a, b, c)$ antallet priser som ikke kan veksles.

- i. Eksperimenter med tre myntsorter, f.eks. ved å la a og b være som i første del av oppgaven og variere c . Bestem $g(a, b, c)$ og $n(a, b, c)$ i disse tilfellene.
- j. Kan Du ved å starte med tilfellet med 2 myntsorter vise for hvilke myntsorter a, b og c tallet $g(a, b, c)$ eksisterer?
- k. Kan Du av eksperimentene i del i gjette en øvre grense for $g(a, b, c)$ når denne eksisterer? F.eks. kan Du undersøke om $g(a, b, c) \leq abc$ når $g(a, b, c)$ finnes. Kan Du finne bedre grenser?
- l. Her har Du en tabell over noen kjente øvre grenser for $g(a, b, c)$ når vi har $a < b < c$:

T. Skolem (1930) $(a - 1)(b + c - 1) - 1$

I. Schur (1935) $(a - 1)(c - 1) - 1$

A. Brauer (1942) $ab/d + dc - a - b - c,$
der d er største felles divisor til a og b

M. Lewin (1972) $[(c - 2)^2/2] - 1$

M. Lewin (1973) $[\frac{1}{2}(c-2)(b-2)] - 1$

J. Roberts (1956) $a(c-a-2 + [a/(c-a)]) + (b-a-1)(c-a-1)$

Y. Vitek (1975) $(c-2)[\frac{a}{2} - 1]$ for a, b, c inkongruente (mod a)

Sett inn i formlene for en del verdier av a, b og c og sammenlign med den virkelige verdien for $g(a, b, c)$. Grensene ovenfor og referenser til originalarbeidene kan Du finne i [3].

m. Kan Du finne noe samband mellom tallene $g(a, b, c)$ og $n(a, b, c)$?

Det er ganske komplisert å finne eksplisitte uttrykk for $g(a, b, c)$ og $n(a, b, c)$ når disse finnes. Slike uttrykk ble først funnet for noen år siden og er ganske involverte. I [2] i litteraturlisten er endel av dette arbeidet beskrevet og Du kan der finne videre referenser til annen litteratur om Du er interessert.

n. Kan Du si noe om eksplisitte uttrykk for $g(a, b, c)$ og $n(a, b, c)$.

Kanskje Du kan bestemme eksplisitte uttrykk for spesielle verdier av a, b og c , som f.eks. $b = a + d$ og $c = a + 2d$ for noe heltall d ?

For fire og flere myntsorter er Du ved *forskningsfronten*. Kan Du si noe i dette tilfellet?

Litteratur

[1] Gardiner, A., *Discover Mathematics*. Oxford Science Publ. 1987.

[2] Selmer, E.S., To populære problemer i tallteorien. I Myntveksling, II Frankering. *Normat* 29 (1981).

[3] Smoryński, C., Skolem's solution to a problem of Frobenius. *The mathematical intelligencer* 3 (1981), s 123–132.

[4] Vitek, Y., *Bounds for a linear diophantine problem of Froebenius, II*. *Canad. J. Math.* 28 (1976), s 1280–1288.

Fibonaccis talföljd

BERNT LINDSTRÖM

KTH, Stockholm

Leonardo av Pisa, även kallad Fibonacci, var en berömd matematiker under medeltiden. Han lärde sig aritmetik av araber i Nordafrika, introducerade det arabiska talsystemet i Europa och skrev 1202 en epokgörande lärobok i algebra *Liber Abaci*. Han är också känd som upphovsmannen till *Fibonaccis talföljd*:

1, 1, 2, 3, 5, 8, 13, ...

i vilken varje tal är summan av de två närmast föregående talen. Denna talföljd har många underbara egenskaper, som inte bara är talkuriosa: Fibonaccis talföljd spelade en avgörande roll när Matijasievič 1970 lyckades lösa ett berömt problem om diofantiska ekvationers lösbarhet: *Hilberts 10:e problem*.

Om man sätter $F_1 = 1$, $F_2 = 1$, så gäller för det n :te talet i talföljden att

$$(1) \quad F_n = F_{n-1} + F_{n-2}, \text{ när } n \geq 3.$$

Vad man framför allt bör studera är algebraiska relationer mellan Fibonaccitalen samt delbarhetsegenskaper. Man har därvid nytta av att kunna matematisk induktion och kongruensaritmetik. Boken *Talteori för alla* av Ogilvy och Anderson (Prisma, 1968) har ett kapitel om kongruensaritmetik och även ett kort kapitel om den här aktuella talföljden. (Fråga efter boken på biblioteket!)

UPPGIFT. Visa att man kan definiera F_0, F_{-1}, F_{-2} etc. så att man får ett *Fibonacci-tal* F_n för varje heltal n så att dessa tal uppfyller relationen (1) utan inskränkning på n .

UPPGIFT. Studera resterna modulo N av Fibonacci-talen, då $N = F_n$ ($n \geq 3$) är ett Fibonacci-tal. När blir resten 0?

EXEMPEL. Om vi väljer $N = 3 = F_4$ blir resterna av F_1, F_2, \dots :

$$1, 1, 2, 0, 2, 2, 1, 0, 1, 1, 2, 0, \dots$$

SLUTSATS. F_m är jämnt delbart med F_n när m är jämnt delbart med n . (m är ett heltal, som är positivt eller negativt.)

UPPGIFT. Skriv ned ett bevis för slutsatsen ovan! Att ett heltal m är jämnt delbart med ett annat heltal n brukar skrivas $n|m$. Slutsatsen ovan kan nu skrivas kortare: $n|m$ implicerar $F_n|F_m$.

Den största gemensamma divisorn till två heltal m och n brukar skrivas (m, n) . Av det föregående inses att $F_{(m,n)}|F_m$ och $F_{(m,n)}|F_n$. Alltså gäller $F_{(m,n)}|(F_m, F_n)$. Man kan bevisa ännu mer:

SATS. $F_{(m,n)} = (F_m, F_n)$, där $m, n \geq 1$.

HJÄLPSATS 1. Det finns heltal a och b sådana att $(m, n) = am + bn$.

HJÄLPSATS 2.

$$F_{m+n} = F_{m-1}F_n + F_mF_{n+1},$$

$$F_{m-n} = (-1)^{n+1}(F_{m-1}F_n - F_mF_{n-1}).$$

Den senare hjälpsatsen kan visas med induktion över $n \geq 1$ när m är ett godtyckligt heltal. Det följer, att om både F_m och F_n är jämnt

delbara med ett naturligt tal N , så är F_{m+n} och F_{m-n} jämnt delbara med N . Man inser sedan att alla Fibonacci-tal av formen F_{am+bn} är jämnt delbara med N . Eftersom F_m och F_n är jämnt delbara med (F_m, F_n) ser man nu, om man använder Hjälpsats 1, att $F_{(m,n)}$ är jämnt delbart med (F_m, F_n) . Vi har tidigare sett att (F_m, F_n) är jämnt delbart med $F_{(m,n)}$. Alltså måste $F_{(m,n)} = (F_m, F_n)$. Satsen är visad.

Hjälpsats 1 brukar bevisas med användande av Euklides algoritm. Kanske Du kan hitta ett bevis i någon bok, t.ex. i Hans Riesels *En bok om primtal* (Studentlitteratur, 1968).

Av den bevisade satsen följer lätt följande egenskaper hos de Fibonaccis tal. De kom till användning i Matijasievič's lösning av Hilbert's 10:e problem (se Fenstads artikel!).

FÖLJDSATS 1. F_n och F_{n+1} saknar gemensamma delare ($n \geq 1$).

FÖLJDSATS 2. $F_m | F_n$ om och endast om $m | n$ ($m, n \geq 1$).

Här följer ytterligare några relationer, som Du kan försöka bevisa (de två första bevisas i boken av Ogilvy och Anderson).

$$F_{n+1}^2 = F_n F_{n+2} + (-1)^n$$

$$F_1^2 + F_2^2 + \cdots + F_n^2 = F_n F_{n+1}$$

$$F_1 F_2 + F_2 F_3 + \cdots + F_{2n} F_{2n+1} = F_{2n}^2, \quad F_n^2 + F_{n+1}^2 = F_{2n+1},$$

$$F_1 + 2F_2 + 3F_3 + \cdots + nF_n = nF_{n+2} - F_{n+3} + 2.$$

Det finns en liten bok av N.N. Vorobyov, *The Fibonacci Numbers*, som innehåller en hel del om Fibonacci-talens delbarhetsgenskaper. Den publicerades först på ryska, men har översatts både till engelska och tyska.

Litteratur

- [1] Fem festliga formler för Fibonacci-fantaster. *Elementa* 63 (1980), s 199.
- [2] Fenstad, J.E., Hilberts 10. problem. *Nordisk Matematisk Tidsskrift*, 19 (1971), s 5–14.
- [3] Gardner, M., *Mathematical Circus, Chapter 13: Fibonacci and Lucas Numbers*. Alfred A. Knopf, New York 1979.
- [4] Ogilvy, C. Stanley och Anderson, John T., *Talteori för alla*. Prisma 1968.
- [5] Vorobyov, N.N., *The Fibonacci Numbers* (i serien Topics in Mathematics). Heath and Company, Boston.

Permutationer med paritet

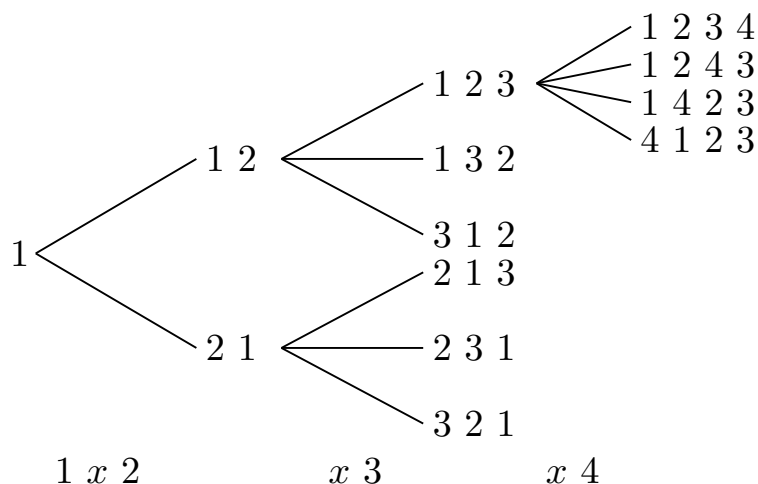
BERNT LINDSTRÖM

KTH, Stockholm

Uppgift. Att studera permutationerna av talen $1, 2, \dots, n$ och indelningen i udda och jämna permutationer ur olika aspekter. Permutationer är särskilt viktiga inom kombinatoriken, sannolikhetsläran och i algebran. I kombinatoriken betraktas en permutation som en uppräkningslista av talen $1, 2, \dots, n$. I gruppteorien (en gren av algebran) studeras permutationer som avbildningar eller operationer.

1. Permutationer ur kombinatorisk synvinkel.

PROBLEM 1. Hur många permutationer finns det av talen $1, 2, \dots, n$? Här är en metod att generera alla permutationer:



SVAR. Det finns $1 \times 2 \times 3 \cdots \times n$ permutationer av talen $1, 2, \dots, n$.

NOTATION. $1 \times 2 \times 3 \times \cdots \times n$ skrivs kortare $n!$, vilket utläses n -fakultet.

EXEMPEL. trefakultet $3! = 6$, fyrfakultet $4! = 24, \dots$

INVERSIONER. Varje gång ett större tal står före ett mindre talar man om en inversion.

EXEMPEL. Permutationen 3 4 1 2 innehåller 4 inversioner: 3 1, 3 2, 4 1 och 4 2.

DEFINITION. Antalet inversioner i en permutation kallas inversionstalet för permutationen. En permutation är *jämn* (*udda*) om inversionstalet är jämnt (resp. udda).

UPPGIFT. Bestäm inversionstalet och pariteten för några permutationer.

EXEMPEL.

<i>Permutation</i>	<i>Inversionstal</i>	<i>Paritet</i>
1 2 3	0	jämn
1 3 2	1	udda
3 1 2	2	jämn
2 1 3	1	udda
2 3 1	2	jämn
3 2 1	3	udda

PROBLEM. Hur många inversioner innehåller permutationen $n \ n - 1 \ n - 2 \ \dots \ 3 \ 2 \ 1$.

SVAR. Permutationen innehåller $n(n - 1)/2$ inversioner (kan visas med induktion över $n \geq 2$).

IAKTTAGELSE. Det finns lika många jämna och udda permutationer av talen 1, 2, 3. Gäller detta mer generellt för permutationerna av $1, 2, \dots, n$?

UPPGIFT. Låt två tal byta plats in en permutation. Påverkas pariteten?

EXEMPEL.

*Jämn paritet**Udda paritet*

1 2 3

1 3 2

2 3 1

2 1 3

3 1 2

3 2 1

REGEL 1. När två intilliggande tal byter plats i en permutation ändras pariteten.

SLUTSATS. Av regel 1 följer att det finns lika många permutationer av de båda slagen jämna och udda.

UPPGIFT. Låt två tal ”på avstånd” byta plats i permutationer. På verkas pariteten?

EXEMPEL.

*Jämn paritet**Udda paritet*

1 2 3

3 2 1

2 3 1

1 3 2

3 1 2

2 1 3

REGEL 2. När två tal byter plats i en permutation ändras pariteten.

Man kan bevisa regel 2 med hjälp av regel 1. Bevisidén illustreras av ett exempel.

EXEMPEL. Permutationen 3 4 1 2 kan överföras i permutationen 2 4 1 3 med hjälp av platsbyten av intilliggande tal på följande sätt

$$\underline{3} \underline{4} 1 2 \rightarrow 4 \underline{3} \underline{1} 2 \rightarrow 4 1 \underline{3} \underline{2} \rightarrow 4 \underline{1} \underline{2} 3 \rightarrow \underline{4} \underline{2} 1 3 \rightarrow 2 4 1 3.$$

Understrukna tal byter plats. Regel 1 användes ett *udda* antal gånger. Pariteten ändras varje gång.

Femtonspelet. Femton brickor, som är numrerade från 1 till 15, placeras slumpmässigt inom en kvadratisk ram. Det gäller nu att

återställa den naturliga ordningen genom att skjuta brickor till den tomma platsen. Det är inte alltid möjligt att återställa ordningen. Man kan visa att *permutationen* som bildas när man läser numren på vanligt sätt från vänster till höger uppifrån och ned *måste vara jämn* för att man skall kunna återställa ordningen.

Brickor i naturlig ordning:

1	2	3	4
5	6	7	8
9	10	11	12
13	14	15	

En (udda) permutation vars ordning inte kan återställas:

7	12	1	10
4	3	8	11
2	13	9	15
6	14	5	

Hur inser man att pariteten måste vara jämn för att ”operationen” skall lyckas?

Enklarest kanske genom att ge den tomma rutan nr 16. Varje gång en bricka skjutes in på den tomma platsen byter talet 16 plats med ett annat tal och pariteten ändras enligt regel 2. När ”bricka 16” återfått sin naturliga plats (längst ned till höger) så har det skett ett *jämnt* antal platsbyten. Detta inses t.ex. om man färgar rutorna vita och svarta som på ett schackbräde. Vid varje drag flyttas ”bricka 16” från svart till vitt eller tvärtom. Begynnelsen och änden är vit; alltså ett jämnt antal drag. Permutationen måste alltså ha summa paritet som den naturliga ordningen, alltså jämn.

Bestämningen av en permutationsparitet med hjälp av inversions-talet kräver att man jämför $n(n-1)/2$ tal. Det finns en snabbare metod att bestämma pariteten kommer vi att finna.

Innan vi lämnar den kombinatoriska avdelningen skulle vi kanske nämna att det finns goda asymptotiska formler som uppskattar talet $n!$ när n är mycket stort. Sådana formler är av intresse i sannolikhetskalkylen. En elementär uppskattning, som man kan visa med hjälp av integralkalkyl, är (e är basen för de naturliga logaritmerna 2,718...)

$$\left(\frac{n}{e}\right)^n < n! < \left(\frac{n+1}{e}\right)^{n+1}.$$

Beviset bygger på att man uppskattar integralen $\int_1^{n+1} \ln x \, dx$ med översummor och undersummor. Observera att integralen kan beräknas exakt med hjälp av en partialintegration.

A propos inversionstal och sannolikhetskalkyl så kan det nämnas att inversionstalen är approximativt normalfördelade. Om detta kan man läsa i Feller: Probability Theory and Its Applications (s.205).

UPPGIFT. Gör ett histogram för inversionstalen till permutationerna av talen $1, 2, \dots, 5$.

Eulers formel $\int_0^\infty e^{-t} t^n dt = n!$ är ett annat roligt sidospår. Om man ersätter n med $x - 1$ i integralen får man en integral som konvergerar för alla reella tal $x > 0$. Den ger *gammafunktionen* $\Gamma(x)$, som man kan läsa mer om i en intressant bok av Emil Artin: The Gamma Function.

2. Permutationer ur algebraisk synvinkel.

Permutationer uppträder i algebran på flera områden. Ett exempel är linjär algebra, där definitionen av determinanter beror på indelningen i udda och jämna permutationer. Ett annat område där permutationer spelar en viktig roll är gruppteorien och teorien för polynomekvationer (s.k. algebraiska ekvationer).

EXEMPEL. Den allmänna lösningen till ekvationssystemet

$$a_1x + b_1y = c_1$$

$$a_2x + b_2y = c_2$$

kan skrivas $x = (b_2c_1 - b_1c_2)/(a_1b_2 - a_2b_1)$, $y = (a_1c_2 - a_2c_1)/(a_1b_2 - a_2b_1)$, när $a_1b_2 - a_2b_1 \neq 0$. Om man skriver upp formlerna för lösning av ekvationssystemet

$$a_1x + b_1y + c_1z = d_1$$

$$a_2x + b_2y + c_2z = d_2$$

$$a_3x + b_3y + c_3z = d_3$$

finner man i nämnarna uttrycket $a_1b_2c_3 + a_2b_3c_1 + a_3b_1c_2 - a_1b_3c_2 - a_2b_1c_3 - a_3b_2c_1$. Observera permutationerna av 1, 2, 3. Termer som svarar mot jämna permutationer har plustecken, termer som svarar mot udda permutationer har minustecken.

PERMUTATIONER SOM FUNKTIONER.

En 1 – 1-avbildning av talen $\{1, 2, \dots, n\}$ på sig själv kallas permutation. Om man skriver bildelementet under varje tal kan man skriva de 6 permutationerna av mängden $\{1, 2, 3\}$:

$$\begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix}, \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix},$$

$$\begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix}, \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix}, \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix}.$$

Man kan definiera produkten av två sådana permutationer f och g som funktionen $f \circ g$ (sammansättningen av funktionerna), $f \circ g(x) = f(g(x))$.

EXEMPEL. Låt $f = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}$ och $g = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix}$. Då blir $f \circ g = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix}$.

En viktig klass av permutationer är de *cykliska*

$$\begin{pmatrix} 1 & 2 & 3 & \dots & n-1 & n \\ 2 & 3 & 4 & \dots & n & 1 \end{pmatrix}.$$

För denna permutation har man infört ett enklare skrivsätt:

$$(1\ 2\ \dots\ n).$$

Varje permutation kan skrivas som en produkt av cykliska permutationer.

EXEMPEL. Betrakta permutationen

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 3 & 4 & 5 & 2 & 1 \end{pmatrix}.$$

Vi finner att $1 \rightarrow 3 \rightarrow 5 \rightarrow 1$ (cykel) och $2 \rightarrow 4 \rightarrow 2$ (cykel). Man kan då skriva

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 3 & 4 & 5 & 2 & 1 \end{pmatrix} = (1\ 3\ 5)(2\ 4).$$

Ordningen mellan faktorerna saknar betydelse när cyklerna saknar gemensamma element (men endast då).

Pariteten för permutationen $\begin{pmatrix} 1 & 2 & \dots & n-1 & n \\ a_1 & a_2 & \dots & a_{n-1} & a_n \end{pmatrix}$ definieras lika med pariteten för $a_1\ a_2\ \dots\ a_n$ och beror alltså på pariteten hos inversionstalet för denna följd.

ÖVNING. Visa att den cykliska permutationen $(1\ 2\ \dots\ n)$ har jämn paritet när n är udda (sic!) och udda paritet när n är jämn. Man kan visa att alla cykliska permutationer av n element har denna egenskap.

PARITETEN FÖR PRODUKTER AV PERMUTATIONER. Man kan visa att pariteten för en produkt av två permutationer är *summan* av

pariteterna för de båda permutationerna. Denna regel kan generaliseras till produkter av flera permutationer. Summan av pariteter beräknas enligt reglerna: jämn + jämn = jämn, jämn + udda = udda, udda + udda = jämn.

EN SNABB METOD ATT BESTÄMMA PARITETEN FÖR PERMUTATIONER. Skriv permutationen som en produkt av cykler. Antalet cykler av jämn längd bestämmer pariteten. Om detta antal är jämnt är pariteten jämn, om det är udda blir pariteten udda.

Denna regel är en följd av föregående sats om pariteten för produkter och övningen strax före.

EXEMPEL. Låt oss använda denna metod för att bestämma pariteten i exemplet från avsnittet om femtonspelet. Permutationen är

$$\begin{pmatrix} 1 & 2 & 3 & \dots & 14 & 15 \\ 7 & 12 & 1 & \dots & 14 & 6 \end{pmatrix} = (1\ 7\ 8\ 11\ 9\ 2\ 12\ 15\ 5\ 4\ 10\ 13\ 6\ 3)(14).$$

Vi får alltså en cykel av jämn längd (cykeln (14) innehåller bara ett tal och behöver inte skrivas ut, om man inte vill). Permutationen är därför udda.

Litteratur

- [1] Lindström, B., Perspektiv på permutationer och paritet. *Elementa* 59 (1976), s 81–83.
- [2] Nagell, T., *Lärobok i algebra*. Almqvist & Wiksell, Uppsala 1949.

Matrisavbildningar

KIRSTI MATTILA

K T H

1. Inledning. I denna uppgift betraktas matriser som avbildningar på planet \mathbf{R}^2 ; speciellt betraktas projektioner och isometrier. En projektion är en avbildning P som uppfyller villkoret $P(P(x, y)) = P(x, y)$ för varje vektor (x, y) . Till exempel är $P(x, y) = (x, 0)$ en projektion som projicerar punkterna i planet på x -axeln och $Q(x, y) = (x, x/2)$ en projektion som projicerar punkterna på linjen $y = x/2$.

Vi betraktar två olika normer (längder) i planet, den euklidiska normen $\|(x, y)\| = \sqrt{x^2 + y^2}$ och maximumnormen $\|(x, y)\| = \max\{|x|, |y|\}$. De punkter i planet som har längden ett är punkterna på enhetscirkeln om normen är den euklidiska normen och punkterna på en kvadrat med hörn $(\pm 1, \pm 1)$ om normen är maximumnormen.

I uppgiften skall bestämmas de projektioner som inte förlänger vektorer (kontraktiva projektioner). Om normen i planet är maximumnormen, så gäller för projektionen Q ovan att

$$\|Q(x, y)\| = |x| \leq \|(x, y)\|,$$

alltså normen ökar inte. Men om normen är den euklidiska normen, så är till exempel

$$\|Q(1, 0)\| = \|(1, 1/2)\| = \sqrt{5}/2 > 1 = \|(1, 0)\|.$$

En isometri är en avbildning som förvarar längden av varje vektor. Till exempel $T(x, y) = (-y, x)$ är en isometri med avseende på båda normerna. Om normen är den euklidiska normen, så finns det en

kontraktiv projektion på varje endimensionellt delrum och det finns oändligt många isometrier. Dessa resultat gäller inte om normen är maximumnormen.

2. Om normer och matriser. En *norm* (eller längd) av en punkt eller vektor X är ett tal $\|X\|$ som uppfyller de följande villkoren:

$$(1) \quad \begin{cases} \|X\| \geq 0 \text{ och om } \|X\| = 0, \text{ så är } X = \bar{0}. \\ \|\alpha X\| = |\alpha| \|X\| \text{ för varje reellt tal } \alpha. \\ \|X + Y\| \leq \|X\| + \|Y\| \quad (\text{triangelolikhet}). \end{cases}$$

Den vanligaste normen i planet är den euklidiska normen

$$\|(x, y)\| = \sqrt{x^2 + y^2} .$$

UPPGIFT 1. Visa att den euklidiska normen uppfyller villkoren (1).

Maximumnormen är

$$\|(x, y)\| = \max\{|x|, |y|\} .$$

UPPGIFT 2. Visa att maximumnormen är en norm.

I det följande talar vi om *rummet* E när normen i \mathbf{R}^2 är den euklidiska normen och om *rummet* Y när normen är maximumnormen.

En *matris* är ett rektulångulärt schema av reella tal. Vi behöver kvadratiska matriser $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ och kolonnmatriser $\begin{bmatrix} a \\ b \end{bmatrix}$. Vi nämner här några egenskaper av matriser. Mera om matriser kan man läsa i referens [2].

Matriserna $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ och $\begin{bmatrix} a' & b' \\ c' & d' \end{bmatrix}$ är *lika*, d v s

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} a' & b' \\ c' & d' \end{bmatrix} \quad \text{om} \quad a = a', \quad b = b', \quad c = c' \quad \text{och} \quad d = d' .$$

Motsvarande gäller för kolonnmatriser. Matriser av samma form kan adderas enligt följande

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} + \begin{bmatrix} a' & b' \\ c' & d' \end{bmatrix} = \begin{bmatrix} a + a' & b + b' \\ c + c' & d + d' \end{bmatrix}$$

och

$$\begin{bmatrix} a \\ b \end{bmatrix} + \begin{bmatrix} a' \\ b' \end{bmatrix} = \begin{bmatrix} a + a' \\ b + b' \end{bmatrix} .$$

Multiplikationen av matriser definieras på följande sätt:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} a' & b' \\ c' & d' \end{bmatrix} = \begin{bmatrix} aa' + bc' & ab' + bd' \\ ca' + dc' & cb' + dd' \end{bmatrix}$$

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} e \\ f \end{bmatrix} = \begin{bmatrix} ae + bf \\ ce + df \end{bmatrix} .$$

Vi skriver $AA = A^2$ om A är en kvadratisk matris. Matriserna

$$I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \text{och} \quad \bar{0} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

kallas för *identitetsmatrisen* respektive *nollmatrisen*. För varje kvadratisk matris B gäller att

$$IB = BI = B \quad \text{och} \quad \bar{0}B = B\bar{0} = \bar{0} .$$

En kvadratisk matris P är en *projektion* om $P^2 = P$.

UPPGIFT 3. a) Visa att identitetsmatrisen och nollmatrisen är projektioner.

b) Visa att om P är en projektion, så är $I - P$ också en projektion.

UPPGIFT 4. Bestäm alla projektioner.

En matris $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ kan uppfattas som en avbildning (funktion) på \mathbf{R}^2 som avbildar punkten (x, y) på punkten $(ax + by, cx + dy)$.

Vi skriver

$$A(x, y) = (ax + by, cx + dy) \quad \text{för varje talpar} \quad (x, y)$$

eller på matrisform

$$A \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \quad \text{för varje kolonnmatris } \begin{bmatrix} x \\ y \end{bmatrix}.$$

Om $\| \cdot \|$ är en norm i \mathbf{R}^2 , så kan man definiera en *matrisnorm* genom

$$(2) \quad \|A\| = \max\{\|(ax + by, cx + dy)\| : \|(x, y)\| = 1\}$$

där $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$.

Vi kan också skriva

$$\|A\| = \max\{\|AX\| : \|X\| = 1\}$$

där normen av en kolonnmatris $X = \begin{bmatrix} x \\ y \end{bmatrix}$ är $\|X\| = \|(x, y)\|$. Speciellt gäller att $\|I\| = 1$.

UPPGIFT 5. Visa, att om $\| \cdot \|$ är en norm i \mathbf{R}^2 , så uppfyller motsvarande matrisnorm (2) villkoren (1).

UPPGIFT 6. Visa, att om P är en projektion, så gäller det att

$$\|P\| \geq 1 \quad \text{om} \quad P \neq \bar{0}.$$

En *isometri* på planet \mathbf{R}^2 med en norm $\| \cdot \|$ är en matris T som uppfyller villkoret

$$\|TX\| = \|X\| \quad \text{för varje kolonnmatris } X.$$

3. Maximumnormen. Om normen i \mathbf{R}^2 är maximumnormen $\|(x, y)\| = \max\{|x|, |y|\}$, så kan motsvarande matrisnorm (2) av en matris $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ uttryckas på ett enkelt sätt som en funktion av matriselement a, b, c och d .

UPPGIFT 7. Bestäm matrisnormen i rummet Y .

UPPGIFT 8. a) Bestäm alla de projektioner på rummet Y som uppfyller villkoret (villkoren)

- (i) $\|P\| = 1$ (kontraktiva projektioner)
 (ii) $\|P\| = \|I - P\| = 1$ (bikontraktiva projektioner).

b) Beräkna $\|I - 2P\|$ för de bikontraktiva projektionerna.

UPPGIFT 9. Bestäm alla isometrier på rummet Y .

4. Den euklidiska normen.

UPPGIFT 10. Bestäm matrisnormen i rummet E .

(Ledning: Om $\|(x, y)\| = \sqrt{x^2 + y^2} = 1$, sätt $x = \cos t$, $y = \sin t$ och bestäm det största värdet av funktionen

$$f(t) = \|A \begin{bmatrix} x \\ y \end{bmatrix}\|^2 = \|A \begin{bmatrix} \cos t \\ \sin t \end{bmatrix}\|^2, \quad 0 \leq t \leq 2\pi.)$$

Om A är matrisen $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$, så kallar vi mängden

$$V(A) = \{(ax + by, cx + dy) : x \text{ och } y \text{ är reella tal}\}$$

värdemängden för A .

UPPGIFT 11. Visa att $V(I) = \mathbf{R}^2$ och

$$V\left(\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}\right) = \{(0, y) : y \text{ är reellt tal}\} = y\text{-axeln}.$$

UPPGIFT 12. a) Bestäm alla projektioner P på E som uppfyller villkoret $\|P\| = 1$. Beräkna sedan $\|I - P\|$ för dessa projektioner.

b) För varje reellt tal k bestäm projektionen P_k så att $\|P_k\| = 1$ och att värdemängden för P_k är linjen $y = kx$.

c) Vilken mängd är $V(I - P_k)$?

UPPGIFT 13. Bestäm alla isometrier på E .

Egenvärden till en matris $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ är nollställen till polynomet

$$p(x) = (a - x)(d - x) - bc,$$

d v s egenvärden är lösningar till ekvationen $p(x) = 0$.

Matrisen $A^t = \begin{bmatrix} a & c \\ b & d \end{bmatrix}$ är den till A transponerade matrisen.

UPPGIFT 14. Beräkna egenvärdena till matrisen $A^t A$. Om λ är den större av egenvärdena, visa att

$$\|A\| = \sqrt{\lambda},$$

där $\|A\|$ är matrisnormen i E (Uppgift 10).

Litteratur

Matrisnormen i Y (Uppgift 7) finns i boken

[1] Faddeev, D.K. och Faddeeva, V.N., *Computational Methods of Linear Algebra*. Freeman 1963.

[2] Lang, S., *Linear Algebra*. Addison Wesley 1972.

Om matrisavbildningar i (oändligtdimensionella) normerade rum kan man läsa i

[3] Taylor, A.E., *Introduction to Functional Analysis*. Wiley 1958.

Volymer av n–dimensionella klot

MIKAEL PASSARE

Stockholms universitet

Ett klot med radien r är mängden av punkter vars avstånd till en given punkt (medelpunkten) är högst r . Låt oss skriva $B^3(r)$ för det klot i (x, y, z) -rummet som har sin medelpunkt i origo (det vill säga $(0, 0, 0)$) och vars radie är r . (Bokstaven B står för *boule* (*fr*) eller *ball* (*eng*) och trean betecknar dimensionen.)

UPPGIFT 1. *Bevisa att*

$$B^3(r) = \{J(x, y, z); x^2 + y^2 + z^2 \leq r^2\}.$$

LEDTRÅD. Anta att $(x, y, z) \in B^3(r)$ och använd Pythagoras sats, först på triangeln med hörn i $(0, 0, 0)$, $(x, 0, 0)$ och $(x, y, 0)$, sedan på triangeln med hörn i $(0, 0, 0)$, $(x, y, 0)$ och (x, y, z) .

Låt nu $B^2(r)$ vara cirkelskivan i (x, y) -planet som har sin medelpunkt i origo (det vill säga $(0, 0)$) och vars radie är r . Då kan vi skriva

$$B^2(r) = \{(x, y); x^2 + y^2 \leq r^2\}.$$

UPPGIFT 2. *Vad blir $B^1(r)$?*

Vi ska också studera klot i högre dimensioner och börjar med det fyrdimensionella klotet $B^4(r)$, som har sin medelpunkt i origo (det vill säga $(0, 0, 0, 0)$) och vars radie är r .

DEFINITION.

$$B^4(r) = \{(x, y, z, w); x^2 + y^2 + z^2 + w^2 \leq r^2\}.$$

UPPGIFT 3. *Ge en definition av $B^n(r)$.*

FÖRSLAG. Eftersom alfabetet är så kort och för att få ett enhetligt skrivsätt är det praktiskt att skriva x_1 -rummet istället för x -axeln, (x_1, x_2) -rummet istället för (x, y) -planet och (x_1, x_2, x_3) -rummet istället för (x, y, z) -rummet. Det n -dimensionella rummet kallar vi då (x_1, x_2, \dots, x_n) -rummet.

Nu inför vi beteckningen $\text{Vol } B^3(r)$ för volymen av $B^3(r)$, det vill säga

$$\text{Vol } B^3(r) = \frac{4\pi r^3}{3}.$$

Om vi låter tvådimensionell volym betyda area och endimensionell volym betyda längd så får vi också

$$\text{Vol } B^2(r) = \pi r^2$$

och

$$\text{Vol } B^1(r) = 2r.$$

För att kunna gå vidare och hitta en formel för $\text{Vol } B^n(r)$ måste vi först tänka efter vad vi ska mena med volymen av en mängd i (x_1, x_2, \dots, x_n) -rummet.

Hur definierar man egentligen vanlig area? Jo, först bestämmer man att en kvadrat med sidan s ska ha arean s^2 och att arean hos en union av disjunkta kvadrater ska vara summan av deras respektive areor. Arean av andra mängder får man sedan genom att approximera dessa med unioner av kvadrater.

UPPGIFT 4. *Låt (x_1, x_2) -rummet vara indelat i ett fint rutnät där varje liten kvadrat har sidan s . Skriv ett datorprogram som approximerar*

$$\text{Vol } B^2(1) = \pi$$

genom att lägga samman areorna hos alla små kvadrater som åtminstone delvis är innehållna i cirkelskivan.

FÖRSLAG. För att få en hygglig approximation bör s inte vara större än 0,0001. Försök gärna förbättra approximationen, till exempel genom att bara ta med hälften av de kvadrater som inte ligger helt innanför cirkeln.

Vanlig tredimensionell volym definieras på liknande sätt genom att man först stipulerar att en kub med sidan s ska ha volymen s^3 och sedan approximerar man med unioner av kuber.

En typisk n -dimensionell kub med sidan s är

$$\{(x_1, x_2, \dots, x_n); a_1 \leq x_1 \leq a_1 + s, \\ a_2 \leq x_2 \leq a_2 + s, \dots, a_n \leq x_n \leq a_n + s\},$$

och vi sätter dess volym (som naturligtvis mäts i m^n om s uttrycks i meter) till s^n . Volymen av andra n -dimensionella mängder som till exempel $B^n(r)$ definieras vi sedan genom approximation som tidigare.

UPPGIFT 5. *Bevisa att om en n -dimensionell mängd förstoras med skalfaktorn k så multipliceras dess volym med k^n .*

LEDTRÅD. Det räcker att göra det för kuber, sedan approximerar man ju.

Nu ser vi att för att finna $\text{Vol } B^n(r)$ så räcker det att räkna ut $\text{Vol } B^n$ (vi skriver bara B^n för enhetsklotet, istället för $B^n(1)$) och sedan använda sambandet

$$\text{Vol } B^n(r) = \text{Vol } B^n \cdot r^n.$$

Innan vi på allvar tar itu med de högre dimensionerna ska vi beräkna $\text{Vol } B^3$. (Vi vet förstås redan att svaret är $4\pi/3$ men nu

ska vi *bevisa* detta.) Ett sätt vore naturligtvis att approximera med kuber direkt och göra en tredimensionell version av Uppgift 4, men man kan också approximera med andra mängder vars volym redan beräknats, exempelvis cylindrar.

UPPGIFT 6. *Bevisa att en cirkulär cylinder med radien r och höjden h har volymen $\pi r^2 h$.*

LEDTRÅD. Arean hos basytan, alltså cirkelskivan, fick vi ju fram genom approximation med små kvadrater. Visa att rätblocket med en sådan kvadrat som bas och med höjden h har volymen $s^2 h$. (Använd en stapel av $\approx h/s$ stycken kuber med sidan s .)

Vi delar in intervallet $[-1, 1]$ (det vill säga B^1) i $2N$ intervall av längd $1/N$. En typisk ändpunkt blir alltså k/N för något k mellan $-N$ och N . Om nu $x_1 = k/N$ så visar Pythagoras sats att $(x_1, x_2, x_3) \in B^3$ precis om $x_2^2 + x_3^2 \leq 1 - (k/N)^2$. För stora värden på N får vi en god approximation av $\text{Vol } B^3$ genom att ta summan av volymerna hos cylindrarna med höjd $1/N$ och radie $\sqrt{1 - (k/N)^2}$ då k går från $-(N-1)$ till $N-1$. (Vi skivar klotet i tunna skivor och approximerar varje skiva med en något större cylinder - rita en figur!) Om vi slår samman de båda identiska cylindrar som hör till $\pm k$ till en enda cylinder med höjden $2/N$ så kan vi skriva

$$\begin{aligned} \text{Vol } B^3 &= \lim_{N \rightarrow \infty} \left(\pi \left(1 - \left(\frac{0}{N}\right)^2\right) \frac{2}{N} + \pi \left(1 - \left(\frac{1}{N}\right)^2\right) \frac{2}{N} + \right. \\ &\quad \left. \dots + \pi \left(1 - \left(\frac{N-1}{N}\right)^2\right) \frac{2}{N} \right) \\ &= \lim_{N \rightarrow \infty} \pi \sum_{k=0}^{N-1} \left(1 - \left(\frac{k}{N}\right)^2\right) \frac{2}{N} \\ &= 2\pi \left(1 - \lim_{N \rightarrow \infty} \sum_{k=0}^{N-1} \left(\frac{k}{N}\right)^2 \frac{1}{N}\right). \end{aligned}$$

UPPGIFT 7A). *Bevisa att*

$$\lim_{N \rightarrow \infty} \sum_{k=0}^{N-1} \left(\frac{k}{N}\right)^n \frac{1}{N} J = \frac{1}{n+1}.$$

LEDTRÅD. Gränsvärdet är $\int_0^1 x^n dx$, enligt definitionen av integralen.

UPPGIFT 7B). *Bevisa att*

$$\lim_{N \rightarrow \infty} \sum_{k=0}^{N-1} \left(\frac{k}{N}\right)^2 \frac{1}{N^m} = 0, \quad \text{om } m > 1.$$

LEDTRÅD. Gränsvärdet kan skrivas

$$\lim_{N \rightarrow \infty} \sum_{k=0}^{N-1} \left(\frac{k}{N}\right)^2 \frac{1}{N} \cdot \lim_{N \rightarrow \infty} \frac{1}{N^{m-1}}.$$

Således har vi

$$\lim_{N \rightarrow \infty} \sum_{k=0}^{N-1} \left(\frac{k}{N}\right)^2 \frac{1}{N} = \frac{1}{3}$$

och följaktligen

$$\text{Vol } B^3 = 2\pi \left(1 - \frac{1}{3}\right) = \frac{4\pi}{3},$$

vilket ju stämmer väl med vad vi redan visste.

Nu är vi mogna att ta språnget in i det fyrdimensionella rummet. Vi ska beräkna $\text{Vol } B^4$ helt i analogi med hur vi nyss fann $\text{Vol } B^3$. Vi börjar med att dela in enhetsskivan B^2 i N stycken ringar genom att rita upp koncentriska cirklar med radie k/N , där k går från 0 till $N-1$. Om nu (x_1, x_2) ligger på en sådan cirkel, det vill säga

$$\sqrt{x_1^2 + x_2^2} = \frac{k}{N},$$

så ser vi att $(x_1, x_2, x_3, x_4) \in B^4$ precis om

$$x_3^2 + x_4^2 \leq 1 - \left(\frac{k}{N}\right)^2.$$

Vi approximerar $\text{Vol } B^4$ med summan av volymerna hos *cylindrar* med samma basarea som tidigare, alltså

$$\pi\left(1 - \left(\frac{k}{N}\right)^2\right),$$

men nu med *höjdarean*

$$\pi\left(\left(\frac{k+1}{N}\right)^2 - \left(\frac{k}{N}\right)^2\right),$$

det vill säga arean hos en typisk ring i vår indelning av B^2 . En stunds eftertanke visar att detta exakt motsvarar vad vi gjorde tidigare, men då hade vi en indelning av B^1 istället.

UPPGIFT 8. *Motivera varför den fyrdimensionella volymen av cylindern ovan måste vara basarean gånger höjdarean, alltså*

$$\pi\left(1 - \left(\frac{k}{N}\right)^2\right) \cdot \pi\left(\left(\frac{k+1}{N}\right)^2 - \left(\frac{k}{N}\right)^2\right).$$

LEDTRÅD. Minns att vi definierade fyrdimensionell volym via approximation med fyrdimensionella kuber och jämför sedan med Uppgift 6.

Vi har alltså (i perfekt analogi med det tredimensionella fallet)

$$\text{Vol } B^4 = \lim_{N \rightarrow \infty} \sum_{k=0}^{N-1} \pi\left(1 - \left(\frac{k}{N}\right)^2\right) \pi\left(\left(\frac{k+1}{N}\right)^2 - \left(\frac{k}{N}\right)^2\right).$$

Men

$$(k+1)^2 = k^2 + 2k + 1$$

så

$$\left(\frac{k+1}{N}\right)^2 - \left(\frac{k}{N}\right)^2 = \frac{2k+1}{N^2}$$

och det följer att

$$\begin{aligned} \text{Vol } B^4 &= \pi^2 \left(1 - \lim_{N \rightarrow \infty} \sum_{k=0}^{N-1} \frac{k^2}{N^2} \cdot \frac{2k+1}{N^2}\right) \\ &= \pi^2 \left(1 - \lim_{N \rightarrow \infty} 2 \sum_{k=0}^{N-1} \left(\frac{k}{N}\right)^3 \frac{1}{N} - \lim_{N \rightarrow \infty} \sum_{k=0}^{N-1} \left(\frac{k}{N}\right)^2 \frac{1}{N^2}\right). \end{aligned}$$

Enligt Uppgift 7 är dessa båda sista gränsvärden lika med $2 \cdot \frac{1}{4}$ respektive 0, så vi får

$$\text{Vol } B^4 = \pi^2 \left(1 - \frac{1}{2} - 0\right) = \frac{\pi^2}{2}.$$

Mera allmänt har således $B^4(r)$ volymen $\pi^2 r^4/2$.

UPPGIFT 9. *Bevisa formeln*

$$\text{Vol } B^n = \frac{2\pi}{n} \cdot \text{Vol } B^{n-2}$$

för varje $n \geq 3$.

LEDTRÅD. Dela först in B^{n-2} i N stycken skal genom att införa koncentrisk klot med radie k/N , $0 \leq k \leq N-1$. Approximera som förut $\text{Vol } B^n$ med summan av volymerna hos *cylindrarna* med basarea

$$\pi \left(1 - \left(\frac{k}{N}\right)^2\right)$$

och *höjdvolym*

$$\text{Vol } B^{n-2} \cdot \left(\left(\frac{k+1}{N}\right)^{n-2} - \left(\frac{k}{N}\right)^{n-2}\right).$$

Dessa *cylindrars* volym ges som vanligt av basarean gånger *höjdvolymen*. Använd också det faktum att

$$(k + 1)^{n-2} = k^{n-2} + (n - 2)k^{n-3} + \text{lägre potenser av } k.$$

Nu är det lätt att räkna ut $\text{Vol } B^n$:

För jämnt n , säg $n = 2m$, får vi

$$\begin{aligned} \text{Vol } B^{2m} &= \frac{\pi}{m} \cdot \text{Vol } B^{2(m-1)} = \frac{\pi^2}{m(m-1)} \cdot \text{Vol } B^{2(m-2)} = \dots \\ &= \frac{\pi^{m-1}}{m!} \cdot \text{Vol } B^2 = \frac{\pi^m}{m!}. \end{aligned}$$

För udda n , säg $n = 2m - 1$, blir formeln inte fullt lika elegant. Vi får

$$\begin{aligned} \text{Vol } B^{2m-1} &= \frac{2\pi}{2m-1} \cdot \text{Vol } B^{2m-3} = \dots \\ &= \frac{(2\pi)^{m-1}}{(2m-1)(2m-3)\dots 3} \cdot \text{Vol } B^1 \\ &= \frac{2^m \pi^{m-1}}{1 \cdot 3 \dots (2m-3)(2m-1)}. \end{aligned}$$

Vi ska nu se att genom att införa beteckningen $x!$ också för andra tal än heltal, exempelvis för $x = 1/2$, så kan man knyta ihop formelerna för det jämna och det udda fallet till en enda formel. Låt oss diskutera hur en sådan utvidgning av fakultetsfunktionen kan göras. Ett första naturligt krav är ju att sambandet

$$x! = (x - 1)! \cdot x$$

ska fortfarande att gälla även för x som inte är heltal. Men detta räcker förstås inte - det finns massor av funktioner som uppfyller detta. Vi kan ju definiera $x!$ helt efter behag i intervallet mellan 0 och 1 och sedan använda den rekursiva formeln för att få en funktion för alla positiva x .

UPPGIFT 10. Låt p och q vara två heltal ≥ 0 sådana att $p < q$. Skriv upp ekvationen (på formen $y = ax + b$) för linjen som förbinder punkterna $(p, \log p!)$ och $(q, \log q!)$. Visa sedan att för varje heltal n mellan p och q så ligger punkten $(n, \log n!)$ under linjen, det vill säga

$$\log n! \leq an + b.$$

LEDTRÅD. Anta först att $q = p + 1$. Då räcker det att kontrollera de bägge fallen $n = p$ och $n = q$. Bevisa därefter att om q ersätts med $q + 1$ så får man en brantare linje. Kom ihåg att

$$\log(q + 1)! = \log q! + \log(q + 1),$$

och rita en figur.

DEFINITION. En funktion f kallas konvex om punkten $(x, f(x))$ ligger under linjen genom $(p, f(p))$ och $(q, f(q))$ för alla tal (ej nödvändigtvis heltal) sådana att $p \leq x \leq q$. (Grafen buktar med andra ord nedåt.)

Ett andra naturligt villkor är alltså att $\log x!$ ska vara en konvex funktion. Märkvärdigt nog bestämmer detta, tillsammans med den rekursiva formeln ovan, $x!$ entydigt, och vi kan nu beräkna vad till exempel $(1/2)!$ måste vara.

Först observerar vi att för $k = 1, 2, 3, \dots$ så ger oss konvexiteten olikheterna

$$\log(k - \frac{1}{2})! \leq \frac{1}{2}(\log(k - 1)! + \log k!)$$

och

$$\log k! \leq \frac{1}{2}(\log(k - \frac{1}{2})! + \log(k + \frac{1}{2})!),$$

det vill säga

$$(k - \frac{1}{2})!^2 \leq (k - 1)!k! \quad \text{och} \quad k!^2 \leq (k - \frac{1}{2})!(k + \frac{1}{2})!.$$

Eftersom nu

$$\left(k - \frac{1}{2}\right)! = \left(\frac{1}{2}\right)! \frac{3}{2} \frac{5}{2} \cdots \frac{(2k-1)}{2}$$

så får vi

$$\left(\frac{2^{k-1}k!}{3 \cdot 5 \cdots (2k-1)}\right)^2 / \left(k + \frac{1}{2}\right) \leq \left(\frac{1}{2}\right)!^2 \leq \left(\frac{2^{k-1}k!}{3 \cdot 5 \cdots (2k-1)}\right)^2 / k.$$

Detta kan också skrivas

$$\begin{aligned} (1+1)^2 \left(1 + \frac{1}{3}\right)^2 \cdots \left(1 + \frac{1}{(2k-1)}\right)^2 / \left(k + \frac{1}{2}\right) &\leq 4\left(\frac{1}{2}\right)!^2 \\ &\leq (1+1)^2 \left(1 + \frac{1}{3}\right)^2 \cdots \left(1 + \frac{1}{(2k-1)}\right)^2 / k. \end{aligned}$$

UPPGIFT 11. *Skriv ett datorprogram som räknar ut höger- och vänsterledet ovan för stora värden på k .*

Man drar slutsatsen att det måste gälla att

$$4\left(\frac{1}{2}\right)!^2 = \pi,$$

det vill säga

$$\left(\frac{1}{2}\right)! = \frac{\sqrt{\pi}}{2}.$$

UPPGIFT 12. *Visa att det n -dimensionella klotets volym kan skrivas*

$$\text{Vol } B^n = 2^n \left(\frac{1}{2}\right)!^n / \left(\frac{n}{2}\right)!.$$

LEDTRÅD. Eftersom vi vet vad $(1/2)!$ är så känner vi också $(n/2)!$ för varje n .

Till sist kan nämnas att man kan generalisera kloten ytterligare genom att för varje positivt tal p definiera

$$B_p^n(r) = \{(x_1, x_2, \dots, x_n); |x_1|^p + |x_2|^p + \dots + |x_n|^p \leq r^p\}.$$

Observera att $B_2^n(r) = B^n(r)$, alltså det vanliga runda klotet.

Då får man formeln

$$\text{Vol } B_p^n = 2^n \left(\frac{1}{p}\right)!^n / \left(\frac{n}{p}\right)!$$

för det n -dimensionella p -klotet.

UPPGIFT 13. *Rita upp B_p^2 för några olika värden på p . Vad bör B_∞^2 vara?*

UPPGIFT 14. *Kontrollera giltigheten hos formeln ovan för $\text{Vol } B_p^n$ i så många fall som möjligt.*

Lycka till!

Mönster

JOHAN PHILIP

K T H

Överallt i naturen och på konstruerade föremål ser man reguljära figurer. Blomblad sitter systematiskt i en ring och på hyreshusen sitter fönstren i rader och kolumner. Trästavarna i parkettgolvet, tänderna på ett kugghjul, blommorna på en tapet eller ett tyg upprepas på ett systematiskt sätt.

Vi skall studera plana (tvådimensionella) ornament, vilka ha något slag av symmetri och/eller är upprepning av en viss ”figur”. Vi skall ej bry oss om vilken ”figur” det är som upprepas utan endast hur den upprepas. Vid studiet av tapetmönster t ex, vilket följer nedan, bryr vi oss alltså ej om vilken blomma eller vilka ränder som upprepas, endast på vilket sätt det sker.

Isometrier. En isometri är en avbildning av en figur som ej ändrar några avstånd inom figuren. Vi studerar avbildningen av en triangel i planet. Bilden är alltså en triangel på ett annat ställe i planet med lika långa sidor som den ursprungliga. Avbildningen kan vara av två slag:

direkt om orienteringen ej ändras, dvs om trehörningen A , B och C ligger medurs kring figuren så gör deras bilder också det, se fig. 1 och 2,

omvänd om orienteringen ändras, dvs om A , B och C ligger medurs ligger deras bilder A' , B' och C' moturs, se fig. 3 och 4.

Vi kommer att studera tre slag av isometrier:

1. *Translation.* Figuren förflyttas utan att vridas.
2. *Rotation.* Figuren vrides kring en punkt O utan att translateras.

Punkten O kan ligga långt från figuren.

3. *Reflektion* i en linje m . Varje punkt A "speglas" i en punkt A' på andra sidan linjen så att AA' är vinkelrät mot m och avståndet från A' till m är lika med avståndet från A till m .

Vi definierar också:

Glidreflektion, vilket är en sammansättning av en reflektion i en linje m och en translation längs m .

SATS 1. *En direkt isometri är en translation eller en rotation.*

BEVIS. En translation är beskriven i Fig. 1. (Den kan eventuellt uppfattas som rotation noll grader kring en punkt oändligt långt borta.) Det som verkligen kräver ett bevis är att varje direkt isometri som ej är en translation är en rotation, dvs att det finns en punkt (O i fig. 2) sådan att isometrin verkligen är en rotation kring den. Låt alltså ABC vara ursprungstriangeln och $A'B'C'$ dess bild. Låt O vara skärningen av mittpunktsnormalerna till sträckorna AA' och BB' . Vi skall visa att mittpunktsnormalen till CC' går genom O . Enligt konstruktionen är $OA = OA'$ och $OB = OB'$. Då vi har en isometri är $AB = A'B'$ så trianglarna OAB och $OA'B'$ är kongruenta av vilket följer att $\angle AOB = \angle A'OB'$. Vi lägger $\angle BOA'$ till var och en av dessa och ser att $\angle AOA' = \angle BOB' =$ rotationsvinkeln. Vi har $\angle OAC = \angle OAB - \angle BAC = \angle OA'B' - \angle B'A'C' = \angle OA'C'$. Eftersom $OA = OA'$ och $AC = A'C'$ följer därav att OAC är kongruent med $OA'C'$. Detta ger direkt $OC = OC'$, och beviset är klart.

SATS 2. *En omvänd isometri är en glidreflektion.*

BEVIS. Studera först fig. 3 vilket är en ren reflektion i en linje m (translationen är noll). Definitionsmässigt går m genom mittpunkterna M_1 , M_2 och M_3 till AA' , BB' och CC' . Man inser lätt (se fig.

4) att om man förskjuter $A'B'C'$ utefter m så kommer mittpunkterna M_1, M_2 och M_3 fortfarande att ligga på linjen m . Omvänt, om ABC och $A'B'C'$ är två isometriska trianglar med olika orientering. Drag sammanbindningslinjerna AA', BB' och CC' . Högst två av dessa kan skära varandra om orienteringen är olika. Mittpunkterna på sammanbindningslinjerna kan ej alla sammanfalla. Lägg en linje genom två olika mittpunkter. Detta är reflektionslinjen m . Man kan translatera $A'B'C'$ längs m så att den blir reflektionen av ABC , enligt fig. 4, vilket avslutar beviset.

Vi har alltså funnit att den direkta isometrin har en *fixpunkt*, dvs en punkt som ej flyttas vid isometrin medan en omvänd isometri har en *fixlinje*.

Vi skall studera vilka avbildningar man får vid sammansättning och upprepning av translationer, rotationer och reflektioner och inför symboler för dessa operationer.

S_i rotation med viss bestämd vinkel kring en punkt O_i .

T_i translation en bestämd sträcka utefter en linje m_i .

R_i reflektion i en linje m_i .

Med S_i^{-1} menar vi rotationen kring O_i med samma vinkel som S_i men åt motsatt håll. Att först använda S_i och sedan S_i^{-1} på en figur innebär alltså att låta den vara still. Vi skriver detta

$$S_i^{-1}S_i = I,$$

där I är identitetsoperatoren som överför en figur i sig själv. På samma sätt definieras R_i^{-1} och T_i^{-1} så att

$$R_i^{-1}R_i = I, \quad T_i^{-1}T_i = I.$$

Med litet eftertanke inser man att

$$S_iS_i^{-1} = I, \quad R_iR_i^{-1} = I, \quad T_iT_i^{-1} = I, \quad R_iR_i = I.$$

Den sista relationen som även skrives $R_i^2 = I$ medför $R_i^{-1} = R_i$. Den matematiska formuleringen av *symmetri* är: En figur F är symmetrisk om det finns en reflektion R sådan att $F = RF$. Linjen m som hör till R kallas symmetrilinje till F .

För två godtyckliga translationer T_1 och T_2 gäller

$$T_1T_2 = T_2T_1 = T_3$$

där T_3 är en translation. Enligt sats 1 gäller för två godtyckliga S_1 och S_2 att

$$S_1S_2 = S_3$$

där S_3 är en rotation men i allmänhet har vi

$$S_1S_2 \neq S_2S_1$$

ty studera hur S_1S_2 och S_2S_1 avbildar O_1 . Enligt definitionen är $S_1(O_1) = O_1$ så $S_2S_1(O_1) = S_2(O_1)$ vilken punkt är skild från O_1 om $O_2 \neq O_1$. Därför är $S_1(S_2(O_1))$ ej $S_2(O_1)$ och olikheten följer.

SATS 3. Om R_1 och R_2 är två reflektioner så är $R_2R_1 = T$ eller $R_2R_1 = S$, där T är en translation och S en rotation.

BEVIS. Vi har $R_2R_1 = T$ när reflektionslinjerna m_1 och m_2 till R_1 och R_2 är parallella (se fig. 5).

Antag nu att m_1 och m_2 skär varandra i O och låt P och Q vara två punkter på m_1 resp. m_2 (se fig. 6).

Låt $A' = R_1A$ och $A'' = R_2A' = R_2R_1A$. Vi har

$$2\angle POA' = \angle AOA'$$

$$2\angle A'OQ = \angle A'OA''.$$

Addera:

$$2\angle POQ = \angle AOA'',$$

dvs R_2R_1 är en rotation med en vinkel dubbelt så stor som vinkeln mellan m_1 och m_2 .

KOROLLAR. *En rotation (vridning) ett halvt varv kring en punkt är ekvivalent med reflektion i två vinkelräta linjer. En glidreflektion är enligt definition en reflektion och translation, och man inser lätt att deras inbördes ordning ej spelar någon roll (fig 7). Om glidreflektionen betecknas G så har vi*

$$G = RT = TR.$$

Symmetrigrupper i planet. Av sats 3 framgår att de tre typer av transformationer man har i planet ej är artskilda. Man kan beskriva rotation och translation som upprepade reflektioner. Symmetri är reflektion. När man vill konstruera mönster har man därför ej så stora variationsmöjligheter som man skulle kunna tro. Man vill ju att mönsterfiguren skall upprepa sig på ett regelbundet sätt.

Vi skall göra följande indelning av tvådimensionella mönster:

- 1) De två punktgrupperna vilka är mönster som erhålles med rotation och reflektion men utan translation.
- 2) De sju frisgrupperna erhållna med rotation, reflektion och med translation i en riktning.
- 3) De sjutton tapetgrupperna erhållna med rotation, reflektion och translation i två riktningar.

Den matematiska definitionen av *grupp* finns i Appendix.

De två slagen av punktgrupper. Den första sortens punktgrupp, den cykliska, innehåller de mönster man får genom rotation av en figur kring en fix punkt. För att figuren skall komma tillbaka till sitt ursprungsläge så att man får ett mönster måste rotationsvinkeln vara $360/n$ grader, där n är ett heltal. Beteckna en sådan rotation S_n . Vi har $S_n^n = I$, ty vridning n gånger med $360/n$ grader överför en figur i sig själv. Grupperna som genereras av S_n brukar betecknas C_n . I fig. 8 har vi avbildat $C_1 \dots C_6$. För grupperna C_2 och C_3 har

vi givit två figurer för att illustrera hur mönstrets utseende kan bero på var rotationscentrum ligger i förhållande till den figur vi roterar.

Gruppen C_n har n element (=figurer). Sålunda har t ex C_5 precis 5 element, nämligen $I, S_5, S_5^2, S_5^3, S_5^4$ ($S_5^5 = I$).

Kontrollera att C_5 är en grupp enligt Appendix genom att göra en multiplikationstabell. Man har t ex

$$S_5^3 \cdot S_5^4 = S_5^7 = S_5^5 \cdot S_5^2 = I \cdot S_5^2 = S_5^2.$$

Kontrollera i multiplikationstabellen att varje S_5^i har invers.

Den andra sortens punktgrupp kallas *dihedral*. I en sådan grupp ingår förutom rotation kring en fix punkt även reflektion kring linjer M_i genom fixpunkten. (Eftersom en rotation kan beskrivas som två reflektioner enligt sats 3 kan de dihedrala grupperna beskrivas med enbart reflektioner.) För att det skall bli ett mönster måste, liksom vid de cykliska grupperna, rotationen vara $360/n$ grader och gruppen med den rotationen betecknas D_n . I fig 9 är $D_1 - D_6$ avbildade. Som exempel skriver vi upp de åtta elementen i D_4 , I, R, R^2, \dots, R^7 , eller om två reflektioner skrives som en rotation ($R^2 = S_4$):

$$I, R, S_4, RS_4, S_4^2, RS_4^2, S_4^3, RS_4^3.$$

De cykliska och de dihedrala grupperna är de enda isometrigrupperna om man bara använder rotation och reflektion.

De sju frisgrupperna. De mönster det här är fråga om är sådana, som upprepas vid translation i *en* riktning som t ex mönstret på en kaminfris. Frisgrupperna har oändligt många element, ty man studerar mönster, som varken börjar eller slutar på visst ställe, utan mönsterfiguren upprepas i det oändliga. Den enklaste frisgruppen, F_1 , är just upprepad translation av en figur (se fig 10).

Om translationen betecknas T så är elementen i F_1 :

$$\dots T^{-2}, T^{-1}, I, T, T^2, \dots$$

Om ett mönster skall upprepas i en riktning så är den enda rotation som kan komma ifråga den med ett halvt varv. Vi får gruppen F_2 (fig 11).

Om vi använder reflektion R_1 i translationsriktningen m_1 får vi ur gruppen F_1 , gruppen F_1^1 (fig 12).

Gruppen F_2 ger på samma sätt gruppen F_2^1 (fig 13). Om vi även betraktar reflektionen R_2 i riktningen m_2 vinkelrätt mot m_1 , har vi

$$R_1 R_2 = R_2 R_1.$$

Utan translation har vi därför endast fyra olika gruppelement

$$I, R_1, R_2, R_1 R_2$$

i F_2 . Kombinationen av dessa fyra element ger inga nya.

Med hjälp av reflektionen R_2 i en linje vinkelrät mot translationsriktningen får vi gruppen F_1^2 (fig 14). (Enligt sats 3 är $R_2^2 = T$.)

Om man försöker få ett nytt mönster ur F_2 genom reflektion R_2 i linje m_2 vinkelrät mot translationsriktningen m_1 får man om m_2 går genom en symmetripunkt i fig 11 återigen gruppen F_2^1 . Genom att låta m_2 gå genom en punkt mitt emellan symmetripunkterna i fig 11 får man en ny grupp F_2^2 (fig 15).

En reflektion R_2 kring en linje som ej går igenom eller mitt emellan symmetripunkterna ger ej något mönster, så vi kan ej få fler mönster ur F_1 genom enbart reflektion.

Möjligheten med glidreflektion G återstår. Eftersom $G^2 = RTRT = TR^2T = T^2$ är en translation, får vi en ny grupp F_1^3 (fig 16).

Försök att generera fler mönster ur F_1^3 genom reflektioner R_1 och R_2 ger tillbaka gamla mönster. Med formellt skrivsätt får vi $R_1 F_1^3 = F_1^1$ och $R_2 F_1^3 = F_2^2$.

Vi har nu konstruerat alla möjliga frisgrupper. De är sju stycken.

De sjutton tapetgrupperna. Vi skall nu studera mönster med translation i två olika riktningar. Beteckna dessa translationer T_1 och T_2 . Mönstren skall alltså vara sådana att om man applicerar $T_1^m T_2^n$ (m och n heltal) på dem så får man tillbaka samma mönster. Det enklaste mönstret av detta slag W_1 är avbildat i fig 17.

Punkterna T_1^m, T_2^n, O (m och n heltal, O origo) säges bilda ett gitter. Vissa gitter är invarianta (ändrar ej) vid rotation vissa vinklar, sålunda är t ex ett kvadratisk gitter invariant vid rotation ett kvarts varv. Vi har

SATS 4 (KRISTALLOGRAFIVILLKORET). *De enda rotationer i planet som lämnar gitter invarianta är de med vinklarna $360/2$, $360/3$, $360/4$ och $360/6$ grader.*

BEVIS. (Jfr fig 18.) Låt P vara en gitterpunkt och låt Q vara en av gitterpunkterna närmast P . Låt P' vara bilden av P vid rotation $360/n$ grader kring Q . Om vi har rotationsinvarians vid denna rotation så är P' gitterpunkt. Låt Q' vara bilden av Q vid rotation $360/n$ grader kring P' . Även Q' är gitterpunkt. Om $n = 6$ sammanfaller Q' och P (triangulärt gitter). Om $n = 5$ och om $n > 6$ ligger Q' närmare P än Q (se fig 18), vilket strider mot antagandet att Q var en av gitterpunkterna närmast P . Gitter med sådana rotationssymmetrier finns alltså ej. För $n = 4$ har vi kvadratisk och för $n = 3$ hexagonalt gitter. För $n = 2$ har vi ren translation av gittret och för $n = 1$ ligger gittret still.

De gitter (mönster) som är invarianta vid rotation $360/n$ grader betecknas W_n . De grupper som finns är alltså W_1, W_2, W_3, W_4 och W_6 . I figurerna 17 och 19-22 är mönstren avbildade. Bredvid varje mönster är en bild av dess enhetscell, dvs den minsta delfigur av mönstret som man kan använda för att bygga upp det. De små cirkulerna i enhetscellerna är de centra kring vilka man roterar cellen vid

uppbyggnad av mönstret och siffrorna vid cirklarna den bråkdel av ett helt varv som rotationen skall ske.

De fem nu beskrivna mönstren har erhållits med operationerna translation och rotation. Genom att även använda reflektion och glidreflektion kan man skaffa sig nya mönster. Vi beskriver dessa med figurer av mönstren och tillhörande (glid-)reflektionslinje (streckad i enhetscellen). Ur W_1 kan man få mönstren W_1^1 och W_1^2, W_1^3 ur W_2, W_2^1, W_2^2, W_2^3 och W_2^4 etc. På detta sätt får vi totalt sjutton mönster som är invarianta vid translation i två olika riktningar och några fler finns ej. Beviset av detta, nämligen att man ej kan finna fler (glid-)reflektionslinjer utelämnas här. (Se referens [1].) Några andra isometriska avbildningar än de använda finns enligt kapitlet om isometrier ej.

Enhetscellerna för W_3 och W_4 har vardera två symmetrilinjer som kan användas som reflektionslinjer. Försök beskriva de erhållna grupperna W_3^1, W_3^2, W_4^1 och W_4^2 . W_6 -cellen har en symmetrilinje att reflektera i vilket ger gruppen W_6^1 . Beskriv den. Alla sjutton tapetgrupperna är nu beskrivna.

Några historiska notiser. Redan Euklides (300 f.Kr.) studerade isometrier. Vetskapen att varje isometri är en rotation eller glidreflektion stammar troligen från Euler (1770-talet). Leonardo da Vinci (omkr 1500) förstod idén med grupperna C_k och D_k . Grupperna D_k verkar förekomma oftare både i naturen och i konsten än C_k -grupperna. Den första matematiska (=systematiska) beskrivningen av de 17 tapetgrupperna finns hos Fedorov (1891), som alltså visade att det ej finns fler grupper. Empiriskt var dessa grupper dock kända av antikens egyptier, greker och kineser. Morerna (1300-talet) kände till alla 17 tapetgrupperna ty de finns på väggarna i Alhambra i Granada.

Litteratur

- [1] Toth Fejes, L., *Regular Figures*. Pergamon Press 1964.
 [2] March, L. & Steadman, P., *The geometry of environment*. RIBA, London 1971.

Appendix.

En grupp är ett matematiskt system av element som kan kombineras med en operation vilken man ofta kallar multiplikation. Operationen skall till varje par av element i gruppen ge ett element i gruppen.

- 1) Om man tar tre element a, b och c i gruppen skall det gälla

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

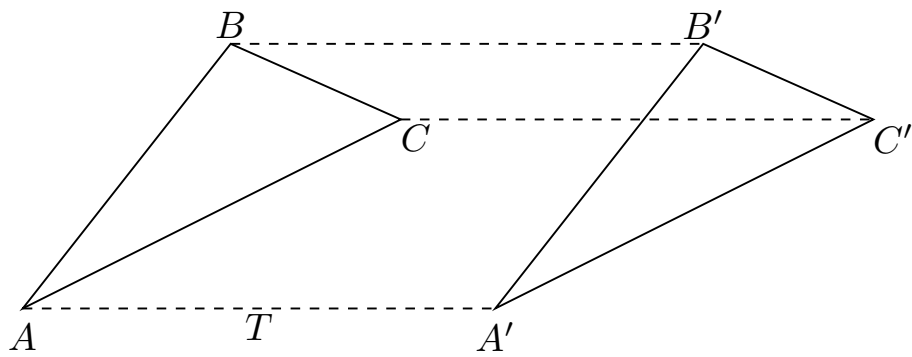
- 2) Varje element a har en invers a^{-1} så att $a^{-1} \cdot a \cdot b = b$, dvs verkan av a^{-1} tas ut av a (och tvärtom). Man skriver $a^{-1} \cdot a = a \cdot a^{-1} = I$ och kallar I enhetselement eller enhet. *Att multiplicera med enheten ändrar ej ett element.*

Jämför en grupp med de vanliga talen där man har *två* operationer, multiplikation och addition och deras inverser: division och subtraktion.

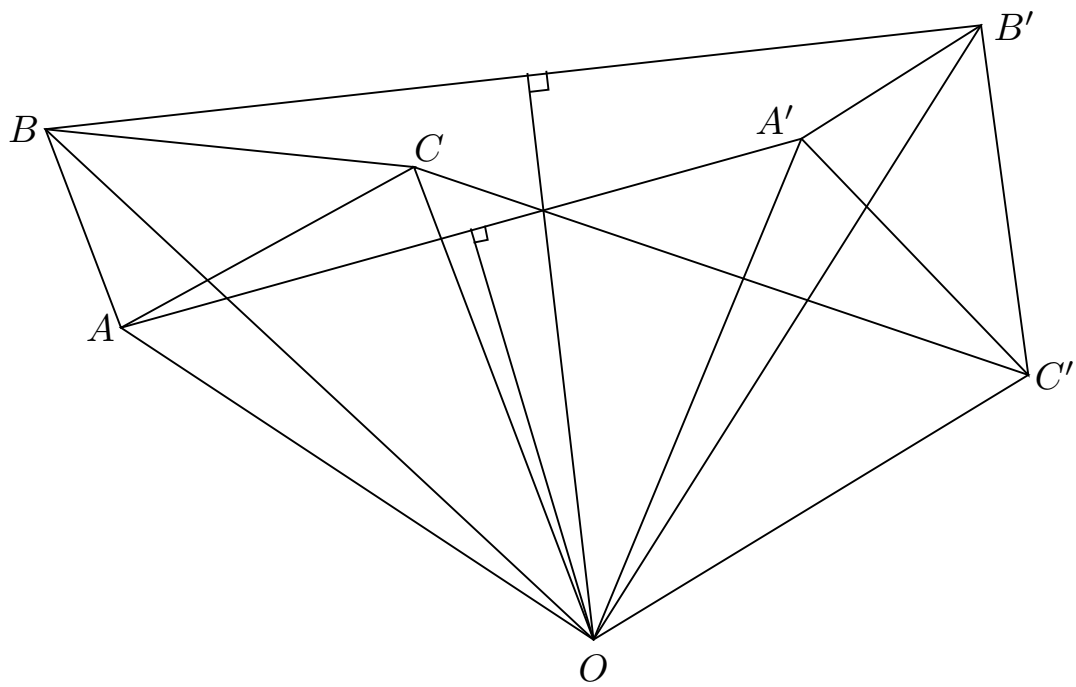
EXEMPEL 1 (PÅ GRUPP). De hela talen med addition som gruppoperation och 0 som enhet.

EXEMPEL 2 (PÅ GRUPP). De vanliga talen utom 0 med multiplikation som gruppoperation och 1 som enhet.

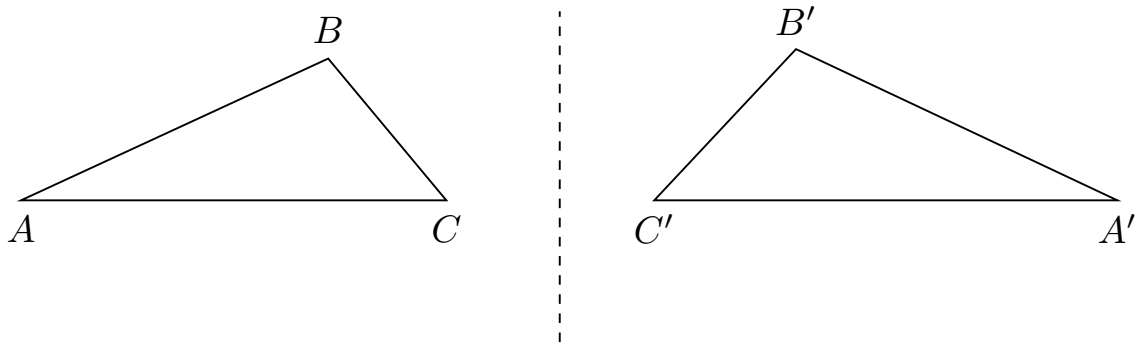
Kontrollera att grupperna 1) och 2) ovan är grupper enligt den matematiska definitionen.



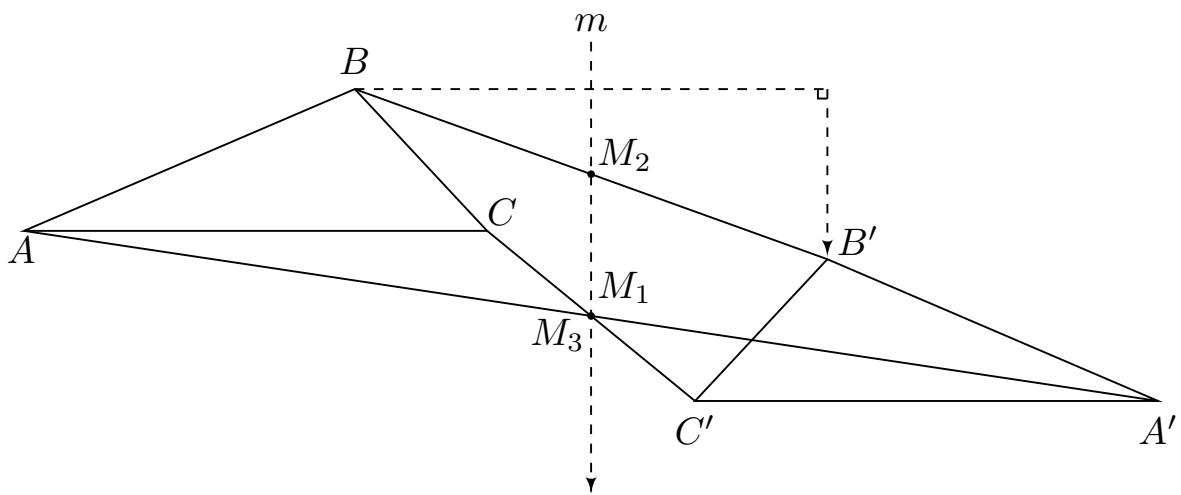
Figur 1



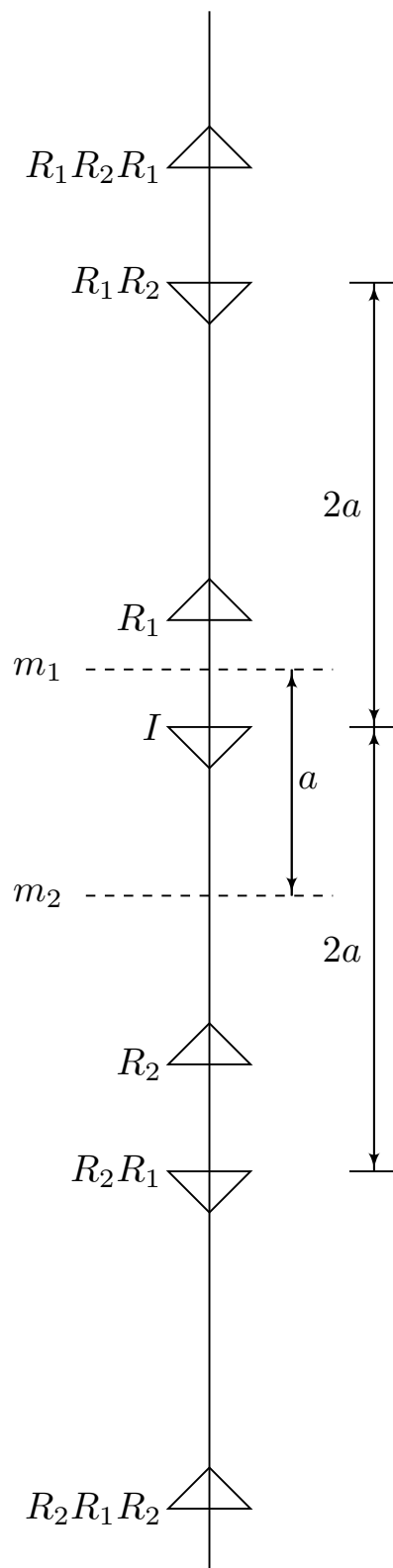
Figur 2



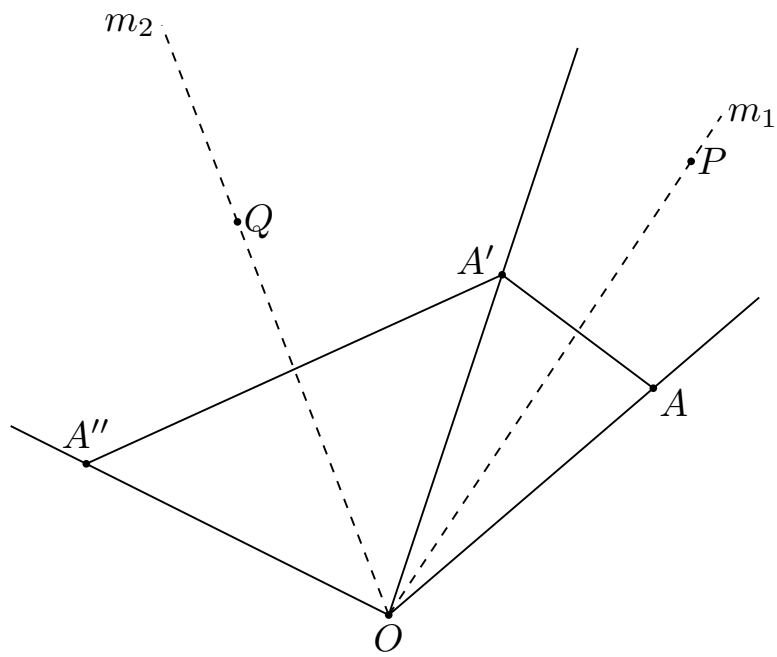
Figur 3



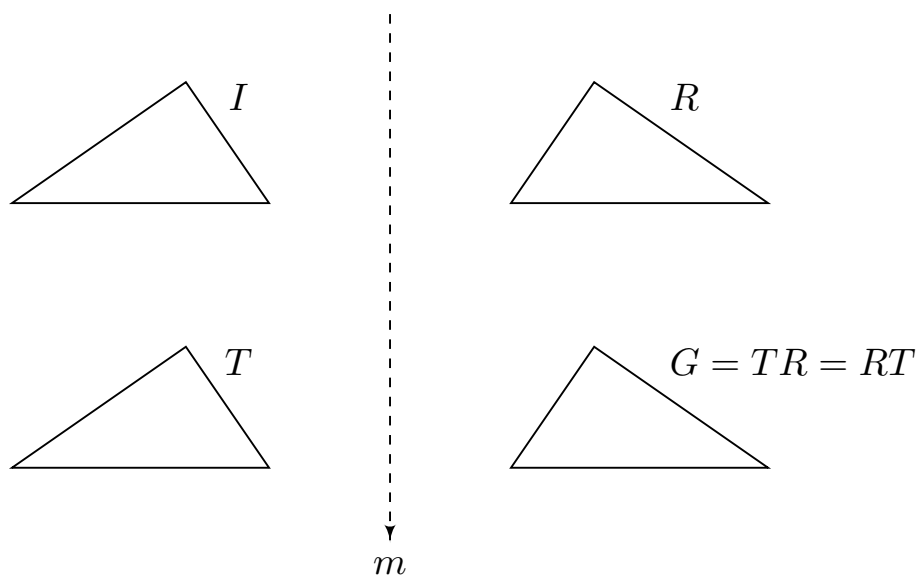
Figur 4



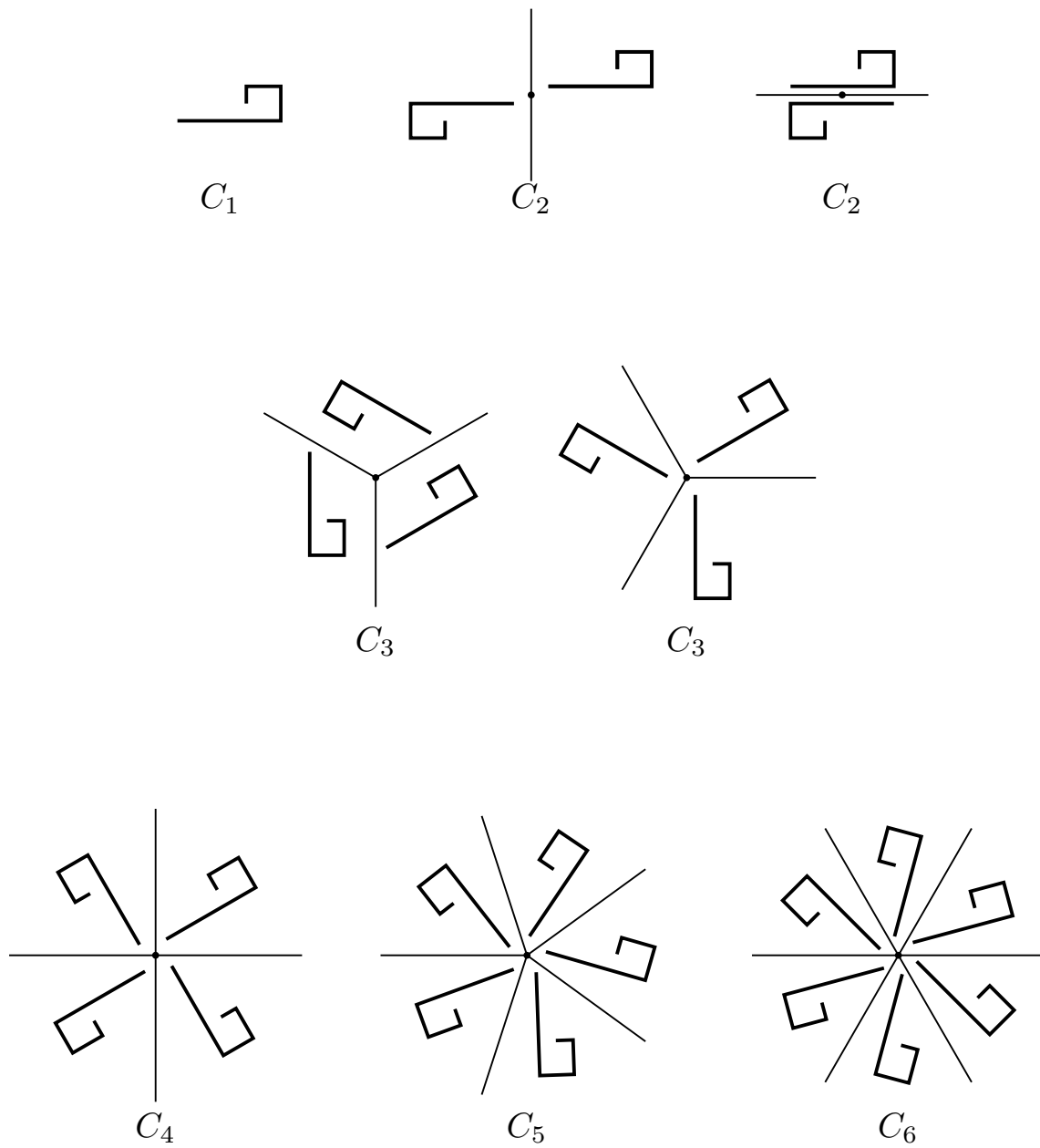
Figur 5



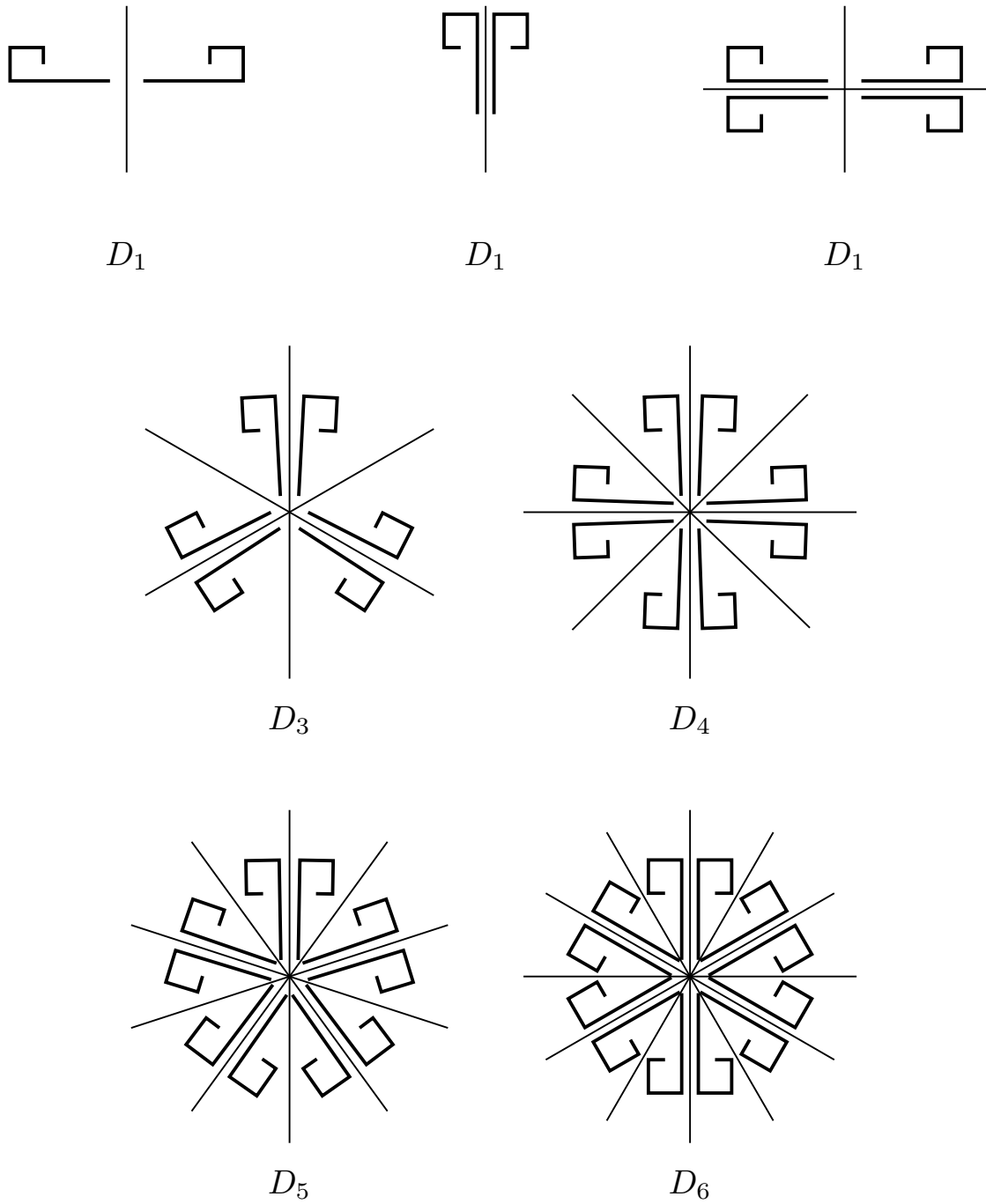
Figur 6



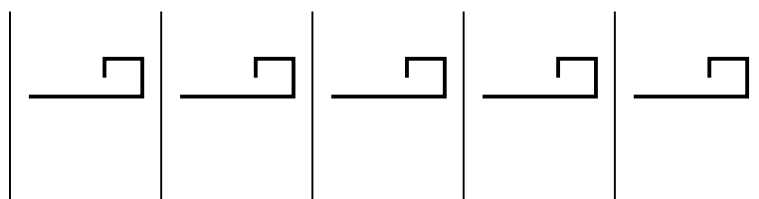
Figur 7



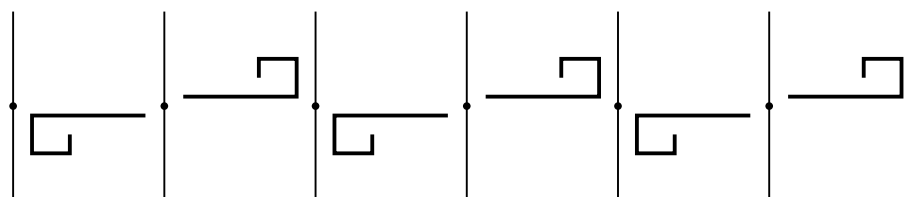
Figur 8



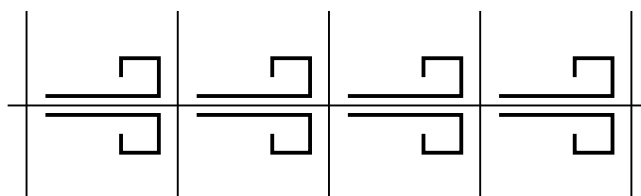
Figur 9

 F_1

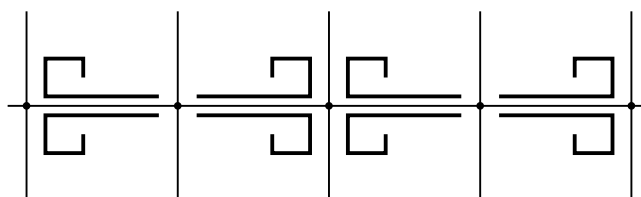
Figur 10

 F_2

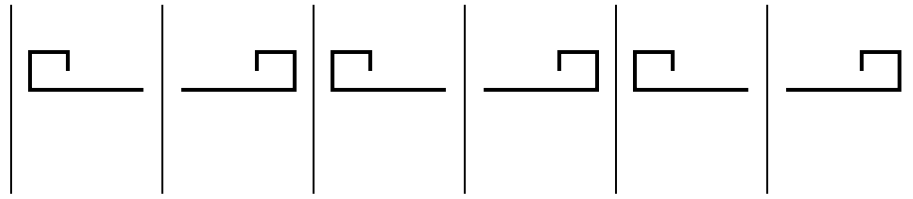
Figur 11

 F_1^1

Figur 12

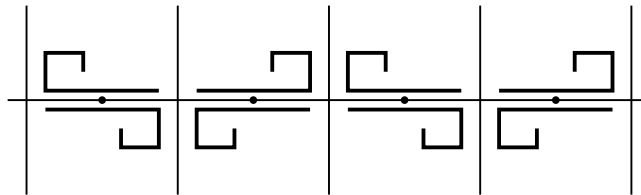
 F_2^1

Figur 13



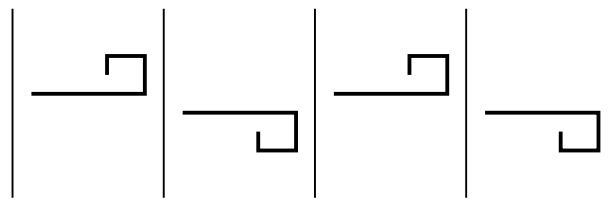
$$F_1^2$$

Figur 14



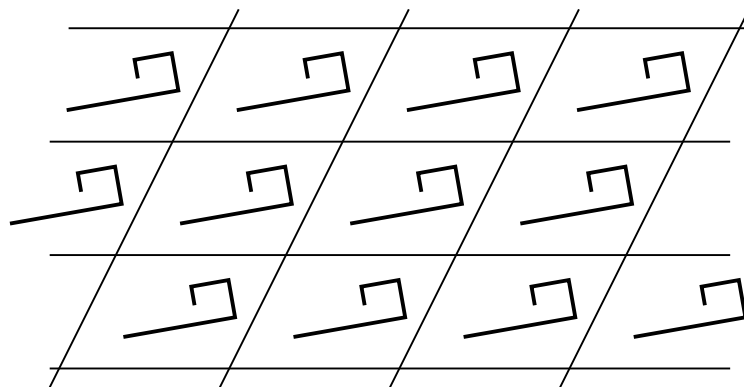
$$F_2^2$$

Figur 15



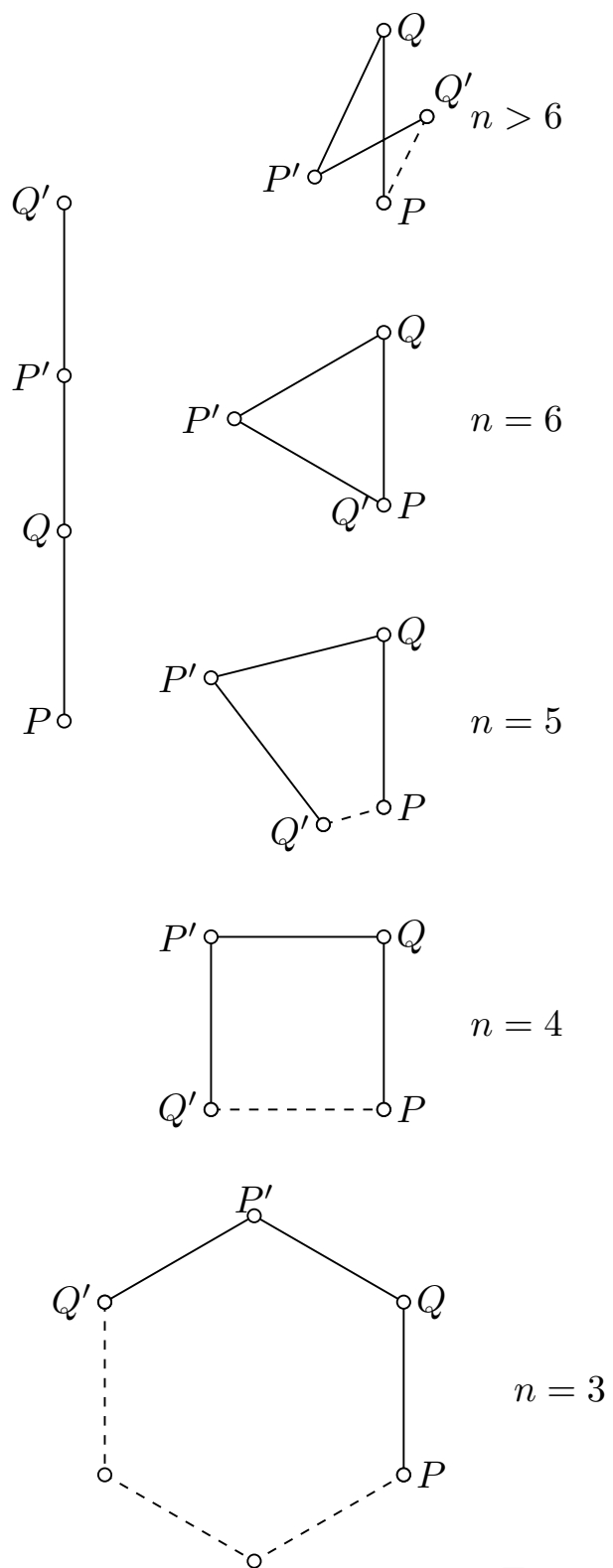
$$F_1^3$$

Figur 16

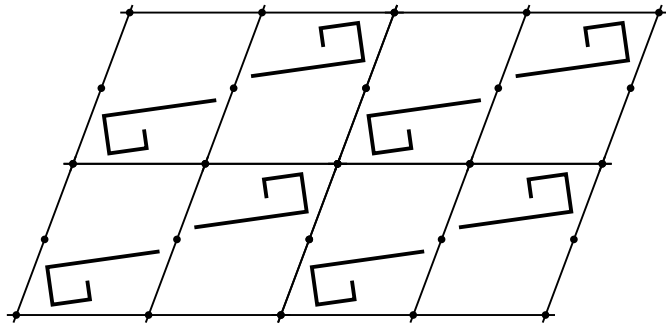


$$W_1$$

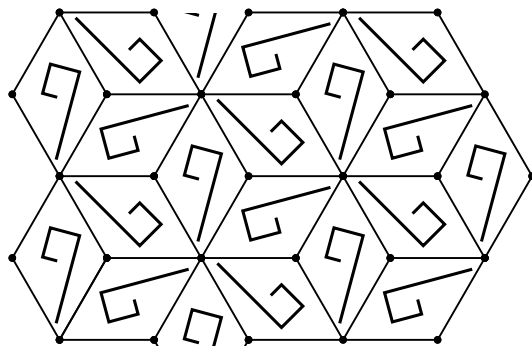
Figur 17



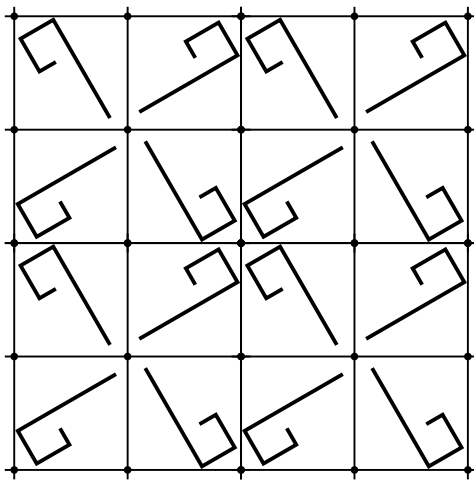
Figur 18



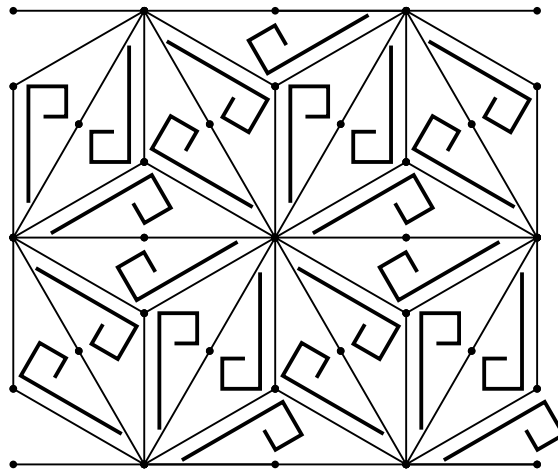
W_2
Figur 19



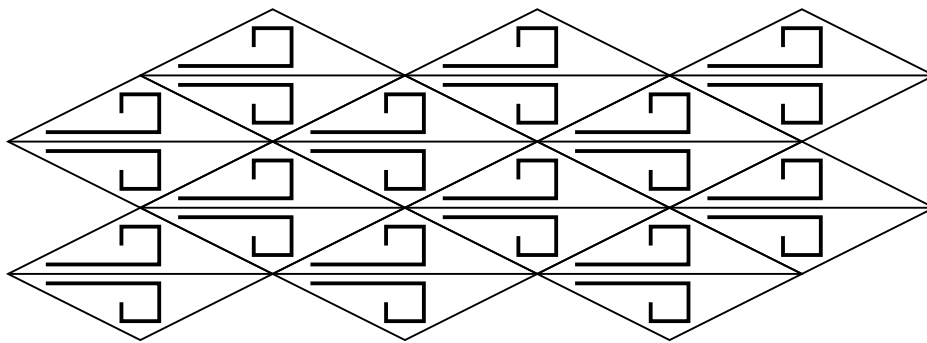
W_3
Figur 20



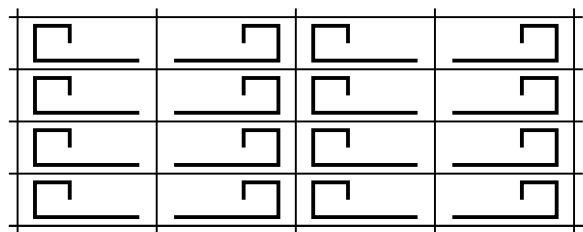
W_4
Figur 21



W_6
Figur 22

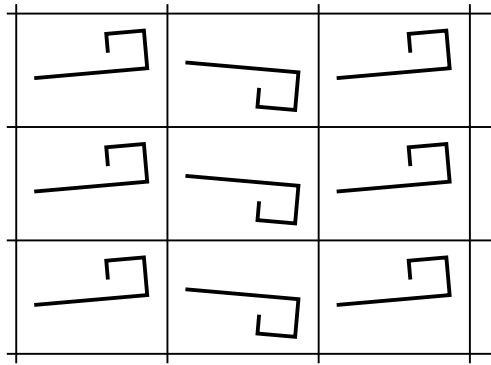


W_1^1
Figur 23

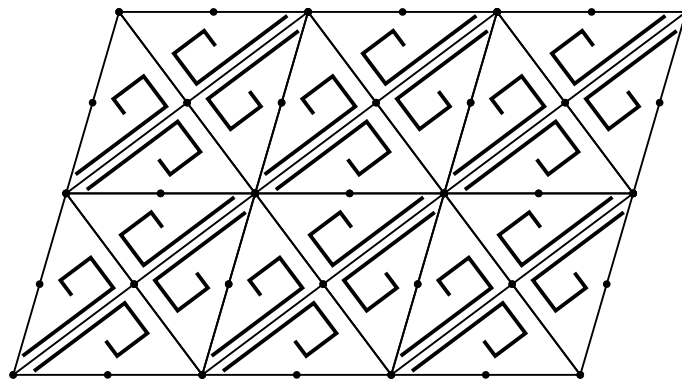


Figur 24

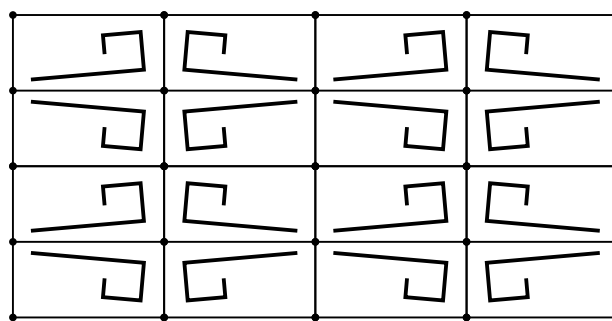
W_1^2


 W_1^3

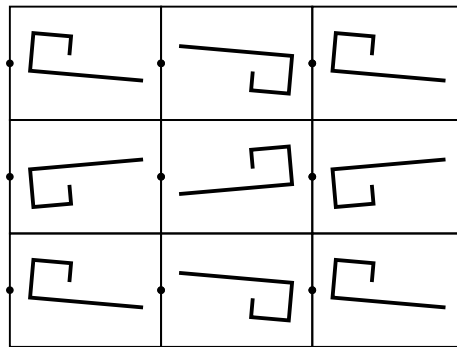
Figur 25


 W_2^1

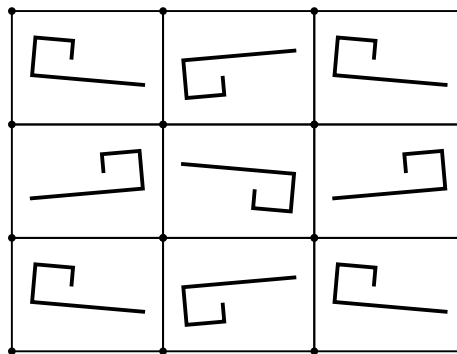
Figur 26


 W_2^2

Figur 27


 W_2^3

Figur 28


 W_2^4

Figur 29

Gör Din egen kurvkatalog

HANS RIESEL

KTH

Krav på utrustning. För denna uppgift måste du ha tillgång till en grafisk dataterminal, så att Du kan rita kurvor på dataskärmen. Du behöver inte ha tillgång till färggrafik. Dessutom måste Du kunna rita ut kurvorna på papper, antingen med en särskilt kurvskrivare, eller med en laserskrivare eller en mekanisk punktmatrisskrivare med hygglig upplösning. — Dessutom bör Du ha tillgång till programvara, som underlättar att rita kurvor på skärmen och sedan matar ut dem på papper (såvida Du inte vill göra sådana program själv).

Allmänt om kurvritning. När man har en kurva definierad genom någon geometrisk egenskap, måste denna först kläs i lämpliga formler, d.v.s. samband mellan koordinaterna för punkterna på kurvan. Antingen arbetar man i rätvinkliga koordinater (x, y) eller i polära koordinater (r, v) . Man kan också arbeta i s.k. parameterform, där x och y (eller r och v) båda är givna som funktioner av en hjälpvariabel t : $x = x(t)$, $y = y(t)$ respektive $r = r(t)$, $v = v(t)$. I samtliga fall måste man *begränsa området* i vilket man önskar få kurvan ritad. Vilken begränsning man väljer beror på hur kurvan ser ut. Ritar man t.ex. olika avsnitt av parabeln $y = x^2$, ser figurerna mycket olika ut, om man väljer $-1 \leq x \leq 1$, $0 \leq y \leq 1$ eller om man väljer $-10^6 \leq x \leq 10^6$, $0 \leq y \leq 10^6$. (Observera, att man bör välja intervall av jämförbar storlek för x och y , annars blir skalan konstig. Ganska vanligt hos program för kurvritning är, att programmet undersöker $x_{\max} - x_{\min}$ och $y_{\max} - y_{\min}$ och självt väljer skalan på x - och y -axlarna, så att kurvan fyller ut hela papperet. Denna teknik

har emellertid den nackdelen, att t.ex. alla ellipser ritas ut som cirklar, oavsett hur runda eller avlånga de är. Det är därför, som man i nyss nämnda parabel låter x ligga mellan -10^6 och 10^6 , trots att alla x -värden som används, bara kommer att ligga mellan -10^3 och 10^3 .)

Ritning av en kurva. För varje kurva skall Du göra följande: Programmera in formlerna som ger kurvan. Hur detta skall göras beror på, hur det allmänna program för kurvritning, som Du har tillgång till är uppbyggt. Antagligen får Du skriva ett par funktioner i Pascal, som beräknar x - och y -koordinaterna för parametervärdet t , och sedan lägga in dessa i kurvritningsprogrammet. Sedan skall Du köra programmet med utmatning på dataskärmen, och kontrollera att allting ”ser bra ut”, d.v.s. att formlerna är rätt inprogrammerade och verkar ge riktiga värden. När detta är klart kör Du ut kurvan på papper med lämplig text, nämligen kurvans namn (om den har något) och dess ekvation (om Du har möjlighet att mata ut formler).

Kurvorna. Följande kurvor är lämpliga att låta ingå i kurvkatalogen. Har Du andra kurvor, som Du vill ta med, går det naturligtvis bra.

Polynom

$$y = \frac{(x^2 - 1)(x^2 - 4)(x^2 - 9)}{36}, \quad -3.2 \leq x \leq 3.2.$$

Ellips

$$\frac{x^2}{16} + \frac{y^2}{4} = 1, \quad \text{eller}$$

$$\begin{cases} x = 4 \cos t \\ y = 2 \sin t, \end{cases} \quad 0 \leq t \leq 2\pi.$$

Hyperbel

$$\frac{x^2}{4} - y^2 = 1, \quad \text{eller}$$

$$\begin{cases} x = \pm 2 \cosh t \\ y = \sinh t, \end{cases} \quad -1 \leq t \leq 1.$$

Superellips

$$\left(\frac{|x|}{3}\right)^{2.5} + \left(\frac{|y|}{2}\right)^{2.5} = 1, \quad \text{eller}$$

$$\begin{cases} x = 3 \operatorname{sign}(\cos t) |\cos t|^{0.8} \\ y = 2 \operatorname{sign}(\sin t) |\sin t|^{0.8}, \end{cases} \quad 0 \leq t \leq 2\pi.$$

Tredjegradskurva

$$y^2 = x^3 + x^2, \quad \text{eller}$$

$$\begin{cases} x = t^2 - 1 \\ y = t(t^2 - 1), \end{cases} \quad -1.6 \leq t \leq 1.6.$$

Cartesii blad

$$x^3 + y^3 = 3xy, \quad \text{eller}$$

$$\begin{cases} x = \frac{3t^2(1-t)}{t^3 + (1-t)^3} \\ y = \frac{3t(1-t)^2}{t^3 + (1-t)^3}, \end{cases} \quad -2 \leq t \leq 3.$$

Konkoid

$$r = \frac{1}{\sin v} + 2, \quad \begin{cases} 0.164 \leq v \leq 2.976 \\ 3.24 \leq v \leq 6.18. \end{cases}$$

Kissoid

$$x(x^2 + y^2) = 2y^2, \quad \text{eller}$$

$$\begin{cases} x = \frac{2t^2}{1+t^2} \\ y = \frac{2t^3}{1+t^2} \end{cases} \quad -3 \leq t \leq 3.$$

Pascals snäcka

$$(x^2 + y^2 - 2x)^2 = x^2 + y^2, \quad \text{eller}$$

$$\begin{cases} x = \cos t(2 \cos t + 1) \\ y = \sin t(2 \cos t + 1), \end{cases} \quad 0 \leq t \leq 2\pi.$$

Kedjelinje

$$y = 0.4 \cosh x, \quad -3 \leq x \leq 3.$$

Släpkurva

$$x = 3 \ln \frac{3 + \sqrt{9 - y^2}}{y} - \sqrt{9 - y^2}, \quad 0.1 \leq y \leq 3.$$

Astroid

$$x^{\frac{2}{3}} + y^{\frac{2}{3}} = 1, \quad \text{eller}$$

$$\begin{cases} x = \cos^3 t \\ y = \sin^3 t, \end{cases} \quad 0 \leq t \leq 2\pi.$$

Booths lemniskata

$$(x^2 + y^2)^2 = 4x^2 + y^2, \quad \text{eller}$$

$$\begin{cases} x = \frac{2 \cos t}{\cos^2 t + 4 \sin^2 t} \\ y = \frac{4 \sin t}{\cos^2 t + 4 \sin^2 t}, \end{cases} \quad 0 \leq t \leq 2\pi.$$

Bernoullis lemniskata

$$(x^2 + y^2)^2 = x^2 - y^2, \quad \text{eller}$$

$$\begin{cases} x = \frac{\sin t}{1 + \cos^2 t} \\ y = \frac{0.5 \sin 2t}{1 + \cos^2 t}, \end{cases} \quad 0 \leq t \leq 2\pi.$$

Hypocykloid med 5 spetsar

$$\begin{cases} x = 4 \cos t + \cos 4t \\ y = 4 \sin t - \sin 4t, \end{cases} \quad 0 \leq t \leq 2\pi.$$

Epicykloid med 5 spetsar

$$\begin{cases} x = 6 \cos t - \cos 6t \\ y = 6 \sin t - \sin 6t, \end{cases} \quad 0 \leq t \leq 2\pi.$$

Cykloid

$$\begin{cases} x = t - \sin t \\ y = 1 - \cos t, \end{cases} \quad -2\pi \leq t \leq 2\pi.$$

Rullkurva

$$\begin{cases} x = 6 \cos t - 1.6 \cos 6t \\ y = 6 \sin t - 1.6 \sin 6t, \end{cases} \quad 0 \leq t \leq 2\pi.$$

Cirkel i polära koordinater

$$r = 2 \cos v, \quad -\frac{\pi}{2} \leq v \leq \frac{\pi}{2}.$$

Ellips i polära koordinater

$$r = \frac{2}{1 - 0.5 \cos v}, \quad 0 \leq v \leq 2\pi.$$

Hyperbel i polära koordinater

$$r = \frac{2}{1 - 2 \cos v}, \quad 1.3 \leq v \leq 5.$$

Kardioid

$$r = 1 - \cos v, \quad 0 \leq v \leq 2\pi.$$

Epicykloid med 15 spetsar

$$\begin{cases} x = 16 \cos t - \cos 16t \\ y = 16 \sin t - \sin 16t, \end{cases} \quad 0 \leq t \leq 2\pi.$$

Rosenkurva

$$r = \sin 5v, \quad 0 \leq v \leq 2\pi.$$

Arkimedes' spiral

$$r = 0.2v, \quad 0 \leq v \leq 20.$$

Logaritmisk spiral

$$r = e^{0.1v}, \quad -20 \leq v \leq 20.$$

Cirkelevolvent

$$\begin{cases} x = \cos t + t \sin t \\ y = \sin t - t \cos t, \end{cases} \quad 0 \leq t \leq 10.$$

Cirkelevolvent

$$\begin{cases} x = \cos t + t \sin t \\ y = \sin t - t \cos t, \end{cases} \quad 0 \leq t \leq 500.$$

Periodisk funktion

$$y = 10 \sin \frac{x}{2} + 5 \sin \frac{x}{3}, \quad 0 \leq x \leq 300.$$

Nästan periodisk funktion

$$y = 10 \sin \frac{x}{2} + 5 \sin \frac{x}{\sqrt{8}}, \quad 0 \leq x \leq 300.$$

Kvadratrötter och kedjebråk

HANS RIESEL

KTH

Regelbundna kedjebråksutvecklingar. Varje reellt tal $x > 0$ kan utvecklas i ett s.k. regelbundet kedjebråk:

$$(1) \quad x = b_0 + \frac{1}{b_1 + \frac{1}{b_2 + \frac{1}{b_3 + \dots}}}$$

Ett mera kompakt skrivsätt för (1) är

$$(2) \quad x = b_0 + \left| \frac{1}{b_1} \right| + \left| \frac{1}{b_2} \right| + \left| \frac{1}{b_3} \right| + \dots$$

Här är b_0 heltal ≥ 0 och övriga b_i heltal > 0 . Talen b_i kallas utvecklingens *delnämnamre*. Att finna framställningen (2) är mycket enkelt. Med beteckningen $[x]$ för den s.k. heltalsdelen i x , det största heltalet $\leq x$, finner man

$$(3) \quad \begin{aligned} b_0 = [x], \quad x_1 = \frac{1}{x - b_0}, \quad b_1 = [x_1], \quad x_2 = \frac{1}{x_1 - b_1}, \\ b_2 = [x_2], \quad x_3 = \frac{1}{x_2 - b_2}, \quad \dots, \quad b_n = [x_n], \quad x_{n+1} = \frac{1}{x_n - b_n}, \quad \dots \end{aligned}$$

Beräkningen avslutas, så snart som ett av talen x_i blir ett heltal. Detta inträffar, om det ursprungliga talet x är ett rationellt tal. I annat fall säger man att kedjebråksutvecklingen blir *oavslutad* eller *oändlig*.

EXEMPEL 1. Den regelbundna kedjebråksutvecklingen för $\sqrt{2}$ beräknas på följande sätt:

$$b_0 = [\sqrt{2}] = 1, \quad x_1 = \frac{1}{\sqrt{2} - 1} = \sqrt{2} + 1$$

$$b_1 = [\sqrt{2} + 1] = 2, \quad x_2 = \frac{1}{\sqrt{2} + 1 - 2} = \sqrt{2} + 1.$$

Härifrån upprepas kalkylen periodiskt. Därför kommer b_2 och alla efterföljande b_i att bli = 2, och utvecklingen, i detta fall oändlig, blir

$$(4) \quad \sqrt{2} = 1 + \frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \dots$$

EXEMPEL 2. Basen för det naturliga logaritmsystemet, talet e , får följande utveckling. Talvärdet på $e = 2.71828182\dots$ ger

$$\begin{aligned} b_0 &= 2, & x_1 &= 1/0.71828182\dots = 1.39221119\dots \\ b_1 &= 1, & x_2 &= 1/0.39221119\dots = 2.54964677\dots \\ b_2 &= 2, & x_3 &= 1/0.54964677\dots = 1.81935024\dots \\ b_3 &= 1, & x_4 &= 1/0.81935024\dots = 1.22047928\dots \\ b_4 &= 1, & x_5 &= 1/0.22047928\dots = 4.53557347\dots \\ b_5 &= 4, & x_6 &= 1/0.535573\dots = 1.867157\dots \\ b_6 &= 1, & x_7 &= 1/0.867157\dots = 1.153193\dots \\ b_7 &= 1, & x_8 &= 1/0.153193\dots = 6.527707\dots \\ b_8 &= 6, & x_9 &= 1/0.5277\dots = 1.8949\dots \\ b_9 &= 1, & x_{10} &= 1/0.8949\dots = 1.1173\dots \\ b_{10} &= 1, & x_{11} &= 1/0.1173\dots = 8.5226\dots \\ b_{11} &= 8, & \dots & \end{aligned}$$

Denna kalkyl låter oss ana Eulers berömda utveckling av e :

$$(5) \quad e = 2 + \frac{1}{1} + \frac{1}{2} + \frac{1}{1} + \frac{1}{1} + \frac{1}{4} + \frac{1}{1} + \frac{1}{1} + \frac{1}{6} + \frac{1}{1} + \frac{1}{1} + \frac{1}{8} + \dots$$

Regelbundna kedjebraåksutvecklingar för kvadratrötter. Man kan visa, att varje kvadratisk irrationalitet x (alltså ett irrationellt

tal, som satisfierar en andragradsekvation $Ax^2 + Bx + C = 0$ med heltalskoefficienter A , B och C) får en regelbunden kedjebråksutveckling, som är *periodisk*. Om x väljs till \sqrt{D} , där D är ett positivt heltal, som inte är en jämn kvadrat, föregås perioden av $b_0 = [\sqrt{D}]$. Periodens sista delnämnare visar sig vara $2b_0$, medan alla andra delnämnare är mindre än $2b_0$. Detta kan användas för att upptäcka, när perioden är slut.

Ett datorprogram för beräkning av kedjebråksutvecklingen för kvadratrötter. Med nedanstående Pascal-program kan Du låta datorn beräkna kedjebråksutvecklingen för \sqrt{D} . När den första perioden i utvecklingen har genomlöpts, avbryts beräkningarna, och resultatet skrivs ut.

```

Program Kedrot(input,output);
(*Beräknar första perioden i den regelbundna
  kedjebråksutvecklingen av sqrt(D) *)

Label 1;
Var D,p0,p1,q0,rot,b0,i : integer;
    sqrtD                : real;

Begin
write('Mata in D för kedjebråksutv. av sqrt(D):');
read(D); writeln;
sqrtD:=sqrt(D); rot:=trunc(sqrt(D+0.5));
write('Den regelbundna kedjebråksutvecklingen av sqrt(');
writeln(D:1,') är:',rot:1,', följt av perioden');
i:=0; p0:=0; q0:=1; b0:=rot;

1: i:=i+1; p1:=b0*q0-p0; q0:=(D-sqr(p1))div q0; p0:=p1;

```



```

b0:=(rot+p0)div q0;
if b0<>*rot then begin write(b0:1,','); goto 1 end;
writeln(b0:1,',...'); writeln('Perioden genomlöst! i=',i:1);
end.

```

Idén bakom datorprogrammet. När man hunnit ett stycke på väg i utvecklingen (2), kan situationen beskrivas på följande sätt:

$$(6) \quad x = b_0 + \frac{1}{b_1} + \dots + \frac{1}{b_{n-1}} + \frac{1}{x_n},$$

där

$$(7) \quad x_n = b_n + \frac{1}{b_{n+1}} + \frac{1}{b_{n+2}} + \dots$$

Antag nu, att vi vid utvecklingen av $x = \sqrt{D}$ kommit fram till att $x_n = (\sqrt{D} + p_n)/q_n$. (För $n = 0$ har man ju $x_0 = x = \sqrt{D} = (\sqrt{D} + 0)/1$, alltså $p_0 = 0$ och $q_0 = 1$.) Då blir $b_n = [x_n]$, som beräknas som $(\text{rot}+pn)\text{div } qn$. Vidare blir

$$(8) \quad \frac{\sqrt{D} + p_n}{q_n} = b_n + \frac{\sqrt{D} - (b_n q_n - p_n)}{q_n} = b_n + \frac{\sqrt{D} - p_{n+1}}{q_n},$$

som ger

$$(9) \quad x_{n+1} = \frac{q_n}{\sqrt{D} - p_{n+1}} = \frac{q_n(\sqrt{D} + p_{n+1})}{D - p_{n+1}^2} = \frac{\sqrt{D} + p_{n+1}}{q_{n+1}},$$

där $q_{n+1} = (D - p_{n+1}^2)/q_n$. Den springande punkten är tydligen, att q_n går jämnt upp i talet $D - p_{n+1}^2$. Försök att bevisa detta! Om Du studerar programmet, skall Du finna att det är ovanstående formler för p_{n+1} och q_{n+1} , som är återgivna för $n = 0$.

Några frågor. Provkör datorprogrammet för $D = 18$. Kontrollera att programmet svarar "4, följt av perioden 4,8,..." Detta betyder, att

$$(10) \quad \sqrt{18} = 4 + \frac{1}{4} + \frac{1}{8} + \frac{1}{4} + \frac{1}{8} + \frac{1}{4} + \frac{1}{8} + \dots$$

Kör nu programmet för alla $D \leq 24$, som inte är jämna kvadrattal. Försök svara på följande frågor:

1. Kan Du finna *någon* regelbundenhet i resultaten? Vad blir utvecklingen om $D = x^2 + 1$, där x är heltal? Utvecklingen om $D = x^2 - 1$? Utvecklingen om $D = x^2 + 2$?
2. Försök komma underfund med, vilka D -värden som ger periodlängden 1 och vilka som ger periodlängden 2.
3. Kör några större värden på D , och försök uppskatta, hur lång perioden kan bli, när den är som längst!

Kedjebråk och approximationer. Den viktigaste tillämpningen av kedjebråk är inom approximationsteorin. Om man avbryter utvecklingen (2) efter n delnämnamre, och beräknar

$$(11) \quad b_0 + \frac{1}{b_1} + \frac{1}{b_2} + \dots + \frac{1}{b_n} = \frac{A_n}{B_n},$$

visar det sig, att de rationella talen A_n/B_n utgör goda approximationer av x . Man kan visa, att

$$(12) \quad \left| x - \frac{A_n}{B_n} \right| \leq \frac{1}{b_{n+1} B_n^2}.$$

EXEMPEL. Om utvecklingen (5) av talet e ovan avbryts omedelbart före delnämnamren 6, fås

$$(13) \quad 2 + \frac{1}{1} + \frac{1}{2} + \frac{1}{1} + \frac{1}{1} + \frac{1}{4} + \frac{1}{1} + \frac{1}{1} = \frac{193}{71} = 2.718310.$$

För denna approximation gäller att

$$(14) \quad |e - 2.718310| = 0.000028 < 0.000033 = \frac{1}{6 \cdot 71^2}.$$

Ekvationerna $x^2 - Dy^2 = \pm 1$. Utgående från (12) visar man lätt, att approximationen A_n/B_n till \sqrt{D} för vissa, lämpligt valda värden på n , är så god, att $A_n^2 - DB_n^2 = \pm 1$. (Eftersom \sqrt{D} är irrationellt, kan aldrig $A_n^2 - DB_n^2$ bli $= 0$, och eftersom uttrycket är ett heltal, är de till absoluta beloppet minsta värdena, som det kan anta, ± 1 .) Man kan med andra ord lösa de diofantiska ekvationerna

$$(15) \quad x^2 - Dy^2 = \pm 1$$

med hjälp av kedjebråksutvecklingen för \sqrt{D} . (Diofantiska ekvationer är ekvationer, där *heltalsvärden* på de obekanta söks.) Det återstår att finna de ovan omtalade värdena på n , för vilka A_n/B_n utgör en tillräckligt god approximation av \sqrt{D} för att $|A_n^2 - DB_n^2|$ skall anta ett så lågt värde som 1. Detta kan ske på följande sätt:

Antag, att x och y satisfierar $x^2 - Dy^2 = \pm 1$. Då finner man i tur och ordning att

$$(16) \quad \begin{aligned} \left(\frac{x}{y}\right)^2 - D &= \pm \frac{1}{y^2}, \\ \left(\frac{x}{y} - \sqrt{D}\right) \left(\frac{x}{y} + \sqrt{D}\right) &= \pm \frac{1}{y^2}, \\ \frac{x}{y} - \sqrt{D} &\approx \pm \frac{1}{2\sqrt{D}y^2}, \end{aligned}$$

eftersom $x/y \approx \sqrt{D}$. Om alltså för något n delnämnen b_{n+1} i kedjebråksutvecklingen av \sqrt{D} blir så stor som $\approx 2\sqrt{D}$, kan villkoret (12) uppfyllas. Men $2\sqrt{D} \approx 2b_0$, och vi har tidigare sett, att *sista delnämnen i varje period är just $2b_0$* . Om vi alltså avbryter kedjebråksutvecklingen omedelbart före en av dessa maximala delnämnen och beräknar motsvarande rationella approximation A_n/B_n till \sqrt{D} , satisfierar A_n och B_n en av de diofantiska ekvationerna (15).

EXEMPEL. $D = 2$ ger de diofantiska ekvationerna $x^2 - 2y^2 = \pm 1$. Kedjebråksutvecklingen (4) för $\sqrt{2}$ ger, om den avbryts strax före alla ställen, där delnämnameren $2b_0 = 2$ står, d.v.s. var som helst, följande approximationer till $\sqrt{2}$:

$$(17) \quad \frac{1}{1}, \quad \frac{3}{2}, \quad \frac{7}{5}, \quad \frac{17}{12}, \quad \frac{41}{29}, \quad \frac{99}{70}, \quad \frac{239}{169}, \quad \frac{577}{408}, \quad \frac{1393}{985}, \quad \dots$$

Dessa approximationer ger i tur och ordning

$$(18) \quad 1^2 - 2 \cdot 1^2 = -1, \quad 3^2 - 2 \cdot 2^2 = 1, \quad 7^2 - 2 \cdot 5^2 = -1, \quad 17^2 - 2 \cdot 12^2 = 1, \quad \dots,$$

alltså varannan gång en lösning till $x^2 - 2y^2 = -1$ och varannan gång en lösning till $x^2 - 2y^2 = +1$. — Man kan bevisa, att alla lösningar till $x^2 - 2y^2 = \pm 1$ kan fås på detta sätt.

Några uppgifter. Med hjälp av datorprogrammet kan Du angripa följande frågeställningar:

1. Beräkna de minsta lösningarna till de diofantiska ekvationerna $x^2 - Dy^2 = \pm 1$ för alla $D \leq 24$, som inte är jämna kvadrattal. Har alltid ekvationen $x^2 - Dy^2 = -1$ någon lösning? Kan Du bevisa något i den ena eller andra riktningen?
2. Redan för måttliga värden på D blir minsta lösningen till $x^2 - Dy^2 = \pm 1$ väldigt stor. Kan Du beräkna minsta lösningen för $D = 94$?
3. Försök bevisa påståendet som gjordes i samband med formel (9) ovan, att q_n alltid går jämnt upp i talet $D - p_{n+1}^2$. Ledning: Ett induktionsbevis, som utnyttjar att $q_0 = 1$ går jämnt upp i $D - p_1^2$ är inte särskilt komplicerat!
4. Om Du klarat punkt 3 ovan, är Du nära ett bevis för att utvecklingen av \sqrt{D} i regelbundet kedjebråk alltid blir *periodisk*. Bevisa

först att talen p_n och q_n i formel (8) är begränsade. Sedan följer att antalet kombinationer av p_n och q_n som kan förekomma i utvecklingen är ändligt. Alltså måste, förr eller senare, för några värden på m och n , $p_m = p_n$ och $q_m = q_n$, varvid kalkylen upprepas från denna punkt, och utvecklingen alltså blir periodisk!

5. Använd kedjebråksutvecklingarna, som Du fått fram med hjälp av datorprogrammet, till att ange rationella approximationer med en noggrannhet motsvarande 3 korrekt avrundade decimaler för alla \sqrt{D} för $D \leq 24$.

Litteraturen om kedjebråk är tyvärr knapp, men en trevlig liten bok är dock

Schmidt, A.L., *Kædebrøker*. Gyldendal, Köpenhamn 1967.

Periodiska decimalbråk

HANS RIESEL

K T H

Inledning. Du har säkert lagt märke till att decimalerna i vissa enkla tal såsom $1/3 = 0.333\dots$ eller $1/11 = 0.0909\dots$, består av en upprepning av en ständigt återkommande följd av samma siffror. Sådana tal kallas *periodiska decimalbråk*. Det visar sig, att varje rationellt tal p/q , förvandlat till decimalbråk, ger antingen ett *avslutat decimalbråk* (såsom $5/8 = 0.625$), ett *rent periodiskt decimalbråk* (såsom $1/37 = 0.027027\dots$) eller ett *orent periodiskt decimalbråk* (såsom $1/6 = 0.1666\dots$), där perioden föregås av en eller flera siffror, som bildar den s.k. *aperiodiska delen*.

Uppgiften. Du skall med hjälp av nedanstående datorprogram försöka komma underfund med lagarna efter vilka de olika fallen uppträder.

```

Program Perdec(input,output);
Label 1;
Var p,q,t,i,j,s : integer;
    rester      : array[0..2000] of integer;

Begin write('Ge q för ber. av p/q: '); read(q); writeln;
for p:=1 to q-1 do begin
    write(p:1,'/',q:1,'=0. '); for j:=0 to q do rester[j] :=-1;
    t:=p; rester[p] :=0;
    for i:=1 to q do begin
        t:=10*t; s:=t Div q; t:=t Mod q; if rester[t] =-1 then

```

```

begin
write(s:1); rester[t] :=i; end
else
  begin if t=p then begin writeln(s:1,'...');
writeln(
  'Rent periodiskt decimalbråk.Antalsiffror=',i:1);
goto 1 end;
if t=0 then begin writeln;
writeln(' Avslutat decimalbråk!');
goto 1 end;
writeln(s:1,'...');
write(' Örent periodiskt bråk. Aperiodiska delen har ');
write(rester[t] :1,' siffr');
if (rester[t] )>1 then write('or')
  else write('a'); write('. Periodlängden
är ');
writeln(i-rester[t] :1);
goto 1 end
end;
1: writeln; end
end.

```

Programmet gör följande: Efter inmatning av nämnaren q beräknas decimalbråksutvecklingarna för alla tal p/q , där $1 \leq p \leq q - 1$. Dessa utvecklingar skrivs ut, jämte en karakterisering av utvecklingens typ. Mata in programmet och kör det för $q = 6$ och verifiera att Du får

$$1/6 = 0.1666\dots, \quad \text{orent periodiskt}$$

$$2/6 = 1/3 = 0.3333\dots, \quad \text{rent periodiskt}$$

$$3/6 = 1/2 = 0.5, \quad \text{avslutat}$$

$$4/6 = 2/3 = 0.6666\dots, \quad \text{rent periodiskt}$$

$$5/6 = 0.8333\dots, \quad \text{orent periodiskt.}$$

Vilka tal p/q ger *avslutade* decimalbråk? Vilka tal ger *orena, periodiska* decimalbråk?

Rena, periodiska decimalbråk. Du kan använda datorprogrammet till att undersöka de rena periodiska decimalbråken. Försök att svara på följande frågor:

1. Antag att nämnaren q inte innehåller några faktorer 2 eller 5. Antag vidare att täljaren p inte har några faktorer gemensamma med q , så att bråket p/q alltså inte kan förkortas. Jämför periodlängderna för de olika bråken p/q , då $1 \leq p \leq q-1$. Vad finner Du för resultat?
2. I vissa fall är perioden för $1/q$ ovanligt lång, $q-1$ siffror. Kan Du karakterisera de q -värden som ger denna maximala periodlängd?
3. För de fall att $1/q$ har maximal periodlängd enligt punkt 2 ovan, står de olika perioderna som fås för alla talen p/q i en enkel relation till varandra. Hur?
4. Det finns något som i förströelsematematiken brukar kallas cykliska tal. Ett sådant är 142857. Ett sådant tals egenskaper visas enklast med följande uppställning:

$$1 \times 142857 = 142857$$

$$2 \times 142857 = 285714$$

$$3 \times 142857 = 428571$$

$$4 \times 142857 = 571428$$

$$5 \times 142857 = 714285$$

$$6 \times 142857 = 857142$$

$$7 \times 142857 = 999999$$

En (liten multiplikator) \times (ett cykliskt tal) ger alltså ett resultat, som är det cykliska talet med någon eller några siffror flyttade från början till slutet av talet. Försök att känna igen det cykliska talet 142857. Kan Du förklara talets egenskaper mot bakgrund av periodiska decimalbråksutvecklingar? Var kommer nästa cykliska tal,

0588235294117647,

ifrån?

5. För de *primtal* q , för vilka periodlängden l i decimalbråksutvecklingen av $1/q$ inte är maximal, vad kan sägas om l ? Lämpliga testfall är $q = 13$ och $q = 37$. Är de perioder, som uppträder, också cykliska tal? Kan Du bringa något system i de olika perioderna?
6. Om $1/q_1$ har periodlängden l_1 och $1/q_2$ har periodlängden l_2 , vilken periodlängd får då $1/q_1q_2$? Lämpliga testfall är $q_1 = 37$, $q_2 = 41$ och $q_1 = 7$, $q_2 = 11$. Kan Du formulera en allmän regel, som har att göra med faktorerna i q_1 och q_2 ?
7. Kan Du finna något samband mellan periodlängden för $1/q$, där q är en faktor i $10^n - 1$ och n ? Ange alla tal, som har en 7-siffrig, periodisk decimalbråksutveckling.

Talsystem med annan bas än 10. Genom några enkla ändringar i datorprogrammet kan Du lätt utföra ovanstående undersökning för ett talsystem med en annan bas B än 10. Det som skall ändras är faktorn 10 i satsen $t:=10*t$, som ändras till $t:=B*t$. (Det kan vara praktiskt att deklarerar variabeln B , och läsa in den i början av körningen.) Om $B > 10$, måste också utskriften av "B-alerna" (motsvarigheten till decimalerna) ändras. Du måste införa symboler för siffrorna motsvarande $10, 11, \dots, B - 1$, lämpligen A, B, C, \dots . I det hexadecimala talsystemet t.ex., där $B = 16$, får man sätta $10 = A, 11 = B, 12 = C, 13 = D, 14 = E$ och $15 = F$. Sedan får man ändra utskriftssatsen `write(s:1)` (och motsvarande satser som använder `writeln`) för siffrorna i perioden till

```
if s<10 then write(s:1) else write(chr(s+55):1)
```

Efter att ha infört dessa ändringar i programmet, provkör och försök att svara på följande frågor:

1. Undersök periodlängden för olika baser för t.ex. bråket $1/7$. Hur går det, om Du väljer baser som är kongruenta mod 7, t.ex. baserna 3, 10 och 17? Eller baserna 2, 9 och 16?
2. Låt q vara ett primtal och undersök periodlängden för $1/q$ för olika baser B . Kan Du säga något om de periodlängder som uppstår? Finns det för ett valt värde på q alltid baser B , som ger den maximala periodlängden $q - 1$? (Sådana tal B kallas i talteorin för *primitiva rötter* till primtalet q .)
3. Om Du har hittat en primitiv rot till q , kan Du härleda andra? Kan Du finna alla primitiva rötter, utgående från en viss av dem? Hur många primitiva rötter $< q$ har primtalet q ?

Om Du vill veta, hur en stor matematiker har behandlat problemet, kan Du läsa i

Gauss, C.F., *Untersuchungen über höhere Arithmetik*. Chelsea Publishing Company 1965, s 366–373 samt s 453.

Summan av två heltalskvadrater

HANS RIESEL

K T H

Problemställning. *Vissa tal kan skrivas som summan av två heltalskvadrater, andra inte!* Så är t.ex. $13 = 2^2 + 3^2$, $9 = 0^2 + 3^2$ medan $n = 11$ inte kan skrivas som $x^2 + y^2$. Hur kan man övertyga sig om detta senare? Jo undersök vad $n - x^2$ blir för $x = 0, 1, 2, \dots$. Uttrycket blir (fortfarande för $n = 11$) 11, 10, 7, 2, $-5 \dots$, alltså aldrig en jämn kvadrat. Eftersom y^2 alltid skall vara ≥ 0 , behöver vi tydligen inte gå längre. Egentligen hade vi bara behövt gå tills $n - x^2 < n/2$, ty om $n = x^2 + y^2$, är minst ett av talen x^2 och $y^2 \geq n/2$. (Både x^2 och y^2 kan inte vara $< n/2$, ty då skulle ju summan bli $< n$.)

Antalet framställningar. Du kanske har lagt märke till, att vissa tal kan skrivas som $x^2 + y^2$ på flera olika sätt, såsom $25 = 3^2 + 4^2 = 0^2 + 5^2$ eller, tydligare, $65 = 1^2 + 8^2 = 4^2 + 7^2$. Man kan fråga sig, på hur många olika sätt ett givet tal kan skrivas som $x^2 + y^2$. I detta sammanhang är Lagranges identitet

$$(a^2 + b^2)(c^2 + d^2) = (ac \mp bd)^2 + (ad \pm bc)^2$$

av betydelse.

En geometrisk tolkning. En annan fråga är, hur många tal $\leq N$, som kan skrivas som $x^2 + y^2$. Titta på figur 1. Där ser Du att varje punkt med heltalskoordinater (x, y) , en s.k. gitterpunkt, i eller på cirkeln $x^2 + y^2 = N$ med radien \sqrt{N} motsvarar en framställning av något heltal $\leq N$ som $x^2 + y^2$.

De flesta gitterpunkterna i eller på cirkeln faller i grupper om 8. Om nämligen $x \neq y$ och även $xy \neq 0$, får man ju framställningarna

$$n = (\pm x)^2 + (\pm y)^2 = (\pm y)^2 + (\pm x)^2,$$

som ger sammanlagt 8 olika möjliga framställningar av talet n . Är $y = \pm x$ eller $x = 0$ eller $y = 0$, får man endast 4 möjligheter, nämligen

$$n = (\pm x)^2 + (\pm x)^2 \quad \text{resp.} \quad n = (\pm x)^2 + 0^2 = 0^2 + (\pm x)^2.$$

För punkten $(0, 0)$ slutligen har man den enda framställningen $0 = 0^2 + 0^2$.

Nu är antalet gitterpunkter i cirkeln med radien \sqrt{N} ungefär lika med cirkelns yta $= \pi N$. Därför måste det totala antalet framställningar av alla heltal $\leq N$ som $x^2 + y^2$ vara ungefär $= \pi N$, med ett fel som är av högst samma storleksordning som cirkelns omkrets $2\pi\sqrt{N}$. I figuren kan Du se hur enhetskvadraterna, en för varje gitterpunkt i eller på cirkeln, nära täcker över cirkeln.

UPPGIFTEN. Du skall med hjälp av dator undersöka ovan relaterade problemställningar. Skaffa Dig först en överblick över vilka tal n , som överhuvudtaget kan skrivas som $x^2 + y^2$. Till hjälp har Du nedanstående Pascal-program, som gör följande beräkningar: För varje tal n i ett givet intervall $(nstart, nslut)$ skrivs talet n och antalet framställningar $r(n)$ ut. $r(n)$ beräknas på samma sätt som i ovan givna gitterpunktsbeskrivning av framställningarna. Vidare beräknas ytan av cirkelringen i vilken gitterpunkterna finns, $\pi(n + 1 - nstart)$, vilket skall jämföras med $\sum_{nstart}^n r(i)$, för att ge en uppfattning om hur väl gitterpunktsmodellen ovan stämmer. För att Du skall kunna uppskatta avvikelserna från gitterpunktsmodellen, beräknas även differensen, dividerad med \sqrt{n} . I programmet beräknas vidare, hur stor bråkdel f av talen i intervallet, som

överhuvudtaget kan framställas som $x^2 + y^2$. Slutligen skrivs primfaktoruppdelningen av talet n ut. — För att spara utskriftsutrymme, hoppas de tal över, som inte kan skrivas som en summa av två kvadrater.

Några detaljer i datorprogrammet. Ett sätt att känna igen en jämn kvadrat z är att undersöka om

$$z = \text{sqr}(\text{trunc}(\text{sqr}(z) + 0.000001)).$$

(Tillägget 0.000001 har gjorts för att klara av fallet då $\sqrt{x^2}$ i datorns aritmetik råkar bli en aning under x , varvid $\text{trunc}(x)$ skulle ge $x - 1$ i stället för x .)

Att hitta alla framställningar av n som $x^2 + y^2$, där $0 \leq x \leq \sqrt{n/2}$ och $\sqrt{n/2} \leq y \leq \sqrt{n}$, görs snabbast genom att skriva

```
for y:=trunc(sqr(n)+0.000001) downto ymin do ...,
```

där $ymin = \sqrt{(n/2)}$, och inte genom satsen

```
for x:=0 to ymin do ...
```

Antalet x -värden är nämligen ca. $0.7\sqrt{n}$, medan antalet y -värden är endast ca. $0.3\sqrt{n}$ varför det första sättet går mer än dubbelt så snabbt som det senare.

Uppdelning av talet n i primtal och tryckning av faktorerna sker med hjälp av proceduren faktor i programmet, som utför ovan beskrivna beräkningar för alla tal n då $nst \leq n \leq nsl$.

```
Program x2y2(input,output);
```

```
Const pi=3.1416;
```

```
Var n,x2,x,r,y,ymin,sum,vx,sn,nst,ns1,n1 : Integer;
```

```
Procedure faktor(n:integer);
```

```

Var g,p,vg : integer;
Begin write(' ');
  while n Mod 2 = 0 do begin write(2:1); n:=n Div 2;
    if n>1 then write('*') end;
  while n Mod 3 = 0 do begin write(3:1); n:=n Div 3;
    if n>1 then write('*') end;
  p:=5; g:=trunc(sqrt(n)+0.000001); vg:=0;
  while p<=g do begin
    while n Mod p = 0 do begin write(p:1); n:=n Div p;
      vg:=1;
      if n>1 then write('*') end;
    while n Mod(p+2) = 0 do begin write(p+2:1);
      n:=n Div(p+2);
      vg:=1; if n>1 then write('*') end;
    p:=p+4; if vg=1 then begin g:=trunc(sqrt(n)+0.000001);
      vg:=0 end;
    end;
    if n>1 then write(n:1)
  end;

Begin
Write('Ge start- och slutvärden för tabellen: ');
read(nst,nsl); writeln;
Writeln(
'   n   r   sum pi(n-n0+1) diff/sqrt(n)   f   n=pqr...');
Writeln; sn:=0; sum:=0;
for n:=nst to nsl do begin
  r:=0; ymin:=trunc(sqrt(n div 2)+0.000001);
  if 2*sqr(ymin){n then ymin:=ymin+1; vx:=0;
  for y:=trunc(sqrt(n)+0.000001) downto ymin do begin

```

```

x2:=n-sqr(y); x:=round(sqrt(x2)+0.000001);
  if x2=sqr(x) then begin vx:=1;
    if (x=0) or (x=y) then r:=r+4 else r:=r+8 end end;
sum:=sum+r; if vx=1 then begin sn:=sn+1; n1:=n-nst+1;
  write(n:7,r:4,sum:7,pi*n1:7:1,(sum-pi*n1)/sqrt(n):12:2);
  write(sn/n1:8:2); faktor(n); writeln; end
end
end.

```

Provkör programmet, och se om Du kan finna några lagbundenheter i de värden som datorn matar ut. Lämpliga testintervall: (1, 400) och (1000, 1200).

Försök svara på följande frågor:

1. Eftersom Lagranges identitet visar, att en produkt av tal, som vart och ett kan skrivas som summan av två kvadrater, självt kan skrivas på detta sätt, kan Du börja med att titta på, vilka *primtal*, som kan skrivas som summan av två kvadrater. Ledning: Om det *udda* talet $n = x^2 + y^2$, så måste ett av talen x och y vara jämnt och det andra talet vara udda. Antag, att det är $x = 2x'$ som är jämnt och $y = 2y' + 1$ som är udda. Vad blir då $x^2 + y^2 \pmod{4}$? Kan Du verifiera detta genom datorkörningar?
2. Trots att primtalen av formen $4k + 3$ tydligen inte kan skrivas som en summa av två kvadrater, visar datorkörningarna att faktorerna 3, 7, 11, ... ibland förekommer i utskriften. Är det något som skiljer förekomsten av dessa faktorer i utskriften från förekomsten av primfaktorer av formen $4k + 1$? Kan Du förklara fenomenet?
3. Försök nu komma underfund med, hur $r(n)$ beror på antalet primfaktorer av formen $4k + 1$ i talet n . Lämpliga testvärden: 5, 65, 1105, 32045 och dessa tal tagna gånger 2, 4 och 8. Prova sedan multipla primfaktorer. Testa t.ex. talen 65, 325, 1625 och 8125.

4. Försök att formulera regler som ger $r(n)$, om primfaktoruppdelningen av n antas känd!
5. Som Du kan se av datorkörningarna, ligger värdet av $\sum r(n)$ över en cirkelring nära cirkelringens yta. Man kan också uttrycka detta så, att den *genomsnittliga storleksordningen* hos funktionen $r(n)$ är $= \pi$, i följande mer precisa formulering:

$$\lim_{n \rightarrow \infty} \frac{r(1) + r(2) + \dots + r(n)}{n} = \pi.$$

Om Du studerar den kolumn i datorutskriften, där skillnaden

$$\left(\sum_1^N r(n) - \pi N \right) / \sqrt{N}$$

har skrivits ut, vågar Du kanske gissa något om storleksordningen på avvikelserna mellan $\sum r(n)$ och πN ?

6. Undersök nu hur *antalet tal*, som kan skrivas som en summa av två kvadrater, i ett litet intervall kring N ändrar sig, när N växer. Lämpliga testintervall: (50, 150), (950, 1050), (9900, 10100), (99800, 100200) o.s.v., så högt upp som Du kan köra. (Hur högt det går beror på vilket värde som `maxint` i Pascal har i Din dator.)
7. Som Du kan se på datorutskriften, så avtar medelvärdet f för antalet tal långsamt, när N växer. Kan Du finna någon lag för hur f avtar? Ledtråd: Eftersom alla tal, som överhuvudtaget kan skrivas som en summa av två kvadrater, har med primtalen av formen $4k+1$ att göra, kanske det är den minskade primtalstätheten högre upp i talserien, som orsakar att f minskar, när N ökar? Avtar möjligen f i samma takt som primtalen?

Om Du vill veta mer om detta problem kan Du t.ex. läsa

Hardy, G.H., & Wright, E.M., *An Introduction to the theory of numbers*. Fifth edition, Oxford Univ. Press, Oxford 1979, s 241–243, 270–271 och 299–302.

Chandrasekharan, K., *Introduction to Analytic Number Theory*. Springer 1968, s 29.

Stokastisk geometri

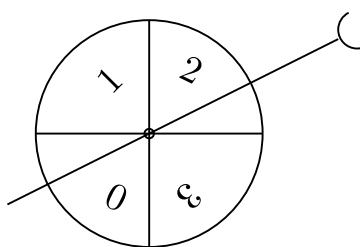
LENNART RÅDE

Chalmers Tekniska Högskola och Göteborgs Universitet

Inledning. I geometrin studerar man geometriska objekt och deras inbördes relationer. Exempel på geometriska objekt är räta linjer, sträckor, trianglar, cirklar osv. Även i den stokastiska geometrin studerar man geometriska objekt men nu är dessa slumpmässiga (stokastiska), dvs de genereras av en slumpmekanism. Den stokastiska geometrin har stor aktualitet bl a på grund av en mångfald tillämpningar t ex inom fysik, biologi, teknik osv.

Det är ofta svårt att lösa problem i den stokastiska geometrin analytiskt dvs med penna och papper. Man använder därför ofta simulering, i allmänhet med hjälp av en dator. Man utgår härvid från *slumptal* mellan 0 och 1, som datorn bildar med hjälp av en slumptalsgenerator. Man behöver i allmänhet inte programmera datorn att bilda slumptal. De flesta programmeringsspråk innehåller nämligen en instruktion, som bildar slumptal mellan 0 och 1 med en slumptalsgenerator. I t ex BASIC är instruktionen "RND" en sådan instruktion, och vidare gäller att instruktionen "RANDOMIZE" ger slumptalsgeneratorn en slumpmässig start. För att genomföra det här specialarbetet behöver Du ha tillgång till en dator. Det går utmärkt med en fickdator (programmerbar miniräknare).

Inledande uppgifter. Som en inledning ges här några uppgifter, som egentligen ej hör hemma i den stokastiska geometrin, även om de har geometrisk anknytning. Men de ger en nyttig bakgrund till de följande uppgifterna.



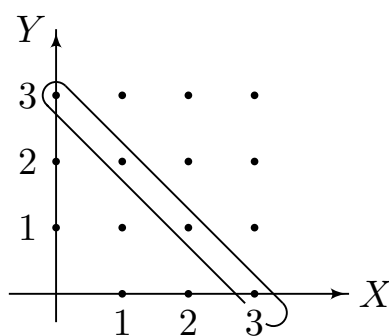
Figur 1

Man kan i BASIC simulera spel på lyckohjulet i figur 1 med hjälp av instruktionen $INT(4 * RND)$, där INT betecknar heltalsdel, t ex $INT(\pi) = 3$, $INT(6,51) = 6$.

UPPGIFT 1. Utforma ett program, som t ex 1000 gånger upprepar försöket att spela två gånger på lyckohjulet i figur 1 och varje gång beräknas summan $X + Y$ av de båda erhållna poängantalen X och Y . Beräkna också medelvärdet av de erhållna observationerna av $X + Y$ och sammanställ dessa i en frekvenstabell. Givetvis kan programmet utformas så att datorn beräknar medelvärde och gör sammanställningen i frekvenstabell.

UPPGIFT 2. Beräkna väntevärde för det simulerade försöket i uppgift 1. Med väntevärde för ett slumpmässigt försök menas summan av produkterna av försökets utfall och sannolikheterna för dessa utfall. För t ex försöket att kasta en symmetrisk tärning är väntevärdet

$$\frac{1}{6} \cdot 1 + \frac{1}{6} \cdot 2 + \frac{1}{6} \cdot 3 + \frac{1}{6} \cdot 4 + \frac{1}{6} \cdot 5 + \frac{1}{6} \cdot 6 = 3.5$$



Figur 2

Av figur 2 framgår att t ex sannolikheten är $4/16$ för att summan av X och Y är 3.

UPPGIFT 3. Upprepa gärna uppgifterna 1 och 2 men för symmetriska lyckohjul med utfallen $\{0, 1, 2, \dots, N - 1\}$ med t ex $N = 5, 6$ och 7. Härled eventuellt en formel, som ger väntevärdet för godtyckligt värde på N .

Slumpmässiga punkter på sträcka.



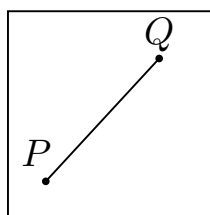
Figur 3

På en sträcka av längden 6 har 7 punkter markerats symmetriskt enligt figur 3. Betrakta försöket att slumpmässigt välja två av dessa punkter och bestämma avståndet L mellan de valda punkterna. Man förutsätts ej kunna välja samma punkt båda gångerna så att avståndet L kan inte vara 0.

UPPGIFT 4. Simulera t ex 1000 upprepningar av detta försök och ge frekvensfördelning och medelvärde för de observerade avstånden. Beräkna också väntevärdet.

UPPGIFT 5. Upprepa gärna uppgift 4 för en sträcka av längden N med $N + 1$ symmetriskt markerade punkter för t ex $N = 7, 8, 9, 10$.

Slumpmässiga punkter i kvadrat.



Figur 4

Betrakta försöket att slumpmässigt välja två punkter P och Q i en kvadrat med sidan 1 och sedan beräkna längden av sträckan PQ .

UPPGIFT 6. Simulera t ex 1000 gånger detta försök och beräkna medelvärde av längderna av de observerade sträckorna PQ . Med integralkalkyl kan man visa att väntevärdet av PQ är

$$\frac{2 + \sqrt{2}}{15} + \frac{1}{3} \ln(1 + \sqrt{2}).$$

Jämför det observerade medelvärdet och väntevärdet.

UPPGIFT 7. Upprepa uppgift 6 men nu under förutsättning att punkterna P och Q väljs slumpmässigt i en kub. I detta fall är väntevärdet

$$\frac{17\sqrt{2}}{105} - \frac{2\sqrt{3}}{35} + \frac{4}{105} + \frac{2}{5} \ln(2 + \sqrt{3}) + \frac{1}{5} \ln(1 + \sqrt{2}).$$

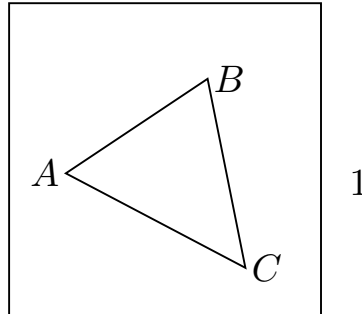
Slumpmässig triangel.



Figur 5

Två punkter P och Q väljs på måfå på en stav av längden 1. Staven bryts i dessa punkter. Vad är sannolikheten att man kan bilda en triangel av de tre erhållna bitarna.

UPPGIFT 8. Utforma ett datorprogram som genomför detta försök t ex 1000 gånger och som bestämmer antalet gånger som man kan bilda en triangel av de tre erhållna sträckorna. Den observerade relativa frekvensen är en skattning av sannolikheten för triangel.

Tre punkter i kvadrat.

Figur 6

Betrakta försöket att slumpmässigt välja tre punkter A, B och C i en kvadrat med sidan 1 och bestämma arean T av triangeln ABC .

UPPGIFT 9. Simulera detta försök t ex 1000 gånger och beräkna medelvärdet av de erhållna triangelareorna.

UPPGIFT 10. Betrakta det slumpmässiga försöket ovan. Skatta med simulering sannolikheten att den erhållna triangeln ABC är trubbvinklig.

Litteratur

Råde, L., *Simulering*. Studentlitteratur, Stockholm 1987.

Frankering og computer-nettverk

ØYSTEIN J. RØDSETH

Universitetet i Bergen

Beskrivelse av oppgaven. I denne oppgaven vil du bruke kombinatorikk, tallteori og muligens også litt analyse. Oppgaven er delt i to. Første delen tar for seg et problem som man ofte finner i *underholdningsmatematikken*, og som vanligvis er beskrevet som et problem med frimerker og konvolutter. I den andre delen skal du benytte resultater fra første del til å konstruere effektive computer-nettverk av en viss type.

Notasjon. Små latinske bokstaver vil i denne oppgaven betegne hele tall, og det vil fremgå fra sammenhengen at noen av disse alltid er ≥ 0 , mens andre alltid er > 0 . Men i punkt 6 (og eventuelt også i punkt 13) kan det kanskje lønne seg å betrakte en av de variable som vilkårlig *reell* fremfor heltallig.

Videre, for et reelt tall α , betegner vi med $\lfloor \alpha \rfloor$ det største hele tall $\leq \alpha$, mens $\lceil \alpha \rceil$ betegner det minste hele tall $\geq \alpha$.

Frankering. I et land har man bare frimerker pålydende 1 krone og 5 kroner. Når en konvolutt har plass til høyst 4 frimerker, kan man frankere beløpene 0,1,2,3,4,5,6,7,8, mens 9 kan ikke frankeres. Vi setter $n_4(5) = 8$.

Mer generelt, hvis en konvolutt har plass til høyst h frimerker, la $n_h(5) + 1$ være det minste positive heltallige beløp som *ikke* kan frankeres. — Med $n = n_h(5)$, kan vi altså frankere beløpene 0, 1, \dots , n , mens konvolutten kan ikke frankeres med beløpet $n + 1$.

1. Lag en tabell over funksjonsverdiene $n_h(5)$ for $h = 0, 1, \dots, 8$.

2. Forsök å gjette deg til en formel for $n_h(5)$, gyldig for $h \geq 3$. — Bevis formelen.

Vi skifter nå ut frimerkeverdien 5 med vilkårlig heltallig verdi $a > 1$. Vi har altså nå frimerker med verdi 1 og verdi a , mens konvolutten har plass til høyst h frimerker. Det minste positive belöp som ikke kan frankeres betegner vi med $n_h(a) + 1$. — Med $n = n_h(a)$, kan vi altså frankere belövene $0, 1, \dots, n$, mens belöpet $n + 1$ ikke kan frankeres.

3. Gi en formel for $n_h(a)$, gyldig for $h < a - 1$.

4. Gitt et belöp m , kan konvolutten frankeres med dette belöpet? Vis at dette spørsmålet kan besvares slik: Foreta heltallsdivisjonen $m : a$. Dette gir oss da en kvotient q og en rest r ,

$$m = qa + r, \quad 0 \leq r < a.$$

Nå kan belöpet m frankeres hvis og bare hvis $q + r \leq h$.

5. Finn en formel for $n_h(a)$, gyldig for $h \geq a - 2$.

Vi antar nå at konvoluttstörrelsen i landet er fullstendig standardisert, og at en konvolutt har plass til høyst h frimerker. Fremdeles skal det bare være to gyldige frimerkeverdier. Men nå ønsker myndighetene å velge disse frimerkeverdiene slik at vi kan få frankert et lengst mulig intervall av suksessive belöp $0, 1, \dots, n$.

Idet den ene frimerkeverdien åpenbart må være 1, står vi nå overfor følgende optimeringsproblem: Gitt h . Bestem $a_0 > 1$ slik at

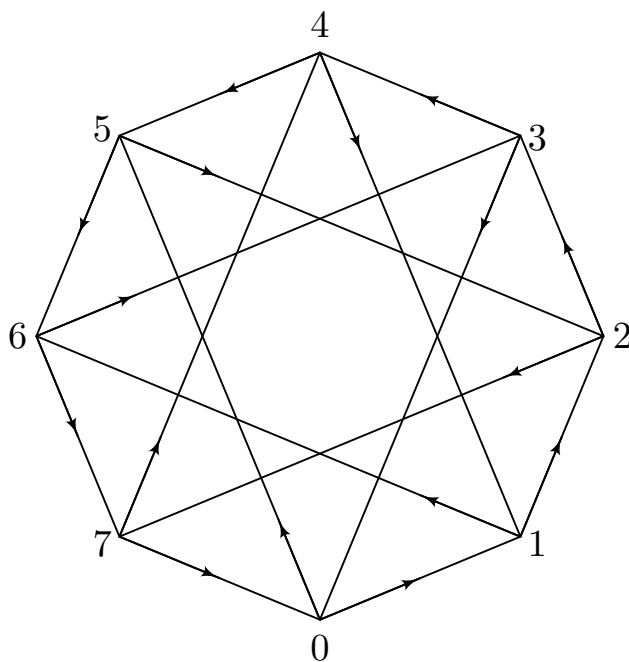
$$n_h(a_0) = \max_{a>1} n_h(a).$$

6. Lös optimeringsproblemet, og vis at

$$n_h(a_0) = \left\lfloor \frac{h^2 + 6h + 1}{4} \right\rfloor.$$

Computer-nettverk. Vi skal nå se på *double loop* computer-nettverk. Disse representerer en pålitelig og attraktiv nettverkarkitektur. Nettverk som dette forekommer gjerne i parallell-prosessor designs, og da særlig i såkalte multicomputere. I disse binder da nettverket sammen et stort antall mikroprosessorer, som har hvert sitt lokale minne.

Følgende figur representerer et *double loop* computer-nettverk med 8 stasjoner. Stasjonene er nummererte med tallene $0, 1, \dots, 7$.



Figur 1

I nettverket sendes det meldinger i pilenes retninger. Et krav til slike nettverk er at vi fra en vilkårlig stasjon kan sende meldinger (eventuelt via andre stasjoner) til enhver annen stasjon. Dette er oppfylt for nettverket over.

Nettverket i Fig.1 er satt sammen av to typer *kanter*: Fra stasjon

nr. i har vi kantene

$$i \longrightarrow i + 1 \quad \text{og} \quad i \longrightarrow i + 5.$$

I denne beskrivelsen vil vi f.eks. for $i = 6$, ha kanten $6 \longrightarrow 6+5 = 11$. Men hvilken stasjon har nummeret 11? — På Fig.1 ser vi at dette må være stasjon nr. 3.

To tall i og j angir samme stasjon dersom differensen $i - j$ er et multiplum av 8. Vi sier i dette tilfellet at i er kongruent med j modulo 8. Vi kan således si at *stasjonsnumrene er angitt modulo 8*.

Vi kan sende meldinger mellom to stasjoner på flere måter. La oss si at vi på Fig.1 ønsker å sende en melding fra stasjon 2 til stasjon 5. Noen muligheter er

$$\begin{array}{ccccccccccc} 2 & \rightarrow & 7 & \rightarrow & 4 & \rightarrow & 1 & \rightarrow & 6 & \rightarrow & 3 & \rightarrow & 0 & \rightarrow & 5, \\ & & & & & & 2 & \rightarrow & 3 & \rightarrow & 4 & \rightarrow & 5, \\ & & & & & & 2 & \rightarrow & 7 & \rightarrow & 0 & \rightarrow & 5. \end{array}$$

Der finnes også mange andre muligheter.

Avstanden $d(i, j)$ mellom de to stasjonene $i \neq j$, er det minste antall kanter man trenger å passere for å komme fra stasjon i til stasjon j . Naturlig nok setter vi også $d(i, i) = 0$.

7. Vis at $d(2, 5) = 3$ for nettverket i Fig.1. — Hva blir $d(5, 2)$?

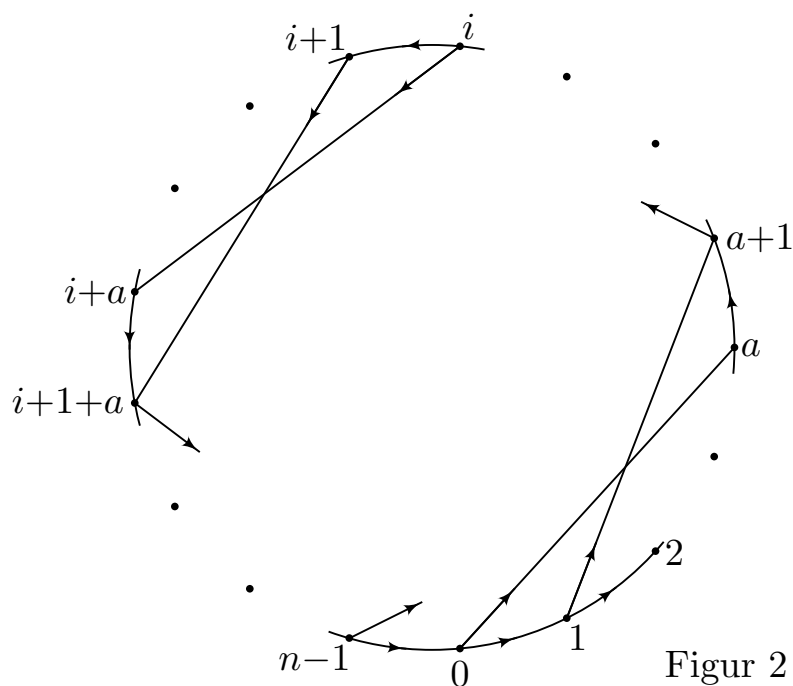
Diameteren $D_8(5)$ til nettverket over, er gitt som

$$D_8(5) = \max d(i, j)$$

hvor max taes over alle par i, j av stasjoner.

8. Bestem $D_8(5)$.

Vi går nå over til å se på den generelle situasjon med et *double loop* computer-nettverk med n stasjoner og *sprangavstand* a , $1 < a < n$. Vi kan antyde situasjonen ved følgende figur:

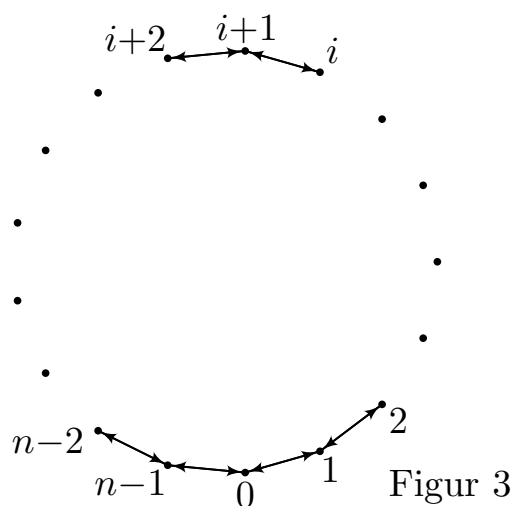


Figur 2

9. Forklar hva som nå menes med at *stasjonene er nummererte modulo n* . Definér også begrepene *avstanden $d(i, j)$ fra stasjon i til stasjon j* og *diameteren $D_n(a)$ til nettverket*.

Av et effektivt nettverk kreves det at diameteren er *liten*. Det gjelder altså å få den største avstanden i nettverket minst mulig.

En vanlig *ingeniørløsning* som ikke er særlig god m.h.p. dette ønsket, er å velge $a = n - 1$. Denne løsningen kan da tegnes slik:



Figur 3

hvor vi mellom en stasjon i og den påfølgende stasjon $i + 1$ har en kant hvor meldingene kan sendes i begge retninger.

10. Bestem diameteren $D_n(n - 1)$ til dette nettverket.

Vi skal nå benytte resultatene om frankering til å få frem bedre løsninger.

11. Bruk resultatet for $n_4(5)$ til å forklare at vi har $D_8(5) \leq 4$.

12. Forklar at hvis $n_h(a) \geq n - 1$, så er $D_n(a) \leq h$.

13. Forsök nå å bruke resultatene om frankering til å finne en a_0 (uttrykt ved n), slik at

$$D_n(a_0) < 2\sqrt{n}.$$

14. For $n = 66 \uparrow 156$, sammenlign $D_n(n - 1)$ fra punkt 10 med $D_n(a_0)$ fra punkt 13.

Det skulle nå være klart at svaret ditt under punkt 13 representerer en vesentlig forbedring av *ingeniörlösningen* under punkt 10.

Resultatet du fant under punkt 13 begynner faktisk å nærme seg det beste man i det hele tatt kan håpe på. Det går nemlig an å vise at vi alltid har

$$\min_{1 < a < n} D_n(a) \geq \lceil \sqrt{3n} \rceil - 2.$$

Man kjenner visse uendelige klasser av tall n , for hvilke likhetstegnet gjelder. F. eks. gjelder dette når n er av formen $n = 3w(w + 1)$.

15. Bruk dette til å bestemme $\min_{1 < a < n} D_n(a)$ for $n = 66 \uparrow 156$.

16. Forsök å planlegge et bevis for at

$$D_n(3w + 2) = 3w \quad \text{når} \quad n = 3w(w + 1).$$

Det er et *ulöst problem* å bestemme lignende resultater for (minst) en a_0 med korresponderende $D_n(a_0)$, slik at

$$D_n(a_0) = \min_{1 < a < n} D_n(a),$$

når n er et *vilkårlig* helt tall ≥ 1 . — Så nå er du ved forskningsfronten!

Litteratur

Hwang, F.K. and Xu, Y.H., *Double loop networks with minimum delay*. Discrete Math. 66(1987), s 109–118.

Selmer, E.S., *To populære problemer i tallteorien I. Myntveksling*. Normat 29(1981), s 81–87.

Selmer, E.S., *To populære problemer i tallteorien II. Frankering*. Normat 29(1981), s 105–114.

Juni-nummeret for 1987 av bladet *Computer* er i sin helhet viet computer-nettverk. Noen av artiklene i dette bladet vil nok være vanskelige å forstå. Men den innledende artikkelen av L. N. Bhuyan og begynnelsen på artikkelen av D. A. Reed & D. C. Grunwald skulle du kunne lese. Hvis du har lyst å se eksempler på andre typer nettverk-arkitektur enn den vi har sett på i denne oppgaven, vil du finne slike i dette bladet. Du vil da finne det spesielt interessant å se nærmere på den mye benyttede *hyperkube*-arkitekturen.

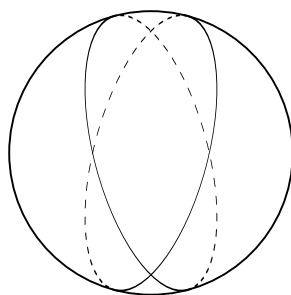
Kurvlängd och geometri på en sfärisk yta

PETER SJÖGREN

Göteborgs Universitet

1. Inledning. Geometrin på en sfärisk yta liknar planets geometri, med flera intressanta skillnader. Som vi skall se nedan, är kortaste vägen på sfären mellan två givna punkter en storcirkelbåge. (En storcirkel är skärningen mellan sfären och ett plan genom medelpunkten.) Därför är det naturligt att betrakta storcirkelarna som sfärens motsvarighet till de räta linjerna i planet. Många begrepp från planets geometri går då att överföra till sfären. Till exempel kan storcirkelbågar bilda trianglar, som har väldefinierade vinklar i hörnen. Vi kommer att se att vinkelsumman i en sådan triangel alltid blir större än två rätta. Cirkelar finns också på sfären, men sambandet mellan radie och omkrets är inte som i planet.

Genom två punkter på sfären kan man alltid dra en storcirkel. Den är entydigt bestämd utom då punkterna är antipoder. En viktig skillnad mellan sfären och planet är att två storcirklar alltid skär varandra. Någon motsvarighet till parallella linjer finns alltså inte på sfären. Två olika storcirklar delar in sfären i fyra områden, se fig. 1. Varje sådant



Figur 1

område begränsas av bara två storcirkelbågar - det är alltså en

tvåhörning eller *biangel*.

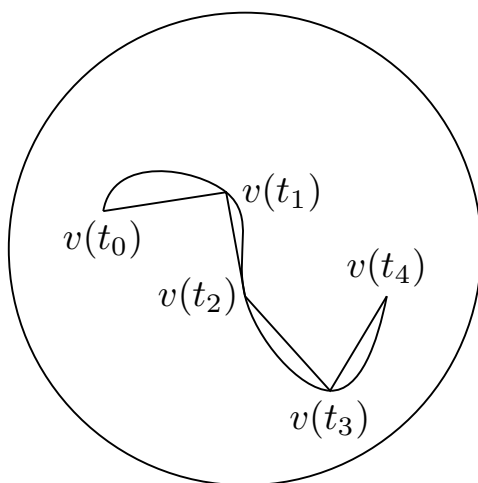
Den här uppgiften går först ut på att definiera kurvlängd och verifiera att de kortaste vägarna på sfären är storcirkelbågar. Därefter undersöker vi cirklar, tvåhörningar och trianglar på sfären.

2. Definition av kurvlängd. Vi väljer sfärens radie som längdenhet, och låter S beteckna en sfärisk yta med radie 1 i det tredimensionella rummet. För att kunna jämföra längden av storcirkelbågar och andra kurvor, måste man först definiera kurvlängd. Vi föredrar att tala om vägar. En väg v är en kontinuerlig funktion $v(t)$, $a \leq t \leq b$, med värden i S . Här är a och b reella tal med $a < b$. Det bör påpekas att nedanstående definition av längd fungerar lika bra om v har värden i det tredimensionella rummet eller planet. Låt

$$a = t_0 < t_1 < \dots < t_n = b$$

definiera en indelning av $[a, b]$ i n små intervall $[t_{i-1}, t_i]$, $i = 1, \dots, n$. Vi skiljer inte på en punkt $v(t)$ och den tredimensionella vektorn till $v(t)$ från sfärens medelpunkt. Då kan vi skriva det rätlinjiga avståndet mellan punkterna $v(t_{i-1})$ och $v(t_i)$ som $|v(t_i) - v(t_{i-1})|$, längden av vektordifferensen mellan $v(t_i)$ och $v(t_{i-1})$. Summan

$$\sum = |v(t_1) - v(t_0)| + |v(t_2) - v(t_1)| + \dots + |v(t_n) - v(t_{n-1})|$$



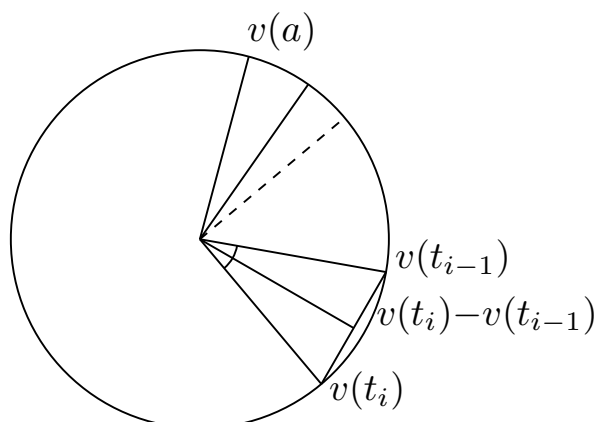
Figur 2

är sammanlagda längden av en följd av kordor till vägen. Detta visas i fig.2. Då indelningen görs fin, dvs alla kordorna görs korta, bör \sum approximera väglängden vi är ute efter. Man säger att v har längden $L \geq 0$ om \sum närmar sig L då indelningen blir fin. Mera precis skall man för varje godtyckligt litet tal $\varepsilon > 0$ ha $|\sum - L| < \varepsilon$ så snart indelningen är tillräckligt fin. Med "tillräckligt fin" menar vi här att alla skillnaderna $t_i - t_{i-1}$ skall vara mindre än något tal $\delta > 0$, som beror av ε . Man säger att längden är oändlig och skriver $L = \infty$, om på samma sätt för varje godtyckligt stort $M > 0$ man har $L > M$ så snart indelningen är tillräckligt fin.

[Man kan visa att ett av dessa båda fall alltid inträffar, så att $L = \lim \sum$ alltid existerar, ändlig eller oändlig: Låt t_0, t_1, \dots, t_n vara en indelning med summan \sum . Om t'_0, t'_1, \dots, t'_m är en förfining av den, dvs $m \geq n$ och varje t_i finns med bland t'_0, t'_1, \dots, t'_m , så ger t'_0, t'_1, \dots, t'_m en summa $\sum' \geq \sum$. Om nu i stället t'_0, t'_1, \dots, t'_m antas mycket fin, måste det mycket nära varje $v(t_i)$ finnas en punkt $v(t'_j)$. Därför ser man att \sum' inte kan vara nämnvärt mindre än \sum , i följande precisa mening: Hur litet talet $\varepsilon > 0$ än är, blir $\sum' > \sum - \varepsilon$ bara t'_0, t'_1, \dots, t'_m är tillräckligt fin. Härav följer nu att det ena eller andra fallet inträffar beroende på om \sum kan ta godtyckligt stora värden eller inte.]

3. Storcirkelbågar. Som exempel beräknar vi längden av en storcirkelbåge. En storcirkelbåge kan göras till en väg $v(t)$, $a \leq t \leq b$, på så sätt att centrumvinkeln mellan $v(t)$ och startpunkten $v(a)$ är t radianer för alla $t \in [a, b]$. Med centrumvinkeln menar vi vinkeln mellan radierna genom två punkter på S . Då väntar vi oss att längden blir $L = b - a$. Vägen och sfärens medelpunkt ligger i ett plan. Figur 3 visar detta plan.

Figur 3



ÖVNING 1. Visa att

$$|v(t_i) - v(t_{i-1})| = 2 \sin(t_i - t_{i-1})/2,$$

lämpligen genom att dra bisektrisen mellan radierna till $v(t_{i-1})$ och $v(t_i)$.

ÖVNING 2. Verifiera att $\sin \alpha \leq \alpha$ för alla $\alpha \geq 0$, och att om $0 < c < 1$ så är $\sin \alpha \geq c\alpha$ för alla tillräckligt små $\alpha \geq 0$. Ledning: Man kan göra detta geometriskt. Ett annat sätt är att skriva sinusfunktionen som integralen av sin derivata cosinus,

$$\sin \alpha = \int_0^\alpha \cos t \, dt.$$

För att uppskatta integralen uppåt och nedåt kan man utnyttja dels att $\cos t \leq 1$ för alla t , dels att $\cos 0 = 1$ så att $\cos t \geq c$ då t är nära 0.

Om man nu kombinerar övning 1 med den första olikheten i övning 2 och summerar, får man

$$\sum \leq (t_1 - t_0) + (t_2 - t_1) + \cdots + (t_n - t_{n-1}) = b - a.$$

Den andra olikheten i övning 2 ger på samma sätt för varje $c < 1$ att

$$\sum \geq c(t_1 - t_0 + t_2 - t_1 + \cdots + t_n - t_{n-1}) = c(b - a),$$

om indelningen är tillräckligt fin. Alltså är $L = b - a$.

SATS 1. Låt x och y vara två olika punkter på S . Om x och y inte är antipoder, går det precis en storcirkel genom x och y , och alltså två storcirkelbågar mellan x och y . Den kortare av dessa bågar är, efter lämplig parametrisering, en väg vars längd är kortast av alla vägar mellan x och y . Om x och y är antipoder, finns oändligt många storcirkelbågar mellan x och y . De är alla vägar av kortast möjliga längd mellan x och y .

BEVIS. Påståendena om antalet storcirkelbågar är uppenbara. Låt v vara en väg från x till y , och ta en indelning t_0, \dots, t_n av dess parameterintervall $[a, b]$. Centrumvinkeln mellan $x = v(t_0)$ och $v(t_i)$ kallar vi θ_i . Vi påstår nu att

$$(1) \quad |v(t_i) - v(t_{i-1})| \geq 2 \sin |\theta_i - \theta_{i-1}|/2.$$

Observera att likhet här gäller då $v(t_i)$ och $v(t_{i-1})$ ligger på en storcirkelbåge genom x . Före beviset av (1) ska vi se hur (1) medför satsen. Välj $c < 1$. Enligt övning 2 blir $2 \sin |\theta_i - \theta_{i-1}|/2 \geq c|\theta_i - \theta_{i-1}|$ för alla i , bara indelningen är tillräckligt fin. Nu kan vi summera och får

$$\begin{aligned} \sum &\geq c(|\theta_1 - \theta_0| + |\theta_2 - \theta_1| + \dots + |\theta_n - \theta_{n-1}|) \\ &\geq c(\theta_1 - \theta_0 + \theta_2 - \theta_1 + \dots + \theta_n - \theta_{n-1}) \\ &= c\theta_n. \end{aligned}$$

Alltså är $L = \lim \sum \geq \theta_n$. Eftersom θ_n är centrumvinkeln mellan x och y , dvs just längden av den kortare storcirkelbågen mellan x och y , följer satsen. Man behöver här inte särbehandla det antipodala fallet.

ÖVNING 3. Bevisa (1), t ex på följande sätt. Låt x vara nordpolen på sfären, och betrakta parallellcirkeln som går genom $v(t_i)$. Den

kan också beskrivas som mängden av punkter på S som bildar centrumvinkel θ_i med x . Projicera punkten $v(t_{i-1})$ vinkelrätt på det plan som bestäms av denna parallellcirkel. Använd nu Pythagoras sats för att se att avståndet mellan punkterna $v(t_i)$ och $v(t_{i-1})$ är som kortast då de *har samma longitud*, dvs ligger på samma storcirkel genom x .

Därmed är satsen bevisad. Med det sfäriska avståndet mellan x och y menar vi i fortsättningen längden av storcirkelbågen i satsen, vilket är detsamma som centrumvinkeln mellan x och y . Detta avstånd blir aldrig större än π , och är π då x och y är antipoder.

4. Figurer på sfären. Man kan tala om cirklar på sfären: Med den sfäriska cirkeln med medelpunkt $x \in S$ och radie $r > 0$ menar vi mängden av punkter på S med sfäriskt avstånd r till x . Observera att detta är en cirkel i det tredimensionella rummet, med en radie som är mindre än r . Motsvarande sfäriska cirkelskiva är mängden av punkter på S med sfäriskt avstånd högst r till x . En sådan brukar kallas en kalott i tredimensionell geometri. För $r > \pi$ består cirkelskivan av hela S , medan cirkeln inte innehåller några punkter.

ÖVNING 4. Ge formler för längden av en sfärisk cirkel och ytan av en sfärisk cirkelskiva med radie r . Vad ger de för $r = \pi/2$ och $r = \pi$?

Betrakta en av de tvåhörningar som bildas av två storcirklar, jämför fig. 1. Den har samma vinkel α i båda hörnen, och α är också vinkeln mellan storcirkelarnas plan. Tvåhörningens yta B är proportionell mot α , och $\alpha = 2\pi$ motsvarar hela S , med ytan 4π . Därför får vi

$$(2) \quad B = 2\alpha,$$

med andra ord är tvåhörningens yta lika stor som dess vinkelsumma.

Vi skall nu härleda en analog formel för en sfärisk triangel. Låt A, B och C vara tre punkter på S med inbördes avstånd mindre än π ,

dvs inte antipoder. De tre korta storcirkelbågarna mellan punkterna delar S i två delar, varav den ena uppenbart är mindre än den andra. Den större delen innehåller t ex till skillnad från den mindre många par av antipoder. Med den sfäriska triangeln ABC menar vi den mindre av de två delarna.

SATS 2. Ytan T av den sfäriska triangeln ABC uppfyller

$$T = A + B + C - \pi,$$

där A, B, C betecknar triangelns vinklar vid resp hörn.

Båda leden i ekvationen är positiva. Därför är vinkelsumman i en sfärisk triangel alltid större än π . Observera att då triangeln är mycket liten, är ytan nära noll och vinkelsumman alltså nära π . Detta stämmer med att en mycket liten del av sfären nästan är plan.

ÖVNING 5. Bevisa Sats 2, t ex på följande sätt. Rita ut triangeln på en boll eller något liknande. Markera de tre punkternas antipoder A', B' och C' . Förläng sidorna i ABC så att man får sfäriska trianglar $A'BC$, $AB'C$ och ABC' , vars ytor vi kallar T_A, T_B resp T_C . Drag också sidorna i triangeln $A'B'C'$ som förstas är kongruent med ABC . Observera nu att hela S är indelad i 8 sfäriska trianglar, som är kongruenta två och två. Därför är $T + T_A + T_B + T_C$ precis hälften av ytan av S , så att

$$(3) \quad T + T_A + T_B + T_C = 2\pi.$$

Använd nu (2) för att få uttryck för $T + T_A$, $T + T_B$ och $T + T_C$. Kombinera med (3), så följer satsen.

5. Det sfäriska sinusteoremet. Vi avslutar med sinusteoremet för sfäriska trianglar. Låt ABC vara en sfärisk triangel som förut. Kalla sidornas sfäriska längder för a, b resp c , så att a är längden av storcirkelbågen BC osv.

SATS 3.

$$\frac{\sin a}{\sin A} = \frac{\sin b}{\sin B} = \frac{\sin c}{\sin C}.$$

ÖVNING 6. Bevisa ett specialfall av denna formel: om vinkeln B är rät gäller $\sin A = \sin a / \sin b$. Man kan göra så här.

Anta först att b och c är mindre än $\pi/2$. Låt O vara sfärens medelpunkt och A_1 en punkt på radien OA som lämpligen ligger nära O . Ett normalplan till OA genom A_1 skär strålarna OB och OC i punkter B_1 resp C_1 . (Rita, eller bygg hellre en modell.) Planen OAB och OBC är vinkelräta eftersom triangelvinkeln B är rät. Därför är sträckan B_1C_1 vinkelrät mot planet OAB , så att vinkeln $A_1B_1C_1$ är rät. I den rätvinkliga (plana) triangeln $A_1B_1C_1$ känner vi också vinkeln $B_1A_1C_1$, som är lika med A . Detta ger $\sin A = B_1C_1/A_1C_1$. Triangeln OB_1C_1 är rätvinklig vid B_1 och dess vinkel vid O är a . Alltså fås $\sin a = B_1C_1/OC_1$. Genom att betrakta triangeln OA_1C_1 får man på samma sätt $\sin b = A_1C_1/OC_1$. Kombinera nu dessa tre likheter, så följer den sökta formeln.

Ovanstående kräver bara små ändringar då b och c är godtyckliga. Punkterna B_1 och C_1 hamnar ibland på förlängningarna bakåt av strålarna OB och OC . De vinklar vi fann vara A och a kan i stället bli $\pi - A$ respektive $\pi - a$.

ÖVNING 7. Bevisa Sats 3 med hjälp av Övning 6, genom att fälla en höjd i den sfäriska triangeln.

Litteratur

Kulczycki, S., *Non-Euclidean Geometry*. Pergamon Press, Oxford 1961.

Kampen om sista stickan

KRISTER SVANBERG

KTH

1. Förberedande exempel. Tänk dig följande enkla spel mellan två spelare kallade A och B:

En hög med ett stort antal stickor läggs på ett bord. Spelare A börjar och får ta 1 eller 2 eller 3 stickor ur högen. Därefter är det B:s tur att ta 1 eller 2 eller 3 stickor ur högen, varefter det är A:s tur osv. Vi säger (influerade av schack-terminologin) att A får börja att ”göra ett drag”, varefter B ”gör ett drag” osv. Det är tillåtet att räkna stickorna på bordet innan man gör sitt drag.

Den spelare vinner som tar den sista stickan ur högen.

FRÅGA. Hur beter man sig för att vinna ett sådant spel?

ANALYS. En spelare som lämnar 0 stickor på bordet efter att ha gjort ett drag har vunnit, enligt reglerna ovan. Däremot kommer en spelare som lämnar kvar 1 eller 2 eller 3 stickor på bordet efter sitt drag att förlora, eftersom motspelaren då kommer att vinna med sitt nästa drag. En spelare som lämnar kvar 4 stickor på bordet kommer att vinna, eftersom motståndaren med sitt nästa drag tvingas lämna kvar 1 eller 2 eller 3 stickor på bordet. Däremot kommer en spelare som lämnar kvar 5 eller 6 eller 7 stickor på bordet att förlora, eftersom motspelaren då med sitt nästa drag kan se till att det blir 4 stickor kvar på bordet.

Sådär kan man fortsätta att resonera, och den slutsats man bör kunna dra är följande:

SATS 1: (= SVARET PÅ FRÅGAN OVAN). *Den spelare som efter att ha gjort sitt drag lämnar kvar ett antal stickor som är jämnt delbart*

mer 4 (dvs 0 eller 4 eller 8 eller 12 osv) kommer att vinna spelet (om hon inte senare gör något allvarligt misstag förstås).

Alternativt kan vi formulera Sats 1 på följande vis:

De "tillstånd" man ska lämna efter sig för att vinna, s k "V-tillstånd", respektive de som leder till att man kommer att förlora (om man lämnar ett sådant efter sig), s k "F-tillstånd", ges av följande tabell:

V-tillstånd: "0 stickor", "4 stickor", "8 stickor", ...

F-tillstånd: "1 sticka", "5 stickor", "9 stickor", ...
 "2 stickor", "6 stickor", "10 stickor", ...
 "3 stickor", "7 stickor", "11 stickor", ...

Att bevisa Sats 1 (i den alternativa formuleringen med V- och F-tillstånd) kan gå till så här:

Antag att spelare A lämnar efter sig ett V-tillstånd (dvs "0 stickor" eller "4 stickor" osv). Då gäller antingen att A redan har vunnit (om det var "0 stickor" hon lämnade efter sig) eller också måste spelare B med sitt nästa drag lämna efter sig ett F-tillstånd, eftersom det *inte* går att komma från ett V-tillstånd till ett annat V-tillstånd med *ett* drag. Men om B lämnar efter sig ett F-tillstånd, kan A med sitt nästa drag lämna efter sig ett V-tillstånd, eftersom det från *varje* F-tillstånd går att komma till ett V-tillstånd med *ett* drag.

Om A en gång har lämnat efter sig ett V-tillstånd, så kommer hon alltså att kunna fortsätta att bara lämna V-tillstånd efter sig (tills spelet är slut), medan stackars B hela tiden tvingas lämna F-tillstånd efter sig.

Antalet stickor på bordet minskar strikt vid varje drag. Alltså kommer spelet förr eller senare att ta slut, och segrare blir den som lämnar efter sig V-tillståndet "0 stickor". Eftersom B hela tiden

tvingas lämna efter sig F-tillstånd kan det inte vara B som vinner.

Alltså vinner A.

2. Ett svårare spel. Ovanstående var förberedelser till det spel vi ska ägna oss åt fortsättningsvis.

Liksom ovan deltar två spelare, kallade A och B.

Minst 2 högar med stickor läggs ut på bordet (t ex 5 högar med respektive 12, 7, 9, 14 och 6 stickor).

Den ene spelaren, säg A, börjar och får ta hur många stickor hon vill ur en och endast en hög. Hon får ta ur vilken hög som helst och hur många stickor som helst, dock minst 1 sticka och högst hela det antal stickor som finns i den hög hon väljer att ta ur.

Därefter är det B:s tur att göra motsvarande, osv.

Det är tillåtet att räkna stickorna i de olika högarna innan man gör sitt drag.

Den spelare som tar den sista stickan från bordet vinner.

3. Specialarbetet. Specialarbetet går ut på att du skall resonera dig fram till hur man beter sig för att vinna nyss beskrivna spel.

Lämpligen sker detta genom följande steg:

(i) Spela spelet några gånger med kamrater, tills du fått litet känsla för det. Till att börja med kan det räcka med att man startar med 2 högar. Därefter kan man utvidga till 3 högar osv.

(ii) Antag att man startar spelet med två högar. Resonera i analogi med ovan (avsnitt 1) och försök lista alla V-tillstånd och F-tillstånd. Ledning: Representera tillstånden med två tal (x, y) , där x = antal stickor i hög 1 och y = antal stickor i hög 2. Då är exempelvis $(1, 1)$ ett V-tillstånd medan $(2, 1)$ är ett F-tillstånd, eller hur?

(iii) Antag nu att man startar spelet med tre högar. Representera tillstånden med tre tal (x, y, z) och försök lista alla V- och

F-tillstånd. $(1, 1, 1)$ är ett F-tillstånd, medan exempelvis $(3, 2, 1)$ är ett V-tillstånd, eller hur?

(iv) Fortsätt med 3-högarsfallet och försök hitta någon regel för när ett tillstånd (x, y, z) är ett V-tillstånd.

Ledning: Binärutveckla x, y och z , dvs uttryck x, y och z med hjälp av enbart 0:or och 1:or i talsystemet med basen 2.

Exempel: $(9, 12, 7) = (1001, 1100, 111)$.

(v) Undersök om den regel du (förhoppningsvis) upptäckte under steg (iv) ovan gäller även om man har fler än 3 högar.

(vi) Formulera en metod att vinna spelet.

(vii) Bevisa att din spelmetod fungerar. Använd lämpligen ett resonemang i analogi med det som användes i beviset av Sats 1 ovan.

(viii) (Frivilligt) Programmera upp din metod på en liten dator och låt dina kamrater, och din mattelärare, försöka slå programmet.

Lycka till!

Ett belysande exempel

LASSE SVENSSON

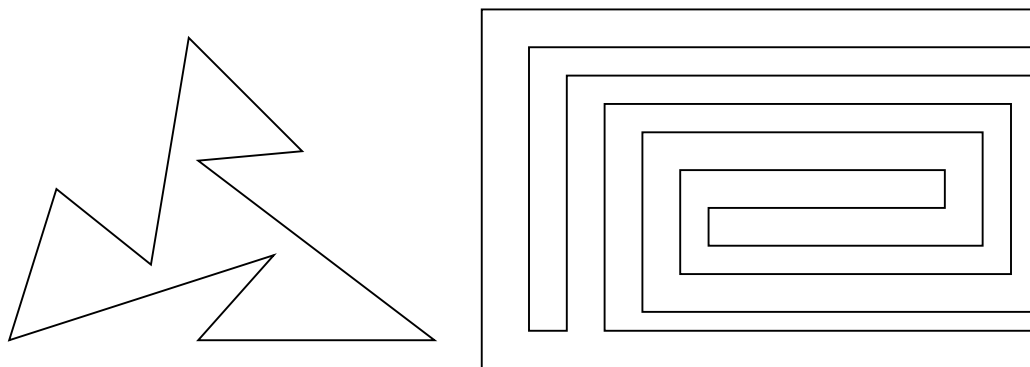
K T H

En lokal med plant golv och tak begränsas av n plana vertikala väggar. (n är ett naturligt tal.) Hur många lampor måste placeras ut i lokalen för att helt upplysa den?

Som du nog observerar så är problemet egentligen plant, dvs handlar om n -hörningar i planet. Vidare noterar du nog att lösningen beror på hur n -hörningen ser ut och vilka lampor vi har. Lamporna idealiserar vi till punktformiga ljuskällor som kastar sitt ljus i alla riktningar.

För att göra det lättare att komma igång har vi *snitslat* en bana som leder i riktning mot en förståelse av problemet.

1. Bästa sättet att börja är att experimentera en smula med enkla konkreta situationer. Hur många lampor krävas t ex i följande fall:



Räcker det alltid att placera ut lamporna i hörnen eller tjänar man i vissa situationer på att placera ut lampor i det *inre* av n -hörningen?

2. I det följande kommer vi att ha nytta av två begrepp: *Korda* och *Korsande kordor*. Genom att ge exempel på dessa tänkte vi att du skulle formulera en precis definition.

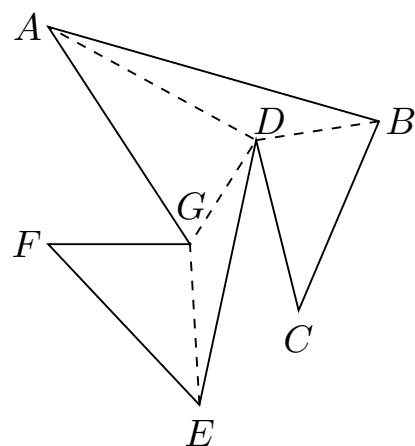
Betrakta n -hörningen nedan.

Kordor

AD
BD
DG
EG

Icke kordor

AB
AC
AF
EC
⋮

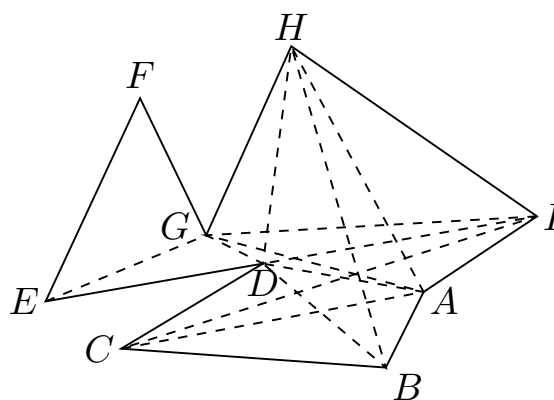


Korsande kordor

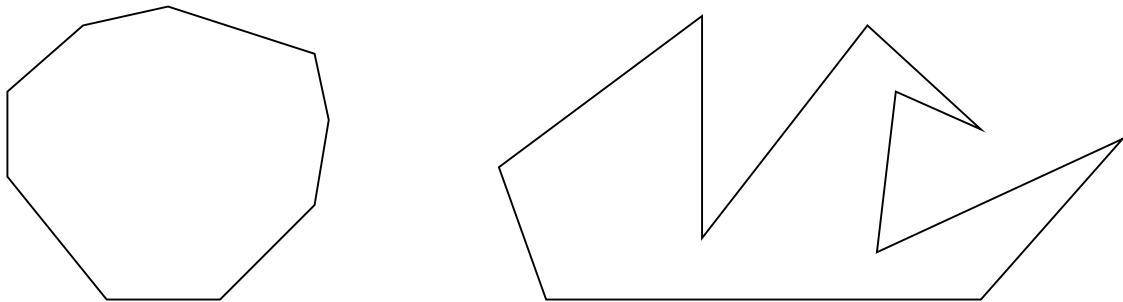
AC, BH
AC, BD
GI, AH
⋮

Icke korsande kordor

AC, CI
EG, DH
AH, BD
⋮



Hur många kordor kan du t ex finna som parvis är icke-korsande i n -hörningen nedan?



Nedan låter vi P beteckna en godtycklig n -hörning.

3. Visa att varje n -hörning innehåller minst en korda om $n > 3$. (Detta kan verka självklart men kräver faktiskt ett bevis.) Försök också att finna en algoritm (\approx automatisk metod) som givet en n -hörning P plockar fram en korda. Här kan du även fundera på, om det som är relevant i P , när det gäller att finna en korda, möjligen kan beskrivas på annat sätt än att ge koordinaterna för varje hörn. (Du behöver inte kunna något programspråk.)

Vad ett induktionsbevis är kan du t ex läsa om i Karl-Johan Bäckströms bok *Diskret Matematik*. Studentlitteratur 1986.

4. Visa med hjälp av induktionsbevis att varje n -hörning P innehåller $n-3$ icke-korsande kordor. Observera också att dessa kordor delar in P i $n-2$ trianglar på ett sådant sätt att om två av dessa trianglar möts i mer än en punkt så är deras skärning precis en gemensam sida. Detta brukar kallas en *triangulering* av P . (Gör gärna en algoritm för detta.)

5. Betrakta en triangulerad n -hörning P . Visa att det går att färga varje hörn i P på ett sådant sätt att

- (a) Endast tre färger förekommer
 (b) Två hörn som är förbundna med en korda eller en sida (i den givna trianguleringen)
 har olika färger.
 Finn också en algoritm för denna färgning.

6. Visa att någon färg förekommer i högst $\lfloor n/3 \rfloor$ hörn. ($\lfloor x \rfloor$ betecknar det största heltal som är mindre än eller lika med x , heltalsdelen av x .)

Vad händer om en lampa placeras ut i varje hörn där denna färg förekommer?

7. Ge för varje n ett exempel på en $3 \cdot n$ -hörning som inte kan belysas med färre än n lampor.

8. Visa hur man nu kan dra slutsatsen att $\lfloor n/3 \rfloor$ lampor alltid räcker för att belysa en n -hörning.

9. För speciella n -hörningar kan man naturligtvis klara sig med betydligt färre lampor än $\lfloor n/3 \rfloor$. Kan du finna en algoritm som givet en n -hörning placerar ut ett minimalt antal lampor som ger full belysning? (Förmodligen svårt)

10. Kan du generalisera problemet t ex:

Vad händer i flera dimensioner?

Vad händer om vissa sidor i n -hörningen är speglar?

...

....

Här kan det vara intressant med problemformuleringar bara. Du behöver inte lösa dem nödvändigtvis.

ANMÄRKNING: Punkterna 9 och 10 är att betrakta som mini-forskningsproblem. Problemförfattaren känner själv ingen lösning på

dessa problem. Förmodligen kan resultaten här vara publiceringsbara i någon tidskrift.

Litteratur

Bäckström, K.-J., *Diskret matematik*. Studentlitteratur 1986.

Explorativ dataanalys (EDA)

KERSTIN VÄNNMAN

Högskolan i Luleå

1. Inledning. I många situationer stöter man på siffror, ofta samlade i en stor hop. Det kan vara i tidningar eller böcker i form av tabeller eller diagram. Det kan vara mätvärden som man själv samlat in. Meningen är att man ska kunna läsa ut något från siffrorna. Eftersom vi människor tänker i bilder är det viktigt att man omformar information serverad i siffror till lättförståeliga bilder. Det gäller att skaffa sig enkla metoder som får informationen hos siffrorna att träda fram och som gör siffrornas budskap synligt.

John W. Tukey lanserade under 1970-talet EDA (explorativ dataanalys), som innehåller många nya och lättfattliga sätt att hantera siffermaterial och upptäcka mönster och samband. EDA-metoderna kompletterar de traditionella sätten. De lyfter ofta fram andra intressanta egenskaper hos material än vad de traditionella metoderna gör. Tukey beskriver själv EDA som ett numeriskt eller grafiskt detektivarbete, en jakt efter ledtrådar för att kunna upptäcka viktiga samband och strukturer. Tukey betonar vikten av grafiska metoder som gör att informationen tränger sig på.

Några av de enkla EDA-metoderna är stam-bladdiagram och låddiagram, användbara just när man vill få en samling siffror att berätta något. Dessa finns beskrivna i [1]. Metoderna går ut på att ordna siffror och rita okomplicerade figurer. De bygger inte på krångliga formler eller besvärliga uträkningar.

Utgående från [1] kan ett flertal specialarbeten göras. Här följer två förslag.

2. Att utforska en del av verkligheten med stam-bladdiagram, lådagram och jämförelser. Läs och arbeta igenom kapitel 1 - 3 i [1] för att få bakgrundskunskaper. Välj sedan den del av verkligheten som känns intressant att studera, t ex returpapper, skolmaten, bekämpningsmedel, vattenförbrukning, fel på bilar, kroppsmått förr och nu, idrottsresultat, etc. Med hjälp av de enkla EDA-metoderna gör du sedan ditt detektivarbete. Här får du träning i att arbeta med siffermaterial, se och upptäcka likheter och olikheter i form av läge och variation, göra jämförelser och presentera information överskådligt.

Om du vill ha tips på uppgifter att arbeta med så gå till kapitel 5 i [1]. Där finns förslag på många olika uppgifter att arbeta med. Dels finns uppgifter med givet siffermaterial, dels uppgifter till vilka man själv producerar eller söker material.

3. Att transformera till insikt. Läs och arbeta igenom kapitel 1 - 4 i [1]. I kapitel 4 lär du dig hur man med hjälp av logaritm- och potensfunktionen kan vrida och vända på sitt siffermaterial för att bättre kunna dra slutsatser. Att arbeta med transformationer på detta sätt ger fördjupade insikter om logaritm- och potensfunktionen. Lös övningarna i kapitel 4 och uppgifterna 521 och 522 om du inte har några intressanta material som behöver transformeras.

Litteratur

[1] Vännman, K. och Dunkels, A., *Boken om kreativ statistik med EDA*. Förlagshuset Gothia, Göteborg, 1984.

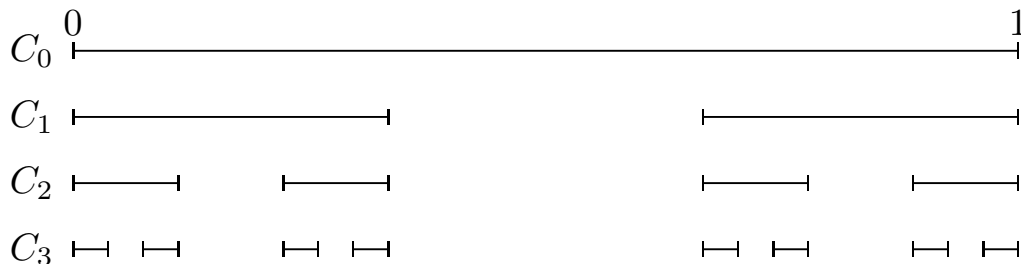
Fraktaler och iteration av funktioner

HANS WALLIN

Umeå Universitet

1. Inledning. Georg Cantor (1845-1918), mängdlärans skapare, konstruerade en mängd som numera kallas *Cantormängden* och som visat sig ha många intressanta egenskaper. I våra dagar har Cantormängden uppmärksammats även utanför matematikerketsar därför att den bidragit till att ge grunden för en beskrivning av många fenomen i naturen med hjälp av fraktaler (se nedan). Uppgiften i detta specialarbete är att konstruera Cantormängden och att lära känna en del av dess egenskaper samt att se hur detta kan läggas till grund för beskrivning av delar av naturen med fraktal geometri (se nedan).

2. Geometrisk konstruktion av Cantormängden C . Vi utgår från det slutna intervallet $[0, 1]$ som vi kallar C_0 . I första steget av konstruktionen delar vi $[0, 1]$ i tre lika långa intervall och tar bort det mellersta, öppna intervallet. Den mängd som återstår kallar vi C_1 som alltså består av 2 slutna intervall av längd $1/3$ vardera. I andra steget av konstruktionen delar vi varje intervall i C_1 i tre lika långa delar och tar bort det mellersta, öppna intervallet. Vi får då kvar en mängd C_2 som består av $4 = 2^2$ slutna intervall av längd $1/9 = 3^{-2}$ vardera, o s v. Efter k steg har vi fått C_k som består av 2^k slutna intervall av längd 3^{-k} vardera. Vi får en oändlig följd av mängder C_0, C_1, C_2, \dots . Figuren visar C_0, C_1, C_2 och C_3 .



Cantormängden C är den mängd som återstår av $[0, 1]$ efter hela denna process, dvs C består av de punkter som tillhör C_k för alla $k = 0, 1, 2, \dots$. Man skriver detta

$$C = \bigcap_{k=0}^{\infty} C_k.$$

Observera att C är en delmängd av C_k för alla k . Cantormängden C är alltså en starkt sönderskuren mängd; den är ett typexempel på det som kallas en fraktal. Exempel på punkter som tillhör C är ändpunkterna av de 2^k intervallen i C_k , för $k = 0, 1, \dots$, t ex punkterna $0, 1, 1/3$ och $2/3$.

- a) Ange ändpunkterna till intervallen i C_2, C_3 och C_4 .
- b) Försök göra samma sak för C_k för ett godtyckligt k . Ledning: Betrakta ändliga summor av typen $\sum a_j/3^j$ där a_j antar värdena $0, 1$ och 2 .

3. Beteckningen $f(A)$. Innan vi går vidare skall vi införa beteckningen $f(A)$ där f är en funktion och A är en mängd. Med $f(A)$ betecknar vi helt enkelt mängden av funktionsvärden $f(x)$ för vilka $x \in A, f(A) = \{f(x) : x \in A\}$.

Exempel: Om $f(x) = 2x + 1$ och A består av punkterna $1, 2$ och 3 , d v s $A = \{1, 2, 3\}$, så är $f(A) = \{f(1), f(2), f(3)\} = \{3, 5, 7\}$.

4. Konstruktion av C med iteration. Låt f_1 och f_2 vara funktioner definierade genom

$$f_1(x) = \frac{x}{3} \text{ och } f_2(x) = \frac{x}{3} + \frac{2}{3}.$$

Låt x_0 vara ett godtyckligt reellt tal som vi kallar för *startvärdet*. Vi definierar en oändlig följd av mängder A_0, A_1, A_2, \dots genom

$$(1) \quad \begin{cases} A_0 = \{x_0\}, \\ A_n = f_1(A_{n-1}) \cup f_2(A_{n-1}), \text{ för } n = 1, 2, \dots \end{cases}$$

Det betyder att $A_1 = f_1(A_0) \cup f_2(A_0) = \{f_1(x_0), f_2(x_0)\}$, $A_2 = f_1(A_1) \cup f_2(A_1) = \{f_1(f_1(x_0)), f_1(f_2(x_0)), f_2(f_1(x_0)), f_2(f_2(x_0))\}$, osv. Om t ex $x_0 = 1$ så är $A_1 = \{1/3, 1\}$ och $A_2 = \{1/9, 1/3, 7/9, 1\}$.

Följden A_0, A_1, \dots är konstruerad utgående från startvärdet x_0 och funktionerna f_1 och f_2 genom *iteration*; vi kallar $\{f_1, f_2\}$ ett *iterativt funktionssystem*.

- c) Beräkna A_3 om $x_0 = 1$ och rita ut A_3 på talaxeln.
- d) Gör samma sak för A_1, A_2 och A_3 om $x_0 = 2$.
- e) Övertyga dig om att $C_n = f_1(C_{n-1}) \cup f_2(C_{n-1})$ för $n = 1, 2, \dots$.

5. Ett datorprogram för iterationen. Om man vill beräkna och rita ut punkterna i A_n för stora värden på n är det lämpligt att använda dator.

f) Gör ett datorprogram som beräknar och ritar ut A_n . Använd programmet för att rita ut A_n för $n = 5$, $n = 7$ och $n = 10$ för några olika startvärden x_0 . Jämför figuren med Cantormängden C .

6. För stora n är A_n ungefär lika med C . Du skall nu försöka bevisa det du sett i figuren i uppgift f .

g) Övertyga dig först om att varje $x \in C_k$ ligger på avstånd mindre än 3^{-k} från en punkt i C .

h) Använd uppgift g för att visa att (då n går mot oändligheten) punkterna i A_n närmar sig punkterna i C , oberoende av valet av x_0 . Mer exakt: För varje $r > 0$ (r får vara godtyckligt litet) finns ett heltal $n(r)$ så att, för $n \geq n(r)$,

1° varje punkt i A_n ligger på avstånd mindre än r från en punkt i C och

2° för varje punkt $x \in C$ finns en punkt $y \in A_n$, så att $|x - y| < r$.

Egenskapen i uppgift h uttrycker matematikerna genom att säga att avståndet mellan mängderna A_n och C går mot noll. Detta är den matematiska förklaringen till att figuren i uppgift f för våra ögon ser ut som C , d v s att vi för stora n (och t o m för ganska små n) inte ser någon skillnad mellan C och A_n .

Det är alltså så att C drar till sig (attraherar) punkterna i A_n . Man kallar därför C för *attraktor* till det iterativa funktionssystemet $\{f_1, f_2\}$.

i) Övertyga dig om att $C = f_1(C) \cup f_2(C)$.

j) Försök att hitta andra mängder E av reella tal så att $E = f_1(E) \cup f_2(E)$.

7. Ett annat sätt att konstruera C genom iteration. Låt f_1 och f_2 vara samma funktioner som förut. Vi konstruerar x_0, x_1, x_2, \dots där x_0 är startvärdet och $x_1 = f_1(x_0)$ eller $x_1 = f_2(x_0)$ och valet mellan $f_1(x_0)$ och $f_2(x_0)$ sker slumpvis med sannolikhet $1/2$ för varje val. Välj sedan $x_2 = f_1(x_1)$ eller $x_2 = f_2(x_1)$ med sannolikhet $1/2$ för vardera valet, o s v.

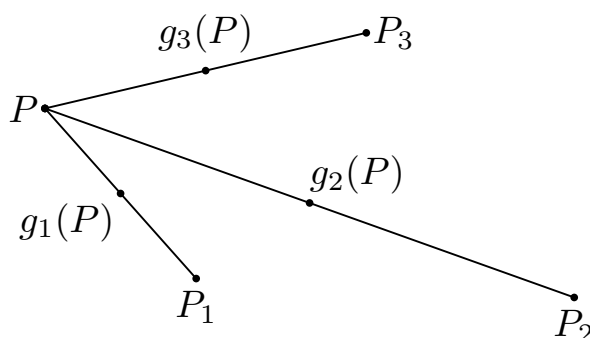
k) Skriv ett datorprogram för detta (du behöver en slumpgenerator).

l) Låt datorn rita ut x_0, x_1, \dots, x_{100} om $x_0 = 1$ (eller om x_0 är en annan punkt i C). Rita ut $x_{10}, x_{11}, \dots, x_{100}$ om $x_0 = 2$ (eller om x_0 är en annan punkt utanför C).

m) Kan du ge en matematisk förklaring till det du ser i figuren i uppgift l?

n) Vad tror du händer om man tar andra sannolikheter i valet mellan f_1 och f_2 ?

8. Ett annat iterativt funktionssystem. Vi inför tre funktioner g_1, g_2 och g_3 definierade på följande sätt. Starta från en triangel i xy -planet med hörn P_1, P_2 och P_3 (t ex $P_1 = (-1, 0)$, $P_2 = (1, 0)$ och $P_3 = (0, \sqrt{3})$). För varje punkt P i xy -planet och för $j = 1, 2, 3$, definierar vi $g_j(P)$ som den punkt i planet som ligger mitt emellan P och P_j (se figuren).



o) Gör om så mycket som möjligt av steg 4 - 7, där f_1 och f_2 ersätts av g_1, g_2 och g_3 . Ledning: (1) ersätts av $A_0 = \{P_0\}$, där P_0 är en punkt i planet och $A_n = g_1(A_{n-1}) \cup g_2(A_{n-1}) \cup g_3(A_{n-1})$, för $n = 1, 2, \dots$. Sannolikheten $1/2$ i steg 7 ersätts med sannolikheten $1/3$.

p) Avsluta med att göra en geometrisk beskrivning av attraktorn till det iterativa funktionssystemet $\{g_1, g_2, g_3\}$ då $P_1 = (-1, 0)$, $P_2 = (1, 0)$ och $P_3 = (0, \sqrt{3})$. (Denna attraktor brukar kallas Sierpinskis triangel.)

9. Fraktal geometri. I princip kan man starta med två eller flera godtyckligt valda funktioner $\{g_1, g_2, \dots, g_N\}$ och hoppas att man skall kunna göra en liknande teori som ovan. För många val av

$\{g_1, g_2, \dots, g_N\}$ går detta och istället för Cantormängden C och Sierpinski's triangel S får man andra typer av attraktorer som, liksom C och S , är starkt sönderskurna mängder som kallas *fraktaler*. Man har funnit att lämpligt valda funktioner $\{g_1, \dots, g_N\}$ ger attraktorer som ser ut som löv, träd och andra geometriska objekt i naturen. Genom iteration kan man återskapa starkt sönderbrutna och oregelbundna former i naturen. Formerna beskrivs matematiskt som attraktorer till iterativa funktionssystem. Dessa attraktorer är fraktaler och fraktal geometri är studiet av sådana formers geometri.

Litteratur

Gleick, J., *Kaos - vetenskap på nya vägar*. Bonniers, 1988.

Innehåller en del populärvetenskaplig fysik och matematik med anknytning till detta specialarbete.

Peitgen, H.O. & Saupe, D., (redaktörer), *The science of fractal images*. Springer-Verlag, 1988.

Kapitel 5 innehåller en del bilder och matematik om iterativa funktionssystem.

Wallin, H. & Fällström, A. & Wallin, M., *Matematiska bilder av fraktaler och kaos*. Matematiska Institutionen, Umeå Universitet, 1989.

Innehåller ett bildmaterial och datorprogram för gymnasiet.

Något om medelvärden

PEPE WINKLER

Uppsala Universitet

Om a_1 och a_2 är två reella, positiva tal så kallas talet $A = \frac{a_1 + a_2}{2}$ för det *aritmetiska medelvärdet* och talet $G = \sqrt{a_1 \cdot a_2}$ för det *geometriska medelvärdet*.

UPPGIFT 1. Gör ett dataprogram som beräknar A och G för givna talpar (a_1, a_2) . Välj några talpar (a_1, a_2) och beräkna för dem A och G .

Troligen har du observerat att

$$(1) \quad G \leq A.$$

UPPGIFT 2. Gäller den här olikheten för godtyckliga par (a_1, a_2) ? Försök visa att så är fallet.

LEDNING. Tänk på den nästan triviala olikheten $0 \leq (x - y)^2$ som gäller för godtyckliga reella tal x och y .

Vad skall x och y ersättas med för att erhålla (1) ?

UPPGIFT 3. För vilka par (a_1, a_2) gäller likheten i (1) ?

UPPGIFT 4. Bestäm bland alla rektanglar med en given omkrets, den som har största arean.

En annan typ av medelvärde är det så kallade *harmoniska medelvärdet* som definieras genom
$$H = \frac{2a_1a_2}{a_1 + a_2} = \frac{2}{\frac{1}{a_1} + \frac{1}{a_2}}.$$

Verifiera att H kan definieras genom

$$\frac{1}{H} = \frac{1}{2} \left(\frac{1}{a_1} + \frac{1}{a_2} \right),$$

dvs inversen till det harmoniska medelvärdet är lika med det aritmetiska medelvärdet av inverser till a_1 och a_2 .

UPPGIFT 5. Beräkna H för samma talpar som du valt i uppgift 1. Undersök hur H förhåller sig till A och G . Ställ upp en förmodan. Bevisa en sats.

UPPGIFT 6. En bil kör från staden **A** till staden **B** med medelhastigheten a_1 km/tim och återvänder omedelbart med medelhastigheten a_2 km/tim. Vad blir medelhastigheten för hela resan? Vilket slags medelvärde är den erhållna medelhastigheten?

Låt oss nu anta att vi har en trippel av reella, positiva tal (a_1, a_2, a_3) . Det är då ganska naturligt att definiera det algebraiska medelvärdet för dem som $A = \frac{a_1 + a_2 + a_3}{3}$, det geometriska medelvärdet som $G = \sqrt[3]{a_1 a_2 a_3}$ och det harmoniska medelvärdet genom

$$H = \frac{3}{\frac{1}{a_1} + \frac{1}{a_2} + \frac{1}{a_3}} \quad \text{eller} \quad \frac{1}{H} = \frac{1}{3} \left(\frac{1}{a_1} + \frac{1}{a_2} + \frac{1}{a_3} \right).$$

UPPGIFT 7. Ändra ditt dataprogram från uppgift 1 så att det beräknar A , G och H för taltripplar (a_1, a_2, a_3) . Välj några tripplar (a_1, a_2, a_3) och beräkna för dem A , G och H . Vilket förhållande mellan A , G och H kan vi observera nu?

För talpar (a_1, a_2) har du tidigare visat att $H \leq G \leq A$ alltid gäller (med likheten då och endast då $a_1 = a_2$). Försök visa att samma förhållande gäller för godtyckliga positiva tripplar (a_1, a_2, a_3) .

Troligen misslyckades du i dina försök.

Nu skall jag försöka leda dig fram till ett resonemang som visar att även för positiva tripplar gäller det ovannämnda förhållandet.

Låt oss istället börja med fyra godtyckliga reella positiva tal a_1, a_2, a_3 och a_4 . Analogt till våra tidigare definitioner, definierar vi

$$A = \frac{a_1 + a_2 + a_3 + a_4}{4} \quad \text{och} \quad G = \sqrt[4]{a_1 a_2 a_3 a_4}.$$

Nu kan vi betrakta A som det aritmetiska medelvärdet för $\frac{a_1 + a_2}{2}$ och $\frac{a_3 + a_4}{2}$ ty $A = \frac{1}{2} \left(\frac{a_1 + a_2}{2} + \frac{a_3 + a_4}{2} \right)$. Om man nu använder olikheten (1) så får vi :

$$\begin{aligned} A &\geq \sqrt{\frac{a_1 + a_2}{2} \cdot \frac{a_3 + a_4}{2}} \quad \{ \text{använd (1) igen och vi för} \} \\ &\geq \sqrt{\sqrt{a_1 a_2} \sqrt{a_3 a_4}} = \sqrt[4]{a_1 a_2 a_3 a_4} = G. \end{aligned}$$

UPPGIFT 8. Definiera även H för a_1, a_2, a_3, a_4 och bevisa att $H \leq G$.

Nu skall vi återvända till vår trippel igen. Låt nu åter

$$(2) \quad A = \frac{a_1 + a_2 + a_3}{3} \quad \text{och} \quad G = \sqrt[3]{a_1 a_2 a_3}.$$

Betrakta nu fyran a_1, a_2, a_3 och $a_4 = A$. Ovan har vi visat att det aritmetiska medelvärdet aldrig är mindre än det geometriska för fyra godtyckliga reella positiva tal, dvs

$$\frac{a_1 + a_2 + a_3 + a_4}{4} \geq \sqrt[4]{a_1 a_2 a_3 a_4}.$$

Om $a_4 = A$ så får vi :

$$(3) \quad \frac{a_1 + a_2 + a_3 + A}{4} \geq \sqrt[4]{a_1 \cdot a_2 \cdot a_3 \cdot A}.$$

Ur (2) för vi att $a_1 + a_2 + a_3 = 3A$ och $a_1 \cdot a_2 \cdot a_3 = G^3$. Sätt in det i (3) och du får $A \geq \sqrt[4]{G^3 \cdot A}$ som kan skrivas om till $A^4 \geq G^3 \cdot A$

dvs. $A^3 \geq G^3$ som i sin tur är ekvivalent med att $A \geq G$, dvs den olikhet som vi ville visa får en trippel.

UPPGIFT 9. Generalisera uppgift 5 till tre medelhastigheter a_1, a_2 och a_3 .

UPPGIFT 10. Formulera ett problem (t. ex. i likhet med uppgift 4) som kan lösas med hjälp av olikheten $G \leq A$ för taltripplar.

Nu är det kanske ganska enkelt att definiera det aritmetiska, geometriska och harmoniska medelvärdet för godtyckligt antal reella, positiva tal n .

Vi definierar alltså

$$A = \frac{a_1 + a_2 + \dots + a_n}{n}, \quad G = \sqrt[n]{a_1 a_2 \dots a_n}$$

och $H = \frac{n}{\frac{1}{a_1} + \frac{1}{a_2} + \dots + \frac{1}{a_n}}.$

UPPGIFT 11. Utveckla ditt dataprogram så att det räknar A, G och H för godtyckliga n -tipplar (a_1, a_2, \dots, a_n) .

UPPGIFT 12. Försök nu visa att $A \geq G$ för $n = 8$ och $n = 16$. Gör beviset analogt till fallet då $n = 4$. Skriv A som det aritmetiska medelvärdet till $\frac{a_1 + a_2 + a_3 + a_4}{4}$ och $\frac{a_5 + a_6 + a_7 + a_8}{4}$.

Hur skall nu olikheten visas för $4 < n < 8$ respektive $8 < n < 16$?

Vi börjar på samma sätt som för $n = 3$.

Om $4 < n < 8$ komplettera a_1, a_2, \dots, a_n till $a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8$ genom att välja

$$a_{n+1} = a_{n+2} = \dots = a_8 = A = \frac{a_1 + a_2 + \dots + a_n}{n}.$$

Nu är det

$$\frac{a_1 + a_2 + \dots + a_n + (8 - n)A}{8} \geq \sqrt[8]{a_1 a_2 \dots a_n A^{8-n}}$$

dvs $\frac{nA + (8-n)A}{8} \geq \sqrt[8]{G^n A^{8-n}}$ som medför att $A \geq G$.
(Genomför räkningen.)

UPPGIFT 13. Genomför hela resonemang även för alla $8 < n < 16$.

Allmänt kan, med hjälp av så kallad *matematisk induktion* och samma resonemang som ovan, visas att:

$$\begin{aligned} A = \frac{a_1 + a_2 + \dots + a_{2^m}}{2^m} &\geq G = (a_1 a_2 \dots a_{2^m})^{\frac{1}{2^m}} \\ &\geq H = \frac{2^m}{\frac{1}{a_1} + \frac{1}{a_2} + \dots + \frac{1}{a_{2^m}}}. \end{aligned}$$

Sedan på samma sätt som ovan kan man generalisera resultatet för godtyckligt antal reella positiva tal.

Mer om matematisk induktion kan du läsa t.ex. i avsnitt 4.2 i A. Vretblads bok Algebra och kombinatorik.

UPPGIFT 14. Försök genomföra fullständigt bevis att, för godtyckliga reella positiva tal a_1, a_2, \dots, a_n gäller

$$H \leq G \leq A.$$

Tänk genom den använda bevismetoden. Observera att metoden kan tillämpas så fort man kan visa olikheten $G \leq A$ för godtycklig n -tippel av reella positiva tal. (För $n = 2$ är beviset av olikheten enklast.)

UPPGIFT 15. Visa att ur olikheten $H \leq G \leq A$ för godtyckliga reella positiva tripplar följer samma olikhet för godtyckliga positiva reella talpar.

UPPGIFT 16. Visa att för varje positivt heltal n gäller olikheten:

$$\left(\frac{n+1}{2}\right)^n \geq n!.$$

LEDNING. Tänk på olikheten $A \geq G$.

Litteratur

Vretblad, A., *Algebra och kombinatorik*. Liber 1985.

Adresslista

CTH & Göteborgs Universitet
412 96 GÖTEBORG
Tel. 031-72 10 00 (växel)

Leif Arkeryd, Jöran Bergh, Lennart Råde, Peter Sjögren.

Högskolan i Luleå
951 87 LULEÅ
Tel. 0920-91 000

Kerstin Vännman, Andrejs Dunkels.

Lunds Universitet
Box 118
221 00 LUND
Tel. 046-10 70 00

Lars Gårding, Lars Hörmander.

KTH
100 44 STOCKHOLM
Tel. 08-790 6000 (växel)

*Anders Björner, Lennart Carleson, Jan Grandell, Björn Gustafsson,
Lars Holst, Johan Håstad, Thomas Höglund, Bo Kjellberg, Göran
Kjellberg, Torbjörn Kolsrud, Dan Laksov, Bernt Lindström, Kirsti
Mattila, Johan Philip, Hans Riesel, Krister Svanberg, Lasse Svens-
son.*

Stockholms Universitet
Box 6701
113 85 STOCKHOLM
Tel. 08–16 45 00

Torsten Ekedahl, Ralf Fröberg, Mikael Passare.

Umeå Universitet
901 87 UMEÅ
Tel. 090-16 50 00

Urban Cegrell, Hans Wallin.

Thunbergsvägen 3
752 38 UPPSALA
Tel. 018–18 32 00

*Leif Abrahamsson, Dag Jonsson, Sten Kaijser, Christer Kiselman,
Pepe Winkler.*

Universitetet i Bergen
Allégaten 53–55
N–5014 BERGEN, Norge
Tel. 00947-5-21 30 50

Øystein J. Rødseth.

