Lennart Carleson

Matematik för vår tid

En presentation och ett debattinlägg

PRISMA

Bokförlaget Prisma Stockholm

Omslag av Dan Jonsson

En originalbok i Prismaserien

Berlingska Boktryckeriet, Lund 1968

Innehåll

1. Inledning

Alla matematiker har många gånger mött frågan: Vad gör du egentligen? Den kommer från familjemedlemmar som ser husfadern dagligen sitta vid sitt skrivbord i timmar med papper, penna och kanske ett par böcker. Speciellt tränger frågan på när de efter ett år ser resultatet: kanske 20 tryckta sidor med helt obegripligt utseende. Och det naturliga svaret förefaller att vara: Praktiskt taget ingenting.

Men frågan kan också komma med större tyngd från administratörer och riksdagsmän och då gärna med tillägget "för nytta". Vi anslår i Sverige kanske sju miljoner kronor till högre utbildning och forskning i matematik. Är detta väl använda pengar? Behövs det verkligen så lång utbildning för att lära barn att addera och räkna reguladetri, och vad tjänar matematisk forskning till? Vi matematiker hänvisar till att lärare måste ha en djupare förståelse för vad de lär ut och att man i vår vetenskap som i alla andra vetenskaper söker sanningen, och att detta är skäl nog. Men en skeptiker blir inte övertygad. Kan man ha någon djupare förståelse för att $2 \times 2 = 4$ — det verkar rena dumheterna. Och vilken sanning är det ni söker? Inte ens en skeptiker som tagit studenten på reallinjen och kanske läst en tid vid universitet eller är ingenjör har fått någon egentlig hjälp: det han lärt sig är vetenskapens resultat på 1600- och 1700-talen och liknar inte alls dagens. Här brukar diskussionen sluta med att matematikerna säger: Vi kan tyvärr inte förklara oss närmare. Ni måste tro oss på vårt ord — vi är ju experterna.

Mest alarmerande är kanske att frågan med ökande skepsis kommer från våra närmaste kollegor, fysiker och övriga naturvetare. Att inte ens de förstår oss är något nytt för vår tid och beror på matematikens snabba utveckling och starka specialisering. Med brist på förståelse följer ofta en allmänt negativ inställning, och man ser nu på många håll i världen en tendens att överflytta matematikundervisningen till tillämpade ämnen. Härmed vill man få garanti för att undervisningen blir "nyttig". Det ligger i öppen dag att detta är olyckligt och innehåller början till stagnation.

Grundorsaken till frågorna är överallt densamma: matematikernas oförmåga att tala om vad de gör och att föra ut sina resultat till icke-matematiker. Det brukar tas som ett axiom att matematik inte kan populariseras, men just denna tes bör man ifrågasätta. Vad som kan förklaras och vad som inte kan förklaras beror på vad som kan tas till utgångspunkt. Att förklara för en stenåldersman vad en radio är skulle förmodligen vara mycket tidsödande, liksom att förklara för en romare för 2 000 år sedan vad formeln $1/3 = 0{,}33\ldots$ betyder. Enligt detta sätt att se skulle alltså svårigheten ha sin rot i undervisningen och speciellt i skolundervisningen. Men läroplanerna i gymnasiet har ju nyligen reviderats, grundskolans står inför en reform och universitetens ändras kontinuerligt. Och "the new math" — slagordet för matematikundervisningens reformering i USA — är ju numera ett begrepp som också ligger bakom förändringarna i Sverige. Har man inte därmed gjort allt som kan göras? I själva verket tror jag att "the new math" gjort problemet större. Man har lagt tyngdpunkten på den moderna matematikens formalism och noggranna definitioner och inte på dess idéer och resultat. Det är oundvikligt att detta är tråkigt och ointressant och ger en vrångbild av dagens matematik. I Sverige har man speciellt tagit upp "mängdlära" i kurserna och detta har på många håll missförståtts. Det rör sig inte om någon teori utan om vanligt sunt förnuft, lika enkelt eller kanske enklare än $1+1=2$. Mängdläran kan emellertid göras svår genom invecklad terminologi och många symboler, och detta förefaller i många framställningar vara huvudsaken.

I Sovjetunionen har "the new math" aldrig slagit igenom — i själva verket har man där en ganska reaktionär syn på matematiken — och i USA förefaller man nu att önska modifiera kurserna i mera traditionell riktning. Men om därmed överdrifterna elimineras, har man inte löst grundproblemet att sprida kunskap om den matematiska forskningen på det sätt som gjorts med atomfysiken. Den kritik som här framförs mot många punkter i den reformerade skolmatematiken får därför inte tolkas som en uppmaning att återgå till det gamla "och beprövade". Forna tiders problemexercis inom snäva områden är sannerligen inget att önska tillbaka. Man bör i stället syfta till nya konstruktiva lösningar och inte okritiskt acceptera vare sig bakåtsträvarnas eller revolutionärernas åsikter.

Denna bok har därför två syften. Främst önskar jag så kortfattat och enkelt som möjligt ge en föreställning om vad matematik av idag är och peka på några av de problem som den fortfarande arbetar med. Speciellt har jag önskat belysa matematikens relation till tillämpningarna, inte genom att betona områden där matematik används nu utan tvärtom betona den matematik som är okänd men som kan ha potentiella användningsområden. Givetvis kan man inte göra detta med vetenskaplig exakthet, lika litet som man kan förklara radioaktivt sönderfall eller relativitetsteori på detta sätt. Jag har i stället försökt att exemplifiera, antyda och överförenkla. Mycket av vad som följer står sig därför inte inför alltför noggrann läsning — den som verkligen önskar sätta sig in i ett område hänvisas till de böcker som omnämns i kommentarerna efter varje kapitel — men det är min förhoppning att boken trots allt ger en riktig bild av matematiken. Troligen kommer många ändå att tycka att mycket är svårt. Det kan bero på att jag uttryckt mig dåligt, men det kan också bero på ovana vid matematiskt symbolspråk och vid de idéer som beskrivs. Det lämpligaste är säkert att bara läsa vidare och sedan gå tillbaka till den dunkla punkten; kanske verkar det hela då lättare.

För att underlätta läsning på detta sätt har partier som kanske är besvärliga eller mera detaljrika markerats med ⟨ i början och ⟩ i slutet. Den som önskar hoppa över en sådan bit behöver då bara söka upp närmaste ⟩ och börja läsa där.

Allmänt har jag försökt att inte förutsätta kunskaper utöver studentexamen, men möjligen kan det finnas undantag. Studentexamenskunskaperna har ju också varierat.

Det andra syftet är att söka ge en bild av var skolmatematiken av idag står i relation till matematiken som helhet och se vilka förändringar man kan tänka sig för att minska den matematiska forskningens isolering. Detta sista måste ses på mycket lång sikt, men om man tror att det är möjligt att ändra något härvidlag, är det angeläget att detta sker snart. I första hand skall då universitetskurserna förändras och man skall därefter gå nedåt i skolan. Genom en ny universitetsorganisation har man gjort det mycket svårt att förändra den allmänna synen på ett ämne. Skolöverstyrelsen har numera ett avgörande inflytande över kursernas uppläggning och innehåll, vilket mycket illa rimmar med att kanske 10 % av de universitetsstuderande kommer att bli lärare. Systemet medför att ett litet antal pedagogiska experter styr inriktningen av våra universitetsämnen och härigenom har man byggt in en tröghet, som kan få mycket allvarliga konsekvenser för hela den akademiska utbildningen.

Dessa frågor faller emellertid utanför ramen för denna bok. Det bör påpekas att jag här inte avsett att lägga fram några förslag till innehåll eller disposition utan velat intressera en större grupp för problemet och för matematiken av idag.

Kommentarer

Allmänna populäröversikter över matematiken är ganska fåtaliga. Den i särklass bästa och fullständigaste är

A. D. Alexandrov, A. N. Kolmogorov, M. A. Lavrentev, *Mathematics. Its Content, Methods and Meaning.* Amer. övers. American Mathematical Society, Providence, R. I., 1962.

Den är skriven av de främsta aktiva ryska matematikerna. Överhuvudtaget har man i Sovjetunionen lagt ned mycket större intresse på att sprida riktig kunskap om matematik än i väster. En ambitiös men ganska svår amerikansk framställning är

E. T. Bell, *The Development of Mathematics.* 2nd ed. 1945.

Det bör påpekas att pocketböckerna i ämnet i regel ej är skrivna av aktiva forskare och de ger ofta en ganska skev bild av matematiken genom sin inriktning på kuriosa.

Matematikens arbetsmetoder har inte alls berörts på föregående sidor. I boken

R. Courant, H. Robbins, *What is Mathematics?* 1946,

ges en behandling av specifika problem som är mycket intressant och upplysande. Problemlösningen har behandlats av G. Polya i olika böcker, t.ex. i

G. Polya, *Of Mathematics and Plausible Reasoning*. Vol. I—II. Princeton 1954.

Slutligen bör nämnas det mycket innehållsrika och underhållande verket

Sigma. En matematikens kulturhistoria. Bd 1—6. Utg. J. R. Newman. Stockholm 1959.

De böcker som citeras i det följande är fackböcker men i regel inte så svåra att förstå. Populariserande arbeten har i regel inte medtagits.

2. Vad är matematik?

Matematikens historia är, kan man föreställa sig, lika lång som mänsklighetens. De mest primitiva levnadsförhållanden i social miljö förutsätter tal, och satsen att kortaste förbindelselinjen mellan två punkter är en rät linje måste ha upptäckts för mycket länge sedan. I de kultursamhällen vi känner i antiken har funnits en matematisk vetenskap. Själva begreppet kultursamhälle förutsätter en positiv inställning till spekulation och tänkande, fristående från nytta och primitiva drifter. De män som sysslade med detta känner vi som filosofer. Deras spekulationer innefattade från början moral och livets mening lika väl som vårt tänkandes natur och matematik. Denna var alltså närmast en del av filosofin. Höjdpunkten nåddes i den grekiska antika kulturen, och det är en närmast otrolig prestation att Euklides då skapade en geometri som inte kunde överträffas förrän på 1800-talet och som för bara några få år sedan lärdes ut till alla barn i hela världen som hade förmånen att alls få gå i skolan. Det som då förde matematiken framåt var kraften hos det mänskliga intellektet utan sidosynpunkter på nytta och användbarhet. Detta var emellertid också svagheten. Begreppsbildningen var för torftig för verkligt intressanta teorier, och man får inte underskatta de svårigheter som följde av något så trivialt som att man länge saknade ett lämpligt sätt att skriva hela tal. Faktum är att mycket obetydliga framsteg gjordes fram till 1600-talet.

Vid denna tid inträffade det naturvetenskapliga genombrottet. Man gjorde stora upptäckter inom astronomin, men framför allt lade Newton fram en beskrivning av dessa upptäckter. Dessa formulerades sedan i ett enkelt språk som Leibniz skapade. Detta språk var matematiken, och därmed hade ur filosofin brutits ut en naturvetenskaplig gren, matematiken, medan ordet filosofi reserverades för den mer spekulativa humanistiska grenen. En viktig del av filosofin var länge logiken, reglerna för slutledningar och korrekta resonemang. Kraven på logiken steg så småningom och det blev nödvändigt att skapa ett särskilt symbolspråk även för den. Därmed hade även logiken fått samma karaktär som matematiken och kan nu anses som en gren av denna.

Kan man då uppfatta matematik som naturvetenskap? Vi har alla en känsla av att påståenden som $1+1=2$ är mycket säkrare, i någon mening sannare, än sådana påståenden som att två kroppar graviterar mot varandra eller att vatten består av väte och syre. Men om man försöker göra en detaljanalys visar det sig mycket svårt att precisera denna föreställning. Vi kan sätta upp invecklade system som ger en innebörd åt talet 1, tecknet + osv., helt oberoende av vår vanliga tolkning. (Denna tolkning är för övrigt nog så invecklad, som vi skall se.) Men vi kan inte tala om systemet eller ge det ett sanningsinnehåll utan att falla tillbaka på vår erfarenhet om vardagligt räknande. Sedan slutet av 1800-talet har matematiker och logiker gjort stora ansträngningar att komma ur detta dilemma och ge matematiken ett absolut sanningsinnehåll. Jag tror att man kort kan säga att försöken misslyckats och också att programmet med all sannolikhet är dömt att misslyckas.

Vi skall först se litet närmare på vad som gjort det möjligt att behandla logiska problem på samma sätt som matematiska. Logiska resonemang innehåller en del som på ett slående sätt liknar vanlig algebra. Sammanställningarna ”både — och” samt ”dels — dels” motsvarar då ”gånger” och ”antingen — eller” motsvarar ”plus”. ”Inte” är analogt med minustecken. Tag t.ex. formeln $a \cdot (b+c) = a \cdot b + a \cdot c$. Vi kan få en motsvarighet inom logiken på följande sätt. Låt a betyda ”A är snäll”, b ”A är gammal”, c ”A är ful”. Vänstra ledet av formeln är då ”A är dels snäll och dels antingen gammal eller ful” och högra ledet ”A är antingen snäll och gammal eller också snäll och ful”. Uppenbarligen är detta en språklig omstuvning utan innehåll på samma sätt som den algebraiska relationen inte utsäger något om talen a, b och c. På samma sätt motsvaras $-(-a)=a$

av att ”A är inte inte-snäll” är detsamma som att ”A är snäll”. Man kan alltså genom ”algebraiska” regler förenkla språkliga satser.

Reglerna för slutsatser kan på liknande sätt formaliseras. Här tillkommer då att man tillordnar satser ett formellt sanningsinnehåll och ger regler för sanningsvärdet hos sammanställda satser. Så är då *a* och (dvs. gånger) *b* sann bara om både *a* och *b* är sanna, medan *a* eller *b* bara kräver endera satsen sann för att vara sann. Det är på intet sätt nödvändigt att förstå eller känna till alla detaljer och symboler i detta mycket invecklade logiska språk för att förstå det väsentliga. Detta består i att man här fått logisk kalkyl som fungerar oberoende av någon tolkning av de symboler som ingår.

Nästa steg i försöket att göra matematikens grunder säkra är att införa grundbegreppen med hjälp av det logiska symbolspråket. Vi skall som hastigast syssla med heltalen 0, 1, 2, ... Dessa skall alltså definieras utan någon hänvisning till intuition eller till erfarenhet. Man utgår från abstrakt mängd *M* föremål som skall uppföra sig som heltalen. Man döper t.ex. ett föremål till ”noll”, talar om att varje föremål har en bestämd granne som man kallar efterföljare. Slutligen ger man vissa regler som att om två föremål har samma efterföljare är de lika. Slutledningsreglerna skall vara abstrakt logiska och får inte vara beroende av någon tolkning av föremålen som tal.

Vad som sker är tydligen att man väljer ut vissa egenskaper hos de välkända heltalen, utnämner dessa till grundbegrepp och definitioner. Detta kan förefalla formalistiskt och egentligen meningslöst och på det hela taget får man nog säga att det är det också. Man bör nämligen begära av det konstruerade systemet två saker. (1) Man måste kunna återfinna *alla* egenskaper hos de hela talen. (2) Man måste kunna *bevisa* att det system som konstruerats inte motsäger sig självt. Just detta är det som misslyckats. Misslyckandet är också så stort som är tänkbart: man kan bevisa att (1) och (2) ej samtidigt kan uppfyllas. Inför man således så många axiom, dvs. spelregler för de abstrakta föremålen, att systemet får alla egenskaper hos de hela talen, har man fått så många komplikationer att (2) är omöjligt.

Vad jag här skildrat är utvecklingen under perioden 1880—

1935. Helt naturligt har matematikerna här resignerat, och man använder nu grundläggande begrepp utan bevis att man inte kan få motsägelser. Givetvis kan man lugnt göra detta. Våra föreställningar om t.ex. heltalen är så intimt förknippade med vårt sätt att tänka och vår erfarenhet om omvärlden att det är närmast meningslöst att försöka tänka bort dem. Det är över huvud taget mycket suggestivt att ett ögonblick reflektera över relationen mellan vårt tänkande och vår omvärld. Vi accepterar väl alla nu att människan utvecklats ur molekyler och urceller under ständig påverkan av omgivningen, dvs. den fysiska världen. Det är därför naturligt att anse att vår uppfattning av logik, av orsakslagen osv. inte alls är något principiellt höjt över naturlagarna utan snarare är ett resultat eller en spegling av dessa. I denna belysning kan man förstå att matematik och logisk analys kan användas till att förutsäga händelser i naturen; de lagar vi under stor möda upptäcker finns så att säga inbyggda i oss själva. Matematik är på detta sätt syntesen av naturvetenskaperna och av organiserat tänkande överhuvudtaget. I det här nämnda perspektivet är det också förklarligt att vissa problem som t.ex. världsalltets oändlighet eller tidens början medför oöverstigliga tankesvårigheter för oss. Det finns ingen anledning att förvänta att vår hjärna skulle kunna förutsäga något i dessa frågor eller att de naturlagar vi observerat skulle gälla för förlopp under tidsrymder av storleksordningar överstigande mänsklighetens egen existensperiod. Det är en fundamental skillnad mellan 10 och 10^{100} år.

De eskatologiska frågor som i förbigående berördes ovan är exempel på nödvändigheten att ha en begränsad ambitionsnivå för problem och lösningsmetoder. Det finns en mycket allmän princip i logiken, som bevisats för heltalen. Den innebär ungefär att man till varje någorlunda komplicerat axiomsystem alltid kan finna ett påstående som formuleras med systemets termer och som man aldrig kan bevisa eller motbevisa med systemets axiom. Man har alltså rättighet att anta att det är sant eller falskt, vilket man vill, och detta kommer inte att medföra någon motsägelse. I denna mening är alltså till och med logiska system relativa och godtyckliga. Ingenting hindrar således att vissa berömda öppna matematiska problem i denna mening saknar sanningsinnehåll: det är en smakfråga om man väljer att anse dem sanna eller falska. Detta skulle kunna gälla för ett så enkelt problem som följande. Finns det oändligt många 9-or i decimalbråksutvecklingen av talet π? Ett problem av denna typ är i en vag mening helt onaturligt — det finns inte något förnuftigt samband mellan 10 och cirkelns omkrets — och man kan inte föreställa sig någon metod att lösa det. Av denna anledning skulle antagandet att frågan besvaras med ja (som ligger närmast till hands) eller nej, aldrig komma i någon relation till traditionell matematik, och det är därför "likgiltigt" vad som gäller.

Det vore nu ett stort misstag att tro att matematikernas arbete med grundfrågorna varit bortkastat. Vad man fått resignera inför är en fullständig utredning av de enklaste begreppen och orsakerna är uppenbarligen att dessa är så nära förknippade med de metoder vi använder för att analysera dem. Man kan jämföra detta med den berömda osäkerhetsprincipen i fysiken. Denna innebär att man inte kan mäta vissa fysikaliska kvantiteter samtidigt med hur stor noggrannhet som helst, eftersom de apparater man utnyttjar själva samverkar med och stör det som skall mätas. För mera invecklade begrepp är emellertid den nya logiken ett mycket effektivt vapen för en korrekt analys. Det är lätt att illustrera behovet av en sådan noggrann behandling inom mängdläran.

Vi har en intuitiv föreställning av begreppet mängd, samling osv. Det innebär att man studerar objekt av en viss typ samtidigt och tar dessa tillsammans till en enhet. Denna idé är troligen den mest fundamentala i det matematiska tänkandet och vi kommer i fortsättningen gång på gång tillbaka till den. Vad vi här bör observera är att begreppet är mycket komplicerat och att man måste iaktta försiktighet då man använder det. Russells berömda paradox illustrerar detta. Om vi nu anser oss veta vad som menas med mängd, så kan vi ju också tala om något sådant som mängden bestående av alla mängder. En mängd kan tänkas ha sig själv som element, även om detta är svårt att föreställa sig. Men låt oss nu tvärtom undersöka mängden av alla mängder som inte har sig själva till element. Om man nu tänker efter, inser man att denna mängd inte kan ha sig själv till element (definitionen förutsätter ju just detta) och också måste ha det (ty motsatsen är orimlig). Något är alltså fel och felet ligger i att vi handskats för vårdslöst med definitioner av mängder. Det är emellertid mycket komplicerat att ställa allt till rätta och här spelar det exakta logiska språket en stor roll.

Grundsvårigheten i Russells exempel ligger i vad som menas med att en mängd är väldefinierad. Vi har sett att det leder till orimligheter om man accepterar hur lösliga definitioner som helst. Den extremt motsatta ståndpunkten är att man fordrar att varje objekt som studeras skall kunna konstrueras med ett ändligt antal steg och med en välpreciserad metod ur det tidigare givna. Denna riktning hade i början av detta sekel många anhängare och principen innebar i matematikens dåvarande situation inte en så allvarlig inskränkning. Huvuddelen av resultaten kunde bevisas på detta konstruktiva sätt. Läget har nu radikalt ändrats och av dagens matematiker kanske närmare hälften bryter mot regeln. Den mest berömda motstridande principen är urvalsaxiomet: om S är ett system mängder M, kan man bilda en mängd M_0 som innehåller precis ett element från varje mängd M i S. Man vet numera att om vi inte redan har motsägelser i traditionell matematik (vilket man alltså inte vet!) så får vi heller inga genom att godta urvalsaxiomet. Det är också känt att axiomet inte är en följd av övriga vanliga axiom. Detta axiom innebär att vidlyftiga icke-konstruktiva metoder blir tillåtna. Det leder till utomordentligt egendomliga konsekvenser — till exempel att det finns lineära funktioner $f(x)$, dvs. sådana att $f(x+y)=f(x)+f(y)$ för alla reella tal x och y, som inte har formen $f(x)=a \cdot x$ — men är trots detta en av de viktigaste grundmetoderna i modern matematik.

En intensiv strid kring dessa grundfrågor fördes i början av 1900-talet. Den slutade i allmän utmattning och utan att klarhet nåtts. Generellt kan man säga att frågan blivit empiriskt löst. Man har inom olika områden olika krav som är avpassade till områdenas naturliga metodik. Härigenom har man så att säga olika hierarkier av matematik med olika underförstådda eller klart utsagda axiomsystem, och man vet också nu rätt väl hur den logiska relationen mellan systemen ser ut. Även om så inte är bevisat och sannolikt aldrig blir bevisat, kan man säga att risken för att någon gren av matematiken skall visa sig innehålla någon motsägelse är utomordentligt liten.

Vad innebär nu vår ofullständiga kunskap om matematikens grunder för skolundervisningen? En rätt självklar konsekvens är att man inte bör krångla till införandet av grundläggande begrepp, och speciellt då de hela talen, genom några invecklade system. En sådan konstruktion leder enbart till att svårigheten flyttas från en punkt till en annan. Detta insåg man beträffande Euklides geometri, som ju helt utmönstrats. Givetvis är det en stor fördel att exercerandet med skenbart logiska resonemang försvunnit ur geometriundervisningen, men som vi skall se har man här gått för långt, så att kunskapen om geometri blivit dålig. I de nya kurserna har emellertid i stället andra grenar formaliserats, varigenom man avser att illustrera matematikens axiomatiska uppbyggnad och öva eleverna i logiskt tänkande. Som det kanske värsta exemplet kan man ta kvasiaxiomatiseringen av exponentialfunktionen och räknelagarna för hela tal. Denna syn på matematikens grunder gör gärna ämnet tråkigt och formellt och förmedlar heller inte en riktig uppfattning om hur matematiken arbetar eller vad axiomatik är. Till sist är det ju också så, som vi här sett, att den intelligente eleven i princip kan göra vilken professor som helst svarslös genom att ifrågasätta sådana saker som resonemangs giltighet och ords betydelse. Man bör således inskränka bevis och definitioner till sådana fall där inte allt är intuitivt klart från början. Det bör dock noteras att till det intuitivt klara hör betydligt mindre än vad forna tiders matematiker och majoriteten av elever anser, men dit hör avgjort hela tal och den del av mängdlära och logik som är aktuell i skolsammanhang.

Under senare år har en ny motsättning uppstått inom matematiken mellan vad som kallas ren och tillämpad matematik. Denna distinktion är något nytt för vårt århundrade. Alla framstående matematiker var tidigare samtidigt astronomer eller fysiker och drömde inte om att det skulle ligga någon motsättning i detta. Tvärtom använde de matematik till att lösa fysikaliska problem och fick inspiration till matematiska frågeställningar och begrepp ur fysikaliska förhållanden. Genom att vår kunskap nu är så omfattande är en specialisering nödvändig, och matematiker saknar numera i regel direkt kunskap om t.ex. fysik. De skapar i stället sina problem själva i relation till det tidigare kända. Detta är på många sätt en försämring och har gjort matematikernas arbete svårare. De problemställningar som problem i naturen ger upphov till har stora utsikter att vara matematiskt givande. När man själv ställer sina problem, vägleds man enbart av sin intuition om sammanhangen, och riskerna att komma in i en återvändsgränd kan då vara stora.

Matematikens problemval är en mycket intressant fråga genom att ämnet i princip är oberoende av yttre förhållanden. Vilka problem är intressanta? Vad menas med att en sats är svår? I stort sett vet man mycket litet faktiskt i frågor som dessa, men man kan finna en närmast förbluffande samstämmighet i uppfattningen hos olika matematiker. Detta är så mycket märkligare som det finns rent emotionella reaktioner med i värderingarna som närmast kan jämföras med estetiska. Man talar om vackra satser och eleganta bevis. Denna enighet tyder på att det skulle finnas en mer objektiv grund. En sats svårighetsgrad med avseende på ett givet kunskapsomfång och med ett givet axiomsystem skulle kunna tänkas definieras som det minsta antal logiska symboler ur axiomsystemet som krävs för beviset. Ett resultats totala svårighet blir då antalet symboler i ett totalt bevis ur axiomen. Givetvis kan detta antal knappast beräknas för konkreta resultat, men uppskattningar av storleksordningen skulle kanske kunna erhållas. Ett resultats "intresse" kanske sammanhänger med detta antal dividerat med antalet symboler som krävs för att formulera resultatet. — I avvaktan på objektiva kriterier kan helt allmänt sägas att en matematikers (och kanske varje vetenskapsmans) storhet bedöms minst lika mycket utifrån de frågor han ställer som från dem han besvarar.

Ur den uppfattning om matematikens natur som vi redovisat följer att skillnaden mellan ren och tillämpad matematik egentligen inte är så stor. De utredningar om logiska sammanhang som matematiker gör är ett djupare inträngande i naturens ordning och kan göra anspråk på att vara naturvetenskap med samma rätt som t.ex. teoretisk fysik. Det har också gång på gång inträffat att matematiska teorier, skapade för sin egen skull, visat sig vara lämpliga modeller för fysikaliska problem. Detta gällde t.ex. differentialgeometrin för Einsteins relativitetsteori, matrisläran för Heisenbergs beskrivning av vågmekaniken, och många tror nu att gruppteorin skall spela en stor roll då fysikerna en dag får ordning på alla de elementarpartiklar som föds i de stora acceleratorerna.

Om splittringen i ren och tillämpad matematik varit skadlig för matematiken har skadan förmodligen varit än större för naturvetenskapen i övrigt. Den moderna matematiken har givit en fördjupad förståelse för de logiska sammanhangen och skapat ändamålsenliga språk. Det är mycket sannolikt att denna kunskap kan utnyttjas för en fördjupad beskrivning av fysikaliska fenomen, men det finns allvarliga brister i kommunikationerna. Man har i regel uppfattat motsättningen som gällande praktiskt räknande å ena sidan och teoretiska (underförstått ointressanta) spekulationer å den andra. Detta är dock ett sekundärt problem. Det som gjort saken akut torde vara att skolmatematikerna — i planeringen av "den nya matematiken" i skolor och på universitetsnivå — överbetonat de formella sidorna av matematiken på bekostnad av det egentliga idéinnehållet. Det är inte avsikten att kritisera kravet på stringens, som var en välbehövlig reaktion mot tidigare generationers lösliga framställning, utan det beskäftiga detaljintresset på sådana områden där studiet i alla fall stannar på ytan. Överbetoningen av detaljerna och det formella leder där bara till att man gör området ointressant. Det är inom undervisningen nödvändigt att skilja på orienterande kunskaper och aktiva kunskaper. Den orienterande undervisningen kan läggas upp på samma sätt som undervisningen i elementarpartikelfysik, medan undervisning i den del av matematiken som behöver användas måste innehålla mycken rutinövning. Det är givetvis helt illusoriskt att man skall kunna ge eleverna en sådan förståelse för ämnet genom omsorgsfull behandling av grunderna att de själva därefter skulle kunna utnyttja metoderna i problemlösning. Först genom en systematisk uppdelning av den antydda typen är det möjligt att realisera skolans mål att samtidigt öka stoffets omfång och behålla ämnets egenart. Genom en sådan uppdelning skulle den nuvarande motsättningen mellan ren och tillämpad matematik minskas, och bredare förståelse för ämnets natur skulle åstadkommas. Vi skall diskutera dessa frågor utförligare i sista kapitlet.

Vad blir då resultatet? Vad är matematik? I grunden är det en naturvetenskap med uppgift att analysera de logiska konsekvenserna av vissa grundläggande empiriska sanningar. Den skapar ett lämpligt språk för detta ändamål, det matematiska symbolspråket. Det är ett misstag att tro att matematiken är absolut sann, och det förefaller vara en djävulscirkel att med logiska metoder analysera alla grundbegrepp i matematiken. Detta hindrar inte att man kan och bör genomföra matematiska bevis med logisk stringens och så att begreppens innehåll och slutledningen framstår som självklara, men man skall inte göra det självklara komplicerat genom beskäftig och missriktad formalisering. Sina problem väljer matematikerna själva på jakt efter ökad förståelse för sammanhangen mellan alla de begrepp som kommer från talens värld, från geometrin, dvs. vår rumsuppfattning, från olika fysikaliska sammanhang, från regelbundenheter i form av statistik, från logiken, och som sedan omstöpts i matematikens smältdegel. Ofta har resultatet blivit oigenkännligt för företrädare för ursprungsområdena, men sammanhanget finns där klart och en betydelsefull uppgift är att se till att kontakten bibehålls. Bilden av matematiken som en spegel av naturen ger en antydan om de möjligheter matematiken erbjuder.

H. Weyl har sammanfattat detta på följande sätt (i fri översättning):

"En verkligt realistisk matematik bör på ungefär samma sätt som fysiken uppfattas som en gren av den teoretiska beskrivningen av den värld vi lever i, och den bör anlägga samma lugna och försiktiga attityd mot utvidgningar av dess grundföreställningar som fysiken traditionellt visar."

Kommentarer

En modern och ganska lättläst översikt av grundfrågorna är

G. T. Kneebone, *Mathematical Logic and the Foundations of Mathematics*. Van Nostrand Ltd, 1963.

Den sats av K. Gödel som avses på s. 14 upptäcktes 1931.

Att urvalsaxiomet inte strider mot övriga axiom i mängdläran bevisades av Gödel 1940 [The Consistency of the Continuum Hypothesis. Princeton, N.J., 1940].

Att urvalsaxiomet är oberoende av övriga axiom — dvs. att dess negation inte heller strider mot dem — visades av P. J. Cohen 1963—64 [The Independence of the Continuum Hypothesis. Proc. Nat. Acad. Sciences 1963, 1964]. Vid de matematiska kongresserna som hålls vart 4:e år utdelas två pris till de främsta yngre forskarna. Detta är de s.k. Fieldprisen. De motsvarar i fråga om ära närmast Nobelpris. P. Cohen fick ett sådant pris vid Moskvakongressen 1966 för detta arbete.

En bok om matematikens historia är

D. J. Struik, *A Concise History of Mathematics*. Vol. I—II. New York 1948. Svensk övers. *Matematikens historia*. Stockholm 1966.

En intressant diskussion om matematiken och verkligheten innehåller

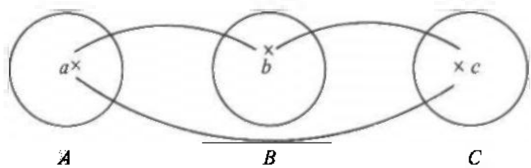
H. Weyl, *Philosophy of Mathematics and Natural Sciences*. Princeton, N.J., 1949,

varur citatet ovan är taget.

3. Tal, mängder, strukturer, avbildningar

Vi är alla inställda på att det mest grundläggande matematiska begreppet är de hela talen. Tag t.ex. följande berömda sentens av Kronecker (som var en av de strängaste anhängarna av idén att endast konstruktiva bevis är tillåtna): "De hela talen skapades av Gud. Allt annat är människoverk." Men är det verkligen så och vad är egentligen tal?

Den idé som är självklar för oss och som vi fordrar att våra sjuåringar skall förstå är följande. Om vi tänker på 2 äpplen, 2 bilar, 2 stjärnor osv. har dessa olika *mängder* något gemensamt, som vi kallar antal och i detta fall kallas två. Det verkar här som om vi använde "2" i definitionen av "2", men så är det inte. Om vi studerar en mängd A och en annan mängd B, vilkas element kan vara vad som helst, äpplen eller bilar, så finns det två (!) möjligheter. Antingen går det att para ihop elementen i A och B så att mängderna precis töms ut, dvs. så att varje a motsvarar ett b och varje b ett a, eller också går det inte. I det första fallet säger vi att A och B har samma antal element, eller har samma mäktighet (h.s.m.). Tydligen är det så att om A h.s.m. som B, och B h.s.m. som C följer att A h.s.m. som C. Vi behöver bara para a med ett c på sådant sätt att de motsvarar samma b. En schematisk figur:



Vidare gäller att A h.s.m. som A (para a med sig själv). Vi för nu ihop alla mängder som har samma mäktighet till en enda mängd. Då får vi en samling supermängder, vilkas element således är mängder. Dessa har egenskapen att varje mängd tillhör en bestämd sådan supermängd. Den egenskap våra 2 äpplen och 2 stjärnor har gemensamt är, att de tillhör samma supermängd och härigenom kan vi säga att de vanliga hela talen *är* de enklaste supermängderna. Låter det här komplicerat? Det är dock precis detta som vi anser självklart och det är den här tankeprocessen våra barn måste göra för sig själva. Den fundamentala idén är att slå samman en grupp föremål till en enhet och studera denna enhet. Enheten tillordnas ett abstrakt begrepp, talet, och man kan sedan laborera med talen och glömma alla de enheter de representerar. Att det är något svårt i denna abstraktion kan varje småskollärare omvittna. Det är mycket lättare att räkna med äpplen och päron, dvs. representanter för supermängden, än med abstrakta tal! Att det så småningom går bra i skolan beror mindre på att eleverna förstått sammanhanget än på att de lärt sig allt utantill. Det är givetvis inget fel i att kunna utantill, men det kan vara intressant att ställa situationen här i relation till vad man högre upp i skolan anser om utantillkunskaper. Ambitionen där är att söka eliminera minneskunskap och basera så mycket som möjligt på deduktion. Att gå långt i den riktningen är med säkerhet mycket ineffektivt. Det är troligt att det skulle gå bra att göra sjuåringar medvetna om vad de egentligen gör och att detta skulle underlätta inlärandet av de grundläggande räknesätten. Innan vi går in på detta, skall vi fortsätta mäktighetsdiskussionen ett steg vidare.

I framställningen om hopparade element var det ändliga mängder vi hade i tankarna. Men ingenting hindrar att man behandlar oändliga mängder på samma sätt. Den supermängd till vilken den enklaste oändliga mängden hör, nämligen den som består av de vanliga talen 1, 2, ... betecknar man alltsedan G. Cantor skrev sitt banbrytande arbete på området, $\aleph_0$ ($\aleph$ är alef — första bokstaven i det hebreiska alfabetet). Nästa fråga blir då: finns det någon oändlig mängd som ej hör till denna klass, dvs. som ej har mäktigheten $\aleph_0$? Det enklaste exemplet på en sådan mängd är mängden av alla reella tal x mellan 0 och 1. ⟨ Antag nämligen att de reella talen kunde paras med de naturliga talen. Det betyder att vi skulle kunna ge dem ordningsnummer 1, 2, 3, ... etc. Utveckla nu varje tal i decimalbråk. Talet nummer 1 har en viss första decimalsiffra. Välj ett heltal a_1 mellan 1 och 8 olikt denna decimalsiffra. a_2 väljs analogt skilt från talet 2:s 2:a decimalsiffra, a_3 från 3:s 3:e siffra osv. Det tal som har decimalutvecklingen 0, a_1 a_2 a_3 ... kan då inte vara lika med något av talen med ordningsnummer 1, 2, ..., eftersom för varje tal någon decimalsiffra skiljer. Alltså har vi hittat ett tal mellan 0 och 1 som inte fanns med i uppräkningen och slutsatsen är oundviklig. Det går inte att ge de reella talen en numrering. ⟩

Mäktigheten hos mängden av reella tal kallas $\aleph_1$. Det är lätt att ge en precis innebörd till att $\aleph_0$ är mindre än $\aleph_1$, nämligen att heltalen kan paras med en del av mängden av reella tal, t.ex. delen $\{1, 1/2, 1/3, \ldots\}$, men inte tvärtom. [Visa det!] Finns det en mängd vars mäktighet i denna mening ligger mellan $\aleph_1$ och $\aleph_0$? Detta har varit ett berömt problem ända fram till för några år sedan då P. Cohen slutgiltigt löste problemet på ett sätt som är karaktäristiskt för problem av denna typ. Vi kan valfritt svara ja eller nej på frågan. Vårt svar kommer inte att strida mot övrig matematik. Problemet är i logisk mening oberoende av sedvanliga axiomsystem.

Det här nämnda belyser hur subtil mängdteorin är. Man får heller inte tro att utvecklingen av denna är något centralt problem i matematiken. Det viktiga är själva begreppet mängd, medan de manipulationer som man utför med mängderna bara utnyttjar enkelt sunt förnuft. I undervisningen bör därför idén föras fram tidigt, redan i de första klasserna, och eleverna bör göras vana vid att föra samman saker i klasser och använda termer som mängd, delmängd, tillhöra en mängd, komplement till en mängd osv. Man kan få många uppslag till arbetsuppgifter av denna typ i s.k. intelligenstest, vilket visar att psykologer funnit begreppsbildningen central i vad man anser vara intelligens. Det förefaller mycket troligt att träning i tidig ålder i denna riktning är en god introduktion till matematiskt tänkande (och ger bra resultat på intelligenstest). Å andra sidan kan den formella träningen med mängdlärans symbolspråk saklöst utgå — när man förstått det intuitiva är allt sådant självklart, och innan man gjort det är det svårt, tråkigt och meningslöst.

Vi skall nu se hur begreppet mängd används i matematiken, och vi börjar med att se hur man logiskt noggrant bygger upp systemet av tal. Tag först de rationella talen 1/2, 2/3, 1/4, 3/4, 1/5, . . ., för att antyda dem mellan 0 och 1. Vi är vana att beteckna dem på detta sätt, alltså två heltal efter eller oftare ovanför varandra med ett streck emellan, men vad 1/2 betyder är att heltalen 1 och 2 i denna ordning bestämmer ett visst rationellt tal. En korrekt procedur går då till på följande sätt. Man utgår från de traditionellt välkända egenskaperna hos de rationella talen. De räknelagar som där gäller tas till definitioner för räknelagar för par av heltal (a, b) där a motsvarar täljaren och b nämnaren och alltså $b \neq 0$. Nu skall ju t.ex. (1, 2) vara detsamma som (3, 6) och därför *identifierar* man paret (a, b) och paret (c, d) om $ad = bc$ $(1 \cdot 6 = 2 \cdot 3)$. Man definierar operationen $+ + : (a, b) + + (c, d) = (ad + bc, bd)$ genom att kopiera formeln $a/b + c/d = (ad + bc)/bd$. På samma sätt gör vi definitionen $(a, b) \cdot \cdot (c, d) = (ac, bd)$. I båda fallen är det viktigt att för heltal gäller att ur $b \neq 0$ och $d \neq 0$ följer $b \cdot d \neq 0$. Vi återfinner alltså de rationella talen i den logiska formen av par av naturliga tal med operationer $+ +$ och $\cdot \cdot$. Vi vill nu att våra gamla naturliga tal skall finnas kvar i systemet och gör en ny definition: naturliga tal kallas nu alla par där $b = 1$. — Uppenbarligen tjänar allt detta inte mycket till för att göra matematikens grunder säkrare: man gör bara en formalisering av vad mänskligheten gjort under tusentals år. Själva konstruktionsidén är emellertid mycket allmän och användbar som vi senare skall se.

På samma sätt kan man (sedan man formellt infört 0) införa de hela talen $(0, \pm 1, \pm 2, \ldots)$ som par (α, a) där α kan ha två betydelser (förtecknet; $+$ eller $-$ är vi vana att skriva) och a är 0 eller ett naturligt tal. Det kan vara en lärorik sysselsättning att göra detta och kontrollera att allt stämmer.

De reella talen är av en annan natur än de rationella. Man kan inte åberopa några iakttagelser i naturen som skäl för att införa dem. Tvärtom, den uppfattning av materien som man numera har innebär att allt iakttagbart förekommer i en minsta enhet som inte vidare kan uppdelas. Detta gäller t.ex. både energi och avstånd. Det skulle alltså gå att göra beskrivningar av naturen under användande av enbart decimalbråk med t.ex. 100 siffror. Man brukar också låta talen motsvara punkter på en linje, dvs. införa koordinater som på en tumstock, och inget strider mot föreställningen att linjen består av punkter på 10^{-100} cm avstånd från varandra. En sådan diskontinuerlig uppfattning skulle genast förklara den berömda paradoxen om Akilles och sköldpaddan: varje gång Akilles sprungit till den plats där sköldpaddan nyss var, har denna hunnit ytterligare en bit och därför hinner Akilles aldrig enligt en kontinuerlig uppfattning ikapp. Finns det emellertid en minsta sträcka och en minsta tid måste Akilles ta ett språng till sköldpaddans plats innan denna hinner därifrån. Det kan här vara intressant att citera skolöverstyrelsens läroplan (s. 261): ”Man utvidgar mängden av rationella tal så att alla punkter på tallinjen tillordnas en koordinat.” Detta är naturligtvis ingen hållbar motivering, eftersom vi inte vet mer faktiskt om punkter på linjer än om reella tal, inte ens om vi accepterar de traditionella euklidiska axiomen för geometrin.

Vad man emellertid skulle förlora i en dylik diskret beskrivning är enkelheten. Formlerna skulle bli invecklade, man skulle få besvärliga regler för avrundning osv. De reella talen kan därför anses vara en matematisk idealisering som införts i förenklande syfte. Geometriskt uttrycker de vår föreställning om att linjen är homogen och sammanhängande.

Logiskt inför man talen enklast genom att resonera baklänges från de sedvanliga räknereglerna för t.ex. decimalbråk. Varje tal x skall kunna skrivas som decimalbråk $x = X, x_1 x_2 \ldots x_n \ldots$, där X är ett naturligt tal och varje x_n är ett heltal mellan 0 och 9. Man observerar också att sådana decimalbråk som 0,1000 ... och 0,0999 ... motsvarar samma tal. Vi kan nu baklänges definiera ett reellt tal som själva följden av hela tal, $\{X, x_1, x_2, \ldots\}$, och införa räkneregler som stämmer om vi skriver upp decimalbråket som förut och räknar som vanligt. Logiskt *är* alltså i denna tolkning ett reellt tal en oändlig heltalsföljd precis som ett rationellt tal var ett heltalspar. De rationella talen återfinns som de följder som är periodiska (t.ex. $53/165 = 0{,}3212121\ \ldots$, där perioden har längden 2). [Bevisa detta, dvs. att rationella tal har periodiska decimalbråk och att periodiska decimalbråk motsvarar rationella tal!]

Vi har alltså sett att mängdbegreppet kommer in på ett avgörande sätt vid noggranna matematiska definitioner. Kanske övertygar detta inte riktigt om att begreppet är viktigt, eftersom det vi definierat är så konkret och välkänt. Vi skall därför ta ytterligare ett exempel av denna typ, nämligen begreppet funktion eller avbildning. Intuitivt uttrycker en funktion att om en kvantitet (i vid mening) är känd kan en annan beräknas ur denna kunskap enligt någon regel. Så är lufttrycket (y) en funktion av höjden (x) över havet, en sträcka (y) en funktion av de båda ändpunkterna (x = ett punktpar), en människas egenskaper (y) en (komplicerad) funktion av arv och miljö (x). Funktionsidén uttrycker således den fundamentala föreställningen om orsak (x) och verkan (y). Den lämpligaste logiska definitionen är att funktionen i fråga är just alla de par (x, y) som kommer i fråga. Varje x förekommer här endast en gång — varje orsak har alltid samma verkan — medan samma verkan kan uppstå ur många orsaker. Då man tänker på funktioner är tolkningen som avbildning från mängden X av orsaker till mängden Y av verkningar oftast den lämpligaste. f fungerar som en maskin:

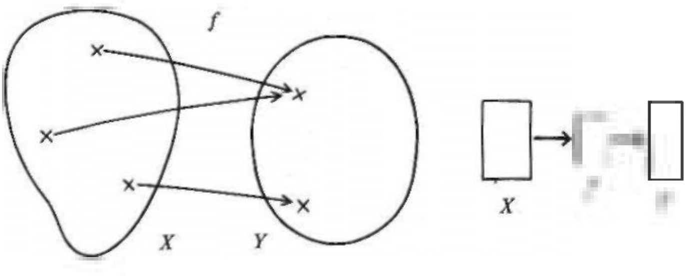


Fig. 1

man stoppar in x och får ut y.

Då vi på s. 23 diskuterade mäktighet hade vi att göra med

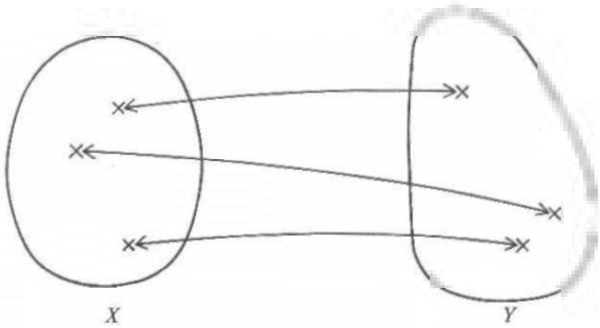


Fig. 2

funktioner f från en mängd till en annan med samma mäktighet. f är här en $1-1$-avbildning, där alltså pilarna i figuren kan gå i båda riktningarna (fig. 2).

Även funktionsbegreppet med sin terminologi skulle man med fördel kunna införa under de första skolåren för att tidigt vänja eleverna vid orden och tankegångarna. Det är inte fråga om att man skulle göra några invecklade konstruktioner utan endast introducera ett matematiskt tänkande.

Vad gör nu matematikerna med mängderna? Man kan säga att de studerar mängder som på olika sätt har en struktur. Vad som kan menas med detta skall vi nu närmast syssla med.

Den enklaste strukturen av en mängd M består i att man har givet ett system S av delmängder av M. Detta system bör uppfylla diverse villkor, av vilka det viktigaste är att föreningen av två mängder i S liksom den gemensamma delen också ingår i S. Detta medger en intressant och oväntad tolkning. Vi tänker oss ett experiment av något slag, fysikaliskt eller för övrigt vad som helst. Det kan utfalla på olika sätt och vi låter M vara mängden av möjliga resultat. En delmängd m av M representerar då en viss del av utfallen och kallas händelse. Föreningen $m_1 \cup m_2$ av m_1 och m_2 representerar då "antingen händelsen m_1 eller händelsen m_2" och den gemensamma delen "både m_1 och m_2". Vi får alltså en spegling av logiska operationer i operationer på mängder. Vi kan nu göra en exakt definition av sannolikhet. Vi antar att P är en funktion från S

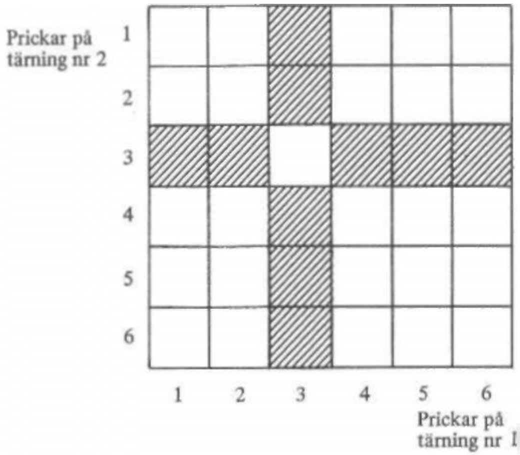


Fig. 3

till intervallet (0, 1) sådan att $P(m_1 \cup m_2) = P(m_1) + P(m_2)$ om m_1 och m_2 saknar gemensam punkt. Vi kan tänka oss situationen så, att man lagt en massa på M av totalvikt 1 och $P(m)$ är den massa som faller på m. Vi kallar nu $P(m)$ "sannolikheten för händelsen m". — Det sagda behöver belysas av ett exempel. Låt experimentet vara att vi kastar två tärningar. Det kan utfalla på 36 sätt, så M består av 36 punkter. S är här alla möjliga delmängder, men så enkel blir situationen egentligen bara då M är ändlig. P lägger massan 1/36 i varje punkt av M. Den skuggade mängden i fig. 3 representerar händelsen "precis en av tärningarna har 3 prickar" vilken alltså har sannolikhet $10 \cdot 1/36 = 5/18$.

Denna mängdstruktur tjänar alltså till att ge en klar matematisk bakgrund till sannolikhetsteorin. Vi skall senare studera detta närmare.

Vi skall nu se hur analysens grunder hänger samman med mängdidén. Det centrala begreppet inom analysen är konvergens. Om $x_1, x_2, x_3, \ldots, x_n, \ldots$ är en följd reella tal, säger man att x_n konvergerar mot x om x_n obegränsat närmar sig x, när index n växer obegränsat. Alternativt kan vi säga att om vi studerar intervall O som innehåller x i sitt inre faller alla x_n

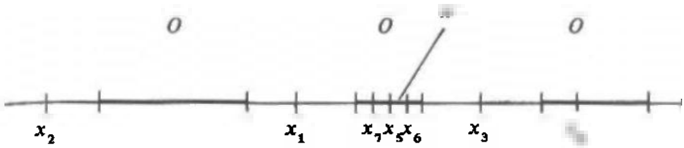


Fig. 4

utom ändligt många i O. Givetvis gäller samma sak om O är vilken mängd som helst som innehåller ett intervall kring x. Med ett suggestivt språkbruk kan vi säga att dessa mängder O fungerar som ett filter, som filtrerar ut alla talföljder utom dem för vilka x_n närmar sig x obegränsat. Varje annan följd kommer nämligen att ha oändligt många punkter utanför ett lämpligt valt O och blir då bortfiltrerat av denna mängd.

Att överföra idén om konvergens till planet är lätt: vi byter bara ut intervall mot cirklar — eller kvadrater, det går lika bra — och på liknande sätt gör vi i 3 dimensioner. Men vi vill också tala om konvergens i sådant sammanhang som

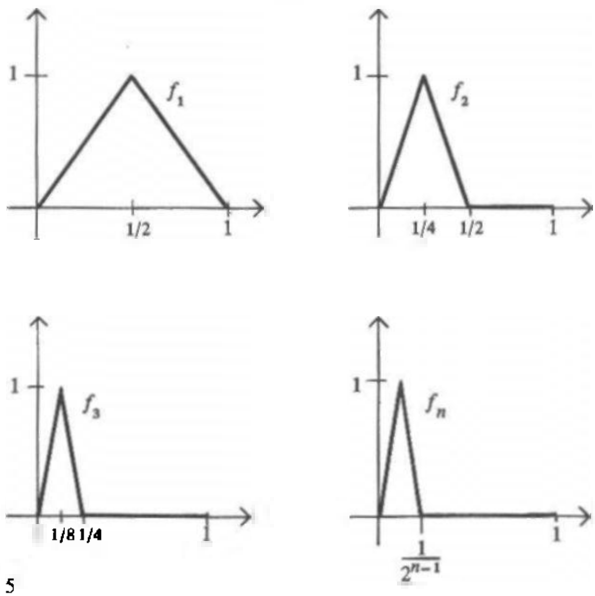


Fig. 5

då vi studerar följder av funktioner. Tag t.ex. de funktioner $f_n(x)$, $n=1, 2, \ldots$, på $(0, 1)$ som är antydda i fig. 5.

Funktionerna liknar mer och mer funktionen $f_0(x)=0$, men å andra sidan antar varje funktion $f_n(x)$ värdet 1. Om vi med utsagan f_n konvergerar mot f_0 menar att maximum av $f_n(x)$ blir litet så konvergerar f_n inte mot f_0, men om vi menar att ytan mellan f_n och x-axeln blir liten så har vi konvergens. Båda begreppen är fullt rimliga och användes i matematiken.

Den allmänna idén om konvergens kan nu preciseras med hjälp av ett system S av delmängder O av en allmän mängd M. S skall ha ungefär de egenskaper vi tidigare införde. De mängder som innehåller en punkt x i M är omgivningar till x och tjänar som filter vid definition av konvergens mot x. x_n konvergerar alltså i "topologin" bestämd av S mot x om utanför varje omgivning av x bara finns ändligt många x_n. I våra exempel är mängderna av funktioner $f(t)$ så att i första fallet $\text{Max}\,|f(t)|<c$ för något c, och i andra fallet $\int_0^1 |f(t)|dt<c$, omgivningarna till "punkten" f_0 representerande funktionen identiskt noll. Vi får alltså inte konvergens i första fallet eftersom för t.ex. $c=1/2$ alla f_n faller utanför omgivningen. I den andra topologin däremot har vi konvergens. Detta är alltså inte ett absolut begrepp utan beror på systemet S. Härmed har man lagt grunden för studiet av funktioner på M på samma sätt som vi är vana att studera funktioner som är definierade på t.ex. intervall. Observera att i våra exempel ett uttryck som $\int_0^1 f(t)^2 dt$ är en funktion på M och således f är variabeln.

Slutligen kan man också införa struktur på en mängd inte genom ett system delmängder utan genom kompositionsregler. På mängden av heltal finns t.ex. de två olika sätten att kombinera heltal till nya heltal + och $\cdot$. Dessa operationer är förbundna med varandra genom regler sådana som $a\cdot(b+c)=$ $=a\cdot b+a\cdot c$. För många objekt som har intresse i matematiken finns kompositionsregler som i större eller mindre utsträckning liknar dem som heltalen eller de rationella talen har. Vi skall här bara nämna några exempel. Om E är en mängd och $f(x)$ avbildar E in i sig själv finns på mängden M av dessa avbildningar en kompositionsregel $*$ som helt enkelt består av att vi utför avbildningarna efter varandra (fig. 6).

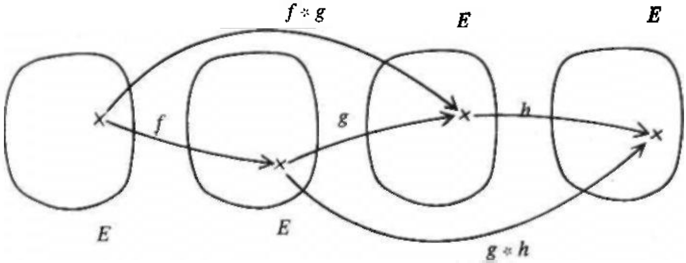


Fig. 6

Tydligen gäller för $(f * g) * h = f * (g * h)$ precis som för $+$ och $\cdot$, men det gäller inte säkert $f * g = g * f$. Om vi antar att avbildningarna är $1-1$ kan man tydligen lösa ekvationer $x * f = g$ precis som man kan för $+$ för heltalen och $\cdot$ för rationella tal (men där inte för heltal). Vad vi då fått kallas en grupp och detta begrepp kan analyseras enbart utgående från de regler som operationen $*$ lyder och alldeles oberoende av de eventuella tolkningar som man kan ge symbolerna f, g osv.

Ytterligare ett exempel. Polynom $P(x) = a_0x^n + a_1x^{n-1} + \ldots + a_{n-1}x + a_n$ där koefficienterna $a_0, a_1, \ldots, a_n$ t.ex. är reella tal, bildar en mängd som i hög grad liknar heltalen. Här finns både en plus- och en gånger-operation definierade av vad vi får genom att på vanligt sätt addera koefficienterna respektive multiplicera ihop och ordna efter potenser av x. Mängden av polynom har alltså en invecklad struktur, och man kan nu utgående från enbart kompositionsreglerna göra en hel teori. Denna gäller då inte bara polynom som vi tog till mönster utan för många andra system där symbolerna har en annan tolkning.

Det matematiska område vi nu sist diskuterat är algebran. Det är således teorin för matematiska strukturer bestående av kompositionsregler. Analysen är analogt en struktur bestående av omgivningssystem. Dessa idéer är centrala i den moderna utvecklingen av den matematiska forskningen. Vad man eftersträvar är att föra samman metoder och begrepp som tidigare uppstått inom olika grenar men som är lika till sin logiska struktur. Man vill förutsätta bara så mycket som verkligen behövs för att man skall kunna bevisa det resultat man eftersträvar. Härigenom vinner man en stor allmängiltighet och också en ökad förståelse för de logiska mekanismer som fungerar i ett visst problem. Det finns naturligtvis många alternativa sätt att införa en hierarki av strukturer — från det strukturfria allmänna mängdbegreppet till det mest specialiserade, som t.ex. de rationella talen. Många definitioner och begrepp har förts fram och sedan sjunkit i glömska, och det system man nu arbetar med måste anses väl förankrat och bör komma att bli bestående.

Man skulle önska att undervisningen i gymnasiet förmedlade något av denna allmänna syn på matematiken och därmed gav en antydan om den rikedom på möjligheter som den moderna matematiken har. Den bild jag här gett har skapats under vårt århundrade och har ännu inte fått någon form som passar elementär undervisning. Man får inte därför tro att det är för svårt. Vad det gäller är att ge de ord som används intuitivt innehåll, att införa dem tidigt och att låta dem komma till användning systematiskt.

Vi skall i följande kapitel försöka göra en översikt av några av matematikens största områden och punktvis föra fram diskussionen till problem som matematiker just nu arbetar med. Detta kan ge en antydan om vad en kunskapskurs i gymnasiet som vi nämnde i förra kapitlet skulle kunna tänkas innehålla. Samtidigt skall vi anknyta till nuvarande kurser och försöka belysa deras relation till modern matematik.

Kommentarer

En utförligare och modern bok om matematikens struktur är

A. M. Gleason, *Fundamentals of Abstract Analysis*, Addison-Wesley 1966.

Beträffande kontinuumhypotesen, att $\aleph_1$ följer närmast efter $\aleph_0$, visades dess relativa motsägelsefrihet och oberoende av Gödel resp. Cohen i de citerade arbetena.

4. De hela talen

Representation och komplexa tal

Vi bekymrar oss nu inte vidare om heltalens logiska bakgrund utan skall här först ägna oss åt deras representation. Att detta inte varit något trivialt problem för mänskligheten illustreras av det romerska siffersystemet. Detta var uppbyggt på ett så opraktiskt sätt att det förhindrade all utveckling av matematiken. Vi är nu så vana vid vårt ursprungligen arabiska siffersystem, att vi aldrig ägnar en tanke åt att det var svårt att hitta på det. Vad värre är, vi har en tendens att blanda ihop siffrorna, dvs. representationen av talen, med talen själva, och härigenom åstadkommer vi med säkerhet förvirring hos våra barn när de skall lära sig räkna.

Till detta kommer en ny omständighet. I framtiden kommer datamaskiner att vara en integrerad del av vår tillvaro, och det kommer att vara betydelsefullt att kunna umgås naturligt med dem och förhindra att de blir våra herrar. Det är tänkvärt att småskolebarnens undervisning bör anpassas till situationen i samhället år 1985. Nu är datamaskinens naturliga talsystem baserat på tvåsystemet, och för att man skall kunna förstå sammanhanget krävs en mindre mekanisk kunskap om talen än den som för närvarande läres ut.

Varför använder vi 10-systemet, och varför används 2-systemet i datamaskiner? Det är ingen tvekan om att våra 10 fingrar spökar bakom 10-systemet, men 10 är också ungefär en lämplig kompromiss mellan två motstridande effekter. I 2-systemet har vi en maximalt enkel multiplikationstabell att lära oss: $0 \cdot 0 = 0$; $0 \cdot 1 = 0$, $1 \cdot 0 = 0$, $1 \cdot 1 = 1$, och få symboler att lära oss att skriva (0 och 1), men å andra sidan skulle ett telefonnummer i Stockholm bestå av 20 siffror, och försök att hitta på ett sätt att uttala det! En större bas än 10 leder till kortare siffergrupper men svårare multiplikationstabell. — En datamaskins normala minnesenhet är en anordning med två tillstånd, och det är lätt att elektroniskt ordna sådana övergångar från det ena läget till det andra som motsvarar addition och multiplikation. Att det behövs många nollor och ettor att representera även måttligt stora tal är här en obetydlig olägenhet. — Det finns stor anledning att tidigt i skolan göra barn förtrogna med 2-systemet, dels för att klargöra representationsmetoden som ju är densamma som för 10-systemet men enklare, dels därför att 2-systemet spelar stor roll både för datatekniken och för matematiken själv.

Ett intresseväckande sätt att sätta sig in i 2-systemet är följande lek. Man tänker på ett tal, t.ex. under 1 000. Hur många frågor, med svar bara ja eller nej, behövs för att identifiera talet? Man ser snart att 10 gissningar räcker och vad man gör är att successivt bestämma a_0, a_1, $\ldots$, a_9 som 0 eller 1 i framställningen $a_0 2^9 + a_1 2^8 + \ldots + a_8 \cdot 2 + a_9$ av talet. Man frågar först: är talet $> 512 = 2^9$? Om svaret blir ja, är $a_0 = 1$, annars är $a_0 = 0$. Om t.ex. $a_0 = 1$, frågar man nästa gång: är talet > 768 $(= 512 + 256)$ och får härigenom reda på a_1. Det är klart hur frågandet fortsätter. [Försök bevisa att 9 frågor inte räcker.] — Denna lek kan anses vara grunden till en ny snabbt expanderande gren av matematiken, informationsteorin, som har stor betydelse även för humanistiska ämnen, t.ex. språkforskningen. Man inför en enhet för information, kallad bit (*bi*nary digi*t* = = siffra i 2-systemet) som har den intuitiva tolkningen av en upplysning om vilkendera av två, från början lika troliga händelser som inträffat. Om en händelse inträffar som från början hade sannolikhet p, har vi fått informationsmängden $-{}^2\log p$. Observera alltså att valet mellan två lika sannolika händelser ger informationsmängden 1, ty $-{}^2\log 1/2 = 1$. Man kan nu t.ex. göra beräkningar över hur stor informationsmängd varje bokstav i skriven svenska innehåller. Om bokstäverna vore oberoende av varandra och lika vanliga skulle informationsmängden vara ${}^2\log 29 = 4{,}8$ (ung.), eftersom det finns 29 bokstäver. Emellertid är det inte så: a är absolut vanligare än b, s följs inte gärna av d osv. Man finner att informationsinnehållet bara är ca 1,5 bit per bokstav. Detta tal beror rätt mycket på vem som talar eller skriver. Informationsinnehållet i Thomas Manns prosa överstiger säkert 2 bits per bokstav, medan informationsinnehållet i ett valtal (som bekant) brukar ligga närmare noll. Man kan själv empiriskt bestämma talet för en viss text genom att pröva hur stor andel av bokstäverna man kan låta ta bort innan det inte längre går att rekonstruera innehållet i texten. Att språket har denna struktur att ge samma information flera gånger är av stor betydelse, eftersom det ger garantier mot missförstånd. Det medför också att det finns en metod att översätta (koda) det skrivna eller talade språket så att antalet bokstäver minskar till ungefär bråkdelen $1{,}5:4{,}8<1/3$ av det ursprungliga antalet. Detta kommer att få en stor ekonomisk betydelse för telegraf- och telefontrafik genom att linjerna efter kodning kan utnyttjas ca 3 gånger effektivare.

Ett analogt problem gäller att bestämma antalet binära operationer som behövs för att lösa ett visst matematiskt problem. Tag t.ex. ekvationen $x^3+ax+1=0$, där a är ett tal mellan 0 och 1 angivet med 20 binära siffror. Hur många additioner, det vill säga hur lång tid, behövs i en metod att bestämma den reella roten x med 15 binära siffror? Dessa problem har stor betydelse för ett ekonomiskt utnyttjande av datamaskiner och är till stor del olösta.

Vi skall här inledningsvis också behandla de komplexa talen. De har givetvis inget att göra med vårt ämne, de hela talen, men vi behöver dem i den fortsatta framställningen.

De komplexa talen är kanske av större betydelse för matematikerna än de reella. De är omgivna av ett visst skimmer av mystik, som så småningom försvinner helt enkelt genom att man blir van vid dem och inte genom att man i någon djupare mening förstår vad som pågår.

Logiskt innebär införandet av komplexa tal inget problem. Vi vill att det skall finnas ett "tal" x så att $x^2+1=0$. När det nu inte finns något, låtsas vi att det finns och kallar det i eller $\sqrt{-1}$. Vi sysslar nu med tal som kan betecknas $a+ib$ där a och b är reella. Observera att + och ib inte är definierade. Om vi räknar på formellt utan att bekymra oss om vad det som vi skriver betyder, finner vi t.ex.

$$(a+ib)(c+id)=ac+i\cdot i\cdot bd+ibc+iad=$$
$$=(ac-bd)+i(bc+ad).$$

Division blir lite besvärligare:

$$\frac{a+ib}{c+id}=\frac{(a+ib)(c-id)}{(c+id)(c-id)}=\frac{ac+bd}{c^2+d^2}+i\cdot\frac{bc-ad}{c^2+d^2},$$

så att alltså till exempel

$$\frac{1}{1+i}=\frac{1}{2}-\frac{1}{2}i\,.$$

Vi gör nu resonemanget baklänges, som vi gjort flera gånger tidigare. Vi *definierar* ett komplext tal som ett par (a, b) av reella tal. Det är ju inget mystiskt med det! Vi definierar + + genom

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

och · · genom samma formel som förut

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

På samma sätt gör vi med division. Systemet av talpar $(a, 0)$ bär sig åt precis som reella tal och kan identifieras med dessa. Man skriver sedan bara ett + och · tecken, och vi har konstruerat de komplexa talen. De är alltså helt enkelt par av reella tal.

Men varför är de så viktiga? Tydligen kan vi nu lösa ekvationen $x^2+1=0$, ty $(a, b)^2+ +(1, 0)=(0, 0)$ har lösningen $(a, b)=$ $=(0, 1)$. Det märkliga är nu att på köpet följer att varje ekvation $P(x)=x^n+a_1x^{n-1}+\ldots+a_{n-1}x+a_n=0$ har en lösning, t.o.m. om vi låter a_j vara komplexa tal. Vi kan också dela upp polynomet i faktorer, $P(x)=(x-\alpha_1)(x-\alpha_2)\cdot\ldots\cdot(x-\alpha_n)$, där α_i är komplexa tal (t.ex. $x^2+1=(x+i)(x-i)$). En annan viktig följd är konstruktionen av exponentialfunktionen e^z för komplexa z. Vi definierar helt enkelt e^z som $e^a(\cos b+i\sin b)$ om $z=a+ib$. Ur additionssatserna för $\cos x$ och $\sin x$ följer då lätt att $e^{z_1+z_2}=e^{z_1}\cdot e^{z_2}$. Varför inte i stället definiera e^z som $e^a(\cos 2b+$ $+i\sin 2b)$ till exempel? Samma räkneregel för $e^{z_1+z_2}$ skulle gälla då också och vi skulle få rätt resultat för reella z $(b=0)$.

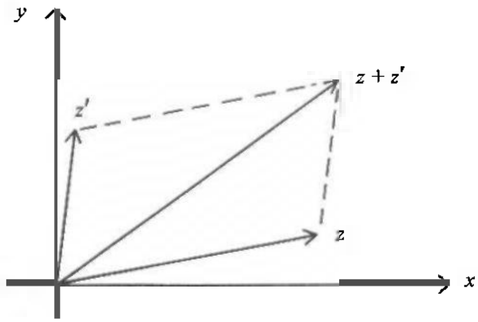


Fig. 7

Orsaken till att den tidigare definitionen är den naturliga är att även deriveringsregeln

[illegible]

fortfar att gälla för komplexa z. [Kontrollera detta!]

Man kan tydligen geometriskt åskådliggöra ett komplext tal $z = x + iy$ som en punkt i ett vinkelrätt koordinatsystem eller som en pil från 0 till punkten (x, y). $z + z'$ illustreras i fig. 7 av diagonalen i den fullbordade parallellogrammen. Punkten e^{ib} ligger tydligen på cirkeln med radien 1 kring origo, och när b beskriver $(0,2\pi)$ går punkten e^{ib} ett varv runt cirkeln (fig. 8). Man får konstiga formler som t.ex. $e^{i\pi} = -1$, som dock bara betyder att $\cos \pi = -1$ och $\sin \pi = 0$. Vi observerar här också att

$$\frac{1}{2\pi} \int_0^{2\pi} e^{inx} \cdot e^{-imx} dx = \begin{cases} 1 \text{ om } n = m \\ 0 \text{ om } n \neq m \end{cases}$$

för hela tal n och m.

Vi har alltså infört en addition av talpar. I det ligger ingenting speciellt. Det går lika bra att på samma sätt addera tre tal samtidigt och studera taltripler (a, b, c). Det märkliga för de komplexa talen ligger i definitionen av multiplikation som alltså kunde införas för talpar så att alla vanliga räknelagar bibehålls. Detta går inte för tre eller flera tal. Däremot går det på ett bestämt sätt att införa multiplikation för uppsättningar x av

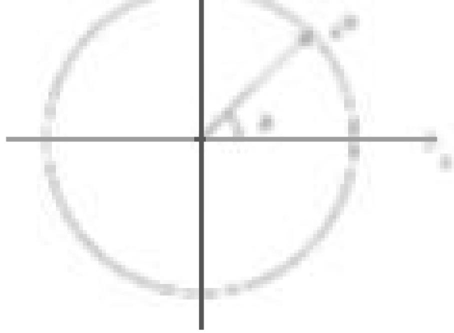


Fig. 8

fyra tal, $x=(a, b, c, d)$, om man avstår från regeln $x \cdot y = y \cdot x$. Man får då de s.k. kvarternionerna. Fyra är här det enda antalet — för tre tal är även detta omöjligt. Det är en intressant men rätt svår uppgift att försöka återupptäcka detta.

Vad vi här sagt ger ingen förklaring till den stora betydelse de komplexa talen har. Denna sammanhänger med teorin för analytiska funktioner, som vi senare skall beröra. Emellertid kommer frågan tillbaka ändå: varför förekommer dessa funktioner så ofta i matematiken? Till sist kan nog ingen lämna en tillfredsställande förklaring. Man brukar hänvisa till det förhållandet att i många ekvationer som beskriver hur naturfenomen beror på tiden förekommer tidsvariabeln t på ett sådant sätt att om man formellt byter t mot en ny variabel $i\tau$ kommer den nya ekvationen att innehålla τ på samma sätt som rumskoordinaterna. Tiden skulle alltså vara en rumsbestämning som sträcker sig in i det komplexa, men vad detta till slut betyder överlåter jag åt spekulativt lagda läsare att fundera över.

Primtal

Det låter kanske paradoxalt, men vi vet faktiskt mindre om de hela talens egenskaper än om andra matematiska begrepps, som ytligt sett ser mycket svårare ut. Som vi skall se kan en lekman här ställa problem som ingen kan besvara. Vi skall börja med frågor rörande primtal.

Primtal är som bekant medlemmarna i följden 2, 3, 5, 7, 11, ... bestående av tal som inte ytterligare kan faktoriseras. De har studerats sedan flera tusen år och är intressanta därför att de är byggstenarna till de hela talen när man studerar multiplikation. Särskilt berömt är Euklides bevis att det finns oändligt många primtal. Antag att $p_1=2$, $p_2=3$, ..., p_N vore samtliga primtal uppräknade i storleksordning. Bilda talet $p_1 \cdot p_2 \cdot \cdot \ldots \cdot p_N+1$. Är detta ett primtal? Nej, ty det är större än det största primtalet p_N. Alltså måste det vara delbart med något primtal. Men vilket tal p_i vi än delar med, får vi resten 1. Alltså har vi fått en motsägelse. Här har vi ett exempel på indirekt bevis. [Försök att hitta på ett direkt bevis!]

Nästa fråga blir: kan man ange alla primtalen? I princip kan man göra det successivt genom att helt enkelt pröva alla heltal 2, 3, 4, 5, 6, ... i ordning och se vilka som kan delas med mindre tal. Det går att hitta på förenklande regler och låta en datamaskin göra arbetet. Härigenom har man skaffat sig mycket omfattande tabeller. Men en fråga sådan som: vilket primtal har ordningsnumret $10^{1\,000}$ kan man inte besvara och det är rätt säkert att man aldrig kan besvara den. Det finns nämligen ingen "formel" som anger p_n som funktion av n. Alternativt skulle man vilja ha en formel som anger antalet $\pi(x)$ av primtal $<x$.

I detta läge har man övergått att studera $\pi(x)$ mera ungefärligt. Vilken storleksordning har t.ex. $\pi(x)$? Gauss gjorde redan som gymnasist iakttagelsen ur stora tabeller att om man tar ett stort tal x på måfå så är chansen att detta skall vara ett primtal $1/\log x$, där log här och i fortsättningen betecknar den naturliga logaritmen, dvs. med basen $e=2{,}7\ldots$ Detta innebär således att bland de 999 talen $10^{100}+1$, $10^{100}+2$, ..., $10^{100}+999$ finns troligen ett primtal. På basis av denna iakttagelse är det inte så svårt att inse att $\pi(x)$ bör växa som $x/\log x$ eller noggrannare

$$\pi(x)=\int_2^x \frac{dt}{\log t}+R(x)=I(x)+R(x),$$

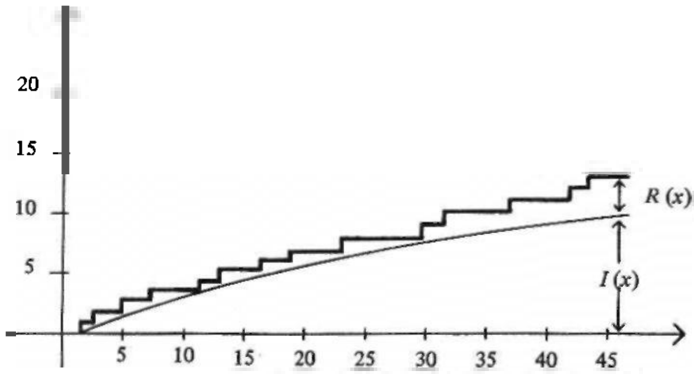


Fig. 9

där $R(x)$ är en rest som är liten i jämförelse med integralen då x växer (fig. 9). Resultatet visades först mot slutet av 1800-talet och numera finns många bevis. Problemet är ju i princip elementärt men alla kända sätt att lösa det var fram till för knappt 20 år sedan mycket avancerade med metoder från analysen. Då fann A. Selberg ett elementärt bevis. Detta kan förstås utan omfattande förkunskaper men är därför inte lätt.

Hur Gauss kom fram till sin hypotes vet man inte. Det är ju knappast möjligt att ur sådana tabeller som han disponerade se att $1/\log x$ är en bättre gissning än $1/(\log x + 0{,}01)$ t.ex., men den *är* bättre. Förmodligen vägleddes han av principen att en enkel formel som stämmer ungefär, är troligen sann.

⟨ Man kan göra storleksordningen $1/\log x$ trolig genom följande suggestiva resonemang. Kalla "sannolikheten" att talet n är ett primtal s_n och sätt t.ex. $s_2 = 1$. Tag nu ett stort heltal N. Sannolikheten att N är ett primtal kan vi få på följande sätt. Det kan i så fall inte vara delbart med något primtal mindre än N. Nu är sannolikheten att N är delbart med "primtalet 2" = (sannolikheten att 2 *är* ett primtal) × (sannolikheten att talet är delbart med 2) = (ung.) $s_2 \cdot 1/2$. För "primtalet 3" blir motsvarande resultat $s_3 \cdot 1/3$ osv. Nu skall inget av detta inträffa. Motsatserna har sannolikheten

$$\left(1 - \frac{s_2}{2}\right), \left(1 - \frac{s_3}{3}\right), \ldots, \left(1 - \frac{s_{N-1}}{N-1}\right)$$

och att dessa alla inträffar har sannolikhet

$$\left(1-\frac{s_2}{2}\right)\left(1-\frac{s_3}{3}\right)\cdot\ \ldots\ \cdot\left(1-\frac{s_{N-1}}{N-1}\right),$$

som alltså är $= s_N$. Vi ser att s_{N+1} är en likadan produkt med en faktor till, $(1-s_N/N)$, varför

$$s_{N+1} = s_N - \frac{s_N^2}{N}$$

Vad vi fått fram är en differensekvation ur vilken vi kan beräkna s_N successivt ($s_2 = 1$, $s_3 = 1/2$, $s_4 = 5/12$, osv.). Vi kan få en uppfattning om s_N för stora N genom att skriva formeln

$$\frac{1}{s_{N+1}} - \frac{1}{s_N} = \frac{1}{N}\frac{s_N}{s_{N+1}} = \frac{1}{N} \text{ (ungefär)}$$

eftersom N och $N+1$ bör ha ungefär lika chans att vara primtal. Lägger vi nu ihop dessa formler får vi

$$\frac{1}{s_{N+1}} - \frac{1}{s_2} = \frac{1}{2} + \frac{1}{3} + \ \ldots \ + \frac{1}{N} \text{ (ungefar)}$$

$$= \log(N+1) \text{ (ungefär)}$$

(eftersom $1/2 + 1/3 + \ldots + 1/N = \int_2^N \frac{dx}{x}$ (ungefär)). Vi får

$$s_N = \frac{1}{\log N} \text{ (ungefär)}$$

dvs. Gauss gissning. 〉

Denna sorts resonemang är naturligtvis inget bevis. Vi har ju laborerat med sannolikheter på ett helt otillåtet sätt. Inte desto mindre används ungefärliga och suggestiva resonemang mycket ofta för att undersöka vilka resultat man kan vänta sig och för att kontrollera om en hypotes har rimliga konsekvenser. Det kräver naturligtvis stor erfarenhet och kunskap att bedöma vilka typer av resonemang som är tillförlitliga och vilka approximationer som får göras.

Om vi fullföljer tankegången om att x är ett primtal med sannolikhet $1/\log x$ så följer ur den s.k. centrala gränsvärdessatsen (se s. 119) att $R(x)$ bör ha storleksordningen ungefär $\sqrt{x}$. Att finna ett bevis att detta verkligen är sant är ett av matematikens mest berömda öppna problem. Det kallas Riemanns hypotes. Det är också ett ovanligt problem därigenom att det saknas säkert stöd för någon uppfattning om vad man skall tro, och det finns ingen bestämd matematikeropinion. Att Riemann uttalade en så bestämd uppfattning beror med största sannolikhet på, att han gjort ett förbiseende i sin undersökning.

Det finns ett stort antal olösta problem beträffande primtal. Jag vill här bara nämna ytterligare ett: Goldbachs problem. Det är flera hundra år gammalt och gäller frågan om varje jämnt tal är en summa av två primtal: $6=3+3$, $8=3+5$, $10=5+5$, $12=5+7$ etc. Man har provat detta långt upp i talserien och varje matematiker torde vara övertygad om att hypotesen inte är falsk (däremot *kan* den vara omöjlig att bevisa (se s. 15)). I vilket fall som helst förefaller ett bevis mycket avlägset. Att man kan vara övertygad beror på följande tankegång. När vi bildar $p+p'$ för alla par (p, p') av primtal $\leq x$ får vi ungefär $x^2/\log^2 x$ tal, alla $<2x$. Varje tal förekommer ungefär $x/2 \log x$ gånger och chansen att ett visst tal inte skulle förekomma är mycket liten. Chansen är störst för små jämna tal, men där stämmer det ju, som prövningar visar. Man drar då slutsatsen att hypotesen bör vara sann. Bakom den ev. giltigheten av detta resonemang ligger tanken att primtalen är utspridda utan något system men just därför att det inte finns någon användbar regelbundenhet är det svårt att bevisa gissningen. Primtalen är ju "naturliga" vid multiplikation men har ingen användbar struktur då man adderar dem. Man har trots allt lyckats visa det närliggande resultatet att varje tillräckligt stort udda tal är en summa av tre primtal. Detta resultat är väsentligt lättare genom att det förväntade antalet representationer av ett tal x är ännu mer överväldigande stort och alltså då rimligen åtminstone positivt. Det är intressant att man inte vet om *varje* tal har denna sista egenskap. Den gräns ovanför vilken beviset gäller är nämligen så fantastiskt stor att t.o.m. datamaskiner står maktlösa.

⟨Hur bevisar man nu resultat som detta sista? Man bildar då en funktion

$$f(z)=z^2+z^3+z^5+\ldots+z^{p_N}=z^{p_1}+z^{p_2}+\ldots+z^{p_N},$$

där z varierar som ett komplext tal. Om vi nu upphöjer $f(z)$ till tredje potensen,

$$f(z)^3=z^6+3z^7+3z^8+4z^9+\ldots+a_n z^n+\ldots+z^{3p_N},$$

så anger a_n, $n<p_N$, på hur många sätt man kan skriva helheten n som en summa av 3 primtal; $a_7=3$ ty $7=2+2+3=2+3+2=3+2+2$ t.ex. Vad man vill visa är att heltalet $a_n\neq 0$. Nu finns det en enkel formel som uttrycker a_n med hjälp av $f(z)$ ty om vi använder formeln på sidan 39 får vi

$$\frac{1}{2\pi}\int_0^{2\pi} f(e^{it})^3 e^{-int}dt=\frac{1}{2\pi}\int_0^{2\pi} e^{6it}\cdot e^{-int}dt+3\frac{1}{2\pi}\int_0^{2\pi} e^{7it}\cdot e^{-int}dt+\ \ldots$$

$$\ldots+a_n\frac{1}{2\pi}\int_0^{2\pi} e^{int}\cdot e^{-int}dt+\ldots=a_n.$$

Huvuddelen av bidraget till integralen till vänster i formeln kommer från närheten av rationella tal $t=r/s$. För rationella tal kan man göra explicita beräkningar genom att man vet hur resterna fördelar sig då man delar primtal p_n med heltal s.⟩

Att genomföra ovanstående är mycket komplicerat (ett fullständigt bevis är 100 sidor långt) och här finns fortfarande betydelsefulla insatser att göra. Poängen är att man studerar talteori med hjälp av funktioner och man talar därför om analytisk talteori. På analogt sätt sammanhänger Riemanns hypotes med hur goda uppskattningar man kan få av summor

$$f(t)=\cos(t\log 2)+\cos(t\log 3)+\ldots+\cos(t\log N)$$

för stora tal N och t. Även här finns många olösta problem.

Diofantiska ekvationer

Det andra stora klassiska området inom talteorin gäller existens av heltalslösningar till ekvationer. Det enklaste fallet är ekvationen $ax+by=c$, där a, b, c är givna heltal och man söker lösningar x och y som också är heltal. Tydligen måste varje tal som går jämnt upp i a och b också gå jämnt upp i c för att det skall kunna existera lösningar ($4x+12y=4(x+3y)=5$ kan inte ha lösningar). Det visades av Diofantos för ca 1 500 år sedan att i så fall finns det också lösningar och det kan vara ett intressant prov på skarpsinne att försöka hitta på ett bevis. Om vi går till ekvationer av andra graden (t.ex. $x^2-5y^2=3$) vet man väl hur det ligger till och här finns en mycket gammal, intressant och inte så svår teori. Går man sedan till ekvationer av högre grad finns endast spridda kunskaper och ingen som helst allmän metodik. Vi skall här nämna endast ett nytt resultat och ett sammanhängande problem.

Vi studerar ett polynom $P(t)=a_0t^n+a_1t^{n-1}+\ldots+a_n$ av grad $n\geq 3$ med koefficienter a_i som är heltal. Vi antar att $P(t)$ inte kan delas upp i faktorer av samma typ som P (som ett exempel kan vi välja $P(t)=t^4-2$). Vi byter nu t mot x/y och multiplicerar med y^n:

$$y^nP\left(\frac{x}{y}\right)=f(x,y)=a_0x^n+a_1x^{n-1}y+\ldots+a_ny^n.$$

Låt $g(x,y)$ vara bildat på analogt sätt ur ett heltalspolynom $Q(t)$, vilket som helst av gradtal $m<n-3$ (i vårt exempel kan vi ta $Q(t)=t+1$). Då gäller att ekvationen $f(x,y)=g(x,y)$ bara har ett ändligt antal heltalslösningar. I exemplet: det finns bara ändligt många hela tal x och y så att

$$x^4-2y^4=x+y.$$

Vad är bakgrunden till detta resultat? Vi delar ekvationen $f(x,y)=g(x,y)$ med y^n och får

$$P\left(\frac{x}{y}\right)=\frac{1}{y^{n-m}}Q\left(\frac{x}{y}\right),$$

där alltså exponenten i nämnaren $n-m$ är >3. Det är lätt att förstå, att x/y för heltalsvärden på x och y inte kan bli mycket stort i en sådan ekvation, för då blir vänsterledet mycket större än högerledet, eftersom P har högre grad än Q. Vi får alltså

$$\left|P\left(\frac{x}{y}\right)\right|\leq \text{Konstant}\cdot y^{-3},$$

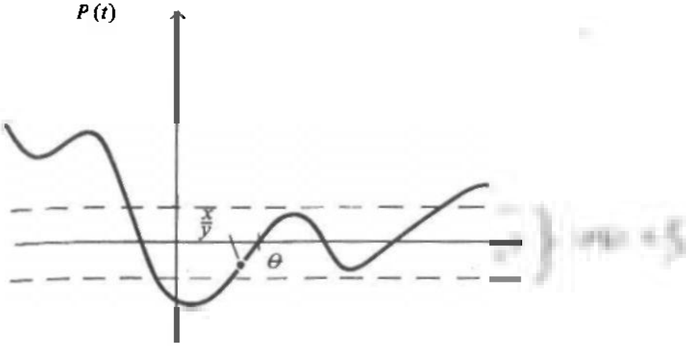


Fig. 10

Om vi nu hade oändligt många lösningar till den diofantiska ekvationen skulle y i regel vara stort, dvs. $P(x/y)$ mycket litet. Det är då lätt att förstå att x/y måste ligga så nära en rot θ till ekvationen $P(t)=0$ att $|\theta - x/y| < \text{Konstant}/y^3$. $P(t)$ är ju nämligen $= a_0(t-\theta)P_1(t)$, där $P_1(t)$ är ett polynom av grad $n-1$ sådant att $P_1(\theta)$ inte är noll (se fig. 10).

Vad man nu kan visa är att för rötter θ till ekvationer med heltalskoefficienter, kan en sådan olikhet bara vara uppfylld för ändligt många par (x, y), t.o.m. om 3 byts mot vilket som helst tal >2. Det är intressant att observera att talet 2 i villkoret >2 är det riktiga. Varje reellt tal θ kan nämligen alltid approximeras så att

$$\left|\theta - \frac{x}{y}\right| < \frac{\text{Konstant}}{y^2}$$

med y hur stort som helst. Ett bevis går till så här. För varje y, $1 < y < N$, bestämmer vi x så att $0 < y\theta - x < 1$. Vi får då N stycken tal $y\theta - x$. Alla ligger inom intervallet $(0, 1)$. Två av dessa måste då ha ett avstånd $< 1/N$. Om motsvarande (x, y) kallas (x_1, y_1) och (x_2, y_2) får vi

$$0 < y_2\theta - x_2 - (y_1\theta - x_1) < 1/N,$$

varav efter division med $y_2 - y_1$ som ju är $< N$

$$0 < \left|\theta - \frac{x_2 - x_1}{y_2 - y_1}\right| < \frac{1}{(y_2 - y_1)N} < \frac{1}{(y_2 - y_1)^2}$$

Talen $x=x_2-x_1$ och $y=y_2-y_1$ ger alltså en lösning. Vi får i detta bevis konstanten i olikheten $=1$. Den bästa konstanten kan visas vara $1/\sqrt{5}$. Det finns bara ett tal θ som inte medger bättre konstant, nämligen $(\sqrt{5}-1)/2$.)

Man vill naturligtvis ha reda på precis hur många lösningar en ekvation som $x^4-2y^4=x+y$ har. Här finns ju $x=1$, $y=0$ men finns det fler? Eftersom vi vet att antalet är ändligt, skulle man kunna tänka sig att helt enkelt pröva igenom, eventuellt med hjälp av datamaskin, de möjliga paren. Men detta förutsätter att man kan ange en gräns, uttryckt i koefficienterna i f och g, ovanför vilken säkert inga lösningar finns. Och detta är just vad man inte kan. Alla bevismetoder, som gett resultat av den typ det här är fråga om, går ut på att antagandet om oändligt många lösningar innebär en motsägelse. Bevisen är således icke-konstruktiva och det kan inte uteslutas att det är logiskt omöjligt att konstruera den gräns man söker som funktion av koefficienterna i P och Q. Man skulle i så fall i detta problem ha det första naturliga exemplet, dvs. som inte konstruerats speciellt för ändamålet, som ej är konstruktivt lösbart.

Sammanfattningsvis är vår kunskap om diofantiska ekvationer i två obekanta synnerligen bristfällig, trots ett intensivt arbete speciellt under tidigare skeden av matematikens utveckling. Orsaken är att man saknar metoder och allmänna teorier. Detta är i än högre grad fallet då antalet obekanta är ≥ 3. Det mest berömda exemplet är Fermats ekvation $x^n+y^n=z^n$ som antages sakna icke-triviala heltalslösningar x, y, z för alla $n>3$. Detta är bevisat med olika metoder för så många speciella n (t.ex. alla $n<2000$), att man kan känna sig helt säker på att hypotesen är sann, men det finns ingen antydan om i vilken riktning ett allmänt bevis vore att söka.

Matematikernas misslyckande med att lösa Diofantos problem innebär inte alls att mödorna varit bortkastade. Ur ansträngningarna har vuxit fram allmänna teorier som belyser de problem det gäller och som fått stor betydelse i andra sammanhang.

Man kallar ett sådant tal θ som $\sqrt{2}$ eller $\sqrt[4]{-5}$ algebraiskt, därför att det är lösning till en vanlig polynomekvation med heltalskoefficienter, $x^2=2$, resp. $x^4+5=0$. Man ser att om man i första fallet bildar alla tal av formen $a+b\sqrt{2}$ och i andra $a+b\sqrt{(-5)}+c\sqrt{25}+d\sqrt{(-125)}$, a, b, c, d heltal, så kan man addera och multiplicera ihop sådana tal och få samma slags tal tillbaka. På detta sätt fungerar alltså dessa tal på samma sätt som heltal. Man kan sedan bilda kvoter och får då något som liknar rationella tal. Sådana talmängder, talkroppar, har nu studerats mycket ingående. Det är lätt att förstå att detta har med diofantiska ekvationer att göra. Tag t.ex. ekvationen $x^2-2y^2=A$. Denna kan skrivas $(x+\sqrt{2}y)(x-\sqrt{2}y)=A$ och om nu tal av formen $x+\sqrt{2}y$ bär sig åt som heltal, måste både $x+\sqrt{2}y$ och $x-\sqrt{2}y$ vara faktorer i A. Man inför motsvarigheter till primtal och kan visa att "tal" entydigt kan uppdelas i "primfaktorer". Svårigheten ligger i att ställa dessa "primtal" i relation till vanliga heltal. — Under senare tid har man gjort rent abstrakta konstruktioner som har egenskaperna hos dessa algebraiska tal och av allt att döma finns där stora arbetsuppgifter.

Det kan här vara av intresse att påpeka att tre klassiska problem, som fortfarande har stor lockelse för amatörer, sedan länge lösts med besläktade metoder. Det gäller lösningen av den allmänna femtegrads- (eller högre) ekvationen genom rotutdragningar, tredelning av en godtycklig vinkel genom ett ändligt antal konstruktioner med passare och linjal samt konstruktion på samma sätt av en cirkel vars yta är lika stor som en given kvadrats.

I skolan lär vi oss formeln för rötterna till ekvationen $x^2+ax+1=0$, $x=-a/2\pm\sqrt{(a^2/4-1)}$. Alldeles analoga formler, fastän mycket mera komplicerade och innehållande tredje- och fjärderötter, finns för ekvationerna $x^3+ax+1=0$ och x^4+ax+ $+1=0$. För 5:tegradsekvationen finns ingen som helst sådan allmän formel innehållande enbart rotoperationer, och beviset av detta beror på att det talsystem som man får då man som ovan lägger till exakta rotuttryck till de rationella talen har en egenskap som talsystemet $a+b\theta+c\theta^2+d\theta^3+e\theta^4$ inte har för lösningen θ till $\theta^5+a\theta+1=0$. Mer om detta på s. 57.

Att tredela en vinkel motsvarar att konstruera en rot θ till en ekvation $4x^3-3x=c$. Man vill nämligen bestämma $\cos v$ ur ekvationen $\cos 3v = c$ och det gäller ju för alla vinklar v att $\cos 3v = 4\cos^3 v - 3\cos v$. Att göra konstruktionen med passare och linjal innebär att vi i varje konstruktionssteg löser en andragradsekvation. Passaren ger cirklar $x^2 + y^2 + Ax + By + C = 0$ (C) och linjalen linjer $Dx + Ey + F = 0$ (L) och sammanställer vi två ekvationer av typ (C), eller en (C) och en (L) får vi, när vi eliminerar y, en andragradsekvation i x. Algebraiskt innebär detta att man successivt lägger kvadratrötter till sitt system. Vad man visar är att roten θ inte kan ingå i ett sådant system annat än för speciella c. T.ex. för $c = 1/2$, $3v = 60°$, går konstruktionen inte.

Frågan om cirkelns kvadratur är det principiellt enklaste problemet. Av de skäl vi förklarat skulle speciellt följa att cirkelns omkrets 2π vore rot till en ekvation med heltalskoefficienter om cirkelns radie kunde konstrueras ur kvadratens sida. Men man kan bevisa (på ett par sidor) att detta inte är fallet.

Det förekommer fortfarande mycket ofta att icke fackmatematiker tror sig ha funnit en konstruktionsmetod för något av dessa problem, trots att man alltså sedan snart 100 år vet att detta är omöjligt. Felet består nästan alltid i att konstruktionen är en så god approximation att man tar för givet att två punkter sammanfaller, den konstruerade och lösningen, därför att man inte kan se skillnaden. Till exempel är ju x och $\sin x$ mycket lika för små värden på x. Psykologiskt är detta intressant som exempel på att man inte accepterar expertutsagor i skenbart enkla sammanhang och också på bristen på kontakt med matematisk forskning.

Även ur andra synpunkter har den allmänna teorin varit av stort intresse. Man har för vissa system kunnat lösa problemet om primtalsfördelningen och bevisa Riemanns hypotes och på detta sätt har teorin gett ökad insyn i de gamla olösta problemställningarna.

Har nu talteorin någon betydelse utanför matematiken? På det hela taget har den inte haft det, utan haft sin lockelse som en utmaning till den mänskliga nyfikenheten. Aktiviteten inom det klassiska området har minskat helt enkelt därför att problemen varit för svåra och man saknar idéer och metoder. Det är emellertid inte uteslutet, att vissa fysikaliska fenomen visar sig följa något mer invecklade heltalslagar än dem man hittills stött på, t.ex. i periodiska systemet och att talteoretikernas mödor kan komma till användning.

Kommentarer

Informationsteorin behandlas ur allmänna aspekter i L. Brillouin, *Science and Information Theory*. New York, Academic Press, 1956.

Beträffande den matematiska teorin, och speciellt kodningsproblemet, dvs. hur mycket signalerna kan kortas ned, se

A. J. Khinchin, *Über grundlegende Sätze der Informationstheorie*. Arbeiten zur Informationstheorie. Berlin 1957.

Det elementära beviset för primtalssatsen upptäcktes av Atle Selberg 1948. För detta erhöll han ett Fieldpris vid kongressen i Cambridge, Mass., 1950. Beviset finns medtaget i

T. Nagell, *Introduction to Number Theory*. Uppsala 1951.

Differensekvationen $s_{N+1}-s_N = -(1/N)s_N^2$ studeras enklast genom att man jämför s_N med den ekvation man får genom att byta differensen $s_{N+1}-s_N$ mot en derivata

$$y' = -\frac{1}{x}y^2.$$

Härav följer

$$\frac{d}{dx}\left(\frac{1}{y}\right) = -\frac{y'}{y^2} = \frac{1}{x} = \frac{d}{dx}(\log x),$$

dvs.

$$\frac{1}{y(x)} - \frac{1}{y(1)} = \log x$$

och om $y(1)=0$

$$s_N \sim y(N) = \frac{1}{\log N}.$$

A. J. Khinchin, *Three Pearls of Number Theory,* Rochester, N.Y., 1952, behandlar tre fascinerande talteoretiska problem på ett elementärt sätt, som ger en god idé om talteorins metoder och svårigheter.

Satsen om representation av tal som summa av tre primtal visades 1937 av Vinogradov. En översikt av dessa problem finns i

L.-K. Hua, *Abschätzungen von Exponentialsummen.* Teubner Verlag 1959. (Enzyklopädie der mathematischen Wissenschaften.)

Gränsen i satsen är oerhört stor, något sådant som $10^{100\,000}$.

Satsen på s. 47 om approximation av algebraiska tal visades 1955 av K. F. Roth [Rational Approximation to Algebraic Numbers. Mathematika 1955]. Han erhöll för detta Fieldpriset vid kongressen i Edinburgh 1958.

Beträffande konstanten $1/\sqrt{5}$ på s. 48 kan nämnas att alla tal utom tal ekvivalenta i viss mening med $(\sqrt{5}-1)/2$ medger konstanten $1/(2\sqrt{2})$ där det åter bara finns ett undantag $\sqrt{2}+1$. Här finns en serie dylika konstanter och denna serie är ännu inte helt konstruerad.

De tre klassiska konstruktionsproblemen, som alltså alla är omöjliga, finns behandlade i följande böcker:

Vinkelns tredelning: I T. Nagell, *Lärobok i algebra.* Uppsala 1949, s. 200; detta problem löstes 1837.

5-tegradsekvationen: I G. Birkhoff, S. MacLane, *A Survey of Modern Algebra.* Macmillan 1947, s. 436; lösningen gavs av Abel 1823. Galois utvecklade som 20-åring 1830 den teori som förklarar sammanhangen.

Cirkelns kvadratur: I G. H. Hardy, E. M. Wright, *Introduction to Number Theory.* Oxford 1938, s. 172; lösningen kom 1882.

5. Algebra

Algebra torde för de flesta vara liktydigt med bokstavsräkning, dvs. identiteter av typen $a^2 - b^2 = (a+b)(a-b)$. Vi har tidigare fört fram som allmän beskrivning att algebran är läran om mängder med kompositionsstruktur. Man har alltså definierat en eller flera operationer som binder samman elementen i en mängd till nya element samt vissa regler hur dessa operationer samverkar. De vanligaste beteckningarna för operationerna är $+$ och $\cdot$ men man måste frigöra sig från de traditionella tolkningarna av dessa tecken. Det betydelsefulla i den axiomatiska metodik man använder är att många olika begrepp, tagna ur skilda delar av matematiken lyder samma lagar om man på lämpligt sätt tolkar de aktuella operationerna. Härigenom kan man således i olika sammanhang använda resultat som härletts en gång för alla. Ju mera invecklad en struktur är, desto mer kan givetvis härledas, men å andra sidan passar därigenom färre matematiska system in i beskrivningen. — Detta är den bild av algebran som bör förmedlas och inte uppfattningen att det handlar om formella omformningar av rationella funktioner. Det sista är naturligtvis relativt betydelselöst och tråkigt.

Vi skall nu se närmare på några av de strukturer som man på detta sätt infört.

Gruppteori

⟨Talen (1, 2, 3, 4, 5) kan ordnas på $120 = 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5$ olika sätt. [Bevisa det!] För en omordning σ kan man lämpligen använda följande beteckning

$$\sigma: \begin{pmatrix} 1 & 2 & 3 & 4 & \text{[illegible]} \\ 2 & 5 & 4 & 1 & \text{[illegible]} \end{pmatrix}$$

som alltså betyder att 1 övertar 2:s plats, $2 \to 5$ osv. Egentligen har vi att göra med en $1-1$-avbildning σ av en mängd bestående av 5 element på sig själv, där $\sigma(1)=2$, $\sigma(2)=5$ osv. Tag nu en annan omordning τ bland de 120, t.ex.

$$\tau\colon \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 3 & 5 & 2 & 4 \end{pmatrix}.$$

Vi kan definiera en produkt $\tau \cdot \sigma$ som den avbildning som består i att först utföra σ och sedan τ ($\sigma(1)=2$ och $\tau(2)=3$ ger alltså $(\tau\sigma)(1)=3$):

$$\tau \cdot \sigma\colon \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 3 & 4 & 2 & 1 & 5 \end{pmatrix}.$$

Vi observerar att $\sigma \cdot \tau$ är något helt annat:

$$\sigma \cdot \tau\colon \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 4 & 3 & 5 & 1 \end{pmatrix}.$$

Den avbildning e som inte ändrar något, fungerar som en etta — enhet— så att $\sigma e = e\tau = \sigma$ för alla σ. Avbildningen baklänges σ^{-1} kallas invers till σ. Den motsvarar division så att $\sigma \cdot \sigma^{-1} = \sigma^{-1}\sigma = e$. I vårt exempel är

$$\sigma^{-1} = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 4 & 1 & 5 & 3 & 2 \end{pmatrix}.$$

De konstruktioner vi här gjort kan lika väl utföras på vilken grundmängd M som helst i stället för de fem elementen och resultatet är ett exempel på en grupp. En grupp har alltså den logiska formen av omvändbart entydiga avbildningar av en mängd M på sig själv. Operationen är att vi utför avbildningarna efter varandra. Om vi betecknar elementen med $a, b, c, \ldots$ och operationen med $\cdot$ innebär detta väsentligen att $(ab) \cdot c = a(bc)$ och att ekvationerna $a \cdot x = b$ och $x \cdot a = b$ alltid är entydigt lösbara. Omvänt kan man genast uppfatta elementen i en sådan *axiomatisk definition* av en grupp som avbildningar. Avbildningarna är då definierade genom operationen $\cdot$, och de är alltså $x \to a \cdot x$. Avbildningen kan sedan identifieras med elementet a.

Man kan genast peka på resultat som lätt följer ur definitionen. Som exempel kan vi nämna att lösningen x till $a \cdot x = a$ är densamma för alla a. Denna kallas enhet och betecknas e. För e gäller också $e \cdot a = a$ för alla a, men detta är ett undantag. Tvärtom är det mycket viktigt att man inte antar $a \cdot b = b \cdot a$, som ju t.ex. inte gällde i vårt exempel. [Försök bevisa de resultat vi nämnde.]

Den enklaste gruppen är den cykliska där man får alla element genom att bilda $e, a, a^2, a^3, \ldots$ för ett lämpligt element a. Då gäller automatiskt $b \cdot c = c \cdot b$. Det finns nu två möjligheter: antingen är alla a^n olika eller också gäller $a^{n+m} = a^n$ för vissa n och m. I det sista fallet följer $a^m = e$ och sedan $a^{m+1} = a$, $a^{m+2} = a^2, \ldots$ $a^{2m} = a^m = e$ etc. Alla element erhålls således i uppräkningen $e, a, a^2, \ldots, a^{m-1}$. Det vi gör är alltså att lägga ihop heltal i exponenter när vi multiplicerar, och varje gång vi passerar m börjar vi om från 0. Detta är i sin tur samma sak som att säga att två tal är lika om de ger samma rest vid division med m. Geometriskt betyder gruppen en vridning av systemet av m punkter på lika avstånd runt en cirkel så att varje punkt övergår i den närmast följande. Vi kan alltså uppfatta det axiomatiska systemet antingen som rester i talteori eller vridning i geometrin. I denna frihet att tolka ligger betydelsen av axiomatisk framställning. Denna grupp brukar betecknas Z_m. — Finns inget tal m och n som ovan ser man lätt att vi kan uppfatta vår grupp helt enkelt som mängden Z av hela tal (= exponenterna).

Ett begrepp till behöver vi införa. Om G' är en delmängd av gruppen G och G' också är en grupp, talar vi om en delgrupp. I t.ex. den cykliska gruppen Z_4 av ordning 4 är elementen 0 och 2 (e och a^2) en delgrupp Z_2 av ordning 2 (observera att 2 går jämnt upp i 4). Vi tar nu två element x och y i G och bildar mängderna X av element $x \cdot a'$, där a' genomlöper G', och mängden Y av element ya'. Om x tillhör G', blir X tydligen precis G' eftersom till varje z' i G' finns ett a' med $xa' = z'$. Om alltså både x och y tillhör G' blir $X = Y = G'$. Vi skall nu se att helt allmänt gäller att antingen är $X = Y$ eller också har X och Y inga gemensamma element. Den matematiska termen för detta är att vi infört ekvivalensklasser X, Y,

Fig. 11

... Antag således att det finns ett gemensamt element så att $x a_0' = y b_0'$. Då följer $x = y b_0' a_0'^{-1} = y c'$ för ett visst c' i G'. Mängden X är alltså lika med mängden $y c' a'$. När a' genomlöper G' gör $c' a'$ det också och alltså är $X = Y$. G blir härigenom uppdelad i delar X, Y osv. så att varje element tillhör exakt en sådan del (fig. 11).

I vårt exempel är delarna $X = G' = \{0, 2\}$ och $Y = \{1, 3\}$.

Man kan nu försöka införa en multiplikation mellan delarna X genom att säga att $X \cdot Y =$ den del som $x \cdot y$ tillhör. För att detta skall vara en riktig definition får resultatet inte bero på hur vi väljer x i X och y i Y. Det är lätt att se att vad som fordras av G' för detta är att elementet $z a' z^{-1}$ tillhör G', om a' gör det, för varje z i G. G' säges då vara en *normal* delgrupp och vi använder beteckningen N för sådana G'. Mängderna X kan nu visas uppfylla axiomen för en grupp [kontrollera detta!] under denna $\cdot$ -operation och den nya gruppen betecknas G/N. Observera att om $a \cdot b = b \cdot a$ för alla a och b i G, så är alla G' normala eftersom $z a' z^{-1} = z z^{-1} a' = ea' = a'$.

Den självklara frågan blir nu: finns det (icke-triviala) normala delgrupper? Det gör det tydligen inte alltid. Ur vår utredning följer nämligen lätt att om G har ändligt många element, n stycken, så har varje X lika många, h stycken vardera, och h måste alltså gå jämnt upp i n. [Visa detta!] Om då n är ett primtal p, t.ex. om $G = Z_p$, finns inga icke-triviala delgrupper G' överhuvudtaget. Att utreda under vilka omständigheter normala delgrupper existerar är ett omfattande och viktigt problem, där nyligen stora framsteg gjorts.)

Den snedstrecksoperation, G/N, vi just definierat är den allra viktigaste i modern algebra och förmodligen i modern matematik också, och vi skall göra oss förtrogna med den genom en serie exempel.

1. De hela talen Z bildar ju på naturligt sätt en grupp under addition. En undergrupp G' är de jämna talen J. Gruppen är kommutativ (dvs. $b+a=a+b$ för alla element a och b i gruppen). Vad är Z/J? Två tal x och y i Z hör till samma klass om $x+$(jämnt tal)$=y+$(jämnt tal), dvs. om $x-y$ är jämnt. Det finns alltså bara två klasser: klassen av jämna tal (j) och klassen av udda (u). Vi har symboliskt $j+j=j$, $j+u=u$ samt $u+u=$ $=j$. Om j motsvarar 0 och u 1 är detta additionsregeln i Z_2. Man ser att $Z/J=Z_2$.

2. De komplexa talen C består av par (a, b) av reella tal R vilket vi kan skriva $R\times R$. Under addition är de reella talen R en delgrupp. Man visar lätt att $C/R=R$, dvs. / är en slags omvänd operation till $\times$ i $C=R\times R$.

3. Låt G vara alla polynom under addition och N alla polynom som är $=0$ i punkterna a_1, a_2, a_3. Två polynom P och Q hör alltså till samma klass om $P-Q$ är $=0$ i punkterna a_1, a_2, a_3. Välj ett polynom A_1 som $=1$ i a_1 och 0 i de andra två punkterna. Välj A_2 så att $A_2=1$ i a_2 och $=0$ i de andra två samt A_3 analogt. Varje P är då $=P(a_1)A_1+P(a_2)A_2+P(a_3)A_3+$(ett polynom som är $=0$ i a_1, a_2, a_3). Talen $P(a_1)$, $P(a_2)$, $P(a_3)$ bestämmer alltså den delmängd som P tillhör varför $G/N=$ alla taltripler under komponentvis addition.

Allmänt innebär /-operationerna att man tar fram det väsentliga i ett problem och de resulterande grupperna blir ofta väsentligt enklare än de ursprungliga. Man kan uttrycka sig allmänt så: bortsett från N är G detsamma som G/N — den matematiska termen är ”modulo” i stället för bortsett från. Bortsett från de jämna talen är de hela talen bara 0 eller 1. Modulo alla polynom som är $=0$ i tre punkter, är polynomen tre talkomponenter.

⟨ Vi kan nu mer exakt förklara hur beviset av olösbarhet av femtegradsekvationen med enbart rotoperationer går till. Till det talsystem T man får genom att till de rationella talen lägga lösningen θ, hör ett antal avbildningar $T\to T$ som bibehåller $+$ och $\cdot$. Dessa avbildningar bildar en grupp G. Om vi lägger till en m:te rot $\sqrt[m]{a}$, blir gruppen cyklisk med m element. Om man kunde få θ genom successiva rotutdragningen skulle det finnas en växande serie undergrupper G_i så att G_i/G_{i-1} vore cyklisk, eftersom steget från G_{i-1} till G_i motsvarar tillägget av en viss m:te rot. Nu är det lätt att se att i allmänhet är G helt enkelt gruppen av alla avbildningar av 5 element på sig själv. Man visar att denna explicita grupp inte kan lösas upp i en sådan växande serie av grupper G_i så att G_i/G_{i-1} är cykliska — detta kallas att vara lösbar. Därav följer beviset. Eftersom 3:e- och 4:egradsekvationen kan lösas, måste således gruppen av avbildningar av 3 och 4 element på sig själva vara lösbar. Här är en konstruktion för 3. De tre avbildningarna

$$(1,2,3) \begin{array}{l} \nearrow (3,1,2)=a \\ \rightarrow (2,3,1)=b \\ \searrow (1,2,3)=c \end{array}$$

bildar en grupp N: $ab=e=ba$ och man ser att den är identisk med Z_3. Den kan kontrolleras vara en normal delgrupp, t.ex. om $y=(2,1,3)$ gäller

$$\begin{aligned} y\,a\,y^{-1} &= (2,1,3)\cdot(3,1,2)\cdot(2,1,3)= \\ &=(2,1,3)\cdot(1,3,2)= \\ &=(2,3,1)=b. \end{aligned}$$

Vi kan nu bilda G/N och man ser att $G/N=Z_2$ är den enklaste cykliska gruppen. [Problemet är svårare för 4 element, men gör ett försök!] 〉

Gruppteorins största betydelse ligger inom topologin, som vi skall syssla med i nästa kapitel. Den har utvecklats för sin egen skull under lång tid. Det är märkligt, att trots mycket arbete och trots att reglerna för grupper är så enkla, har på de allra sista åren stora framsteg gjorts och många betydelsefulla problem är fortfarande olösta. Ett exempel på ett nyligen löst problem är följande: Varje grupp som innehåller ett udda antal, n, element är lösbar. Det är mycket lätt att se att om n är ett primtal är gruppen själv cyklisk. Det är en trevlig sysselsättning att försöka visa satsen då $n=p_1\,p_2$ för två udda primtal p_j. Beviset för det allmänna fallet är oerhört komplicerat — det fyller omkring 200 trycksidor — och är ett exempel på vilka djupa resultat som kan erhållas, inte genom en elegant idé, utan genom ett systematiskt inträngande i de logiska förhållandena. Det finns en allmän princip i matematiken som säger ungefär följande. Ju större allmängiltighet ett resultat har, desto enklare är ett bevis. I algebran innebär detta att mera invecklade strukturer skulle ha intressantare och svårare satser, beroende på att flera axiom har möjlighet att samverka. I stort sett är principen säkert sann, men den nämnda satsen ur gruppteorin visar att den måste användas med försiktighet.

Det är troligt att gruppteorin kommer att spela en betydelsefull roll, då fysikerna någon gång i framtiden får ordning på atomkärnornas möjliga tillstånd. Man vet tidigare att en elektron kan ha två så att säga egentillstånd, som man brukar kalla spin, som kan illustreras genom rotation i en viss riktning eller motsatta. Övergången från det ena tillståndet till det andra kan beskrivas med gruppen Z_2. På liknande sätt kan man tänka sig att atomkärnans mycket mera komplicerade tillstånd kan beskrivas ske genom mera komplicerade grupper av övergångar. Det finns av denna anledning allt skäl för matematiker att fortsätta studera grupperna och för fysikerna att följa med den kunskap som kommer fram.

Vektorrum

Vi har tidigare stött på begreppet vektor när vi sysslade med komplexa tal. Den bild man bör hålla i minnet är den pil som illustrerade det komplexa talet. Klart var att om vi betecknar n reella eller komplexa tal tillsammans med en symbol $x = (x_1, x_2, \ldots, x_n)$ kan vi definiera analogt $x+y=(x_1+y_1, \ldots, x_n+y_n)$. Man utför helt enkelt många vanliga additioner samtidigt. Vi kan också multiplicera med ett reellt tal t enligt regeln $tx=(tx_1, tx_2, \ldots, tx_n)$ och då gäller $t(x+y)=tx+ty$. Man kan också tänka sig alla x_j och t som komplexa tal. Sådana system som har en $+$-operation på detta sätt – bildar en grupp under $+$ – och dessutom har en sådan multiplikationsoperation med t reellt eller komplext tal, bildat *vektorrum* V. Ett exempel utgör alla f, reella funktioner $f(u)$ definierade på vilken mängd som helst, om man definierar $(f+g)(u)=f(u)+g(u)$. (Observera att $+$ här har olika betydelser i vänstra ledet och högra ledet.)

Om V' och V'' är vektorrum kan vi bilda V bestående av alla par $x=(x', x'')$, precis som de komplexa talen bildades ur de reella (fast här finns ingen multiplikation $x \cdot y$ förstås). Vi betecknar detta $V = V' \times V''$ och man ser att den omvända operationen är $V'' = V/V'$, varmed man alltså uppdelar en vektor x i två komponenter.

Med vektorrummen har man alltså utökat strukturen hos grupper genom att antaga att $x+y=y+x$ och genom att införa en ny operation: multiplikation med t. Härigenom har man fått ett mindre allmänt begrepp, som visar sig rätt ointressant ur algebraisk synpunkt, men som spelar en avgörande roll i modern analys.

Ring

Vi fortsätter vårt program att införa i vårt algebraiska system fler och fler axiom som approximerar de egenskaper som de rationella talen har. Vi antar nu att vi har både en $+$ och en $\cdot$-operation. Vi antar *inte* att man alltid kan lösa ekvationer $x \cdot a = b$, och *inte* heller att det finns något som motsvarar 1, dvs. så att $x \cdot 1 = x$. Vi får då begreppet ring. Man observerar att här kan mycket väl hända, att $x \cdot y = 0$ utan att varken x eller $y=0$. Tag t.ex. systemet av alla talpar $x=(x_1, x_2)$ med $x \cdot y = (x_1 y_1, x_2 y_2)$; då gäller $(1, 0) \cdot (0, 1) = (0, 0) = 0$.

En typisk illustration till begreppet är alla polynom eller alla polynom med heltalskoefficienter. Axiomen fungerar tydligen lika bra om polynomen beror av fler variabler än en, dvs. polynom som x^2+y^3 eller $x_1 \cdot x_2 \cdot x_3$, där vi har två respektive tre variabler. ⟨ För att belysa att en ring är något mycket allmännare än polynom, utgår vi från vilken kommutativ ($a \cdot b = b \cdot a$!) grupp som helst. Bilda nu mängden av alla ändliga formella "summor" av element ur G med heltalskoefficienter h, $h_1 a_1 + h_2 a_2 + \ldots + h_n a_n$. Logiskt innebär detta att vi ger en funktion $G \to Z$ (heltalen) som avbildar alla element utom ändligt många på noll. Bilden av a anger då vilken koefficient h vi skall sätta framför a. Vad vi fått betecknar vi med x, y etc. $x+y$ definierar vi genom att lägga ihop koefficienterna framför varje a och $x \cdot y$ genom att multiplicera ihop som vanligt och

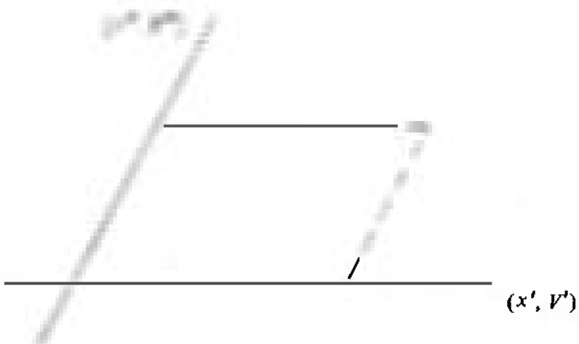


Fig. 12

använda reglerna för multiplikation i G. T.ex. $(2a_1+3a_2)(3a_1-$ $-2a_3) = 6a_1{}^2 - 4a_1a_3 + 9a_1a_2 - 6a_2a_3 = 6b_1 - 4b_2 + 9b_3 + 6b_4$, där $b_1 = a_1{}^2$ osv. Vad vi nu fått är en ring. Om t.ex. $G = Z$ är systemet detsamma som heltalspolynomen [tänk igenom detta!]. Väljer vi $G = Z_2$ får vi en multiplikation av linjära heltalsfunktioner i t enligt regeln

$$(h_0+h_1t)(h_0'+h_1't) = (h_0h_0'+h_1h_1')+(h_0h_1'+h_0'h_1)t$$

som kan jämföras med multiplikation av komplexa tal.⟩

⟨Vi har tidigare sett att man är intresserad av att lösa ekvationer $P(t_1, t_2, \ldots, t_n) = 0$ för polynom P och av deras lösningar, eventuellt bara i heltal $t_1, \ldots, t_n$. Kalla mängden av lösningar M: det är alltså denna mängd som är av intresse. Till M tillordnar vi mängden av alla polynom som är $= 0$ på M. Kalla denna mängd I. Om vi använder beteckningarna x och y för polynom har I egenskaperna: om x och y är i I är även $x-y$ i I och $z \cdot x$ tillhör I för varje z i ringen. Dessa egenskaper kan vi använda som definition i en godtycklig ring av ett system I som vi kallar *ideal*. Detta begrepp är det centrala därför att det ger anknytningen mellan de abstrakta begreppen och de klassiska problemen.

För att illustrera, tag den enklaste ringen, nämligen heltalen Z. Om I är ett ideal, låt a vara det minsta positiva tal som ingår i I. Då ingår tydligen alla tal $n \cdot a$. Om ett tal b i I inte

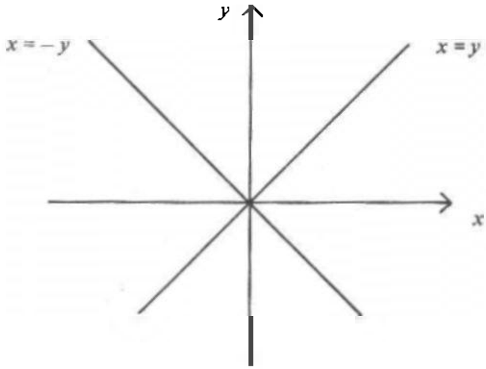


Fig. 13

hade formen $n \cdot a$ skulle man kunna hitta m så att $0<b-ma<a$ men $b-ma$ hör till I och vi har fått en motsägelse. Till varje I hör alltså a så att I = mängden av $n \cdot a$, dvs. vi kan identifiera I med heltalet a.

Nu kan man addera och multiplicera ideal. I_1+I_2 är alla element av formen x_1+x_2 med x_1 från I_1 och x_2 från I_2 och $I_1 \cdot I_2$ alla summor av element $x_1 \cdot x_2$. Det är lätt att kontrollera att dessa mängder är ideal. Om vi tar exemplet Z, är det a som är tillordnat I_1+I_2, inte a_1+a_2 utan största gemensamma faktorn till a_1 och a_2, medan $a_1 \cdot a_2$ hör samman med $I_1 \cdot I_2$. Vi kan nu införa primideal som sådana I som inte kan faktoriseras $I=I_1I_2$. De motsvarar precis primtalen för Z. Detta är grunden till de generaliseringar av primtal som vi talade om i förra kapitlet.⟩

Ringteorin har haft sin största betydelse för studiet av kurvor och ytor bestämda av polynomrelationer, dvs. mängderna M. Speciellt har man varit intresserad av hur dessa mängder förgrenar sig. Situationen kan illustreras av ekvationen $x^2=y^2$ som geometriskt kan åskådliggöras genom två räta linjer (fig. 13).

I origo sker här en förgrening. I flera variabler än två kan situationen vara mycket komplicerad och här pågår en stor forskningsverksamhet, där man alltså studerar den geometriska bilden med metoder ur algebran. Studiet av mångfalderna M har också stor betydelse i analysen. Slutligen finns anknytningar till talteorin som vi nyss nämnde.

Homomorfismer

Det finns många andra system som man sysslar med i algebran: som ett tillägg kan nämnas att själva ordet algebra används för den struktur man får om man har en multiplikation definierad i ett vektorrum, och att en kropp är ett system som fungerar som de rationella talen. Om vi går tillbaka till konstruktionen av rationella tal (s. 26), ser vi att vi kan bilda en kropp ur vilken ring som helst så snart regeln, att $b \neq 0$ och $d \neq 0$ medför $bd \neq 0$, gäller. Vi vet att den inte gäller automatiskt, men på detta sätt uppstår ju t.ex. rationella funktioner ur polynom, och de rotsystem vi diskuterade på s. 49 uppfyller också villkoret.

Vi skall emellertid avslutningsvis diskutera en av de viktigaste idéerna i algebran. Vi utgår från två algebraiska system av samma typ, kalla dem A och B. Vi studerar nu avbildningar från A till B som har den egenskapen att de bibehåller de operationer som finns definierade på A resp. B. Om således $a \to b$ och $a' \to b'$ så skall gälla $a +_A a' \to b +_B b'$ osv., där $+_A$ betyder $+$-operationen på A och analogt för B. Avbildningar av detta slag kallas homomorfismer.

Som enklaste exempel tag som A gruppen av heltal under vanlig addition och som B gruppen T av komplexa tal på enhetscirkeln med gruppoperation vanlig multiplikation av komplexa tal. $0 \to z_0$ varför $0 = 0 + 0 \to z_0 \cdot z_0 = z_0$. Det följer $z_0{}^2 = z_0$ varför $z_0 = 0$ eller $z_0 = 1$. $z_0 = 0$ ligger inte på enhetscirkeln. Alltså gäller $0 \to 1$. Enheten i A avbildas alltså på enheten i B, och detta gäller allmänt för grupphomomorfismer, liksom att inverser övergår i inverser. Antag nu att $1 \to c = e^{ix}$. $2 \to c^2, \ldots,$ $n \to c^n$ följer då. Alla avbildningar har alltså formen $n \to e^{inx}$.

På liknande sätt ser man att gruppen R av reella tal avbildas på T genom $x \to e^{itx}$, t godtyckligt reellt. Här finns även andra homomorfismer med mycket säreget beteende (se s. 17), som kan uteslutas genom vissa tilläggsantaganden om kontinuitet. Avbildningarna $x \to e^{itx}$ kan anses vara grunden till en viktig gren i analysen, Fourierteorin.

⟨Som ett tredje exempel, tag ringen av polynom $P(x)$ i en variabel x med heltalskoefficienter och dess avbildningar på heltalen. Polynomet $x \to$ heltalet m, och det följer att $P(x) \to$ $\to P(m)$. Dessa homomorfismer motsvarar alltså det konkreta: sätt in ett fixt värde på variabeln och detta har uppnåtts genom helt abstrakta definitioner. Ett litet mer komplicerat exempel får man genom att avbilda ett vanligt heltalspolynom $P(t) =$ $= h_0 + h_1 t + \ldots + h_n t^n$ på $(h_0 + h_2 + h_4 + \ldots) + (h_1 + h_3 + \ldots)t$ i vårt exempel på sidan 61.⟩

⟨Tag till sist en homomorfism mellan två grupper G och G_1 och låt N vara den del som avbildas på enheten e_1 i G_1. Man ser lätt att N är en delgrupp av G som dessutom är normal ty om $a \to e_1$ gäller $z\,a\,z^{-1} \to z_1 e_1 z_1^{-1} = z_1 z_1^{-1} = e_1$. Alltså tillhör även $z\,a\,z^{-1}$ N. Det är lätt att se att G_1 kan uppfattas som G/N. Tar vi t.ex. $G = R$ och $G_1 = T$ med homomorfismen $x \to e^{itx}$, t fixt, så är N alla x med $e^{itx} = 1$, dvs. $x = 2\pi/t$ (heltal). Således gäller $R/T = Z$. [Visa att också $R/Z = T$!]⟩

⟨Om vi har en följd grupper G_1, G_2, G_3, ... med tillhörande homomorfismer α_1, α_2, ... av G_1 på G_2, G_2 på G_3 osv.

$$G_1 \xrightarrow[\alpha_1]{} G_2 \xrightarrow[\alpha_2]{} G_3 \xrightarrow[\alpha_3]{} \ldots$$

och man antar att α_1 avbildar G_1 på den del av G_2 som av α_2 avbildas på enheten i G_3 och analogt för alla tre på varandra följande grupper, har vi fått vad som kallas en exakt följd. Vi kan göra analoga definitioner för homomorfismer mellan ringar. Detta begrepp spelar en central roll i de senaste årens utveckling av algebran, speciellt för dess användning i topologin som vi skall behandla i nästa kapitel. I förgrunden står alltså mängden av avbildningar mellan olika system och den struktur *denna* mängd har. Av allt att döma kommer betydelsen av denna del av algebran att växa ytterligare.⟩

Sammanfattningsvis kan man säga att algebrans betydelse i modern matematik är i starkt växande. Vad man lyckats med är att i axiomatisk form och med enbart kompositionsregler bygga upp stora delar av hela matematiken, även sådana som har helt annat ursprung. Den traditionella tillämpningen är talteori och geometri, men man har även kunnat konstruera system, som beskriver vissa delar av analysen. Denna algebraisering av matematiken kommer av allt att döma att fortsätta. Den kommer att ge betydelsefulla resultat, helt säkert, men i dagens läge har den haft den negativa effekten att minska förståelsen för den matematiska forskningen hos fysiker och har frestat undervisningsreformatorerna att överdriva formaliseringen.

Som vanligt vid motsättningar är problemet väsentligen ett kommunikationsproblem. En matematiker talar om "modulo" när han ungefär menar "bortsett från", han säger att två system *A* och *B* är isomorfa när han menar att de är samma sak bara med olika beteckningar. Han säger att *A* är homomorf med *B* om han menar att *A* är mera omfattande än *B* men att systemen annars bär sig åt på samma sätt. En grupp är en samling omordningar av elementen i en viss mängd osv. Matematikern har på så sätt rätt i att hans språk är mycket exakt — han vet precis vad han menar med sina ord — men det vore fel att förneka att formalismen ibland kan bli ett självändamål med mycket invecklad terminologi som används mera för att imponera än för att skapa klarhet. De grundläggande begreppen är inte så många, något eller några tiotal kanske, och de sammanhänger med och beskriver fenomen som man möter i alla möjliga sammanhang.

Algebrans betydelse har helt försummats i de traditionella nya skolkurserna. Precis som i diskussionen om matematikens grundbegrepp är huvudproblemet att göra eleverna förtrogna med vissa enhetliga grundbegrepp och några beteckningar och inte att öva handhavande och problemlösning. Det väsentliga är att orden används regelbundet och att analogier mellan system som ser olika ut betonas. Kan man få eleverna att förstå att operationen att vrida kvadrater 90° egentligen är samma sak som att dela heltal med 4, att multiplikation av positiva tal är likvärdigt med addition av alla reella tal (via exponentialfunktionen), att addition av funktioner lyder samma lagar som addition av tal osv., så förmedlar man en riktig bild av algebra. Det som bör bortarbetas är bilden av algebran som polynomreduktion och uppsökandet av gemensamma faktorer via diverse trick. Man strävar också i skolan att lära eleverna axiomatik. Det naturliga området är då algebran, där man i samma form behandlar många problem. Detta borde vara innebörden i ordet axiomatik, inte det logiska problemet att formalisera det välkända.

Kommentarer

Som allmän elementär översikt av algebran kan rekommenderas den ovannämnda boken av Birkhoff-MacLane. För ett djupare studium av gruppteorin kan man läsa

M. Hall, *The Theory of Groups*. Macmillan 1959.

Resultatet om lösbarhet av grupper med udda antal element bevisades av W. Feit och J. G. Thomson 1963 ["Solvability of Groups of Odd Order". *Pacific Journal of Mathematics*, 1963].

6. Geometri

För 50 år sedan spelade geometrin en mycket framträdande roll i skolan. Det är intressant att se hur djupt rötterna till denna undervisning går i den matematiska tankebyggnaden. Geometriundervisningen innehöll två helt olika komponenter. Å ena sidan betonades starkt Euklides axiomatiska system. Avsikten var sannolikt att öva eleverna i logiskt tänkande och Euklides var här mönstret. Vid kursomläggningarna för några år sedan utmönstrades Euklides helt, förmodligen därför att hela konstruktionen gick långt utöver vad man på ett givande sätt kan lära 10-åringar, men också därför att det hela egentligen var kvasilogik. Det finns nog mycket litet att invända mot den förändring som här skett; den har medfört en stor lättnad både för elever och lärare. Sekelskiftesgeometrin innehöll å andra sidan en svår kurs i konstruktionsgeometri, med invecklade system av cirklar och räta linjer. Även detta är månghundraårig matematik. Den utbyttes sedan mot en detaljerad och mycket åldersdiger kurs i kägelsnittsteori, som varit ett tacksamt mål för kritiker såsom meningslös problemexercis. I den senaste kursomläggningen har även denna geometridel nästan helt mönstrats ut. Resultatet har blivit en mycket dålig utbildning i geometri i vid mening, som står i stark motsättning till områdets stora betydelse för den moderna matematiska forskningen. Hur många av dagens studenter kan bevisa att vinkelsumman i en triangel är 180° och hur många av lärarna vet att detta resultat bara är sant, om man inkluderar parallellaxiomet?

Innan vi går in på den nuvarande situationen skall vi se litet närmare på axiomatiken. Det enklaste sättet att eliminera be-

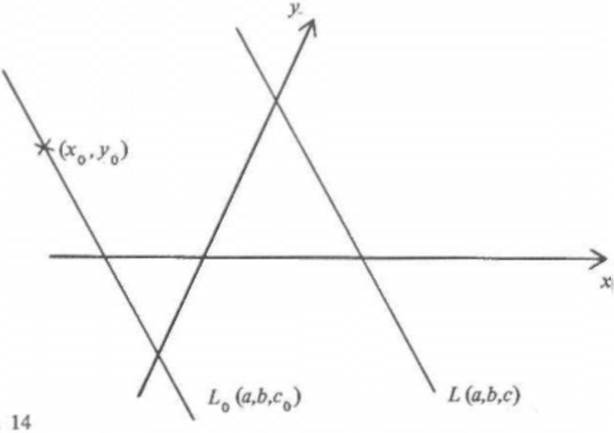


Fig. 14

hovet av axiomatik för euklidisk geometri är följande kringgående manöver.

Vi inför i planet på vanligt sätt ett koordinatsystem x, y, varigenom vi får $1-1$-motsvarighet mellan punkter i ett plan och par av tal (x, y). En rät linje L motsvarar en ekvation $ax+by+c=0$ och omvänt (se fig. 14). På vanligt sätt vänder vi nu på resonemanget. En punkt *är* ett talpar (x, y), en linje en taltrippel (a, b, c) där naturligtvis (a, b, c) och (Ka, Kb, Kc), $K \neq 0$, anses vara samma sak. En punkt (x, y) ligger på en linje (a, b, c) om $ax+by+c=0$ etc. På detta sätt blir geometrin bestämd av teorin för de reella talen och är lika sann som den.

I denna modell absorberas automatiskt det mest berömda av Euklides axiom: parallellaxiomet att man genom en punkt utanför en linje kan dra en entydig linje parallell med den givna. Om linjen är (a, b, c) och punkten (x_0, y_0) bestämmer vi c_0 så att $ax_0+by_0+c_0=0$ och (a, b, c_0) är den sökta linjen L_0. — Man kan också göra formella axiomsystem à la Euklides och i sådana är parallellaxiomet nödvändigt, men man kan även konstruera system som på andra sätt helt liknar den vanliga geometrin, men där parallellaxiomet inte gäller. Låt oss se litet på ett sådant exempel, som alltså är en modell för icke-euklidisk geometri.

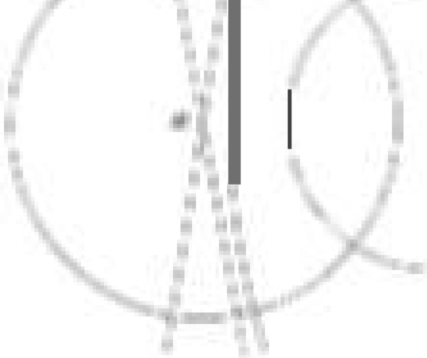


Fig. 15

Tag som värld en cirkelskiva med radie R ($=10^{100}$ mil t.ex.). Vi lever nära dess medelpunkt. En "punkt" i vår modell är en vanlig punkt i cirkeln, men en "linje" L är en cirkel som är vinkelrät mot den stora cirkeln — hit räknas också alla diametrar D (se fig. 15). Det är inte svårt att kontrollera att alla axiom i Euklides geometri är uppfyllda för detta system av "punkter" och "linjer" med undantag för parallellaxiomet.

Vi ser i figuren att diametrarna D och D' inte skär "linjen" L_0 men går genom samma "punkt" M (origo). Om vi inskränker oss till en omgivning av medelpunkten, så att L_0:s närmaste punkt ligger, säg högst 10^{50} mil från M så ser man att variationen på vinkeln v mellan D och D' är av storleksordningen 10^{-50} grader. Det går alltså inte att genom mätningar skilja på D och D' och det följer att det inte finns någon logisk anledning att föredra euklidisk geometri framför icke-euklidisk. Dess fördel är dess enkelhet och att den stämmer tillfredsställande i praktiken.

Logiskt är detta exempel mycket intressant. Vi utnyttjar alltså den euklidiska geometrin för att konstruera en modell av en geometri där de euklidiska axiomen inte gäller. Detta visar att om den euklidiska geometrins axiom inte medför motsägel-

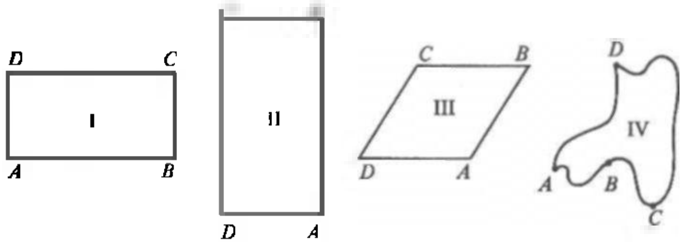


Fig. 16

ser, gör de icke-euklidiska axiomen det ej heller. Alldeles samma logiska struktur har beviset beträffande urvalsaxiomets relation till övriga axiom i mängdläran (se s. 17). Man konstruerar alltså modeller inom traditionell mängdlära och får därför som resultat att om det inte tidigare fanns motsägelser innebär det nya axiomet heller inte motsägelser.

Topologi

Vi har diskuterat två möjligheter att bygga upp geometrin. Den första var att basera den på elementär algebra av reella tal, den andra att skapa ett fristående axiomsystem. Man kan också grunda geometrin på gruppteorin. Euklidisk geometri är då studiet av de egenskaper som inte ändrar sig under gruppen av sådana avbildningar som bibehåller längder (ortogonala gruppen). Man kan också studera egenskaper som inte beror på gruppen av avbildningar som överför räta linjer i räta men som får ändra längder (affin geometri). I första fallet är alltså

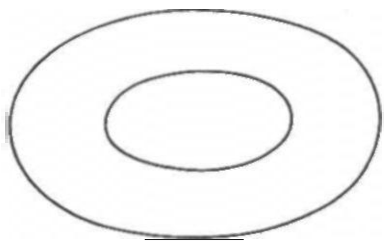


Fig. 17

figurerna I och II "kongruenta", i det andra I, II och III "kongruenta" (se fig. 16). Slutligen kan man studera gruppen av kontinuerliga deformationer överhuvudtaget. Då är alla figurerna I—IV "kongruenta". Detta sista område är topologin och den matematiska forskningen i vad som förr kallades geometri är nu centrerad till detta område. Det är således läran om de geometriska förhållanden som inte beror på begrepp som längd och area, utan som bara kan bero på den geometriska situationen. I exemplet fig. 17 finns då bara följande element: 2 kurvor som ej skär sig själva eller varandra, varav den ena omsluter den andra. (Det visar sig ingalunda självklart vad som menas med "kurvor" och "omsluter".) Precist innebär uttrycket "bara beror på den geometriska situationen" att utsagorna skall vara desamma för alla andra uppsättningar (av kurvor osv.) som kan avbildas omvändbart entydigt och *kontinuerligt* på den ursprungliga; i vår bild t.ex.

Fig. 18

Det första man nu måste göra är att definiera kontinuitet, och detta skall göras så allmänt som möjligt så att vår framställning täcker så stora områden som möjligt. Vi påminner om diskussionen om öppna mängder och konvergens på s. 30. Avbildningen $f\colon A \to B$ är då kontinuerlig om det alltid följer att $f(x)$ närmar sig $f(a)$ när x närmar sig a $(x \to a)$ med iakttagande av de regler om vad ordet "närmar" betyder på mängderna A resp. B — med andra ord, vilken topologi man har, enligt matematisk terminologi. För att man på ett meningsfullt sätt skall kunna arbeta med topologierna i A och B behövs fler antaganden om de öppna mängderna, väsentligen av typen att om $a \neq a'$ kan man hitta en omgivning till a som ej innehåller a'. Detta kan göras på många sätt och leder till en hierarki av topologiska rum med olika axiomsystem av samma typ vi

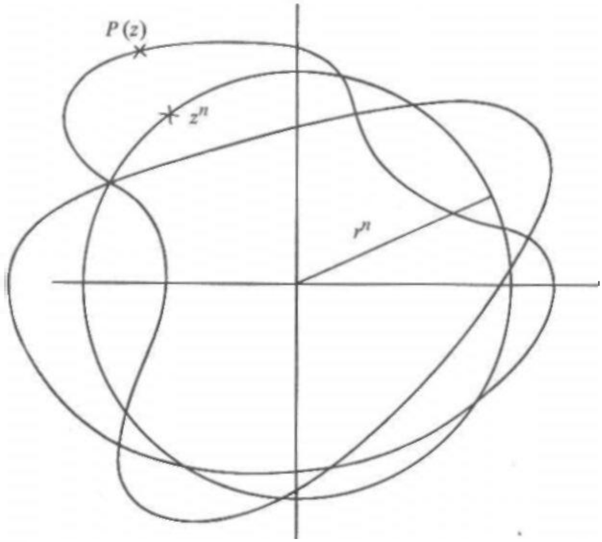


Fig. 19

mötte i algebran. En systematisk undersökning av dessa axiomsystem var en livlig forskningsinriktning under första delen av 1900-talet och kan nu anses väsentligen avslutad.

För att dessa definitioner skall ha mening, måste man föreställa sig mycket mera allmänna situationer än vanliga funktioner $f(x)$ definierade på intervall. ⟨ Tag som ett första exempel A som mängden av kontinuerliga funktioner $x(t)$ på $0<t<1$ och som B mängden av reella tal med vanlig topologi. Avbildningen $f\colon x \to \int_0^1 x(t)^2 dt$ är då kontinuerlig i den första topologin på s. 32 men inte i den andra. Som ett exempel, för att visa det sista påståendet, tar vi funktionerna $x_n(t)=\sqrt{n}$ [illegible] Eftersom $\int_0^1 x_n(t)dt=n^{-1/2}$ går x_n mot noll-funktionen i topologin, när n växer, men $\int_0^1 x_n(t)^2 dt$ närmar sig 1/2 som ju $\neq \int_0^1 0^2 dt=0.$ ⟩

⟨ Låt oss som ett annat konkret exempel se närmare på algebrans fundamentalsats: ett polynom $P(z)=z^n+a_1z^{n-1}+\ldots+$ $+a_{n-1}z+a_n$, $a_n\neq 0$, antar värdet 0 för minst ett värde på z. Vi studerar avbildningen: till varje värde på $r>0$ tillordnar vi den kurva γ_r som beskrivs av z da z går runt cirkeln med radie r kring $z=0$. Mängden A är alltså de positiva talen, B en mängd av kontinuerliga kurvor. Kurvorna i B varierar kontinuerligt med r — detta påstående bör göras precist men är intuitivt klart. För stora r beskriver z^n cirkeln med radie r^n n gånger. Resten av polynomet $a_1z^{n-1}+\ldots+a_n$ är mycket mindre än r^n och alltså beskriver $P(z)$ också en kurva som går n gånger runt. Tydligen kommer γ_r att fortsätta att gå n gånger runt tills vi passerat ett värde på r då kurvan går genom origo. Måste ett sådant värde finnas? Ja, ty då r är mycket litet ligger γ_r mycket nära punkten a_n, som ju inte är noll, och kan då inte gå runt noll. Därmed skulle beviset vara färdigt. γ_r måste passera genom origo för något värde på r, dvs. $P(z)$ måste anta värdet noll. — Svårigheten i att göra beviset fullständigt ligger i att tala om vad som menas med att gå n gånger runt en punkt, vilket är en typisk uppgift för topologin.〉

Genom att låta rymderna A och B bestå av funktioner respektive kurvor finns det alltså möjlighet att utnyttja topologins resultat för studier i analysen. Mest markant blir detta om man kombinerar idén om ett vektorrum med idén om en topologi. Vi skall i nästa kapitel behandla detta närmare.

I matematiken talar man ofta om plan och sfärer i 4, 5 eller n dimensioner och detta kan låta både invecklat och mystiskt. Det är emellertid mycket enkelt. I ett plan R_2 inför vi ett ortogonalsystem (x, y). Ett plan är då en ekvation $x\xi+y\eta=c$ där ξ och η är fixa tal. Om vi kallar $z=(x, y)$ och $\zeta=(\xi, \eta)$ kan vi införa beteckningen $\langle z, \zeta\rangle$ för $x\xi+y\eta$, och $\langle z, \zeta\rangle=c$ är då en linje. En cirkel med centrum i origo är alla z med $x^2+y^2<R^2$ och $x^2+y^2=R^2$ är periferin. Vi generaliserar nu: $x=(x_1, x_2, x_3, x_4)$ är en punkt i ett 4-dimensionellt rum R_4. Mängden av sådana uppsättningar x av fyra tal så att $\langle x, \xi\rangle=x_1\xi_1+x_2\xi_2+$ $+x_3\xi_3+x_4\xi_4=c$ ger för fixt ξ ett "plan"; $x_1{}^2+x_2{}^2+x_3{}^2+x_4{}^2<R^2$ ett klot K_4; $=R^2$ ger dess rand, en "sfär", som vi för $R=1$ betecknar S_4. Två punkter "ligger nära varandra" om alla koordinater är nära i vanlig mening. Utvidgningen till 5 och fler koordinater är klar.

〈Vi kan illustrera hur dessa definitioner fungerar med ett konkret exempel. Tag två hoplänkade ringar som i fig. 20.

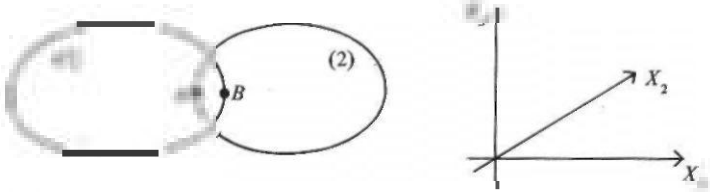


Fig. 20

Det är ju klart att man inte kan ta isär dem utan att klippa upp den ena ringen. Om man emellertid lägger till ytterligare en dimension går det. Det kan man inse på följande sätt.

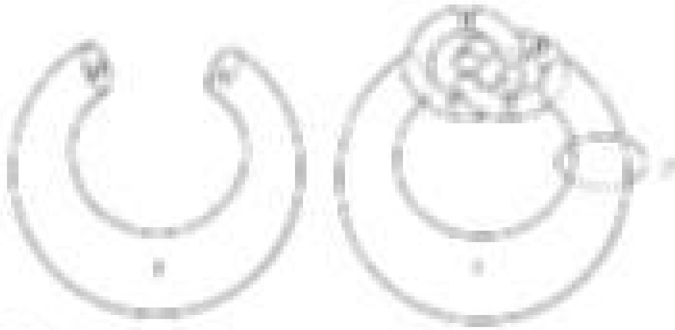
Vi tänker oss (2) vinkelrät mot (1). Vi kan representera (2) med ekvationer

$$x_1 = x_1(t),\quad x_2 = 0,\quad x_3 = x_3(t),\quad 0 \leq t \leq 1.$$

Vi håller nu (1) stilla och drar (2) åt höger. Vi kan representera den förflyttade kurvan $(x_1(t)+s,\ 0,\ x_3(t))$ där s anger förflyttningen. Vi kan anta att för $s=1$ är kurvorna isär. För ett visst värde på s, $s=s_0$, sammanfaller A med B vilket är otillåtet. Vi kan nu undvika detta genom att flytta ut ringen i 4 dimensioner och bildar $(x_1(t)+s,\ 0,\ x_3(t),\ s)$. Denna kurva kan ju aldrig skära (1) eftersom för (1) är fjärde koordinaten 0. När vi fått $(x_1(t)+1,\ 0,\ x_3(t),\ 1)$ flyttar vi bara fjärde koordinaten 1 tillbaka till 0 och har fått tillbaka vår ring (2) i tre dimensioner utan att den någonsin skurit (1).

Låt oss nu återgå till frågan om vad som bör menas med en "sluten kurva som ej skär sig själv". Det enklaste exemplet på en sådan är en cirkel och i någon mening uppfattar vi en dylik kurva γ som en deformerad cirkel. Man kan då definiera begreppet som en omvändbar kontinuerlig bild av en cirkel: $f\colon S_2 \leftrightarrow \gamma$. Det är nu en uppgift för topologin att förklara sådana termer som "omsluter" och "det inre av γ" och bevisa att γ uppför sig på rätt sätt i förhållande till definitionerna. Dessa definitioner gjordes i topologins tidigare skede i slutet av 1800-talet. Att γ sönderdelar planet i exakt två sammanhängande delar är t.ex. sedan en klassisk sats.

Som en illustration av topologiska problem skall vi syssla



(b)

med följande naturliga utvidgningsproblem vid definitionen av kurva. Man vill naturligtvis att *hela* den geometriska situationen för γ skall vara analog med den för cirkeln. T.ex. kan man önska att avbildningen f kan utvidgas till en avbildning $R_2 \leftrightarrow R_2$ med S_2 i ena planet och γ i det andra. Det är en berömd sats i tidig topologi att detta mycket riktigt går.

Generaliseringar är nu självklara. En enkel sluten yta y^3 är en kontinuerlig bild av en sfär: $f{:}S^3 \leftrightarrow y^3$ och på samma sätt i n dimensioner. Här inträffar nu något märkligt. Det är inte säkert att f kan utvidgas till en avbildning $R^3 \leftrightarrow R^3$. Ett exempel kan se ut på följande sätt:

⟨ Utgå från en böjd yta som i fig. 21a. Delen I i fig. 21a avbildas på delen I på sfären i fig. 23a så att alltså de två streckade områdena motsvarar att man skurit av den böjda ytan vid B och A. Sätt nu på handtag som i fig. 22, så att

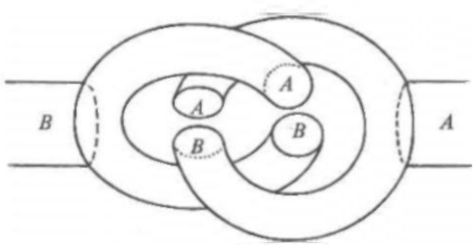


Fig. 22

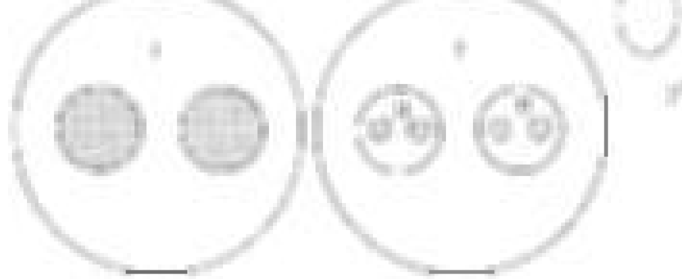


Fig. 23

(a) (b)

ytorna B motsvarar varandra och ytorna A motsvarar varandra. Vi får då något som ser ut som i fig. 23b. De två påsatta handtagen kan avbildas på delar av de streckade områdena så att II motsvarar II. Det blir nu fyra streckade områden på sfären. Vi sätter nu på 4 parvis hopflätade handtag och fortsätter på detta sätt i all oändlighet. Den resulterande ytan kan då avbildas omvändbart entydigt och kontinuerligt på sfärens yta. Den uppkomna ytan brukar kallas sfär med horn.

En avbildning mellan denna yta och sfären kan nu inte utvidgas till en avbildning $R^3 \leftrightarrow R^3$. Då skulle nämligen kurvan γ i fig. 21b motsvara γ' i fig. 23b. γ' kan deformeras till en punkt utan att träffa sfären. Då skulle vi alltså kunna göra samma sak med γ utan att träffa den behornade sfären och det är tydligen omöjligt. ⟩

Ett viktigt problem i topologin är följande: Antag att vi har två mängder A och B med sina topologier och två avbildningar f_1 och f_2 av A på B. Kan vi uppfatta f_2 som en deformation av f_1? Vi skall se litet närmare på detta problem då A och B är n-sfärer och en av avbildningarna är identiteten. Vi kan då uttrycka problemet på följande sätt. f avbildar $S^n \leftrightarrow S^n$. De deformerade avbildningarna uppfattar vi som avbildningar av sfärer med radie r, $1/2 \leq r < 1$, på sfärer med radie r och för $r < 1/2$ skall vi ha identiteten. Frågan är nu om denna utvidgning är möjlig. I planet är det lätt att skriva upp en sådan avbildning f explicit. Vi låter sfären med radie r motsvara sfärer med samma radie. Det går då att beskriva avbildningen med en monoton funktion $\varphi_r(v)$ av centrumvinkeln för varje r, sådan att $\varphi_r(2\pi)-\varphi_r(0)=2\pi$. $\varphi_1(v)$ är då given och vi definierar

$$\varphi_r(v)=\begin{cases} v, \ 0<r<1/2, \ \text{(identiteten!)} \\ (2r-1)\varphi_1(v)+(2-2r)v, \ 1/2<r<1, \\ \varphi_1(v), \ r\geq 1. \end{cases}$$

Motsvarande problem i högre dimensioner är betydligt mycket mer komplicerat och här finns en omfattande teori. Det vore alldeles fel att tro att detta är matematiska spetsfundigheter utan egentlig betydelse. Man kan rätt generellt säga att alla problem som innehåller n parametrar, där alltså n storheter samverkar, är beroende av geometriska förhållanden i motsvarande antal dimensioner. Det brukar här invändas att man i fysikaliska problem på grund av problemets natur vet, att lösningen måste uppföra sig så väl, att egendomligheter av det slag vi här diskuterat inte kan inträffa för dessa lösningar. Enligt detta sätt att resonera skulle bara den del av matematiken som sysslar med "vanliga" funktioner vara av intresse för tillämpningar. Bortsett från att ett problems logiska struktur är av eget intresse, är detta en allvarlig felsyn. För en fysiker är den numeriska konstruktionen av lösningen det väsentliga. Detta sker genom någon form av successiva beräkningar ur det givna. Vid dessa beräkningar finns det i regel ingen möjlighet att skilja mellan "vanliga" funktioner och sådana med patologiskt uppträdande. Existerar således denna senare sorts lösningar, kan de uppträda i beräkningarna och dessa leder då inte till målet. Man vill därför ha garantier att lösningar av den sorten inte existerar, dvs. man måste ta hänsyn till dem vid förhandsberäkningarna och det är just det matematikerna gör.

Algebraisk topologi

Avsikten med denna matematiska gren är att beskriva hur ytor av olika slag och av olika dimensioner kan sättas samman och att göra detta med algebraiska termer. För att man skall kunna

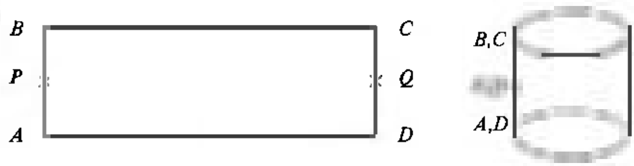


Fig. 24

förstå något av vad det rör sig om måste man först ha en föreställning om några litet komplicerade ytor. Vi nöjer oss av förklarliga skäl med 3 dimensioner men teorin skall sedan vara allmän.

Tag först en rektangel $ABCD$ (fig. 24) som vi tänker oss i papper. Vi kan nu klistra ihop den så att B sammanfaller med C och A med D. Vi får då en cylinder. Man kan uttrycka samma sak med att säga att vi identifierar AB och DC: de markerade punkterna P och Q är alltså lika.

Vrid nu rektangeln 180° innan den klistras ihop, dvs. identifiera AB med CD (i den ordningen). Vi får då en yta som kallas Möbius band (fig. 25) som uppenbarligen inte är en cylinder. Hur skall vi beskriva skillnaden? Detta är en uppgift för den algebraiska topologin.

Tag nu cylindern och böj ihop den till ett runt slutet rör, dvs. identifiera också BC med AD (i den ordningen). Nu är alltså A, B, C, D alla samma punkt. Vi har fått en torus, en ringyta (fig. 26).

Tag till sist cylindern och försök (men det går inte!) böja ihop den till ett rör så att BC sammanfaller med DA. Man ser

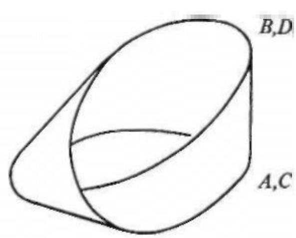


Fig. 25

Fig. 26

att denna yta inte kan realiseras men det går att beskriva den just med de identifikationer vi gjort: detta är Kleins flaska.

Vad vi nu vill beskriva är hur ytorna hänger samman. Det är tydligen en skillnad mellan en sfär och en torus. Ett sätt att ange detta upptäckte redan Euler. Vi delar in de tre ytorna i triangelformade områden. För sfären kan vi t.ex. ta två vinkelräta storcirklar genom en nord- och sydpol samt ekvatorcirkeln. Vi betecknar med α_0 antalet hörn, med α_1 antalet sträckor, begränsade av två hörn, och med α_2 antalet trianglar. För delningen av sfären får man $\alpha_0=6$, $\alpha_1=12$, $\alpha_2=8$. Vi bildar talet

$$g_S=\alpha_0-\alpha_1+\alpha_2=-2.$$

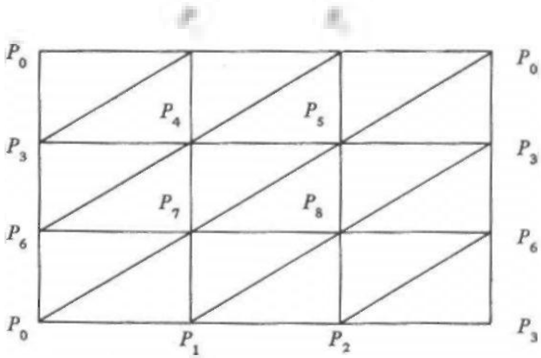


Fig. 27

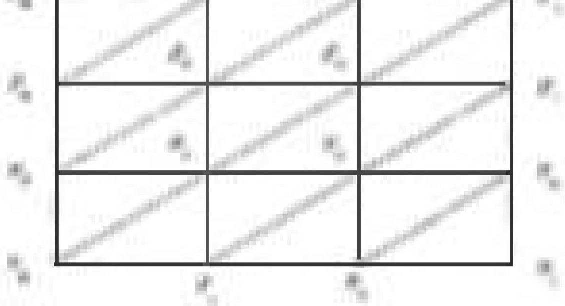


Fig. 28

För torusen T gör vi den uppdelning i trianglar som visas i fig. 27. Vi ser att $\alpha_0=9$, $\alpha_1=27$, $\alpha_2=18$, varför

$$g_T=9-27+18=0.$$

Om vi nu prövar andra indelningar i trianglar så skall vi se att talen inte ändrar sig. De beror bara på den yta vi studerar. På något sätt beskriver g hur sammanhangsgraden av ytan växer. En brist är nu att g i våra exempel hoppar från -2 till 0. Finns ingen yta (utan rand som sfären och torusen) som ger -1? En sådan kan inte realiseras men vi kan konstruera den som vi gjorde med Kleins flaska. Tag en rektangel och identifiera båda motsatta sidorna med omkastad riktning (se fig. 28). Vi får $\alpha_0=9$, $\alpha_1=27$, $\alpha_2=17$ [$P_0P_1P_4$ finns två gånger i figuren], varför $g=9-27+17=-1$.

Dessa exempel innehåller nyckeln både till vad det är geometriskt som skiljer ytorna och till en metod att behandla problemen. Det som skiljer är antalet "olika" sätt det finns att skära upp ytan, dvs. rita slutna kurvor, utan att ytan efter uppklippningen faller sönder. För sfärens del är det omöjligt: så fort vi klipper runt, lossnar en bit. På torusen kan vi göra det på två sätt; dels kan vi klippa av röret och dels kan vi klippa runt så att vi får något som liknar en cirkelring. g skiljer sig också med 2 enheter och antal olika sätt är helt allmänt $g+2$. — Den metod att studera ytors sammanhang som suggereras är att dela upp i triangelformade områden och sedan manipulera med trianglarna så att resultatet inte beror på hur vi delade. Detta leder till rent algebraiska begrepp och metoden har nu vuxit till en av de mest aktiva grenarna av modern matematik, som nu lever sitt eget liv och där de ursprungliga geometriska förhållandena spelar mycket liten roll. Vi skall här antyda metoden i 2 dimensioner, högre dimensioner är analogt men möjligheten till komplikationer växer starkt.

⟨Tag alltså en triangel $T=(P_0P_1P_2)$. Genom att skriva upp hörnen P i denna ordning har vi även angivit en omloppsriktning; triangeln $(P_0P_2P_1)$ betecknar vi formellt med $-T$. På samma sätt är sträckan $S=(PQ)=-(QP)$. Vi tillordnar även formellt en punkt ett tecken; detta saknar geometrisk tolkning.

Vi inför nu en randoperation ∂ som tillordnar till trianglar och sträckor deras rand (med ett formellt tecken ur omloppsriktningen):

(1) $\quad \partial(P_0P_1P_2)=\{(P_0P_1),\ (P_1P_2)\ \text{och}\ (P_2P_0)\};$

(2) $\quad \partial(PQ)=\{(P),\ -(Q)\}.$

Om vi nu tar randen av varje sträcka i (1) får vi (P_0), $-(P_1)$; (P_1), $-(P_2)$; (P_2), $-(P_0)$. Om vi nu slår ihop resultatet ser vi att varje punkt förekommer en gång med + och en gång med $-$. I den meningen gäller alltså att ∂ två gånger, ∂^2, ger noll.

Vi tar nu de trianglar $T_1, T_2, \ldots, T_t$, som hör till vår uppdelning av den givna ytan. Varje triangel T_i kopplar vi samman med ett heltal τ_i, en slags multiplicitet, som får variera fritt. Vi får då ett system k av par av trianglar och heltal

$$\left.\begin{matrix} T_1, T_2, \ldots, T_t \\ \tau_1, \tau_2, \ldots, \tau_t \end{matrix}\right\} k.$$

På samma sätt kan vi göra med sträckorna S_i och punkterna P_j i vår uppdelning. Vi får "randen" till k, ∂k, genom att ta randen till varje T_i, "multiplicera" med motsvarande τ_i och slå ihop sträckor från olika trianglar under hänsynstagande till tecken. Vi får då ett analogt system sträckor

$$\left.\begin{matrix} S_1, S_2, \ldots, S_s \\ \sigma_1, \sigma_2, \ldots, \sigma_s \end{matrix}\right\} l=\partial k.$$

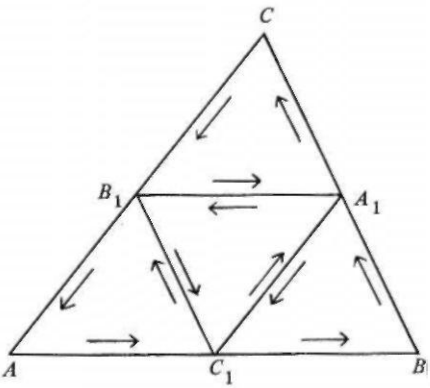


Fig. 29

Om vi slår samman två k, k^1+k^2, genom att addera motsvarande τ_i^1 med τ_i^2, gäller tydligen $\partial(k^1+k^2)=l^1+l^2$. Systemet k är en grupp under denna $+$-operation, liksom systemen l, och $\partial: k \to l$ är en grupphomomorfism. Eftersom ∂^2 ger noll gäller $\partial l = \partial^2 k = \{0\}$, dvs. gruppen som bara består av noll.

Vi kan se vad operationen betyder i följande enkla exempel (fig. 29).

$$\begin{aligned} l = \partial(T_1+T_2+T_3+T_4) &= (C_1A)+(B_1C_1)+(AB_1)+ \\ &+(BC_1)+(A_1B)+(C_1A_1)+(CA_1)+(B_1C)+(A_1B_1)- \\ &-(B_1C_1)-(C_1A_1)-(A_1B_1)= \\ &=(C_1A)+(BC_1)+(A_1B)+(CA_1)+(B_1C)+(AB_1) \end{aligned}$$

och

$$\partial l = (C_1)-(A)+(B)-(C_1)+(A_1)-(B)+\ldots=0.$$

Det kan nu finnas andra uppsättningar sträckor l sådana att $\partial l=0$ än de som uppstår genom ∂k. $\partial l=0$ anger att motsvarande sträckor bildar en sluten kontur, men denna behöver inte vara rand till något "område" k. I vårt exempel torusen är t.ex. sträckan $P_1P_4P_7P_1$ ett sådant l. De bildar en grupp L och vi kan bilda $L/\partial K=H_1$. Denna kallas nu homologigruppen av ordning 1. Vad som gör H_1 intressant är att den kan visas vara oberoende av hur vi delade in den givna ytan i trianglar. Som ett exempel kan nämnas att H_1 (torus) $= Z \times Z =$ mängden av heltal (n, m) under vektoraddition. Just detta att det ingår *två* grupper Z på detta sätt i H_1:s struktur, ger oss ett *tal*: 2, som hör till torusen. Detta betecknas p_1. Gruppen H_2 är helt enkelt den undergrupp av K för vilka $\partial k = 0$ och H_0 definieras analogt med H_1 utgående från sträckor. Vi får tre tal p_0, p_1, p_2 (för torusen $p_0 = 1$, $p_1 = 2$, $p_2 = 1$) och $p_0 - p_1 + p_2 = g$. Detta är en relation som gäller för varje yta när vi genomför motsvarande konstruktion.〉

Vad som här gjorts kan generaliseras på många sätt: till m dimensioner, till andra grupper än heltalen (som τ, σ, π), till allmänna topologiska rum. Man har även infört en axiomatik för hela teorin, varigenom det blivit en gren av algebran. Det är denna riktning som är den nu dominerande och som sannolikt på längre sikt kommer att starkt påverka hela det matematiska tänkandet.

Operatorn ∂ liknar på många sätt derivering. Den tar fram den marginella ändringen i triangeln, som ju är randen, på samma sätt som derivatan mäter funktionens ändring. Man kan då fråga sig om det inte här finns en omvänd operation som skulle motsvara integration. Denna kan mycket riktigt definieras: till en sträcka i vår uppdelning tillordnas de trianglar till vilka den är sida, naturligtvis kompletterat med regler om tecken. Vi får då en motsvarande teori, cohomologi, som befinner sig i snabb utveckling. Det är i denna teori som begreppet exakt följd spelar sin största roll.

Ett enda slående resultat i den algebraiska topologin skall nämnas: om $f(x)$ är en kontinuerlig avbildning av ett n-dimensionellt klot in i sig själv, finns det minst en punkt a så att $f(a) = a$. Om man således rör om i en burk med färg, så finns det alltså efter omröringen en punkt som ligger i samma läge som från början. — Detta kan nu generaliseras till funktionsrum och är då ett mycket användbart sätt att visa att invecklade relationer mellan funktioner har en lösning. Tag som ett exempel en avbildning som $f(t) \to c \cdot \int_0^1 K(t, x) \cdot f(x)^2 dx + 1/2$, där K är en kontinuerlig funktion. Vi studerar avbildningen för sådana $f(x)$ att $\int_0^1 f(x)^2 dx < 1$. Om då c är tillräckligt liten, gäller samma olikhet för bilden av f, vilket motsvarar att "klotet" bestående av alla f med $\int_0^1 f(x)^2dx<1$ avbildas på sig själv. Det följer då att det finns en lösning till $f(t)=c\int_0 K(t,x)f(x)^2dx+$ $+1/2$. Detta ger tydligen ett rent existensbevis; vi har inte angivit någon metod att verkligen konstruera lösningen, inte heller vet vi om det finns många lösningar. För tillämpningarna är därför resultatet otillfredsställande, men dess betydelse ligger i att säkerställa existens av lösningar med en enhetlig metod. När man så vet att de finns, kan man med säker bakgrund söka speciella konstruktionsmetoder.

Differentialgeometri

De avbildningar vi hittills studerat har varit allmänna kontinuerliga och kongruensbegreppet blir då det sedvanliga topologiska. Vi kan naturligtvis pålägga våra avbildningar starkare villkor, t.ex. att de skall vara kontinuerligt deriverbara. Man trodde länge att man då skulle få samma kongruensbegrepp. Om t.ex. två enkla kurvor i planet har kontinuerlig tangent kan man avbilda, dvs. deformera planet kontinuerligt deriverbart så att den ena kurvan överföres i den andra. Det var en sensation då Milnor 1956 visade att motsvarande resultat inte gäller på den 6-dimensionella sfären. Man fick härigenom en ny topologi, differential-topologi, som är i snabb utveckling. Den är speciellt betydelsefull på grund av det nära sammanhanget med differentialekvationer.

Vi har hittills sysslat med att beskriva en allmän geometrisk situation. Resultatet skall alltså vara detsamma efter det att vi utfört en avbildning eller med andra ord en deformation av situationen. Vi skall nu i stället försöka beskriva deformationen. Till denna klass av problem hör då t.ex. att beskriva kurvor och ytor, eftersom vi kan uppfatta dem som deformerade sträckor respektive plan.

Den enklaste avbildningen i planet är den lineära — affina — definierad av lineära koordinatsamband.

$$x'=ax+by$$
$$y'=cx+dy.$$

Denna kan skrivas symboliskt $z' = Az$, om z = vektorn (x, y) och z' vektorn (x', y') och A står som förkortning av de fyra talen (a, b, c, d). Om vi har två avbildningar A_1 och A_2 hörande till $(a_1, \ldots, d_1)$ resp. $(a_2, \ldots, d_2)$ kan vi på naturligt sätt bilda $A_1 + A_2$ som hörande till $(a_1 + a_2, \ldots, d_1 + d_2)$. $k \cdot A$ betyder att alla tal $a, \ldots, d$ multipliceras med k. $A_1 \cdot A_2$ slutligen betyder att vi först utför A_2 och sedan A_1; a-talet blir t.ex. $a_1a_2 + b_1c_2$. Symbolerna A bildar nu en algebra (se s. 63), där tydligen i regel A_1A_2 inte är detsamma som A_2A_1, och studiet av dessa *matris*algebror är utomordentligt betydelsefullt, speciellt vid numeriska tillämpningar som vi nu närmare skall förklara.

Om vi nu studerar en allmän avbildning $(x, y) \to (x', y')$ nära en viss punkt, som väljes till origo, kan vi som en första approximation finna en affin avbildning, som approximerar den givna på samma sätt som tangenten till en kurva approximerar kurvan. Vi vet att en av de viktigaste operationerna i analysen är att ur tangentens lutning (i fysikalisk tolkning t.ex. hastigheten) återvinna funktionen (läget). Detta är nämligen integrationen, som vi skall behandla i nästa kapitel. Det motsvarande problemet för avbildningar, dvs. att ur de lokala matriserna A återvinna avbildningen, är utomordentligt svårt och inte alls fullständigt utrett. Speciellt gäller detta för tre och flera variabler, där ju ett precis analogt problem finns.

Vid numeriska beräkningar i flera variabler gör man så att man följer den regel $z' = Az$ som konstruerats nära origo. Man finner att punkten z_1 (nära origo) då motsvarar z_1'. Man tar den regel A_1 som gäller i z_1, följer denna till z_2 osv. och hoppas att det vi på detta sätt konstruerar ligger nära den lösning vi söker. Om vi tänker oss att punkterna $z_1, z_2, \ldots$ ligger 10^{-6} från varandra ser vi att vi behöver 10^6 punkter för att nå fram till $(1, 0)$. Det behövs då bara att felet i varje beräkning i medeltal är större än 10^{-6} för att totalfelet skall kunna bli större än 1. Det krävs alltså bättre teoretiska uppskattningar och att finna dessa i detta och liknande sammanhang är en huvuduppgift för den numeriska matematiken, där många viktiga problem återstår att lösa.

Den klassiska matematiken beskrev kurvors och ytors krökningsförhållanden i tre dimensioner och skapade här en mycket fullständig och tilltalande teori. Man övergick därefter — efter traditionellt matematiskt mönster — att undersöka analoga problem i flera dimensioner än 3. Detta gjordes enbart efter rent matematiska linjer men den teori man skapade bildade grunden för den beskrivning av tid-rum-relationerna, som kallas relativitetsteorin. Man mäter här avstånd med en formel $x^2+y^2+z^2-t^2$ (som ju övergår i den vanliga $x^2+y^2+z^2+\tau^2$ om vi sätter $t=i\tau$! Se s. 40). Denna sorts "avstånd", som inte ens alltid är positivt, kan ju synas rätt osannolikt, men den betydelse det nu fått är ett exempel på värdet av den rena matematiska spekulationen. — Längs denna linje har matematisk forskning arbetat vidare med mycket allmänna rum och med mycket allmänna avståndsbegrepp. Även här har man kunnat axiomatisera stora delar av teorin.

Vi går tillbaka till formeln för affin avbildning på s. 84 Den är ju speciellt enkel att tolka om A kan tolkas som ett komplext tal och Az betyder $A \cdot z$. Då är $a=d$ och $b=-c$. Detta är de berömda Cauchy-Riemannska ekvationerna och avbildningen $z'=f(z)$ kallas analytisk. Det enklaste exemplet är polynom, men alla "vanliga" funktioner som e^z, $\sin z$, $\sqrt{z}$, är analytiska. Denna klass av avbildningar dyker upp överallt i matematiken. I talteorin studerar man t.ex. primtalsfördelningen med hjälp av den analytiska funktionen $\zeta(z)=1^{-z}+$ $+2^{-z}+\ldots+n^{-z}+\ldots$, som har mening om för $z=x+iy$ gäller $x>1$. Algebrans fundamentalsats, att varje polynomekvation har minst en rot, förstår man bäst i detta sammanhang; inom analysen är klassen allestädes närvarande. Orsaken är att funktioner, definierade för reella värden, ofta har naturlig utvidgning till det komplexa planet som i exemplen ovan.

Matematiker har försökt att generalisera begreppet analytisk funktion till flera variabler för att därmed få ett lika bra hjälpmedel att studera vanliga funktioner i flera reella variabler. Den definition vi valde kan naturligtvis inte generaliseras — det finns ju ingen motsvarighet till multiplikation av komplexa tal i flera variabler. Men vi skulle också ha kunnat få samma klass av funktioner genom att kräva att Az skall avbilda cirklar på cirklar; i flera dimensioner skulle AX, $X=(x$ [illegible] $x_n)$ avbilda sfärer på sfärer. Men denna generalisering går heller inte;

man kan visa att då måste avbildningen vara en stel vridning. Det är emellertid trots allt sannolikt att det finns ett annat sätt att se på denna fråga så att man kan få en intressant och givande generalisering av begreppet analytisk funktion till avbildningar mellan rymder av flera variabler, men denna återstår att upptäcka.

Stort intresse har man under senare år lagt ned på avbildningar $(z_1, z_2, \ldots, z_n) \to z$ från n komplexa variabler till en komplex variabel så att varje fix avbildning där bara en variabel z_j varierar $((z_1, a_2, \ldots, a_n) \to z$ till exempel) är analytisk. Hit hör t.ex. polynom i flera variabler, eftersom vi får vanliga polynom om vi håller alla variabler utom en konstanta. Teorin kräver en mycket utvecklad topologisk teori och det har här spelat stor roll att man algebraiserat topologin eftersom det är omöjligt att intuitivt föreställa sig hur "ytor" i det komplexa begränsar och skär varandra, när antalet variabler är större än 1, så att antalet dimensioner blir minst 4. Det finns anledning vänta en fortsatt utbyggnad och än större allmängiltighet i denna teori.

Vad som här framställts avviker på ett drastiskt sätt mot traditionell geometri. Topologin har en växande betydelse inom matematiken och kommer med stor säkerhet att starkt påverka andra grenar. På vilket sätt bör nu skolundervisningen influeras av detta?

Det är klart att i grundskoleundervisningen skall fortfarande triangelgeometri, ytberäkningar o.d. vara viktiga moment, som bör bedrivas med ett minimum av bevis. I de högre klasserna bör även trigonometri övas. Allt detta syftar till direkt användning av matematiken i andra sammanhang och bör undervisas med detta som mål och inte, såsom förr Euklides geometri som exempel på matematiska teorier. Utöver detta borde en undervisning inriktad mot topologin förekomma. Det är t.ex. lätt att experimentellt verifiera och också bevisa Eulers polyederformel (s. 79); man kan diskutera kontinuerliga avbildningar — som exempel på funktioner i något vidare mening än de traditionella reella funktionerna — man kan göra algebrans fundamentalsats trolig som vi gjorde på s. 72 osv. Detta skulle leda till en bättre förståelse av geometriska problem och skulle vara en värdefull information om vad matematiker kan göra.

Kommentarer

För en ganska elementär redogörelse för plan icke-Euklidisk geometri hänvisas till

O. Perron, *Nichteuklidische Elementargeometrie der Ebene*. Teubner 1962.

En intressant diskussion ur allmän synpunkt av den elementära geometrin finns i

E. E. Moise, *Elementary geometry from an advanced standpoint*. Addison-Wesley 1963.

Den klassiska källan för användningen av grupp-betraktelser i geometrin är F. Kleins arbeten. Se

F. Klein, *Elementarmathematik vom höheren Standpunkt aus II Geometrie*. Berlin 1925.

Denna innehåller också en diskussion av skolundervisningen i geometri.

En allmän utmärkt översikt av topologin, som ändå inte är för svår, är

J. G. Hocking—G. S. Young, *Topology*. Addison-Wesley 1961.

Milnors resultat publicerades 1956 i *Annals of Mathematics 64*. Han erhöll ett av Fieldprisen vid Stockholmskongressen 1962.

För numerisk användning av matriser se

V. N. Faddeeva, *Computational methods of linear algebra*. Dover 1959.

En allmän inledning i numerisk analys är

E. L. Stiefel, *An introduction to numerical mathematics*. Academic Press 1963.

Den som vill studera klassisk differentialgeometri, dvs. teori för ytor och kurvor i rymden, bör lämpligen läsa någon av de klassiska franska översikterna i analys, t.ex.

Ch. de la Vallée Poussin, *Cours d'analyse*. Omtryckt Dover 1946.

En bra inledning till teorin för analytiska funktioner är

K. Knopp, *Elements of the theory of functions*. I, II. Dover 1947, 52.

H. Cartan, *Elementary theory of analytic functions of one or several complex variables*. Addison-Wesley, 1963, behandlar också flera komplexa variabler.

7. Analys

Praktiskt taget hela gymnasiekursen i matematik faller inom området analys och där inom vad som brukar kallas det klassiska området. Det som man lär ut är forskningsresultat som initierades på 1600-talet. Givetvis har dessa klassiska områden utvecklats under de gångna 300 åren och den tankeräfst som genomfördes av 1800-talets matematiker har inneburit att begrepp och bevis blivit klara och precisa. I dagens matematiska forskning spelar området emellertid en allt mindre roll, vilket står i stark motsättning till dess betydelse i tillämpningar. Även här finns ett viktigt problem som vi senare skall diskutera.

På grund av analysens betydelse för undervisningen skall vi här rätt utförligt diskutera de centrala momenten i skolkurserna och se hur den senare forskningen belyser dessa.

Det viktigaste grundbegreppet är funktionen, dvs. den matematiska termen för orsaksidén, att kännedom om vissa förhållanden bestämmer vissa andra. I skolan står termen nästan uteslutande för reella funktioner av reella variabler och i fråga om generalitet brukar man inskränka sig till att låta dessa funktioner vara definierade endast på delmängder av reella axeln, såsom t.ex. $x \to 1/x$, $x \neq 0$. Detta ger emellertid en helt missvisande bild av vad som täcks av begreppet. Vi har tidigare här sett exempel som dessa:

$A(x)$ – antalet sätt att skriva x som en summa av två primtal; här är heltalen det naturliga definitionsområdet;

gruppmultiplikation, som alltså är en funktion definierad på gruppen med värden i gruppen;

allmänna homomorfier: funktioner från en algebraisk struktur till annan;

i analytisk geometri, där t.ex. tre tal (a, b, c) bestämmer en linje, alltså en funktion från R^3 till alla linjer i ett plan;

i topologin, t.ex. avbildningar från reella tal till kurvor (s. 73).

Det är utan tvekan mycket lättare att förstå betydelsen av termer som definitionsområde och värdeförråd om begreppen presenteras i allmänna sammanhang som i de nämnda exemplen, givetvis tagna ur områden som eleverna tidigare är förtrogna med, än om man gör konstlade skarvningar eller inskränkningar av funktioner med "naturliga" definitionsområden. En presentation av detta slag har också fördelen att förmedla en modern uppfattning. Ur tillämpningarnas synpunkt kan man även ifrågasätta den exklusiva inskränkningen till funktioner av en variabel. Det är dock relativt få fenomen i den fysiska världen där man på ett ändamålsenligt sätt kan uttrycka beroendet genom en enda variabel. För att ange ett läge fordras ju t.ex. tre tal och med en samtidig tidsangivelse blir det fyra.

Nästa steg är nu att successivt inskränka klassen av reella funktioner av en reell variabel genom begrepp som kontinuitet och deriverbarhet. Detta görs numera med ganska stor ambition på stringens och med precisa definitioner av typen: $f(x)$ är kontinuerlig i $x=a$ om till varje $\varepsilon>0$ det existerar ett tal $\delta(\varepsilon)>0$ så att så snart $|x-a|<\delta(\varepsilon)$ det följer att $|f(x)-f(a)|<\varepsilon$. [Denna definition är för övrigt inte bra eftersom den förutsätter att $f(x)$ är definierad i en hel omgivning av a.] För den som inte är väl förtrogen med dessa ting tidigare måste denna definition låta som kinesiska och det *är* en sorts kinesiska. Det är de matematiska symbolernas sätt att uttrycka att funktionskurvan asymptotiskt hänger ihop i punkten a. Kontinuitet på ett intervall är intuitivt ännu enklare men logiskt betydligt svårare. Det som gör detta till ett dilemma i undervisningen är att det inte går att visa eleverna några exempel där inte kvasidefinitionen "kurvan är sammanhängande" är klart överlägsen.

Man erkänner behovet av en precis definition först då man mot bakgrunden av sitt erfarenhetsmaterial inser att saken är så komplicerad att missförstånd eller fel kan uppstå utan precisa begrepp. Historiskt tog det matematiken hela skedet fram till 1800-talet innan man nådde detta stadium. Det förefaller ju oss numera som ett mysterium att Newton kunde beräkna planetrörelser utan att veta att en derivata är ett gränsvärde och Eulers manipulerande under 1700-talet med relationer som $1-1+1-\ldots=1/2$ eller ännu värre $1-1+1\cdot2-1\cdot2\cdot3+$ $+1\cdot2\cdot3\cdot4-+\ldots=0{,}5963\ldots$ visar att om idéer är riktiga, så kan ett geni arbeta långt utan precisa begrepp. I fråga om kontinuitetsbegreppet torde inte många matematiker ha någon användning för definitionen särskilt ofta. Under sina studier och i sitt arbete har matematikern fått en sådan vana att begreppets funktion blivit självklar. Det väsentliga är alltså erfarenheten: att ha studerat och använt kontinuerliga och diskontinuerliga funktioner. Det är dessa exempel som ger begreppet innehåll, inte abstrakta definitioner. Teori utan verklighetsinnehåll är meningslös. Det finns också en klar tendens i dagens undervisning att underskatta svårigheterna att förstå det matematiska symbolspråket. Även här är erfarenhet och övning enda lösningen, men den är så tidskrävande att det är omöjligt att ge tillräckligt utrymme i skolundervisningen. Det är därför nödvändigt att lägga huvudvikten på idéerna och inte på formen.

Det är klart att matematiken kräver så exakta definitioner som möjligt och så få som möjligt. Man vill sedan härleda satser: i det aktuella fallet t.ex. att kontinuitet i varje punkt på ett slutet intervall medför att funktionen är begränsad. För nybörjaren (och han har många sympatisörer, t.o.m. bland nobelpristagare i fysik) verkar detta sofism. Man skulle lika gärna kunna *anta* att funktionerna hade alla de egenskaper man behöver, eftersom när det kommer till en konkret tillämpning är allt klart och tydligt. Dessutom vet man ju inte att man har minimala antaganden i alla fall, så varför inte lägga till några extra. Och till slut: hur många lärare kan ärligt säga att de själva behärskar begreppsbildningar och bevis av den typ vi här diskuterar?

Enligt detta sätt att se skulle man alltså inskränka på definitionskineseriet och öka erfarenhetsbakgrunden i undervisningen i funktionslära därför att detta är det naturliga inlärningssättet och enda möjligheten att förstå sammanhangen. Detta betyder inte att man skall återgå till den lösliga framställning som var

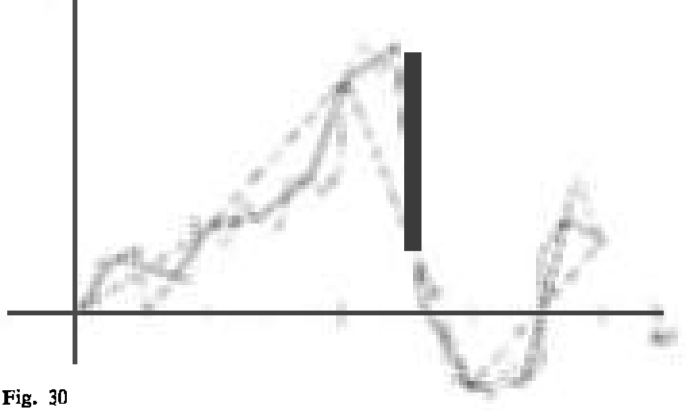


Fig. 30

den vanliga i skolan förr i världen, utan innebär att man på ett realistiskt sätt skall bedöma vilken undervisningsmetod som ger den bästa förståelsen för ämnet. Det vore också ett stort misstag att tro att matematikerna bara behöver syssla med sådana funktioner att man kan hänvisa till intuitionen i alla lägen eller att man inte skulle behöva (ε, δ)-definitioner. Vi skall senare få många exempel på det första förhållandet; här skall vi ta upp ytterligare ett med direkt kontakt med tillämpningarna.

⟨ Låt P vara en partikel som rör sig i ett plan, där vi infört ett koordinatsystem (fig. 30). I x-riktningen låter vi P ha konstant hastighet, medan rörelsen i y-led regleras av slumpen. Lagen för denna y-rörelse är sådan att y under ett tidsintervall av längd t rör sig en sträcka $< \lambda \cdot \sqrt{t}$ med en sannolikhet som bara beror på λ. Väsentligen av denna typ är den s.k. Brownska rörelsen som beskriver vissa fenomen i atomfysiken. Vi får alltså funktioner $y = f(x)$ som beskriver rörelsen. Vad blir det för sorts funktioner? Man kan nu bevisa att i normalfallet är detta en kontinuerlig funktion, men funktionen $f(x)$ är inte deriverbar någonstans. Den väg partikeln P rör sig på hur litet tidsintervall som helst har oändlig längd. ⟩ Vi har alltså det paradoxala förhållandet att fysiker arbetar med modeller där partiklar anses röra sig längs oändligt långa bankurvor utan bestämd riktning vid någon tidpunkt. Inte desto mindre visar exemplet att naturvetenskapen verkligen använder konstiga funktioner och därför behöver precisa begrepp. Det för tilllämpningarna viktigaste skälet att i sina kalkyler även ta hänsyn till funktioner med mycket oregelbundet uppträdande är emellertid, precis som vi förut framhållit, att dessa kommer in i den matematiska teorin helt naturligt och därmed även i numeriska beräkningar.

I matematiken är numera kontinuitetsbegreppet helt centralt och kommer in i alla möjliga sammanhang. Det är därför en självklarhet att man vill analysera det, se efter vilken relation det har till andra närbesläktade begrepp, och att man vill göra detta med så stor allmängiltighet som möjligt. Dessa frågor behandlade vi på s. 71 och har flera gånger återkommit till dem. På det hela taget kan man emellertid säga att någon egentlig forskning rörande reella funktioner av en reell variabel och existens och egenskaper av deras derivator, numera inte bedrives. Under början av 1900-talet var dessa frågeställningar aktuella och fick då lösningar som på det hela taget varit slutgiltiga (se nedan s. 97).

Den naturliga generaliseringen av deriveringsidén till flera variabler är införandet av den partiella derivatan, där man deriverar med avseende på en variabel i taget och låter övriga variabler vara konstanta. Beteckningen är ett krokigt ∂, $\partial f/\partial x$, $\partial f/\partial y$; så blir t.ex. $(\partial/\partial x)(x^2+xy)=2x+y$ och $(\partial/\partial y)(x^2+xy)=$ $=x$. Vi kan givetvis utföra sådana här derivationer efter varandra och man kan visa att det inte spelar någon roll i vilken ordning man utför deriveringen. Vi får då beteckningar som $\partial^2 f/\partial y^2$, $\partial^3 f/\partial x^2 \partial y$ för två y-derivationer resp. två x- och en y-derivation. — Dessa begrepp spelar samma roll vid studiet av funktioner av flera variabler som vanliga derivator vid en variabel och det torde vara ett starkt önskemål ur olika tilllämpade aspekter att skolan sysslade mer med detta, kanske på bekostnad av övningar i konstiga maxima och minima.

Integrationsteori ingår numera i gymnasiekursen, vilket är ett viktigt framsteg. Även här förefaller målsättningen emellertid på tok för ambitiös när det gäller stringensen i definitio-

Fig. 31

nerna. Det är stor risk att denna går ut över förståelsen av sammanhangen, fastän givetvis avsikten är den motsatta.

I integrationsteorin finns ett av de få tillfällen där det matematiska beteckningssystemet är opedagogiskt och vilseledande. Integration i klassisk mening täcker två helt olika problem:

(1) finn en funktion $F(x)$ som har en given funktion $f(x)$ till derivata; en sådan betecknas $\int f(x)dx$;

(2) bestäm arean $A(b, a)$ som begränsas av $y=0$, $y=f(x)$, $x=a$ och $x=b$. (Vi antar $b>a$ och att $f(x)$ är positiv.) Man betecknar $A(b, a)=\int_a^b f(x)dx$.

Sambandet mellan (1) och (2) är nu följande. Man frågar sig naturligtvis om det finns **någon** sådan funktion $F(x)$ som önskas i (1). Svaret är ja — om $f(x)$ är en "rimlig" funktion — ty man visar att $A(x, a)$ är *en* sådan funktion. Nästa steg är att visa att varje $F(x)$ erhålles ur $A(x, a)$ genom addition av konstant. Det betyder att $A(x, a)$ är den enda $F(x)$ som har egenskapen att vara noll då $x=a$. Detta slutligen leder till följande praktiska metod att bestämma areor. Gissa eller konstruera med vilken metod som helst en funktion $F(x)$ som i (1). Lägg till en sådan konstant C att $F(a)+C=0$, dvs. $C=-F(a)$. Då är $F(x)-$ $-F(a)=A(x, a)=\int_a^x f(x)dx$. Först efter denna formel har vi knutit samman de två symbolerna $\int$ och $\int_a^b$. Speciellt olyckligt är att symbolen $\int$ inte står för en bestämd funktion utan för vilken funktion $F(x)$ som helst. Det är alltså inte ett väldefinierat begrepp. Detta är troligen ett unikt fall av begreppsförvil-

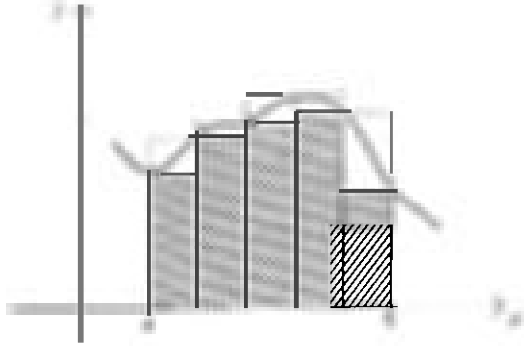
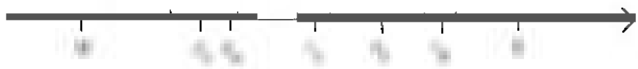


Fig. 32

lelse i matematiken och den enklaste lösningen är att man helt utmönstrar symbolen $\int$.

Om vi nu i undervisningen godtar den intuitiva tolkningen av integralen som en area får vi vissa problem med t.ex. räknelagen $\int^x(f+g)dx = \int^x f\,dx + \int^x g\,dx$. Denna visas enklast genom att man konstaterar att båda leden har samma derivata, och restriktionen till positiva funktioner elimineras genom förskjutning med konstant. Den fundamentala idén, nämligen sambandet mellan (1) och (2) kvarstår givetvis oförändrad. Det som tonas bort är trasslet kring summapolygonerna, som kunde omnämnas som ett sätt att approximera arean och att definiera den exakt ur rektangelytor (fig. 32). Härigenom får man tillbaka summornas fördel t.ex. vid volymsberäkningar.

Ut matematikens synpunkt är emellertid inskränkningen till "intuitiva" funktioner ytterligt otillfredsställande. Dessa visar samma brist vid behandling av problem rörande funktioner som de rationella talen visar för ekvationer. När vi sysslar med integration och vill uttrycka att en följd funktioner konvergerar är det inte funktionernas uppträdande i enstaka punkter vi är intresserade av utan deras allmänna uppträdande. I bilden på s. 31 var det rimligt att anse att följden f_n konvergerar mot noll fastän det alltid finns punkter där funktionerna är 1. I detta exempel var det en punkt som avvek. Tydligen kan vi göra ett likadant exempel med två punkter eller med 10^{10} punkter. Man frågar då: på vilka mängder får vi ändra funktionen utan att det bör påverka integralen? Detta var startpunkten för den lösning integrationsproblemet fick i början av 1900-talet och som på det hela taget varit slutgiltig. Eftersom arean = längd × höjd, och vi inget vet om höjden, är villkoret för att arean inte skall ändra sig rimligen att ”längden av mängden” är noll. Vad menas då med det? En lämplig definition är att man skall kunna täcka mängden med intervall, kanske oändligt många, vilkas totallängd kan göras hur liten som helst. Som ett exempel kan vi ta mängden av rationella tal på (0, 1).



Vi kan ordna dem efter nämnarens storlek och räkna upp dem $r_1, r_2, \ldots, r_n, \ldots$ Vi vill nu täcka dem med intervall vilkas totallängd är liten, kanske högst 10^{-10}. Lägg då kring r_1 ett intervall av längd $1/2 \cdot 10^{-10}$, kring r_2 ett av längd $1/4 \cdot 10^{-10}$ osv. Då får alltså intervallen totalt längden

$$\frac{1}{2}10^{-10}+\frac{1}{4}10^{-10}+\frac{1}{8}10^{-10}+\ldots=\left(\frac{1}{2}+\frac{1}{4}+\frac{1}{8}+\ldots\right)\cdot 10^{-10}=10^{-10},$$

vilket var vad vi önskade.

Om nu en följd ”intuitiva” funktioner $f_n(x)$ är begränsade och konvergerar mot en begränsad funktion $f(x)$ utanför en mängd vars längd är noll, kan man bevisa att integralerna $\int_a^b f_n(x)dx$ också konvergerar mot ett bestämt tal. Detta tal kallar vi nu integralen för $f(x)$ över intervallet (a, b). Som ett exempel kan vi ta den funktion $f(x)$ som $=1$ för rationella x och $=0$ annars. Vi kan välja alla $f_n(x)=0$ överallt och $f(x)$ har alltså samma integral som den funktion som är identiskt $=0$, dvs. 0.

Man kan mycket slående illustrera hur denna utvidgning av de ”intuitiva” funktionerna med starkt oregelbundna funktioner liknar utvidgningen från de ”intuitiva” talen — de rationella — till de reella. Vi väljer emellertid hellre jämförelsen mellan vektorer x i planet, $x=(x_1, x_2)$ och funktioner $f(t)$ på

$0<t<1$. Vi ser först hur vi kan flytta över termer och begrepp från planet till funktioner:

Längd	$\sqrt{x^2_1+x^2_2}$	[illegible]
Avstånd	[illegible]	$\sqrt{\int_0^1 (f(t)-g(t))^2\,dt}$
Rät linje	Alla x med $(a,x)=$	Alla f med $(a,f)=$
eller plan	$=a_1x_1+a_2x_2=c$	$=\int_0^1 a(t)f(t)dt=C$
Normalform	$a^2_1+a^2_2=1$	$\int_0^1 a(t)^2dt=1$

⟨Vi ser här att vi kan i två dimensioner koppla ihop varje plan (= linje) genom origo med punkter (a_1, a_2) som är bestämda sånär som på en faktor k, (ka_1, ka_2). Samma är förhållandet i tre dimensioner där planen har formen $a_1x_1+a_2x_2+$ $+a_3x_3=0$. Vi kan också tolka uttrycken (a,x) som homomorfier från vektorrummen i två och tre dimensioner till de reella talen så att alltså summor avbildas i summor. Vi har exakt samma situation för funktioner. $(a,f)=\int_0^1 a(t)f(t)dt$ är en kontinuerlig homomorfi från vektorrummet av funktioner till de reella talen och man kan visa att varje sådan homomorfi har denna form. Helt allmänt kan man studera kontinuerliga homomorfier från ett vektorrum V, där det finns en topologi till de reella eller komplexa talen; $h\colon V\to R$. Dessa homomorfier bildar själva på naturligt sätt ett vektorrum V^* och de är av central betydelse i analysen. I samtliga våra tre exempel ser vi att V^* kan identifieras med V själv: i två dimensioner fick vi paret (a_1, a_2), i tre (a_1, a_2, a_3) och i funktionsfallet en funktion $a(t)$ med villkor på integralen av $a(t)^2$.⟩

En ytterligare analogi kan vi se i vår tabell. Det avstånd som vi infört uppfyller triangelolikheten

$$\text{Avst. }(f,g)<\text{Avst. }(f,h)+\text{Avst. }(h,g),$$

vilket kan illustreras av en symbolisk figur (fig. 33).

Tag nu en följd vektorer $x^1, x^2, x^3, \ldots, x^n, \ldots$ som närmar sig varandra obegränsat när index n (ej exponent, naturligtvis) växer. Uttryckt med matematiska symboler innebär detta att Avst. (x^n, x^m) närmar sig noll när de två indexen n och m växer

Fig. 33

obegränsat (fig. 34). Då finns det en vektor x som de närmar sig. Detta skulle inte gälla om vi bara sysslade med vektorer med rationella koordinater. Tag t.ex. vektorer längs en koordinataxel och välj koordinaterna som avsnitt i decimalbråksutvecklingen av $\sqrt{2}$:

$$x^1 = 1,\ x^2 = 1{,}4,\ x^3 = 1{,}41,\ x^4 = 1{,}414,\ \ldots$$

Denna följd har den egenskap vi fordrade, nämligen att punkterna närmar sig varandra obegränsat, men de närmar sig inte något *rationellt* tal, utan ett irrationellt, nämligen $\sqrt{2}$. Precis det analoga gäller nu för våra funktioner $f(x)$: för att vi i motsvarande situation skall veta att funktionerna närmar sig någonting, måste vi tillåta de oregelbundna funktionerna. Vi skall senare illustrera betydelsen av detta på sidan 105.

I fysiken har man sedan länge arbetat med ett vagt funktionsbegrepp, som är mycket allmännare än det matematiska. Det mest kända exemplet har namn av Dirac och betecknas $\delta(x)$. Denna "funktion" har egenskapen att

Fig. 34

$$\int_{-1}^{+1} \delta(x)\varphi(x)dx = \varphi(0)$$

för alla ”intuitiva” funktioner φ. Man skall tänka sig detta så att $\delta(x) = 0$ utom i $x = 0$ där den är så ”stor” att ”ytan” mellan $\delta(x)$ och x-axeln är $= 1$. Detta är naturligtvis orimligt men inte desto mindre har fysiken framgångsrikt räknat på med $\delta(x)$ och t.o.m. deriverat den! Hur hänger detta ihop?

Om vi vill fysikaliskt bestämma en funktion $f(x)$ på $(-1, 1)$ gör vi experiment som mäter $f(x)$:s uppträdande med avseende på experimentet. Vi får ett tal som ofta har formen

$$T(\varphi) = \int_{-1}^{+1} f(x)\varphi(x)dx,$$

där alltså $\varphi(x)$ är en ”intuitiv” funktion, bestämd av experimentets uppläggning. Som ett exempel betyder $\int_{-1}^{+1} f(x)x\,dx$ att man beräknar en tyngdpunkt och här är $\varphi(x) = x$; för $\varphi(x) = x^2$ mäter man ett moment. Om vi utför ”alla” experiment beräknar vi ”alla” tal $T(\varphi)$ och det är lätt att se att inte två olika funktioner f kan ge samma $T(\varphi)$ för alla sådana testfunktioner φ. Vi kan alltså identifiera f med integraloperationen T.

Nu vänder vi på synsättet. En operation T, t.ex. definierad för alla funktioner φ som kan deriveras hur många gånger som helst och är $= 0$ nära randen, och med vissa egenskaper som liknar integralens ($T(f+g) = T(f) + T(g)$ och $T(f_n) \to 0$ om $f_n \to 0$ med alla derivator) kallar vi en generaliserad funktion. Vi ser att villkoren är just att T skall vara en homomorfi i den topologi som hör ihop med klassen D av obegränsat deriverbara funktioner; $T: D \to R$. Av denna typ finns många andra operationer än vanliga integraler. Så är t.ex. Diracs ”funktion” δ operationen $\varphi \to \varphi(0)$. I och med denna teori har det fysikaliska betraktelsesättet fått en bestämd innebörd och logisk förankring. Betydelsen av dessa generaliserade funktioner sträcker sig emellertid vida längre, genom att de ger innebörd åt räkningar som tidigare var enbart formella. Till exempel har varje funktion en derivata i denna mening. Derivatan till

$$T(\varphi) = \int_{-1}^{+1} f(x)\varphi(x)dx,$$

som vi ju identifierat med funktionen $f(x)$, är då

$$T'(\varphi) = -\int_{-1}^{+1} f(x)\varphi'(x)dx.$$

För funktioner $f(x)$ som har en riktig derivata $f'(x)$ ger detta rätt resultat, ty

$$\int_{-1}^{+1} f'(x)\varphi(x)dx - T'(\varphi) = \int_{-1}^{+1} (f(x)\varphi'(x) + f'(x)\varphi(x))dx =$$

$$= \int_{-1}^{+1} \frac{d}{dx}(f\varphi)dx = f(1)\varphi(1) - f(-1)\varphi(-1) = 0,$$

eftersom $\varphi(1) = \varphi(-1) = 0$.

Svåra, eller som i exemplet beträffande derivatan eljest omöjliga problem har ofta lösningar i form av generaliserade funktioner. Då man använder de generaliserade funktionerna blir emellertid huvudproblemet i regel inte löst; det blir i stället överfört till att visa att den generaliserade funktion man konstruerat i själva verket är en vanlig funktion.⟩

Den teknik vi använt för att definiera integral av en reell variabel går mycket bra att använda i två variabler (då tolkas integralen som volym i stället för area) och även i tre och flera variabler. Hela proceduren är emellertid oerhört mycket mera allmän och det går egentligen att integrera vilka funktioner som helst (med viss måtta) på vilka rum som helst. När vi i nästa kapitel sysslar med statistik skall vi se exempel på detta.

Derivation är en mycket mera komplicerad operation än integration. Man inser det genast ur sådana exempel som det vi nämnde på s. 93. Det kan vara av intresse att explicit ha sett en funktion som är kontinuerlig utan att vara deriverbar någonstans. Här är en:

$$f(x) = \frac{\sin 3x}{2} + \frac{\sin 9x}{4} + \frac{\sin 27x}{8} + \ldots$$

där lagen är att i sinus förekommer potenser av 3 och i nämnaren potenser av 2. (Serien konvergerar tydligen eftersom termerna är mindre än den geometriska seriens termer 1/2, 1/4, 1/8, ...) Orsaken till de större svårigheterna beträffande derivatan är att man vill uttala sig om funktionens uppträdande i enstaka punkter medan integrationen tar hänsyn till funktionens uppträdande i stort. Det är just detta man kommer förbi med generaliserade funktioner: om f motsvarar $T(\varphi) = \int f(x)\varphi(x)dx$ så motsvaras den eventuella derivatan, vare sig den existerar eller ej, av $-\int f(x)\varphi'(x)dx$. Men naturligtvis är detta en kringgående manöver som hjälper oss mycket litet om vi vill undersöka $f(x)$:s maxima och minima till exempel.

Frågan om existens av vanliga derivator för en funktion av en variabel är mycket ingående studerad. Vi kan här nämna att man tidigt bevisade att en växande funktion har en derivata i alla punkter utanför en mängd vars längd är noll. Exempel visar att just denna sorts mängd nödvändigtvis måste uteslutas. Däremot är vår kunskap fortfarande inte särskilt omfattande för funktioner av två och flera variabler och när det gäller den derivation som sammanhänger med de mycket allmänna integraler vi nyss nämnde, vet man mycket litet.

Varför är nu analysen så betydelsefull för tillämpningar? Orsaken kan rätt generellt sägas vara att våra modeller för naturfenomen beskrives av differentialekvationer. Vi skall ta ett antal exempel som belyser detta.

Om en partikel rör sig längs en rät linje, x-axeln, kan vi ange läget som en funktion $x(t)$ av tiden t. Lägesändringen per tidsenhet, dvs. hastigheten, motsvarar derivatan $x'(t)$ (om den existerar!) och hastighetsändringen, accelerationen $x''(t)$. Om vi utsätter partikeln för en kraft k, så gäller (massan)$\cdot x''=k$. Om nu k bara beror på t kan vi integrera och får (vi sätter massan $=1$)

$$x(t)=\int_0^t \int_0^s k(u)du\, ds + x'(0)t + x(0),$$

dvs. om vi vet kraften i varje ögonblick är det lätt att räkna ut läget. (Detta är principen för tröghetsnavigering som används vid raketskjutningar. Man beräknar hela tiden accelerationen och raketen kan då själv hålla reda på var den är och själv korrigera kursen. Den blir därmed även helt okänslig för störningar i kommunikationerna med basen. De tekniska svårigheterna är att hålla något koordinatsystem fixt; för detta behöver man mycket goda gyrokonstruktioner.) Om emellertid k också beror på läget, om t.ex. partikeln är fäst med ett elastiskt band, får vi en ekvation $x''=k(x, t)$, där alltså det sökta x ingår på båda sidor. Detta är ett exempel på en differentialekvation och det är mycket komplicerat att ange om det finns lösningar eller ej — det beror i hög grad på k:s egenskaper — och problem av denna typ sysselsätter fortfarande många matematiker.

Fenomen som beror på flera variabler beskrivs i regel av ekvationer innehållande partiella derivator, partiella differentialekvationer. Situationen är där oerhört mycket mera komplicerad. ⟨ Tar man som ett exempel en vätska som befinner sig i strömning och där situationen inte ändrar sig med tiden, kan man beskriva vätskans tillstånd med en funktion u som beror av tre rumskoordinater (x, y, z). Vätskans hastighet i de tre axlarnas riktning anges av derivatorna $\partial u/\partial x$, $\partial u/\partial y$, $\partial u/\partial z$. Funktionen u uppfyller följande ekvation som anger att lika mycket vätska strömmar in i en viss del av rummet som det strömmar ut:

$$(E) \qquad \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} = 0.$$

Fastän denna ekvation studerats under mer än 100 år skrivs det fortfarande årligen många avhandlingar och nya fakta kommer i dagen. Som ett något kuriosabetonat exempel kan nämnas att man inte vet om det finns någon strömning så att vätskan är stilla i alla punkter i en mängd av positiv yta på kärlets rand. Motsvarande problem för en plan strömning — eller i en oändligt hög rak tank utan strömning i höjdled — kan man lösa med hjälp av analytiska funktioner (det finns ingen sådan vätskerörelse!) och detta är ett exempel på hur dessa funktioner är hjälpmedel i analysen. ⟩

Om vi byter ett tecken i ekvationen (E), t.ex. framför $\partial^2 u/\partial z^2$, får vi ekvationen

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} - \frac{\partial^2 u}{\partial z^2} = 0$$

vilken beskriver vågutbredning. Man kan lägga märke till att för en godtycklig funktion $F(t)$ är $F(x+y \pm \sqrt{2}z)$ en lösning. Det finns alltså lösningar, som inte kan deriveras hur som helst (eftersom $F(t)$ inte behöver ha dessa egenskaper), men för ekvationen (E) kan man visa att alla lösningar är obegränsat många gånger deriverbara. Vi ser alltså att skenbart obetydliga ändringar i ekvationens utseende radikalt ändrar ekvationens karaktär. I ännu högre grad gäller detta om vi tar med derivator av högre ordning än två, sysslar med system av ekvationer eller låter den obekanta funktionen *u* ingå på ett mera komplicerat sätt. Dessa komplicerade typer av ekvationer är långt ifrån slutstuderade. Det finns många viktiga ekvationer i fysiken där en matematisk behandling fortfarande saknas och, vad värre är, där våra nuvarande metoder verkar klart otillräckliga. Som ett aktuellt exempel kan man nämna de ekvationer som beskriver — eller rättare sagt, anses beskriva — vad som händer när man med magnetfält försöker kontrollera en miljontals grader varm laddad gas. Bakom studiet av detta ligger önskan att tillgodose mänsklighetens energibehov med vätekraft. En riktig matematisk behandling skulle här spela en stor roll, men man måste också vara klar över att det kan finnas två andra skäl varför denna behandling ännu saknas. Det finns nämligen en tumregel i en situation som denna. En komplicerad ekvation tagen på måfå kan inte ha en intressant matematisk teori; en ekvation som beskriver ett intressant fysikaliskt fenomen, är alltid matematiskt intressant. Den första möjligheten är nu att de uppställda ekvationerna kanske inte på ett riktigt sätt beskriver den mycket komplicerade och icke-traditionella fysikaliska situationen. Den andra, fullt tänkbara och dystrare möjligheten är att problemet inte är fysikaliskt intressant, dvs. att det principiellt inte går att kontrollera vätebombreaktionen.

Med en viss överdrift kan man säga att hela den matematiska analysen vuxit fram under strävandena att behandla fysikens differentialekvationer. Detta har skett under ständig växelverkan mellan matematiken och naturvetenskapen, så att alltså naturvetenskapsmännen utnyttjat de framsteg som gjorts i matematiken att beskriva nya fenomen. Det är utomordentligt betydelsefullt att möjligheterna till denna växelverkan bevaras och att förbindelsen mellan ren och tillämpad matematik hålls öppen.

〈 Vi skall konkretisera vad vi hittills sagt med ett speciellt exempel som innehåller många av de viktigare idéerna i hela tankebyggnaden. Det är inte nödvändigt för fortsättningen men utan handfasta illustrationer blir allmänna resonemang gärna innehållslösa.

Alltså: vi har givet en funktion $p(t)$ på $0 < t \leq 1$. Vi antar att $p(t)$ är mycket regulär — t.ex. ett polynom — och positiv. Vi studerar funktioner $f(t)$ så att $f(0) = 0$ och $f(1) = 1$ och vill hitta den eller de funktioner som gör uttrycket $I(f)$

$$I(f) = \int_0^1 [f'(t)^2 + p(t) f(t)^2] dt$$

så litet som möjligt. [Försök visa att om $p = 0$ är $f(t) = t$ enda lösningen.] $I(f)$ har fysikalisk tolkning som energi och problemet gäller alltså att hitta en väg som är så litet energikrävande som möjligt.

Matematiskt är detta ett variationsproblem och behandlingen sönderfaller i följande delar.

1. Finns det *någon* funktion som ger minimum?

I vilket fall som helst finns det en följd $f_n(t)$ så att $I(f_n)$ närmar sig minimivärdet M. Vi bildar nu för två olika index n och m $I(f_n) + I(f_m)$ och skriver om detta:

$$I(f_n) + I(f_m) = 2 \int_0^1 \left[\left(\frac{f_n + f'_m}{2} \right)^2 + p(t) \left(\frac{f_n + f_m}{2} \right)^2 \right] dt +$$

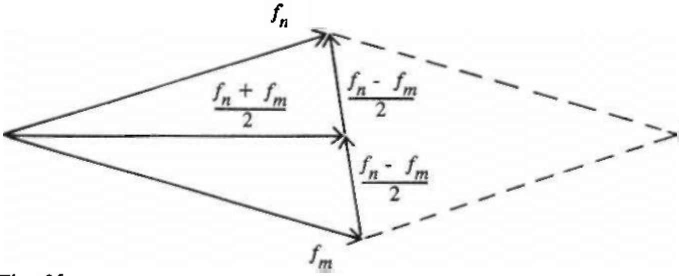


Fig. 36

Att denna relation gäller beror på att de dubbla produkterna tar ut varandra i högerledet. Den första integralen är $I((f_n + f_m)/2)$ och den andra $I((f_n - f_m)/2)$ så relationen kan skrivas

[illegible]

Om vi tolkar $I(f)$ som en kvadrerad längd (jfr tabellen på s. 98) utsäger relationen att summan av kvadraterna av en parallellograms sidor är lika med summan av diagonalernas kvadrater: Vi ser i figuren att när sidorna f_n och f_m blir nästan lika långa som halva diagonalen $(1/2)(f_n + f_m)$ blir längden av andra diagonalen liten. I vår situation fungerar detta på följande sätt. $(1/2)(f_n + f_m)$ antar också de rätta värdena i $t = 0$ och $t = 1$. Eftersom M var minimivärdet gäller $2\,I((f_n + f_m)/2) \geq 2\,M$. Vänsterledet närmar sig $2\,M$ när n och m växer. Alltså närmar sig $I((1/2)(f_n - f_m))$ noll, dvs.

$$\int_0^1 (f_n - f_m)^2 dt \to 0 \quad \text{och} \quad \int_0^1 p(t)(f_n - f_m)^2 dt \to 0.$$

Om vi tolkar dessa integraler som (avstånd)^2 innebär detta att avstånden mellan funktionerna närmar sig noll då index växer. Det allmänna integralbegreppet (se s. 99) garanterar nu att det finns en funktion $g(t)$ som funktionerna närmar sig. För $g(t)$ gäller verkligen $I(g) = M$, som man ganska lätt inser. Vi observerar att det är helt omöjligt på detta stadium att veta något om $g(t)$ är en "intuitiv" eller oregelbunden funktion. Speciellt vet vi ingenting om existens av g:s andra-derivata.

2. Finns det *många* funktioner som ger minimum?

Det är lätt att se att svaret är nej. Om nämligen g och h båda ger minimum så gäller av samma skäl som förut:

$$2\,M = I(g) + I(h) = 2\,I\left(\frac{g+h}{2}\right) + 2\,I\left(\frac{g-h}{2}\right) > 2\,M + 2\,I\left(\frac{g-h}{2}\right).$$

Det sista uttrycket måste alltså vara <0 men består ju av kvadrater och är därför >0. Alltså följer

$$\int_0^1 (g'-h')^2 dt = 0, \quad \int_0^1 p(t)(h(t)-g(t))^2 = 0,$$

som ger $g(t) = h(t)$.

3. Vilka egenskaper har $g(t)$?

Vi skall utföra en operation på $I(f)$ som precis motsvarar vanlig derivering för vanliga funktioner. Om $\varphi(x)$ är en av våra testfunktioner på sidan 100, så uppfyller $g(t) + u \cdot \varphi(t)$ randvillkoren (eftersom $\varphi(0) = \varphi(1) = 0$) för alla tal u. Bilda nu $I(g + u \cdot \varphi)$, som är $\geq M$, och utveckla kvadraterna. Ordna slutligen resultatet efter potenser av u. Då finner vi

$$I(g + u \cdot \varphi) = I(g) + 2\,u \int_0^1 (g'\varphi' + p(t)g\varphi)dt + u^2 I(\varphi).$$

Andragradspolynomet i u måste ha minimum $= M$ för $u = 0$. Alltså måste derivatan med avseende på u vara noll för $u = 0$, dvs.

$$T(\varphi) = \int_0^1 (g'\varphi' + p\,g\,\varphi)dt = 0.$$

$T(\varphi)$ är ”derivatan av I med avseende på φ i punkten g”. Vi finner att i minimipunkten g måste alla derivator vara noll.

Om nu g kan deriveras gäller

$$\int_0^1 (g'\varphi' + g''\varphi)dt = \int_0^1 (g'\varphi)'dt = 0$$

och $T(\varphi) = 0$ ger för alla φ

$$\int_0^1 (-g'' + pg)\varphi\, dt = 0.$$

Parentesen måste då vara noll och g skulle då vara en lösning till differentialekvationen $g'' - pg = 0$. Om $p = 0$ gäller alltså $(g')' = 0$, $g' = \text{konstant} = A$ och $g = At + B$. $g(0) = 0$, $g(1) = 1$ ger $g(t) = t$, precis som vi tidigare nämnde.

När vi nu inte vet om g'' har mening, måste vi tolka ekvationen i de generaliserade funktionernas betydelse och vi ser här i ett konkret fall hur huvudproblemet blir att visa att en viss generaliserad funktion är en vanlig funktion. Detta kan lösas men vi lämnar här problemet. Vad vi skall lägga märke till i ovanstående är hur naturligt det allmänna integralbegreppet kommer in i kalkylen. Vidare ser vi hur fullständig analogin med vanliga funktioner kan göras. Punkter byts mot funktioner, riktningar mot vektorrummen $u \cdot \varphi(x), -\infty < u < \infty$. Samma teknik utnyttjas också för partiella differentialekvationer och är överhuvudtaget fundamental för både matematiken själv och för tillämpningar. $\rangle$

Vi har i exemplet sett att man kan tänka sig något som liknar koordinater även för funktioner. Men hur skall vi införa koordinatsystem? Alla vet att dessa bör anpassas till problemet för att detta skall kunna behandlas framgångsrikt och bli enkelt. Ekvationerna $xy=1$ och $12x^2+7xy-12y^2+2x+11y-27=0$ representerar samma kurva (hyperbel) i olika koordinatsystem, men ingen kan väl tveka om att den första framställningen är enklare och säger mer.

Nyckeln till problemet om koordinatsystem för funktionsrum finns i tabellen på s. 98. Vi fortsätter den:

Två vektorer (funktioner) är vinkelräta eller med andra ord ortogonala om

Ortogonala	$(x, y)=x_1y_1+x_2y_2=0$	$(f, g)=\int_0^1 f(t)g(t)dt=0.$

Om nu a^1 och a^2 är två vinkelräta vektorer av längd 1 i planet kan vi dela upp varje vektor x på följande sätt:

Komposanter	$x=(a^1, x)a^1+(a^2, x)a^2$
Pythagoras sats	(längden $x)^2=(a^1, x)^2+(a^2, x)^2$

I rymden behövs, som man lätt inser, 3 basvektorer a för att man skall kunna dela upp alla vektorer på analogt sätt:

$$x=(a^1, x)a^1+(a^2, x)a^2+(a^3, x)a^3$$

$$(\text{Längden } x)^2=(a^1, x)^2+(a^2, x)^2+(a^3, x)^2.$$

Det hela är bara ett sätt att skriva koordinatsystem. Vad skall vi sätta på ?:s plats?

⟨ Det behövs nu oändligt många koordinataxlar = basvektorer a = funktioner $a(t)$. De skall vara ortogonala, dvs. $\int_0^1 a^i(t)a^j(t)dt = 0$ om $i \neq i$. Dessutom skall de vara **"tillräckligt många" så att** ingen riktning fattas. (Jfr att två vektorer a inte räcker i rymden.) Vi får då bestämt "koordinater till en funktion", nämligen talen

$$\alpha_i = (a^i, f) = \int_0^1 a^i(t)f(t)dt.$$

Dessa tal bestämmer entydigt funktionen och Pythagoras sats gäller. Pythagoras sats:

$$\int_0^1 f(t)^2\,dt = \alpha_1{}^2 + \alpha_2{}^2 + \ldots + \alpha_n{}^2 + \ldots .$$

Det enklaste exemplet på koordinatsystem $a^n(t)$ är **valet**

$$\boldsymbol{a^\circ(t) = 1,\ a^n(t) = \sqrt{2}}\cos(n\,\pi\,t),\ n = 1,\ 2,\ \ldots .$$

Faktorn $\sqrt{2}$ skall vara med så att $\int_0^1 a^n(t)^2dt = 1$. [Kontrollera att $(a^i, a^j) = 0$,

Komposantuppdelningen övergår nu i en oändlig serie

$$\alpha_1 a^1(t) + \alpha_2 a^2(t) + \ldots + \alpha_n a^n(t) + \ldots$$

och frågan är nu i vilken **mening denna serie representerar funktionen f. Man kan bevisa att**

$$\int_0^1 (f(t) - (\alpha_1 a^1(t) + \alpha_2 a^2(t) + \ldots + \alpha_n a^n(t))^2 dt$$

närmar sig noll då n växer. Detta sätt att uttrycka konvergens är här helt naturlig — avståndet mellan f och summan går mot noll — men ur praktisk synpunkt är man mer intresserad av vanlig konvergens. Denna fråga är mycket komplicerad och långt ifrån utredd. Först nyligen har det t.ex. bevisats att om $f(t)$ är en kontinuerlig funktion och $a^n(t)$ är $\sqrt{2}\cos(n\,\pi\,t)$ så konvergerar serien utom på en mängd vars längd är noll. ⟩

⟨ Det trigonometriska systemet är mycket användbart för studiet av differentialekvationer beroende på att när man deriverar, så ändrar sig koordinaterna på ett enkelt sätt. Speciellt lämpligt är att använda de komplexa funktionerna $a(t) = e^{int} =$

$= \cos nt + i \sin nt$, eftersom $a'(t) = -n \sin nt + in \cos nt = in\, e^{int} = in\, a(t)$. Derivering av en funktion motsvarar alltså multiplikation med (in) av n:te koordinaten. Vi skall se litet närmare på motsvarande problem i flera variabler och inskränker oss här till två variabler.

Vi bygger här upp allmänna funktioner med hjälp av de speciella funktionerna $a(t) = e^{i(\alpha_1 t_1 + \alpha_2 t_2)} = e^{i(\alpha, t)}$ där $\alpha = (\alpha_1, \alpha_2)$ är en konstant vektor. Här gäller

$$\frac{\partial^2 a}{\partial t_1^2} = -\alpha_1^2 \cdot a(t), \quad \frac{\partial^2 a}{\partial t_2^2} = -\alpha_2^2\, a(t).$$

Vill man nu lösa en sådan ekvation som exempelvis

$$\frac{\partial^2 u}{\partial t_1^2} + \frac{\partial^2 u}{\partial t_2^2} = f(t_1, t_2),$$

där $f(t_1, t_2)$ är en given funktion, så börjar man med att studera fallet $f(t_1, t_2) = a(t)$. Det allmänna fallet behandlas sedan genom att $f(t_1, t_2)$ sätts samman av funktioner $a(t)$ på ett sätt som är analogt med det vi använde för att bygga upp funktioner på $(0, 1)$ med hjälp av $\sqrt{2} \cos(n \pi t)$. För det speciella valet ligger det nära till hands att prova $u = C \cdot e^{i(\alpha, t)}$. Ekvationen stämmer då om

$$C = \frac{1}{\alpha_1^2 + \alpha_2^2}$$

och komplikationer uppstår bara om $\alpha_1 = \alpha_2 = 0$. Tar vi i stället på motsvarande sätt ekvationen

$$\frac{\partial^2 u}{\partial t_1^2} - \frac{\partial^2 u}{\partial t_2^2} = f$$

måste vi välja

$$C = \frac{1}{\alpha_1^2 - \alpha_2^2}$$

och komplikationer uppstår för alla α med $\alpha_1 = \alpha_2$. Denna skillnad är den grundläggande förklaringen till att den matematiska teorin för de två ekvationerna (E) och (H) som vi införde ovan är så helt olika. Vi ser att andragradspolynomet i nämnaren är bildat mycket analogt till ekvationens utseende. På samma sätt kan man till alla lineära bildningar av derivator tillordna ett polynom och om detta polynom antar värdet noll för $\alpha \neq 0$ eller hur det närmar sig noll för stora α är helt avgörande för ekvationens lösningar. Denna teori har utvecklats starkt under det sista decenniet och här har de generaliserade funktionerna spelat en stor roll, eftersom man för dessa kan tolka vad division med sådant som $(\alpha_1{}^2 - \alpha_2{}^2)$ betyder. Arbetet i denna riktning fortsätter, i syftet att klargöra fallet då ekvationerna inte har konstanta koefficienter och att finna numeriska behandlingsmetoder.⟩

En abstrakt teori, av en utomordentlig betydelse för den matematiska utvecklingen, har också grenat ut sig från differentialekvationerna. Vi kan ju uppfatta differentialuttrycket som en operation — eller en funktion — som avbildar en funktion u på funktionen $\partial^2 u/\partial x^2 + \partial^2 u/\partial y^2$ eller vilken annan ekvation vi studerar. Om vi inför koordinatsystem blir detta operationer som liknar matriser i fallet av n dimensioner och man kan här utveckla teorier som mycket liknar den man har för matriser. Nästa steg är att axiomatisera teorin — funktionerna ersätts av abstrakta vektorrum, vi avbildar inte till samma vektorrum utan mellan vektorrum, avståndet som var definierat genom $\int_0^1 (f-g)^2 dt$ ersätts av en allmän topologi osv. Härigenom har en stark algebraisering skett av analysen och den utveckling som här antytts på några rader har i själva verket varit en dominerande forskningsinriktning under många decennier och är fortfarande mycket betydelsefull.

⟨Särskilt viktiga för tillämpningar är så kallade egenvärdeproblem. Man har då en homomorfi h från ett vektorrum V till sig själv $h\colon V \to V$, och man söker element u så att $u \to \lambda \cdot u$, λ komplext tal. Om t.ex. h är $d^2u/dx^2 + k(x) \cdot u$, beskriver lösningarna u till $h(u) = \lambda u$ svängningstillstånd hos en sträng och motsvarande tal λ beskriver möjliga toner. I flera variabler uppstår motsvarande problem till exempel vid beskrivning av en atoms möjliga tillstånd, och kvantmekanik kan i stor utsträckning anses vara teorin för dessa speciella problem.⟩

⟨Låt oss avsluta även detta avsnitt med att antyda något man inte vet. Man kan beskriva ett egenvärdeproblem genom att säga att man har givet en homomorfi $h: V \to V$, och att man söker en endimensionell rymd L (nämligen tu där t antar alla komplexa talvärden) som avbildas på sig själv av $h: L \to L$. Det är lätt att hitta homomorfier h för vilka någon sådan endimensionell rymd inte existerar. Men om vi har en godtycklig kontinuerlig homomorfi h, existerar det alltid något mindre rum L i V så att $h: L \to L$? Detta är ett berömt öppet problem i detta område. ⟩

Kommentarer

En välbalanserad framställning av den traditionella första årskursen i matematik vid våra universitet finns i följande två böcker

M. H. Protter, C. B. Morrey Jr, *Calculus with Analytic Geometry*. 1963, och *Modern Mathematical Analysis*. 1964 (båda på Addison-Wesleys förlag).

Man kan jämföra dessa med de nyare svenska, mycket abstraktare och mera formalistiska böckerna. De ryska läroböckerna är i regel ännu mer konkreta än dessa båda.

En intressant diskussion om den tidigare synen på divergenta serier finns i de första kapitlen av

G. H. Hardy, *Divergent Series*. Oxford 1949.

Den Brownska rörelsen och dess fysikaliska betydelse diskuteras i

Albert Einstein, *Investigations on the Theory of the Brownian Movement*. Dover 1956.

Den matematiska teorin är svår. Se t.ex. Fellers nedan citerade bok, del II, kap. 8.

En kort framställning av Lebesgueintegralen finns i

J. C. Burkhill, *The Lebesgue Integral*. (Cambridge Tracts in Mathematics and Mathematical Physics, no. 40.) Cambridge 1959.

Det allmänna integralbegreppet behandlas t.ex. i

H. L. Royden, *Real Analysis*. Macmillan. 1963.

Om Hilbertrum kan man läsa i två böcker av

P. R. Halmos, *Finite-dimensional Vector Spaces*. Van Nostrand 1958, och *Introduction to Hilbert Space*. Chelsea 1951.

Generaliserade funktioner — eller distributioner som de i regel kallas — infördes systematiskt av L. Schwartz under senare delen av 1940-talet. Han fick för sin insats ett Fieldpris 1950. På kort tid har dessa blivit matematiskt allmängods. En utförlig presentation finns i

I. M. Gelfand, G. E. Shilov, *Verallgemeinerte Funktionen*. I. Akademieverlag. Berlin 1960.

Ordinära differentialekvationer kan man lämpligen studera i

L. S. Pontryagin, *Ordinary Differential Equations*. Addison-Wesley 1962,

och partiella differentialekvationer i

I. G. Petrovsky, *Lectures on Partial Differential Equations*. Interscience 1954.

Fourieranalysen finns enkelt presenterad i

R. R. Goldberg, *Fourier Transforms*. (Cambridge Tracts in Mathematics and Mathematical Physics, no. 52.) Cambridge 1961.

Att en kontinuerlig funktions Fourierserie i regel konvergerar bevisade förf. 1966 [Acta Mathematica, 1966].

Den första systematiska undersökningen av allmänna partiella differentialekvationer gjordes 1955 av L. Hörmander. Han påvisade betydelsen av uppförandet av det polynom vi nämnde. Han fick för detta arbete ett Fieldpris vid Stockholmskongressen 1962. Teorin är mycket teknisk och finns framställd i

L. Hörmander, *Linear Partial Differential Equations*. Springer Verlag 1963.

8. Sannolikhet och statistik

Sannolikhet

Enligt de nya kursplanerna skall undervisning i statistik bedrivas både i grundskolan och på gymnasiet. Avsikten med detta är i första hand att göra eleverna så förtrogna med statistikens termer och metoder att de sedan som vuxna kan förstå den statistik som möter oss överallt i samhället. Helst skall undervisningen också ge så mycket förståelse för ämnet att eleverna får förmåga att kritiskt bedöma vad de läser.

På gymnasiet har man emellertid drivit ambitionen vida längre. Man vill där ge en matematisk behandling av sannolikhetsbegreppet och behandla så komplicerade begrepp som tidsserier. Man har därmed gett sannolikhetsteorin en särställning bland tänkbara orienteringsområden utanför analysen. Givetvis är detta ett viktigt område med hänsyn till tillämpningarna men det allmänna behovet ligger utan tvekan inom den beskrivande och sociala statistiken. Till detta kommer att sannolikhetsbegreppet är ett av de mest komplicerade i matematiken så snart man kommit förbi så enkla problem som tärningskastning och kortspel.

Vi anknyter till framställningen på s. 29. Vi har alltså bildat en mängd M vars punkter skall motsvara elementära händelser, dvs. ett enstaka möjligt resultat av det experiment vi utför. En allmän händelse är nu en delmängd E av M som vi anser ha inträffat om vilken som helst elementär händelse i E inträffar. (Allt vi nu sysslar med är att anpassa vårt ordval till de abstrakta begreppen, ”resultat” = ”punkt”, ”händelse” = ”delmängd” etc.) Vi skall nu godta så många händelser E att systemet får de egenskaper vi beskrev på s. 29. Vidare skall det finnas givet en massfördelning på M, så att totalmassan är 1, och den massa som faller på E, kallar vi E:s sannolikhet.

Allt detta är nu gott och väl men avviker ju totalt från vår intuitiva idé om sannolikhet. Vi tänker oss där troligen ett försök som kan upprepas obegränsat många gånger. Vi intresserar oss för en viss händelse, dvs. en viss mängd E av resultat av försöket. När vi upprepat försöket n gånger har kanske resultat i E inträffat A_n gånger. Vi anser oss nu böra tillordna E sannolikheten p om A_n/n närmar sig p då n växer, dvs. om alltid i en lång försöksserie E inträffar ungefär bråkdelen p av gångerna. Denna intuitiva idé innehåller ju många frågetecken. Hur kan vi veta att det finns ett sådant tal p? Hur kan vi i så fall bestämma det? Skall verkligen "alltid" gälla att A_n/n närmar sig p? Tar man mycket enkla försök, som att kasta en tärning, säger man att resultaten 1, 2, ..., 6 prickar ju alla är likvärdiga så det är "klart" att 1 prick kommer upp ungefär 1/6 av gångerna. Denna metod att bestämma sannolikheter: "det är ju ingen skillnad på det här och det där och därför är de lika sannolika" är egentligen den enda enkla man kan använda. Denna gör sannolikhetsteorin till en tillämpning av kombinatoriken. I det allmänna fallet är problemet emellertid långt ifrån klart och i själva verket har under hela vårt århundrade pågått en debatt om det logiska sambandet mellan vår intuitiva sannolikhetsuppfattning och den strikt abstrakta vi först nämnde. Formellt är det klart: för att passa in vårt experiment i den abstrakta ramen skall vi lägga massan p på E. Men går det och är p välbestämt? Det ser ut som om detta problem nu skulle få en lösning genom att man fått en hållbar men ändå tillräckligt allmän logisk tolkning av ordet 'försök', som läsaren kanske noterat att vi smusslat in utan definition.

Även ur en annan synpunkt är grundfrågorna i sannolikhetskalkylen besvärliga. Det gäller det intuitiva begreppet "oberoende försök" och "betingad sannolikhet". Med det sista begreppet menas att man studerar sannolikheten för en viss händelse då man vet att en annan händelse inträffat. Den sistnämnda händelsen betingar då den första. Som ett exempel kan vi kasta två tärningar oberoende av varandra. Sannolikheten att få nio prickar finner vi genom att se på antalet kombinationer som ger nio (3+6, 4+5, 5+4, 6+3, dvs. 4 st.) jämfört med totala antalet kombinationer 36. Sannolikheten blir alltså $4/36=1/9$. Om vi emellertid vet att första tärningen visar 2 prickar blir sannolikheten efter denna upplysning 0 — det går nu inte att få nio — medan om den första visar 4 blir den betingade sannolikheten $1/6$ — enbart resultatet 5 på den andra ger nio. I det allmänna fallet betyder betingelsen E att vi enbart studerar delmängder av E och sannolikheten för en händelse F blir förhållandet mellan massan på den del av F som ligger på E och massan på E. Speciellt får man se upp med möjligheten att E har sannolikhet noll — vilket inte får sammanblandas med att händelsen är omöjlig! ”Oberoende försök” innebär nu att man har två olika sannolikhetsfördelningar och om man betingar med den ena skall detta inte påverka den andra. Om detta låter svårt kan man bara konstatera att denna begreppsbildning är en av de besvärligaste i matematiken. Det förekommer ideligen både i vetenskaplig litteratur och i läroböcker att man gör allvarliga fel vid handhavandet av detta begrepp. Som ett exempel kan nämnas att begreppet ”många oberoende försök” är svårare att definiera än då man bara studerar två. Att försöken parvis är oberoende är inte tillräckligt, vilket vore lätt att tro. Å andra sidan är det denna idé om oberoende och beroende som ger sannolikhetskalkylen dess särdrag — bortser man härifrån är allt bara vanlig integrationsteori.

⟨ För att illustrera svårigheterna går vi tillbaka till vår slumpvisa kurva på s. 93. Vad man där gjorde var att i varje punkt på ett intervall göra oberoende försök, sätta ihop alla dessa kontinuerligt många försök till en enda kurva. Denna kurva är alltså totalresultatet av alla försöken och utgör alltså en ”punkt” i mängden M. Denna består då i princip av varenda funktion $y=f(x)$, hur oregelbunden och diskontinuerlig som helst. Att nu precisera tillåtna mängder av funktioner, för vilka vi alltså skall tala om sannolikhet, är givetvis mycket delikat. Vi kanske vill inskränka oss till att studera kurvor genom någon fix punkt. Sannolikheten för en kurva att gå genom en fix punkt (x_0, y_0) är rimligen 0 eftersom från början alla värden på y för $x=x_0$ bör vara i stort sett likvärdiga. Vi har då just en sådan betingelse som ger speciella svårigheter. ⟩

En preliminär slutsats av vad som här kommit fram är att sannolikhetskalkylen som matematisk disciplin och matematisk beskrivning av intuitiva idéer ger ovanligt stora svårigheter. Detta gäller både sambandet mellan verkligheten och beskrivningen och även teorin själv, när man kommit förbi den rent kombinatoriska aspekten.

Vi går emellertid vidare och skall se vad som kan komma fram vid studiet av sannolikhetskalkylen. Man vill först få ut kvantitativa resultat ur beskrivningarna. Om en tärning visar 2 prickar, tillordnar man därför denna elementära händelse talet 2. Om en rekryt är 180 cm lång, är det naturligt att här tillordna talet 180. Matematiskt innebär detta att man studerar funktioner φ definierade på mängden M; φ: $M \to$ reella talen t.ex. men även komplexa funktioner är viktiga. En sådan funktion är vad man kallar statistisk variabel och skall här betecknas symboliskt med Z. Egentligen har man här inte gjort någon ny definition: varenda funktion på M är ju en statistisk variabel. Slutligen ett begrepp till: sannolikheten för att $\varphi(m)$ är $< x$ kallas fördelningsfunktionen till φ och betecknas $F(x)$. $F(x)$ växer alltså från 0 upp mot 1. F:s derivata kallas frekvensfunktion och betecknas $f(x)$. Medelvärdet av den statistiska variabeln definieras så genom integralen

[illegible]

för sådana fördelningsfunktioner $F(x)$ som har kontinuerlig derivata. Om vi symboliskt betecknar sannolikhetsmassan på M med P så skulle vi kunna skriva samma integral på följande sätt

[illegible]

och vi har här ett exempel på integration över en abstrakt mängd. Vi skall uppfatta integralen som en slags summa där vi lägger ihop de olika värdena som φ kan anta var och en med den vikt som massan = sannolikheten anger. I tärningsfallet stämmer denna tolkning precis. Funktionen är φ: (n prickar) $\to$ talet n ($n = 1, 2, \ldots, 6$). Massan är 1/6 i var och en av de sex punkterna i M och medelvärdet blir

$$\varphi(m_1)P(m_1) + \ldots \varphi(m_6)P(m_6) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + \ldots + 6 \cdot \frac{1}{6} = 3{,}5.$$

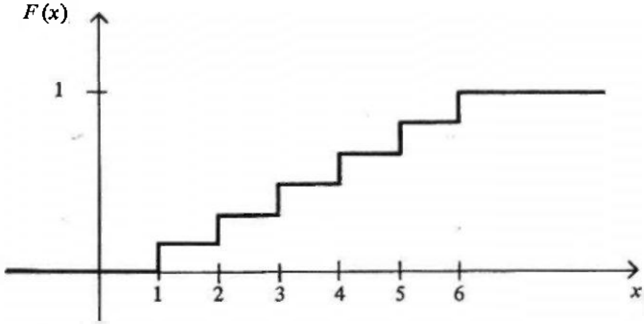


Fig. 37

Motsvarande fördelningsfunktion $F(x)$ ser ut som fig. 37 visar.

Om man ritar upp fördelnings- och frekvensfunktioner för olika iakttagelser i naturen och i samhället får man ofta (men absolut inte alltid!) en frekvensfunktion och fördelningsfunktion som återges i fig. 38. Ekvationen för f är $f(x) = C\,e^{-c(x-\mu)^2}$, **som efter lämpliga variabeltransformationer blir**

[illegible] $\sqrt{\pi}$

Faktorn $1/\sqrt{\pi}$ motiveras av att $\int_{-\infty}^{\infty} f_0(x)dx$ skall vara $= 1$. Denna sorts kurva möter man om man studerar längden hos rekryter, stjärnors ljusstyrka eller de fel som uppstår vid en mätning och det gäller precis vilken mätning som helst där antalet felkällor är stort. Av detta sista skäl kallas kurvan ibland för felkurvan.

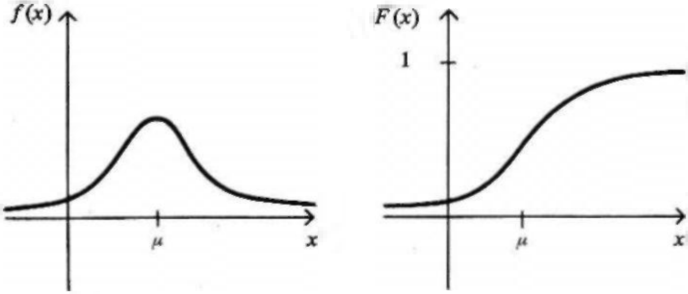


Fig. 38

Vad är orsaken till att denna fördelning beskriver så många olika fenomen?

Följande matematiska sats innehåller förklaringen. Om Z_1, $Z_2, \ldots, Z_n, \ldots$ alla är oberoende var och en med medelvärde 0 och rimligt regelbundna, så är alltid för stora värden på n $Z_1+Z_2+\ldots+Z_n$, efter multiplikation med en lämplig faktor, fördelad nästan som en felkurva. Om alla Z_n är lika, alltså representerar samma försök, blir faktorn ungefär $1/\sqrt{n}$.

⟨ För att belysa detta, tag som varje Z_n att vi kastar krona och klave och räknar krona som $+1$ och klave som -1. $S_n = (1/\sqrt{n})(Z_1+Z_2+\ldots+Z_n)$ representerar då hur mycket man vunnit (eller förlorat), delat med $\sqrt{n}$. Om vi nu bildar medelvärdet av funktionen e^{itS_n}, för ett fixt värde på t med avseende på den fördelning som hör till $(1/\sqrt{n})(Z_1+Z_2+\ldots+Z_n)$ så får vi

$$\left(\frac{e^{i\frac{t}{\sqrt{n}}}+e^{-i\frac{t}{\sqrt{n}}}}{2}\right)^n = \left(\cos\frac{t}{\sqrt{n}}\right)^n.$$

Vi gör detta troligt genom att beräkna fördelningen för S_n för några värden på n.

$n=2.$	S_n	$-\sqrt{2}$	0	$+\sqrt{2}$	
	Sannolikhet	[illegible]/4	[illegible]/2	[illegible]/4	

$$\frac{1}{4}e^{-i\sqrt{2}\,t}+\frac{1}{2}e^{i0}+\frac{1}{4}e^{i\sqrt{2}\,t} \qquad \left(\frac{e^{-i\frac{[illegible]}{\sqrt{2}}t}+e^{+i\frac{[illegible]}{\sqrt{2}}t}}{2}\right)^2$$

[illegible]	[illegible]	$-\sqrt{3}$	$-\frac{1}{\sqrt{3}}$	$+\frac{1}{\sqrt{3}}$	$+\sqrt{3}$
	Sannolikhet	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	[illegible]/8

[illegible]

$$=\left(\frac{e^{-it/\sqrt{3}}+e^{+it/\sqrt{3}}}{2}\right)^3$$

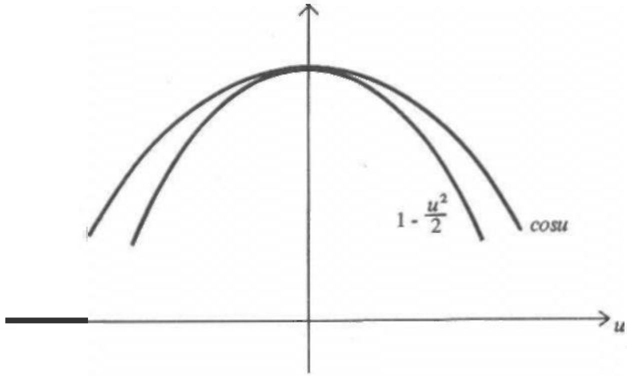


Fig. 39

$n=4$.

S_n	-2	-1	0	$+1$	$+2$
Sannolikhet	$\frac{1}{16}$	$\frac{4}{16}$	$\frac{6}{16}$	$\frac{4}{16}$	$\frac{1}{16}$

[illegible]

Nu antar $\cos u$ i $u=0$ samma värde som $(1-u^2/2)$. Detsamma gäller funktionernas derivator och andra-derivator och funktionen $1-u^2/2$ liknar alltså i hög grad $\cos u$ för små u (se fig. 39). Vi får då

$$\left(\cos\frac{t}{\sqrt{n}}\right)^n = (\text{ungefär}) \left(1-\frac{1}{2}\frac{t^2}{n}\right)^n = \left[\left(1-\frac{1}{\frac{2n}{t^2}}\right)^{\frac{2n}{t^2}}\right]^{\frac{t^2}{2}} =$$

$= (\text{ungefär})\ e^{-t^2/2}$.

Den sista relationen gäller eftersom $[1-(1/x)]^x$ närmar sig $1/e$ då x växer. Medelvärdet för e^{itS_n} är alltså ungefär $e^{-t^2/2}$ när n växer. Nu finns det en formel, nämligen

[illegible]

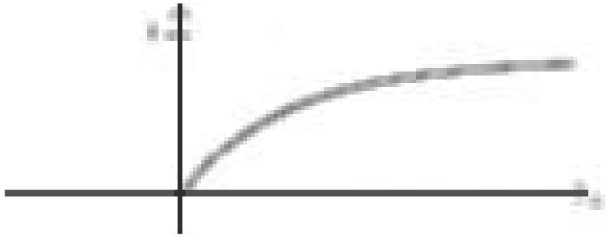


Fig. 40

som visar att felfördelningar ger precis samma medelvärde för e^{itx} för alla t. Detta är ett indicium på vårt påstående och precis den bevisidé vi här använt kan användas för ett riktigt bevis, även i det allmänna fallet.⟩

På vilket sätt förklarar nu denna rent matematiska sats att felkurvan uppträder vid alla de olika fenomen vi nämnde. När vi utför en viss mätning, uppstår ett fel, Z_1, beroende på slarvig inställning, ett annat Z_2 på grund av felaktig avläsning, ett tredje Z_3 på grund av oförutsedda yttre variationer osv. Var och en av dessa kan också delas upp i komponenter, som härrör från olika och oberoende orsaker. Det totala felet Z blir nu $= Z_1 + Z_2 + \ldots$ där varje Z_i är mycket liten, och enligt vår sats skall detta vara approximativt fördelat enligt felkurvan. På samma sätt bestäms en persons längd av en mängd samverkande och oberoende omständigheter, var och en av liten betydelse, och resultatet fördelar sig då också efter felkurvan. Det bör åter betonas att fastän denna kurva är mycket vanlig beskriver den ingalunda fördelningen för alla fenomen. Det kan ju tänkas att en bestämd orsak dominerar över de andra och effekten av denna orsak kommer då att slå igenom i resultatet. Tar man som exempel sannolikheten för att man får vänta en tid $<x$ i en kö har den ofta en fördelning av typen (fig. 40)

$$F(x) = \begin{cases} 0 & , x<0 \\ 1-e^{-x}, & x>0. \end{cases}$$

Det resultat som vi här diskuterat är ett exempel på den viktigaste typen av forskningsresultat i den traditionella sannolikhetsteorin. Man studerar alltså under olika antaganden

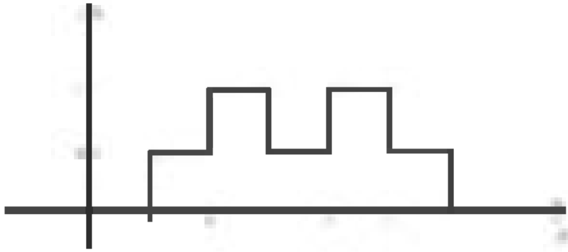


Fig. 41

hur ett stort antal försök samverkar och vilka gränsfördelningar som kan finnas. Denna forskningsriktning ger både betydelsefulla resultat för tillämpningar i statistik och eleganta matematiska resultat.

Vi var nyss intresserade av vilka värden S_n kunde anta och med vilka sannolikheter. Nästa steg är att samtidigt studera hela följden $S_1, S_2, \ldots, S_n$. Om vi tar bort faktorn $1/\sqrt{n}$ igen kan vi åskådliggöra förloppet i ett diagram (fig. 41).

Vi får alltså en slags slumpvis kurva med hörn för heltalsvärden. Ett sådant här förlopp är en stokastisk (=statistisk) process och studiet av dessa är numera den viktigaste grenen av sannolikhetsteorin. De slumpvisa kurvor vi tidigare studerat är tydligen en kontinuerlig motsvarighet till den enkla process vi nämnde. Ofta är det lämpligt att uppfatta n som tid och processer av detta slag beskriver då t.ex. förlopp av följande slag. När vi nu vet att vid tiden $t = n$ ett fysikaliskt system (kanske ett antal atomer eller en kö framför en biljettlucka) befinner sig i ett visst tillstånd T (hastigheter och lägen, eller antal väntande personer) är sannolikheterna kända för att systemet vid nästa avläsningstid, $n+1$, skall vara i ett tillstånd 1, 2, ... Vi vill nu veta t.ex. systemets livslängd i medeltal i atomfallet eller i köfallet medelväntetid för att kunna avgöra om vi skall öppna en biljettlucka till.

⟨ De stokastiska processerna har alltså en stor betydelse i teknik av olika slag. De sammanhänger också på ett mycket intressant sätt med andra matematiska områden. Vi skall illustrera detta med ett exempel. Vi studerar ett område i planet och väljer ut en del av områdets rand (se fig. 42). Tag nu

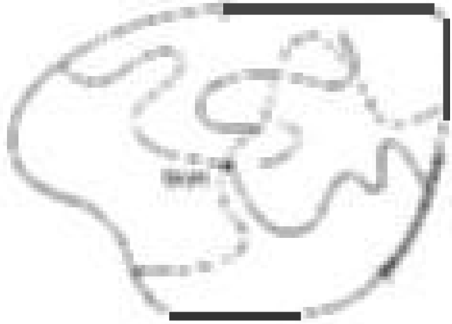


Fig. 42

slumpvisa kurvor som går ut från en viss punkt (x, y) i området när vi låter alla riktningar vara lika sannolika och felkurvan bestämmer fördelningen av avståndet mellan två punkter.

Sannolikheten att vår kurva träffar den utvalda delen av randen före resten av randen blir då en funktion u av (x, y). Denna funktion kan nu visas vara en lösning till ekvationen $\partial^2 u/\partial x^2 + \partial^2 u/\partial y^2 = 0$. Detta resultat kan nu användas till en rätt förbluffande metod att lösa denna differentialekvation numeriskt. Man lottar då en lång svit tal och matar in dem i en datamaskin. Dessa lottade tal får nu styra den process vi beskrev, när vi inskränker (x, y) till att variera i ett finmaskigt kvadratiskt nät. Vi undersöker nu hur ofta vi träffar den aktuella randbiten först och får då en ungefärlig uppfattning om sannolikheten u. Denna numeriska metod kallas Monte Carlo-metoden. Den är inte särskilt effektiv i detta fall (dvs. det fordras för många kalkyler för att få fram ett bra resultat = det tar för lång tid) men i andra sammanhang är metoden av praktisk betydelse.〉

Statistik

Statistiken är en vetenskap med många olika aspekter, där matematiken ingalunda är den viktigaste. Vi särskiljer några olika moment.

Det första stadiet brukar bestå i insamlande av det material man önskar studera och i redovisning av detta material. Poängen med statistiska studier är då att man i regel inte redovisar totalbeståndet. Är vi intresserade av längden hos svenska män mäter vi inte alla manliga innevånare, och vill vi studera kvalitén hos en viss produkt innan den lämnar fabriken gör vi ett urval. Ett annat förfaringssätt skulle bli alldeles för dyrbart och för övrigt behövdes i så fall ingen vetenskap. Ett sådant urval kallas stickprov. Det är nu helt avgörande att stickprovet avspeglar hela beståndet. Om vi mäter längder, måste alla åldrar, socialgrupper osv. vara representerade i urvalet. Ett annat exempel: om vi skickar ut ett frågeformulär och får 80 % tillbaka besvarade är det inte bara möjligt utan till och med troligt att de återstående 20 % representerar vissa bestämda grupper och att därför de 80 % inte är ett representativt urval. I fysikaliska experiment är det avgörande att man kan kontrollera de yttre betingelserna. Av de många fel som görs vid användningen av statistiken är detta med icke-representativa urval ett av de vanligaste. Att komma förbi denna svårighet är en icke-matematisk bedömningsfråga, som alltså är central vid användning av statistik.

Nästa steg är att redovisa materialet och på ett schematiskt sätt beskriva det. Detta sker med tabeller, kurvor och i populära sammanhang numera ofta med gubbar och figurer av olika slag. Även detta är en praktisk fråga utan egentligt matematiskt intresse. Det finns också här stora möjligheter till tendentiös framställning. Ett matematiskt begrepp beskriver inte säkert på ett riktigt sätt en fördelning. Om vi t.ex. anger den mycket höga medelinkomsten i Förenta staterna har vi inte gett en antydan om den mycket utbredda fattigdom som finns där. Vi kunde ju få en medelinkomst på 4 000 dollar om hälften av innevånarna tjänade 8 000 dollar och hälften noll, likaväl som om alla hade 4 000 dollar i inkomst, men de två samhällena vore onekligen rätt olika. Särskilt med grafisk framställning kan man skapa illusion av större eller mindre skillnader, alltefter framställningens syfte, som alla läsare av politiska skrifter kan konstatera.

Matematiken kommer in i sammanhanget först då man vill analysera materialet. I regel är situationen någon av ett fåtal standardtyper. För dessa är den matematiska analysen genomförd en gång för alla, och ett problem får då karaktären av att man stoppar in sitt material i en formelapparat och trycker på knappen. Det är sedan av underordnad betydelse hur maskinen är hopsatt, och i själva verket är konstruktionen så pass komplicerad att även majoriteten av våra utövande statistiker inte känner till den. Det betydelsefulla ligger i verifikationen att förutsättningarna för formeln är uppfyllda och här görs de flesta felen. Några exempel. Är det någon skillnad i längd mellan svenskar och norrmän? Man gör ett urval av vardera och provar med någon standardmetod sannolikheten att båda urvalen har samma fördelningsfunktion. Ett test på att två fördelningsfunktioner är lika har en ganska komplicerad matematisk teori, men alla statistiker klarar lätt problemet (kan man hoppas!). Är det något samband mellan cigarrettrökning och lungcancer? Man gör ett urval av cigarrettrökande personer och observerar lungcancerfrekvens och gör motsvarande med en kontrollgrupp. Man kan sedan med matematiska metoder analysera om den iakttagna skillnaden i frekvens kan bero på slumpen. Det är mycket viktigt att observera att man härmed inte kommer åt den viktigaste frågan, nämligen om cigarretterna orsakar cancer. Det kanske finns fler cigarrettrökare i städer och kanske det är stadens rök eller stress som orsakar ökningen i cancer. Statistiska orsaksanalyser är mycket komplicerade och icke-matematiska moment spelar en avgörande roll.

Sammanfattningsvis är alltså de betydelsefullaste aspekterna på statistiken icke-matematiska. För de flesta av oss är de sociala tillämpningarna de viktigaste och också de svåraste. Givetvis är statistikens språk på ett naturligt sätt matematiken, men de matematiska teorierna har en relativt underordnad betydelse. Vi behöver alla vana att läsa och bedöma tabeller och vi behöver ett kritiskt sinne gentemot slutsatser dragna ur statistiska resonemang. Det är en angelägen uppgift för skolan att ge detta. Däremot är det orimligt att skolan skulle kunna ge en effektiv undervisning i självständigt utnyttjande av statistik. De matematiska grundproblemen är svåra, insamlandet av materialet är fullt av fallgropar och det fordrar stor erfarenhet att veta vilken matematisk-statistisk apparat man skall utnyttja.

Den naturliga förankringen av statistikundervisningen borde därför vara inom ämnet samhällskunskap, kanske under medverkan av matematiklärare, men det bör klart sägas ut att detta ämne till sin natur är något annat än matematik. Statistikens starka ställning i matematikkurserna — som alltså beror på bedömningar av några "experter" i Skolöverstyrelsen — har medfört en stor förändring av universitetskurserna som ur matematikämnets egen synpunkt är helt felaktig. Vi har här ett drastiskt exempel på hur en diskutabel skolreform snedvrider universitetsutbildningen. Sannolikhetsläran hör däremot givetvis hemma i matematiken, men dess betydelse bör bedömas gentemot alla andra områden av ämnet, som det kan vara önskvärt att ge undervisning i.

Kommentarer

För studium av sannolikhetsteori och statistik hänvisas till

H. Cramér, *Sannolikhetskalkylen och några av dess användningar*. Uppsala 1949.

En utförligare men givetvis svårare bok är

H. Cramér, *Mathematical Methods of Statistics*. Princeton 1946.

En rent matematisk framställning är

W. Feller, *An introduction to Probability Theory and its Applications*. I—II. Wiley 1957 resp. 1966.

Relationen mellan frekvenssynen och klassisk sannolikhetskalkyl har belysts i ett nytt arbete av en ung svensk matematiker, P. Martin-Löf [*The Definition of Random Sequences*. Information and Control 1966].

Monte Carlo-metoden finns presenterad i

J. M. Hammersley, D. C. Handscomb, *Monte Carlo Methods*. Wiley 1964.

9. Matematiken i samhället

Vi har på de föregående sidorna upprepade gånger talat om matematikens två roller. Den ena rollen spelas av den "rena" matematiken. Denna ställer själv sina problem, formulerar själv sina definitioner och mål. Dess värderingar är dess egna och de kan närmast jämföras med estetiska. Matematikern talar gärna om vackra satser och bevis; med detta menar han att resultatet ger ny belysning och förklaring till något område och att det är enkelt, överraskande, klart och slutgiltigt. Bevis och slutledningar skall vara logiskt fullständiga ("modulo" matematikens grunder, varom mera strax) och byggda på klara definitioner.

Den tillämpade matematiken å andra sidan utnyttjar matematikernas språk för att beskriva fenomen i naturen. Man är här mindre intresserad av logiska bevis än av resultat. Praktiska metoder att effektivt beräkna de intressanta kvantiteterna är huvudsaken, medan förklaringen till att metoden fungerar (liksom då också dess eventuella begränsningar) får komma i andra hand.

Mellan höger- respektive vänsterflyglarna inom dessa båda grupper riktas ofta skarpa skott. Den renaste rena matematikern betraktar den tillämpade matematikern som en lösaktig släkting som han helt förskjuter. Han föraktar bristen på stringens och anser sig själv representera höjden av intellektuell aktivitet. Han talar gärna om, som något berömvärt, att hans matematiska teorier aldrig kommer att profaneras av någon tillämpning. — Vad han då bortser från (bland annat) är att de matematiska begrepp och föreställningar som han arbetar med i grunden genomgående kommer från verkligheten och direkt från tillämpningar. Även när det gäller detaljer kommer frågeställningarna ofta från fysiken och en av de starkaste indikationerna på ett matematiskt intressant problem är en intressant tillämpning. Att sedan matematiska områden utvecklar sig själva enligt sina egna logiska lagar strider inte mot detta grundförhållande. Förmodligen var denna grupp starkare och mera krigisk tidigare under 1900-talet då man i Russells anda fortfarande trodde på ett slutet logiskt system innefattande hela matematiken. Som vi sett har senare undersökningar visat att de logiska sammanhangen är så komplicerade att någon totallösning troligen inte kan konstrueras. Detta har varit grunden till en viss ödmjukhet och självbesinning.

Aktiviteten på den tillämpade vänsterflygeln har i stället vuxit under senare år. Man anser där den rena matematiken vara en lek med symboler i ett slutet system utan egentligt intresse för världen utanför systemet. Den är alldeles för abstrakt och utnyttjar en så invecklad och till stor del omkonstruerad terminologi, att den är helt obegriplig för icke invigda. Det kritiseras, med ett visst berättigande, att matematiker behärskar invecklade logiska strukturer, men ber man dem numeriskt beräkna en lösning till en viktig differentialekvation, så kan de inte det.

Den viktigaste bakgrunden till aktiviteten har emellertid varit "den nya matematiken" i skolan. Denna sysslar enligt kritikerna alltför mycket med onyttiga begrepp (såsom mängder), stringenta och invecklade definitioner av intuitiva idéer (t.ex. funktioner och reella tal) och subtila diskussioner av självklara resultat (t.ex. att en kontinuerlig funktion på ett slutet intervall är begränsad). Härigenom ger den för litet av aktiv kunskap att verkligen beräkna något konkret, att få fram lösningar till faktiska problem.

Denna aktivitet har på vissa håll i världen lett till en mycket skarp motsättning mellan ren och tillämpad matematik. Detta gäller speciellt i Frankrike och vid vissa universitet i Förenta staterna. På längre sikt kan detta få mycket negativa följder och ansträngningar borde göras av alla berörda parter att analysera orsaken till situationen och framför allt att sätta sig in i motståndarnas argumentering, så att motsättningarna överbryggas.

På det hela taget är den skisserade föreställningen om den rena matematiken knappast rättvis. Det är klart otillfredsställande att acceptera matematiska beskrivningar bara därför att de i praktiken hittills visat sig fungera genom att de ger resultat som överensstämmer med iakttagelser. Inte förrän en beskrivning är matematiskt genomarbetad känner man dess giltighetsområde och förstår sammanhangen. Detta hindrar naturligtvis inte att man kan pröva hållbarheten i en beskrivning utan full matematisk stringens. Detta arbetssätt används alltid även i den rena matematiken för att se vart den inslagna vägen leder. Men målsättningen bör i båda fallen vara en logiskt sammanhängande teori. ”Vad du ej klart kan säga, vet du ej.” Samspelet mellan tillämpningarna och logik är emellertid, som vi tidigare sett, mycket mera komplicerat och dubbelriktat. Vi skall åter ta upp Newtons insats, som är helt mirakulös i sin geniala intuition. När han formulerade sin gravitationslag — att det finns en kraft mellan två kroppar omvänt proportionell mot avståndets kvadrat — innehöll de mätvärden som skulle förklaras med lagen fel på kanske 5 %. Inte desto mindre är lagen sann med 1 000-tals gånger noggrannare mätvärden. På ett underbart sätt innehåller alltså lagen mer sanning än vad den konstruerades för att förklara. Hur är detta möjligt? Enda tänkbara ”förklaringen” är att en enkel lag som denna avspeglar fundamentala, närmast logiska lagar (av typen Newtons lag att en kropp har en massa som är oberoende av dess tillstånd för övrigt eller Einsteins lag att det inte är så men att det inte finns någon större hastighet än ljusets). I sökandet efter dessa lagar spelar den logiska analysen av de matematiska begreppens samspel en avgörande roll, och denna analys har i hög grad utvecklats under de senaste decennierna. Det är mycket sannolikt att i denna utveckling idéer och resultat kommit fram, som skulle kunna medföra stora framsteg i tillämpningar om fysikerna kände till dem. Detsamma gäller de nya stora framstegen på de traditionella områdena, speciellt inom teorin för differentialekvationerna.

Matematikens relation till tillämpningarna kompliceras emellertid av just de förhållanden vi nu berört. Vi har konstaterat att fysikens naturbeskrivningar i regel har formen av differentialekvationer och att den matematiska analysen i stort sett kan anses vara en teori utvecklad kring dessa differentialekvationer. Den matematiska forskningen har under senare år avlägsnat sig från dessa områden mot speciellt topologi och algebra. Det är inom dessa områden de viktigaste resultaten nu uppnås och det är härifrån de starkaste impulserna till förändringar kommer. Dessa förändringar har skett under ett par decennier och har medfört att fysiker och andra naturvetenskapsmän står främmande för dagens matematik. I andra riktningar ser många matematiker ingen användning eller tolkning av sin forskning och förlorar intresset för tillämpningar. Situationen har alltså karaktären av en odämpad svängning som tenderar att öka avstånden mellan ståndpunkterna.

Då det gäller att bedöma om situationen är ohjälplig eller ej är naturligtvis den avgörande frågan om den nya inriktningen av matematiken i objektiv mening saknar betydelse utanför ämnet eller ej. Svaret känner ingen, men man måste bestämt varna för den förutfattade meningen att svaret med säkerhet är att modern topologi och algebra är enbart intellektuella spekulationer. Vid valet av modeller är fysikern och den experimentelle naturvetaren hänvisade till att utnyttja de kunskaper i matematik som de har. Om analysens särställning beror på fysikernas matematiska kunskaper eller på objektiva förhållanden kan vi inte veta. Man får inte tro att våra beskrivningar är objektivt sanna och entydigt bestämda av fenomenen. Man kan erinra om Heisenbergs och Schrödingers formellt mycket olika teorier i kvantmekaniken. Sannolikt är ännu mycket större olikheter i modellerna möjliga. I det osäkra läge vi således befinner oss i är det oförsvarligt att skära av förbindelserna mellan matematiken och tillämpningarna och därmed försvåra ett möjligt utnyttjande av de matematiska forskningsresultaten.

Informationsöverföringen är emellertid en stor och svår fråga. Delvis är problemet här detsamma som på så många andra håll. Den ökade specialiseringen har överallt lett till en isolering av de olika vetenskaperna från varandra helt enkelt därför att ingen människa längre kan behärska mer än till och med en ganska liten sektor av sin egen vetenskap. I många fall har man med framgång löst detta problem genom team-work. Man försöker också lösa det med datateknik, genom att införa kodbeteckningar på innehållet i vetenskapliga uppsatser och därefter lagra informationen i datamaskiner. (Inom parentes sagt förefaller det mycket optimistiskt att tro att man på denna väg skall kunna göra några större framsteg.) När det gäller relationerna mellan matematiken och tillämpningarna är samarbetet särskilt svårt genom matematikens speciella formelspråk och abstrakta terminologi. Det är en mycket tidsödande process att översätta tekniska data till funktioner och operationer och få klarhet i vilka inskränkningar som kan påläggas som följd av den fysiska situationen. Matematiken är här det universella språket och det är självklart att anpassning skall göras till detta språk. Än svårare är det givetvis att utnyttja nya matematiska forskningsområden och någon patentlösning av dessa problem lär knappast finnas.

I Sverige har den tillämpade matematiken en betydligt sämre situation än i både USA och i ännu högre grad än i Sovjetunionen. Endast en handfull matematiker är verksamma i industrin och våra forskningsinstitutioner i fysik, kemi osv. har inga matematiker knutna till sin verksamhet. Det förefaller mycket sannolikt att en aktivare matematisk insats skulle kunna betyda mycket både för industri och forskning. En förutsättning är att matematikerna knyts direkt till det dagliga arbetet så att översättningen från teknik till matematik förenklas. Man bör gå fram på två samtidiga vägar. Ett serviceinstitut bör inrättas, som kan användas som konsultorgan och där samtidigt en efterskolning av matematiker och utveckling av metoder kan ske. Det är knappast lämpligt att inrätta akademiska forskningsbefattningar då ämnet tillämpad matematik är helt obestämt — det rör sig snarare om en inställning till ämnet matematik och om ett syfte än om ett nytt ämne. Vidare bör matematiker aktivt och direkt knytas till större tekniskt komplicerade industrier och till forskningsinstitutioner.

Om då kritiken av matematiken som vetenskap kan avvisas, vad kan man säga om synpunkterna på skolundervisningen? I denna ligger tydligen också grundproblemet då det gäller kommunikationerna. I skolan får man den första kunskapen om vad matematik är och kan användas till, och där skapas också de attityder som bestämmer den framtida uppfattningen om ämnet. På många sätt är den nuvarande situationen otillfredsställande och på många sätt kan man sympatisera med den kritik som framförs.

Matematikundervisningen i skolan har två ändamål. För det första skall den ge en introduktion till matematiken som sådan. Det gäller här ett allmänkulturellt syfte och ämnet bör i detta avseende jämställas med alla andra skolämnen, t.ex. fysik eller historia. För det andra skall den ge sådana färdigheter som behövs för förståelse av andra skolämnen, speciellt fysik, och som eleverna senare har nytta av helt allmänt. Vad som sagts här tidigare antyder dock att det mellan dessa syften inte råder så stark motsatsställning som många tror, men låt oss skärskåda hur syftet realiseras under nuvarande förhållanden.

Vi har konstaterat att orienteringen i skolan om vetenskapen matematik går fram ungefär till de idéer som lanserades kring sekelskiftet 1700. Då hade Newton och Leibniz utvecklat grunderna av differential- och integralkalkylen, Fermat hade infört koordinatsystemet — som numera presenteras i vektorform. Grunderna hade lagts till sannolikhetskalkylen, logaritmräkningen var känd och talet *e* skulle snart introduceras. Dessa föreställningar har sedan kompletterats med 1800-talets strikta krav på konvergensbevis och grundläggande idéer om funktioner och mängder. Det är lärorikt att ställa frågan hur matematiken under nästan två hundra år kunde arbeta med begrepp som inte hade egentlig mening. Svaret är att begreppen motsvarar primitiva intuitiva föreställningar, och ingen tänkte på att inom samma allmänna ram kunde finnas mycket mer komplicerade objekt än de man då kände. Först sedan erfarenheten och kunskapen vidgats, insåg man att det förelåg ett behov av mera precisa definitioner och bevis. Detta historiska resonemang motsvaras i dagens situation av ett pedagogiskt problem, som vi tidigare berört.

Det är ju helt unikt att skolans orientering inom en central vetenskap nöjer sig med att sluta med idéer från 1700-talet. Inom ämnena fysik och kemi hade utvecklingen inte ens börjat och tyngdpunkten i de nuvarande kurserna i dessa ämnen ligger i vårt århundrades vetenskapliga resultat. Som försvar kan nu anföras: (1) Vi har ju moderniserat kurserna med striktare definitioner, axiomatik och mängdlära. (2) Det går inte att ge en orientering om modern matematik. [Det kan noteras att (2) motsäger (1).] (3) De moment som lärs ut är de viktigaste för tillämpningarna. De tar så många schematimmar även på den naturvetenskapliga linjen, att någon utökning inte är möjlig. Det är därför omöjligt att ge en vidgad orientering.

(1) Att mängdläran — eller rättare sagt föreställningen om mängder, ty någon lära om mängder existerar inte utanför logiken — spelar en central roll i modern matematik beror på att idén är en av de mest primitiva i vår föreställningsvärld. Det är med dess hjälp möjligt att på ett enhetligt sätt lägga grunden till skilda områden och att sammanföra likartade tankegångar. Det naturliga är därför att föra in mängder i undervisningen så tidigt som möjligt, dvs. under de allra första skolåren. Görs det senare krävs en motivering varför begreppet är intressant, som är omöjlig att ge med ett så litet bakgrundsmaterial som eleverna nu får. Undervisningen rörande mängder blir därför gärna en ointressant katalog av termer och tecken och en serie övningar med dessa symboler som alla egentligen bara uttrycker logiska trivialiteter.

Utöver mängdläran har man i skolkurserna infört moment av den moderna matematikens symbolspråk. Vidare vill man införa idén om stringens och axiomatik. Som vi tidigare flera gånger framhållit, är det en stor risk att man motverkar sitt syfte genom att göra det självklara och välkända komplicerat. Det är totalt ointressant att betona att additionen av hela tal är kommutativ ($x+y=y+x$) när eleverna aldrig drömt om att en sådan regel inte skulle gälla. Liknande synpunkter gäller beträffande alltför formell behandling av andra klassiska områden (t.ex. logaritmer). Det är troligen denna formalistiska sida hos "den nya matematiken" som mest irriterat de tillämpade ämnena. Den ger heller ingen riktig bild av modern matematik. Tvärtom, i ljuset av modern logik har matematiker numera en mycket moderat uppfattning om stringens in absurdum.

(2) Att högre matematik inte kan populariseras eller på ett

meningsfullt sätt läras ut i orienterande form är en mycket allmän uppfattning. Denna åsikt bör emellertid inte accepteras utan att man prövat den omsorgsfullt. Är det så, är ämnet hänvisat till en kulturell isolering i ett elfenbenstorn, och som vi sett innebär detta kanske en ohjälplig splittring mellan ren och tillämpad matematik. Det är här intressant att jämföra matematikens svårigheter med kärnfysikens.

Atomfysik är till sin natur abstrakt i praktiskt taget samma mening som matematik. Vi sammanfattar vår erfarenhet av atomernas uppträdande i suggestiva bilder som små kulor eller något slags vågrörelse och på ett högre stadium i form av ekvationer och relationer. Psykologiskt fungerar de matematiska begreppen på ungefär samma sätt. En kontinuerlig funktion för en matematiker är en funktion som fungerar på ett visst sätt. Var och en har nog en personlig föreställning som han faller tillbaka på när han arbetar med begreppet. Givetvis kan han de abstrakta definitionerna, men de finns långt tillbaka i medvetandet. Det är således idén och "funktionen" av funktionen som i första hand används, inte den matematiska konstruktionen eller den formella definitionen.

Atomfysiken är numera grunden till hela fysikundervisningen i skolan. Givetvis utreder man ingenting i detalj, utan förklarar och gör sannolikt. Frågan är då: skulle ett sådant förfarande i matematik ge en riktig och lämplig bild av ämnet? Och skulle det gå att genomföra?

Det är inte möjligt att idag svara oreserverat ja på någon av dessa frågor. På samma sätt som dagens skolundervisning inte ger en student sådana kunskaper i atomfysik att han självständigt kan konstruera ett fysikaliskt experiment och värdera resultatet, lika litet skulle en undervisning av den typ vi diskuterar ge eleverna förmåga att lösa annat än mycket enkla problem. Ser man matematiken som en helhet, där alltså bevisens konstruktion oskiljaktigt sammanhänger med resultatet, anser man nog att ämnet inte bör läras ut på detta sätt. Detta är emellertid en extremt estetisk inställning, som de flesta matematiker troligen inte accepterar. Huvuddelen av alla matematiska bevis är nämligen tekniska anpassningar av ett ganska litet antal idéer och utnyttjande av tidigare resultat. Den enda rimliga värderingen i vår vetenskap som i andra bör då baseras på de idéer som förs fram och de resultat som uppnås. Den tekniska sidan är viktig och svår men av underordnad betydelse. Med den utgångspunkten är det tydligen tänkbart att i översiktlig form förmedla det väsentliga. Sedan är det en annan sak att matematikundervisningen också skall ge självständig förmåga att lösa vissa typer av problem (använda logaritmer, rita kurvor, solvera trianglar, beräkna integraler etc.) och här är det tekniska en huvudsak, som förefaller att starkt undervärderas av företrädarna för "den nya matematiken".

Är det möjligt? — Givetvis krävs en omsorgsfull planering och en anpassning av det matematiska språket till elevernas förkunskaper. Men det finns inga bevis för att det inte vore möjligt, och enda sättet att få reda på svaret är att pröva. Föregående kapitel är ett försök att genomföra en dylik framställning, givetvis i betydligt större omfattning och mera skissartat än vad som vore lämpligt i skolan.

(3) En väsentlig nedskärning av nuvarande kursomfång är knappast lämplig. Den är uppbyggd för att motsvara tillämpningarnas behov och de matematiska momenten har byggts på som ett försök att tillgodose ämnets särskilda synpunkter. Med en mera schematisk behandling skulle en väsentlig tidsbesparing kunna göras, och denna tid skulle kunna användas för en översikt av den typ vi här diskuterat.

Hur skulle då en sådan reformerad matematikundervisning kunna se ut? Här skall presenteras ett idealiserat program som bortser från många problem speciellt lärarutbildningen. Det kan kanske i alla fall ge en allmän uppfattning om vad man kan tänka sig. Det har säkert många drag gemensamma med försök som gjorts och görs på många håll i världen. Speciellt har man i USA lagt ned ett stort arbete och även utsatts för mycken kritik. Jag är emellertid inte tillräckligt orienterad om försöksverksamheten och skall inte försöka göra någon redovisning.

Grundskolan

Det man här främst skall ha i minnet är att de skolbarn som börjar undervisas 1970 skall få kunskaper som passar in i samhället 1985. Vi har säkert alla märkt hur betydelsen av en exakt räkneförmåga (långa additioner och stora multiplikationer) blir mindre och att denna förmåga utnyttjas alltmera sällan. Varje snabbköp har en automatisk additionsmaskin och automatiken övertar en allt större sektor av vår tillvaro. Undervisningen bör därför syfta till att göra människan till automatikens herre, inte till att lära oss att överträffa den på dess eget område, vilket för övrigt vore dömt att misslyckas. Dessa synpunkter bör alltså vara vägledande då det gäller den praktiskt syftande matematikundervisningen, som givetvis är totalt dominerande i hela den obligatoriska skolan. Till denna bör dock fogas en förberedande ren matematikundervisning.

Undervisningen bör därför syfta till att ge förståelse för hur räknandet fungerar och hur automatiken arbetar. Det första momentet är då talsystemet och där speciellt de två olika talsystemen med baserna 2 resp. 10. Det första är datamaskinens språk, det andra vårt eget men också snabbköpsmaskinens. Det är troligen inte särskilt många, varken bland skolelever eller vuxna, som har klart för sig varför de automatiska regler vi använder när vi adderar och multiplicerar — för att inte tala om divisionen — ger rätt resultat. Orsaken är enkel: ingen har någonsin förklarat det. Om man från början gav en ordentlig grund kanske först för 2-systemet, där det fordras mindre fingerfärdighet för att skriva talen, så skulle förklaringarna ta mycket liten tid och ge en helt annan stabilitet i undervisningen. Varför inte använda symboler i förklaringarna, $100x + 10y + z$ eller $a + b \cdot 2 + c \cdot 2^2 + d \cdot 2^3$ och logiska symboler som $\Rightarrow$ och $\in$ t.ex.? Är det säkert att barn är främmande för abstraktioner av detta slag? Ju tidigare de införs framgångsrikt, ju enklare blir det när de behövs på allvar senare.

Den värdefullaste kunskapen i räkning är förmågan att uppskatta och bedöma ett svar. Blir det större eller mindre än det givna? Blir det ungefär 10, 100 eller 1 000? Handlar det om km eller mm? Kan en bil gå 100 m/sek? Det är vid förberedande diskussion av denna typ som man tränar matematisk förmåga, inte vid det mekaniska arbetet. Och det är denna slags kunskap och detta sätt att tänka man har störst nytta av både i vardagslivet och i mera vetenskapliga sammanhang.

Inom geometri ges numera i skolan väsentligen en enbart empirisk kurs. Två trianglar är kongruenta om de, när man ritat upp dem på papper och klippt ur dem, kan placeras så att de precis täcker varandra. I stort sett är detta en välgörande reaktion mot Euklides stränga regemente. Risken nu är uppenbarligen att man går för långt. Om man förklarar sammanhangen för litet, blir allt mekanisk minneskunskap som glöms fort. Dessutom blir sådan undervisning tråkig, och intrycket från universitetsundervisningen är att geometrikunskaperna nu är mycket dåliga. Det gäller för övrigt även triangelsolveringsgeometrin på gymnasiet. — Vad man gärna skulle vilja få in i lågstadiekurserna är elementär topologi. Man skulle förklara och diskutera sådant som slutna kurvor, enkla kurvor, rand, sammanhangstal, slutna ytor, genus (sammanhänger nära med g, se s. 80) osv. Allt detta är ytterligt primitiva föreställningar, mycket enklare än t.ex. begreppet rät vinkel (som ju är mycket subtilt, genom sitt samband med parallellaxiomet). Man måste här liksom när det gäller mängdläran se upp för vanföreställningen att det vi tycker verkar främmande, i någon objektiv mening skulle vara svårt. Betydelsen av topologin för modern matematik har här många gånger framhävts och sannolikt har dess begrepp och metoder trängt mycket längre 1980. Det är i detta sammanhang intressant att studera intelligenstest på s.k. visuo-spatial förmåga. De kan ha följande typ: Vilken figur av följande fem passar inte

Fig. 43

ihop med de övriga? Detta är ett typiskt problem i elementär topologi. Att det förekommer i detta sammanhang antyder det grundläggande i frågor av denna art.

Vi har tidigare berört mängdläran och därmed sammanhängande begrepp som funktioner. I Förenta staterna är det numera vanligt att man undervisar i mängdlära i ungefär femte klassen, medan man i Sverige tills vidare har lagt undervisningen i gymnasiet. Det är viktigt att åter betona att mängdläran inte är en teori utan ett sätt att systematisera en del av vårt tänkande. Undervisningen får därför inte bestå i formalistiska övningar och kataloger av nya ord utan bör vara exemplifierande. Begreppen måste användas fortlöpande och införas så tidigt att de blir en del av det naturliga ordförrådet.

I stort sett kan naturligtvis lågstadieundervisningen bara utformas på ett sätt. Det viktiga är dock att lära barnen skriva siffror, addera och minnas multiplikationstabeller och lära sig våra konventioner om mått. De variationer som kan göras inom denna ram är emellertid inte betydelselösa. För den framtida inställningen till ämnet är det viktigaste att man från början ger fullständiga förklaringar. Vi är som vuxna så inlärda och vana vid lågstadiematematiken att vi anser den självklar och oerhört enkel. Vi glömmer då att det tagit mänskligheten lång tid att konstruera systemet och att det kunde ha utformats på många andra sätt. Man får därför inte underskatta svårigheterna eller underlåta att förklara; det "självklara" kanske inte är så självklart och de "dumma" frågorna kanske bara avspeglar bristfälliga förklaringar.

Gymnasiet

Vi skall här bara syssla med den naturvetenskapliga grenen av gymnasiet. Vilka avvägningar som kan tänkas för andra studiemål får lämnas därhän.

Precis som på lågstadiet ger kravet från tillämpningarna och tillgängliga timmar i stort sett ett automatiskt svar beträffande kursinnehåll: differential- och integralkalkyl, logaritmer, exponentialfunktion, triangelsolvering, approximation. Vad som kan diskuteras är sättet för undervisning och relativt obetydliga tillägg.

Den diskussion vi fört på föregående sidor kan nu sammanfattas till följande resultat:

(1) Den traditionella matematikkursen bör ges med mindre

formell stringens men med utökad exemplifiering. Exemplen bör belysa både de fall då satser och resultat gäller och då de inte gäller. Att detta motsvarar önskemål från tillämpat håll är klart. Det är emellertid troligt att denna typ av undervisning bildar en bättre grund för undervisningen i ren matematik. Precis som vi lätt underskattar svårigheterna på lågstadiet med 10-systemet, underskattar man lätt svårigheten att översätta intuition till matematisk formalism och tvärtom. Denna översättning är en ofrånkomlig del av matematiken men är varken den viktigaste eller intressantaste. Det verkar därför naturligt att övning i detta läggs senare, då erfarenheten om vad som täcks av symboler är större. Matematiskt symbolspråk bör systematiskt införas och övas i universitetens lågstadieundervisning. Efter några års övning blir språket naturligt på samma sätt som det talade, men så lång tid tar det (som var och en kan se som rättat t.ex. trebetygsskrivningar vid ett universitet).

Vad man också måste acceptera är att aktiv, användbar kunskap kräver mycket övning av rutinkaraktär. Det är helt orimligt att man genom att ge eleverna förståelse för logiska sammanhang skall kunna kompensera träning, och man skall komma ihåg att denna del av kursen just är motiverad av sin betydelse för tillämpningar. Min tro är att man nu fört in för mycket i kurserna och att resultatet kan bli att eleverna egentligen inte lär sig något alls. Det går heller inte att bortse från att kursernas innehåll måste anpassas till elevunderlagets förutsättningar att tillgodogöra sig innehållet.

(2) Under punkt (1) finns säkert 80 % av undervisningen och denna del syftar praktiskt taget helt till tillämpningar. För att tillgodose de rent matematiska kraven kan nu ges en allmän kurs av den typ som ges i kärnfysik. Denna bör innehålla:

(a) en orientering om matematisk logik i lämplig anslutning till filosofiundervisningen;
(b) en översikt av algebraiska strukturer, speciellt gruppteori;
(c) en relativt omfattande orientering i topologi och särskilt den kombinatoriska topologin;
(d) analys med speciell betoning av de partiella differentialekvationerna i fysiken och numeriska metoder;

(e) sannolikhetsteori;

(f) berömda matematiska problem. Här kommer naturligt en orientering i talteori in.

Under 3 års gymnasiestudier kan kanske knappt en timme per vecka ägnas åt undervisning av denna typ, dvs. cirka 15 undervisningstimmar per område. Det förefaller mycket troligt att man inom denna ram kan ge en bild av modern matematik och på ett riktigt sätt introducera matematiskt tänkande. Inom detta kursavsnitt bör man även ge problem som då på ett riktigare sätt mäter matematisk förmåga än prov av traditionell typ. Man kan här jämföra sedvanliga prov med den typ av prov som ges vid nationella och internationella tävlingar.

(3) Statistikundervisningen görs mera deskriptiv, inriktas på social statistik och lägges inom ramen för samhällskunskap.

Man måste naturligtvis vara medveten om att en reform i den riktning som här antytts är en stor och komplicerad förändring. Men målet är också betydelsefullt. Försummar matematikerna att sprida sina resultat utanför sin egen krets har de svikit ett förtroende. Avsikten med vår forskning kan inte vara att bygga elfenbenstorn åt oss själva utan måste vara att ge bidrag till det totala vetandet i förhoppningen att detta i någon mening skall gagna alla. Det talas idag ofta om att avståndet i tid från forskningsresultat till allmän kunskap och praktisk användning blivit kortare. Vi bör inte acceptera att denna överföringstid för matematikens del skall vara 200 år.

Kommentar

Den tidigare citerade boken av H. Weyl är av intresse även här. Beträffande matematikens framgång i fysiken, se

E. P. Wigner, *The Unreasonable Effectiveness of Mathematics in Natural Sciences*. Communications on Pure and Applied Mathematics, Vol. 13 (1960).

Register

Ett urval Prismaböcker

MATEMATIK, NATURVETENSKAP, TEKNIK

Irving Adler: *Tänk på ett tal* 6: 50
Isaac Asimov: *Kemins historia* 12: 50
Isaac Asimov: *Blodet* 6: 50
Louis J. Battan: *Elementens raseri* 4: 50
Willem A. van Bergeijk—John R. Pierce—Edward E. David Jr: *Ljudvågorna och vi* 5: 50
Francis Bitter: *Magneter* 7: 50
George A. W. Boehm: *Den nya matematiken* 4: 50
R. L. F. Boyd: *Rymdraketer och satelliter* 4: 50
Charles V. Boys: *Underverk med såpbubblor* 4: 50
James A. Coleman: *Hur uppstod universum?* 8: 50
James A. Coleman: *Relativitetsteori för alla* 4: 50
P. J. Davis: *De stora talens värld* 8: 50
K. S. Davis—J. A. Day: *Vatten — vetenskapens spegel* 6: 50
René Dubos: *Louis Pasteur och den moderna vetenskapen* 4: 50
Donald G. Fink: *Datamaskinen och den mänskliga hjärnan* 24: 50
Donald G. Fink—David M. Lutyens: *Bakom TV-rutan* 4: 50
George Gamow: *Trettio år som skakade fysiken* 14: 50
George Gamow: *Tyngdkraften* 5: 50
Tord Hall: *Gauss, matematikernas konung* 11: 50
Robert Hooke: *Vetenskapliga försök* 9: 50
Donald J. Hughes: *Den fantastiska neutronen* 4: 50
Patrick M. Hurley: *Hur gammal är jorden?* 4: 50
Lawrence P. Lessing: *Vad är kemi?* 5: 50
H. W. Newton: *Solens ansikte* 6: 50
Ivan Niven: *Reella tal* 12: 50
C. Stanley Ogilvy: *Matematiska problem som väntar på sin lösning* 8: 50
C. Stanley Ogilvy—John T. Anderson: *Talteori för alla* 12: 50
Alfred Romer: *Den rastlösa atomen* 5: 50
W. W. Sawyer: *Vad är differential- och integralkalkyl?* 11: 50
Bengt Stolt: *Geometri — euklidisk och icke-euklidisk* 11: 50
Dirk J. Struik: *Matematikens historia* 12: 50
Åke Wallenquist: *Astronomiskt lexikon* inb. 16: 50
Åke Wallenquist: *Planeter, stjärnor och kvasarer* 7: 50

Warren Weaver: *Sannolikhet* 13: 50
Victor F. Weisskopf: *Vid kunskapens gränser* 12: 50
R. R. Wilson—R. Littauer: *I kärnfysikerns verkstad* 7: 50

EKONOMI

Erik Bolinder: *Individen och den industriella miljön* 12: 50
Sune Carlson: *Företagsledare i arbete* 9: 50
Sune Carlson—Dick Ramström: *Modern företagsledning* 19: 50
Edmund Dahlström (red.): *Teknisk förändring och arbetsanpassning* 19: 50
Erik Elinder m.fl.: *Företag och marknader* 11: 50
Stig T. L. Ericson: *Vardagens ekonomi* 8: 50
Erik Fältström: *Ledarskap och ledarutveckling* 34: 50
Peter Gorpe: *Modern administration* 13: 50
Harry G. Johnson: *Världsekonomin vid skiljevägen* 7: 50
Koncentrationsutredningen — en sammanfattning 4: 50
Börje Kragh: *Finansieringsproblem och strukturomvandling* 21: 50
Börje Kragh: *Konjunkturbedömning* 14: 50
Carl-Henrik Kreuger—Carl Birger Lovén: *Den moderna konferensen* 34: 50
Erik Ottoson—Jan Stephansson: *Det nya kontoret* 37: 50
Robert Rock: *Företag i förvandling* 15: 50
Leif H. Skare: *Handbok för kontoret* inb. 59: —
Hans Swedberg: *Sveriges utrikeshandel* 8: 50
Erik Westerlind—Rune Beckman: *Sweden's Economy* 9: 50
Krister Wickman m.fl.: *Problem kring inflation och penningvärde* 8: 50
Bo Wickström: *Konsumenten väljer märke* 19: 50
Joan Woodward: *Industriell organisation* 34: 50