Christer Kiselman

,, *I framtiden kanske det kommer att finnas barn som lärt sig digital geometri i skolan och som genom analogin kan lära sig den euklidiska.*

Jag har länge varit intresserad av matematik, men också – och ännu tidigare – av astronomi. Kemi och språk har varit andra viktiga intressen i en tidig ålder.

Min far Sam Svensson (1896–1966) var mycket intresserad av astronomi. Av honom lärde jag mig mycket, bl. a. sfärisk trigonometri. Jag läste, liksom han en gång gjort, astronomi på Stockholms högskola, men det var först min äldste son Dan Kiselman som fullföljde detta tregenerationsprojekt och blev professionell astronom.

Den euklidiska geometrin lärde jag mig i Norra real i Stockholm av min matematiklärare Bertil Broström. Det var en upplevelse, och hans undervisning har satt djupa spår, som nog denna uppsats vittnar om.

Min mest inspirerande lärare i Norra real var Karl Axnäs (1899–1984), som var min lärare i tyska. Jag lyssnade också på hans radiokurs i ryska. Han grundlade mitt livslånga intresse för språk. Detta tidigt väckta språkintresse ledde bl. a. till ledamotskap i Esperantoakademin 1989, där jag var vice preses 1998–2004. Under hösten 2007 läser jag persiska.

På Stockholms högskola (numera Stockholms universitet) studerade jag matematik, astronomi, teoretisk filosofi och ryska; senare meteorologi.

Lars Hörmander blev min handledare i matematik. Han är en stor matematiker och har sedan dess varit mitt föredöme. Genom honom kom jag in på partiella differentialekvationer, ett stort och ständigt aktuellt forskningsområde. Jag var med honom ett år vid Institute for Advanced Study i Princeton. Ett problem som han formulerade kunde lösas fullständigare i det komplexa området än i det reella, och så kom jag in på analys i flera komplexa variabler, ett annat mycket livaktigt forskningsområde, som jag sedan höll på med i många år. En uppsats jag skrivit uppmärksammades av André Martineau (1930–1972), en annan stor matematiker, som omedelbart bjöd in mig till Université de Nice, där jag tillbringade ett år som gästprofessor.

Mina år i USA och Frankrike gav mig viktiga erfarenheter både vetenskapligt och personligt.

Senare kom jag in på komplex analys och geometri även i oändligt många variabler, en fascinerande värld där mycket är annorlunda.

Någon gång under 1990-talet väcktes så mitt intresse för digital geometri. En av dem som inspirerade mig var Gunilla Borgefors, som skriver ett annat kapitel i denna bok (om tessellationer).

Jag organiserade den första Fransk-nordiska sommarskolan i matematik år 2001. En av huvudföreläsarna där var Jean Serra, en av skaparna av den matematiska morfologin. Han har sedan dess varit en ständig inspiration.

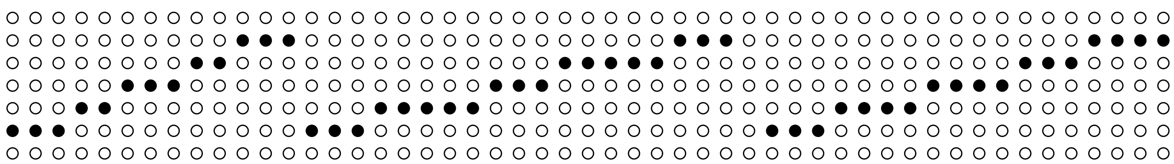
Jämförelserna mellan å ena sidan matematiska modeller grundade på reella eller komplexa tal och å andra sidan diskreta modeller är sedan dess ett tema i min forskning i geometri, optimeringsteori och matematisk morfologi, liksom i partiella differentialekvationer och partiella differensekvationer.

På den vägen är det. I den lilla uppsats som följer försöker jag förena den nya digitala världen med klassisk geometri och topologi.

Datorskärmens geometri

Punkter, räta linjer och plan har människan studerat i mer än två tusen år, och vissa kurvor, som ellipser och hyperblar, har varit föremål för vår nyfikenhet nästan lika länge. Många andra kurvor har studerats i flera hundra år. Studiet av dessa kurvor grundar sig på att vi kan göra teckningar på papper och få idéer från handritade bilder.

Med datorerna har vi nu fått ett nytt sätt att rita. På en datorskärm ser vi bilder, och bilderna består av små bildelement eller pixlar, som ögat sätter ihop till geometriska objekt. (Ordet *pixel* kommer av engelskans *picture element.*) En rät linje blir då inte det som Euklides avsåg med en rät linje, utan en ändlig mängd av små fläckar på skärmen, som ögat ändå fogar ihop till ett sammanhängande linjestycke. En kurva är likaså en ändlig mängd av bildelement. (Se figur 1.)



Figur 1. Tre pixelmängder. Vilka av dem representerar en sträcka?

Modeller grundade på de reella talen kontra diskreta modeller

Matematiska modeller som grundar sig på de reella talen (de tal som kan representeras av decimalutvecklingar) eller de komplexa talen (par av reella tal som sätts samman med den imaginära enheten i) har använts med stor framgång i fyra sekler. Ett paradexempel är Newtons gravitationslag, som säger att gravitationskraften är omvänt proportionell mot kvadraten på avståndet. Från den kan man härleda att två sfäriskt symmetriska kroppar i universum rör sig i elliptiska banor runt den gemensamma tyngdpunkten – om hastigheten inte är alltför stor; vid större hastigheter kan banorna bli paraboliska eller hyperboliska. De får inte heller störas av andra kroppar. Newton kunde från en enda lag härleda Keplers tre lagar, som hade formulerats som empiriska slutsatser från bl. a. Tycho Brahes observationer.

Men om man har mer än två himlakroppar, så kan man inte lösa ekvationerna som beskriver deras rörelse explicit. Man måste då räkna på en dator för att få fram planeternas lägen, och därmed har man, som man säger, diskretiserat problemet: man har infört ändliga tidssteg och även diskretiserat positionerna i rymden – planeterna hoppar fram från ett läge till nästa. Derivator ersätts av differenskvoter. Förresten gör man ju det även i fallet med en enda planet runt solen, trots att detta tvåkropparsproblem kan lösas explicit.

Modellen som beskriver himlakropparnas banor baseras på de reella talen men fordrar alltså ändå lösning med diskreta metoder. I andra sammanhang förekommer det att verkligheten i stället direkt modelleras diskret. Båda dessa fenomen antyder att det kan vara intressant att också göra geometrin diskret, eller digital. (Dessa två ord är i detta sammanhang nästan synonyma.)

1. Datorskärmens geometri

Finns det en geometri för dessa bilder på datorskärmen? Kan man skapa en tankebyggnad som behandlar punkter, linjer och kurvor på skärmen på samma konsekventa sätt som Euklides gjorde när han beskrev punkter, linjer och kurvor i planet, det klassiska planet? Hur skall en sådan tankebyggnad se ut? Svaret är att en sådan tankebyggnad redan finns, åtminstone delvis, även om den ännu inte är lika utvecklad som den euklidiska, som ju haft mer än tvåtusen år på sig att fulländas.

Ett billigt sätt att komma undan är att uppfatta bilderna som mer eller mindre noggranna approximationer av räta linjer eller kurvor som konstruerats med hjälp av de reella talen. Men det är inte tillfredställande. Det man bör göra är att behandla dessa ändliga punktmängder med samma exakthet som Euklides hade i sin geometri. Detta är den digitala geometrin. Den är ung jämfört med Euklides'.

Azriel Rosenfeld (1931–2004) gav år 1974 en definition av begreppet digital rät linje. Erik Melin publicerade år 2005 en annan digitalisering av räta linjer. Den påminner mer om den klassiska i det att den tar hänsyn till egenskaper som närhet. Närhet uttrycks i matematiken inte nödvändigtvis med avstånd, utan man inför något som kallas *topologi*, som uttrycker närhet och som kan tjäna samma syfte som avstånd. Vi skall snart presentera en sådan topologi, kallad Khalimskys topologi.

Vi kan också tala om kurvor i det digitala planet. Vi kan faktiskt ta vilket begrepp som helst i den euklidiska geometrin och försöka översätta det till den digitala geometrin, och se om ett visst resultat i den euklidiska geometrin blir sant i den digitala.

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

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

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

2. Att räkna med kartesiska koordinater

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

En revolution i beräkningshänseende kom med René Descartes (1596–1650). Han representerade en punkt i planet med ett par av tal (x_1, x_2) (talen kallas för punktens *kartesiska koordinater* efter hans latinska namn Cartesius). Det gjorde att han kunde räkna på alla geometriska objekt. När man talar om en punkt, så kan man också räkna på den; det är bara att räkna på de två talen x_1 och x_2.

På liknande sätt kan man räkna på linjer: en linje i planet är mängden av alla par av tal (x_1, x_2) som uppfyller en ekvation $a_1x_1 + a_2x_2 + a_3 = 0$, där inte både a_1 och a_2 är noll.

I det tredimensionella rummet representeras en punkt av en trippel (x_1, x_2, x_3) av tal och ett plan av mängden av alla tripplar (x_1, x_2, x_3) som uppfyller en ekvation $a_1x_1 + a_2x_2 + a_3x_3 + a_4 = 0$, där inte alla tre talen a_1, a_2, a_3 är noll.

Om man har två räta linjer i planet med ekvationerna $a_1x_1 + a_2x_2 + a_3 = 0$ och $b_1x_1 + b_2x_2 + b_3 = 0$, så skär de varandra i en viss mängd, och denna ges av alla lösningar (x_1, x_2) till båda ekvationerna. Kanske finns det ingen lösning (linjerna är skilda och parallella) eller oändligt många lösningar (linjerna sammanfaller) eller så finns det precis en lösning (linjerna skär varandra i en enda punkt). Man kan alltså genom att räkna avgöra vad som gäller och man kan uttrycka svaret med hjälp av de sex talen a_1, a_2, a_3, b_1, b_2 och b_3. På detta sätt kunde Descartes avgöra många geometriska frågor genom beräkningar.

3. Att representera figurer i en dator

Nästa revolution i beräkningshänseende kom med datorerna, som gjorde det möjligt att lagra och bearbeta större informationsmängder än vad man kan göra med papper och penna. Speciellt kan man lagra information om geometriska figurer.

Vi vet sedan Cartesius att två tal räcker för att beskriva en punkt i planet och tre tal för en punkt i rummet. Hur många tal behövs det för att beskriva en cirkel i planet? Jo, tre: mittpunkten ges av två koordinater och radien av ett tal.

Hur många tal behövs det för att beskriva en ellipsoid i det tredimensionella rummet? Mittpunkten ges av tre koordinater; för de tre halvaxlarna behövs ytterligare tre tal. Vidare behöver man ange riktningen hos den största axeln; till detta behövs två vinklar. För att sedan bestämma riktningen hos den näst största axeln behövs ytterligare en vinkel. Alltså kan ellipsoiden beskrivas fullständigt av nio tal. (Om man från början vet att några axlar är lika, så behövs det färre.)

Ett reellt tal kan behöva oändligt många decimaler för att beskrivas fullständigt, så vi måste nöja oss med att lagra ändligt många av dem. Att lagra nio tal med en viss rimlig precision i en dator kräver inte mycket minnesutrymme.

Men hur blir det om man vill beskriva en godtycklig mängd i planet, dvs. en mängd som får se ut precis hur som helst? En regelbunden mängd, som en ellips, kan beskrivas med några få uppgifter: mittpunktens läge, axlarnas längd och huvudaxelns riktning, men det finns många andra mängder, och i allmänhet krävs en stor samling data för att bestämma mängden.

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

Hur blir det nu om vi vill lagra information för att beskriva en godtycklig mängd i planet? Om vi som ett exempel tar en skärm som har 1 024 gånger 768 pixlar, dvs. är indelad i 1 024 små intervall på längden och 768 intervall på höjden, så innebär det att det finns $1\,024 \times 768 = 786\,432$ pixlar. Hur kan en delmängd bestående av dessa pixlar beskrivas? För varje pixel måste man tala om huruvida den ingår i mängden eller ej. Om vi skriver en etta när pixeln är med i mängden och en nolla när den inte är med, behöver vi alltså skriva 786 432 nollor eller ettor för att beskriva en godtycklig delmängd av skärmen. Det innebär att det finns $2^{786\,432}$ olika mängder. Det är ett tal som är ungefär lika med $10^{236\,740}$, dvs. måste skrivas med 236 741 siffror. Detta tal kan jämföras med massan hos all materia i universum, som några astronomer skattar till 10^{53} kg, vilket är 6×10^{79} protonmassor eller 10^{83} elektronmassor – talet 10^{83} skrivs med en etta och blott 83 nollor. Stora tal brukar ju kallas för astronomiska, men det är en metafor som inte bara har bleknat: den är helt missvisande.[1]

4. Jordans kurvsats

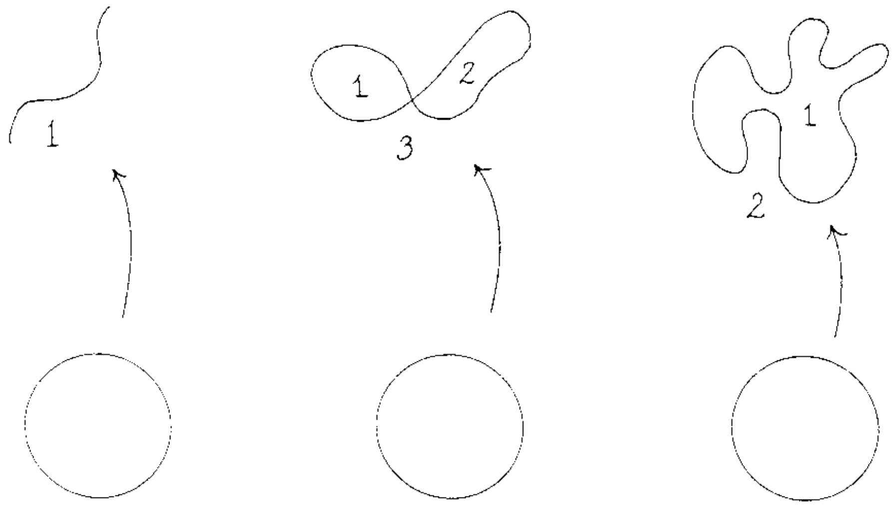
En cirkel i planet delar in detta i två delar: ett inre område, där avståndet till mittpunkten är mindre än radien, och ett yttre område, där avståndet till mittpunkten är större än radien. På samma sätt är vi vana vid att andra, mindre regelbundna kurvor delar in planet i ett inre område och ett yttre område. Vi använder staket för att stänga inne fåren och stänga ute vargarna, men staketet behöver ju inte ha cirkelform.

> [FAKTARUTA: Jordans kurvsats från 1893 säger att en enkel sluten kurva i planet delar in detta i två delar (en yttre del och en inre del som är omsluten av kurvan). Att kurvan är sluten innebär att man kan gå runt längs den. Att kurvan är enkel betyder att det inte finns några dubbelpunkter (som i en åtta).]

Vi kan alltså deformera cirkeln utan att denna egenskap att dela in planet i två delar går förlorad. Camille Jordan (1838–1922) visade 1893 en sats med just denna innebörd. Alla kurvor har inte denna egenskap; vi måste därför lägga på något antagande om kurvans utseende. Om kurvan ser ut som en åtta, så delas planet in i tre delar. Sådana kurvor får därför inte accepteras om vi skall få den önskade uppdelningen av planet. (Se figur 2.) De punkter som inte ligger på kurvan kallas för dennas *komplement*; satsen säger alltså att komplementet består av precis två delar.

Jordans kurvsats gäller alltså för kurvor som är tillräckligt lika cirklar i någon mening. Detta begrepp *tillräckligt lika* måste förstås preciseras, liksom vi bör förklara

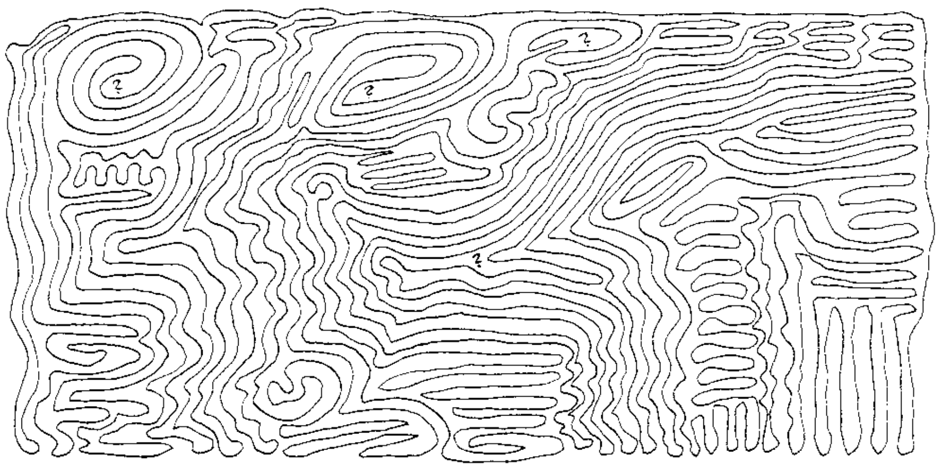
[1]Den som tror att denna slutsats beror på att universum har en ganska låg densitet ombedes beräkna massan hos ett fiktivt universum med radie lika med 14×10^9 ljusår, vilket är ungefär $1,3 \times 10^{26}$ m = 130 Ym (yottameter) och densitet lika med den hos en neutronstjärna, säg 10^{17} kg/m^3 = 100 Eg/m^3 (exagram per kubikmeter). Slutsatsen är densamma.



Figur 2. En kurva som inte är sluten; en kurva som inte är enkel; en enkel och sluten kurva. Kurvornas komplement består av en, tre respektive två delar.

vad som menas med att kurvans komplement består av två delar. För att göra det behöver vi begrepp från en gren av matematiken som kallas topologi – vi skall ta upp det i nästa avsnitt.

Beviset för Jordans kurvsats är svårt. För en cirkel är det ju lätt att avgöra om man ligger innanför eller utanför – man mäter bara avståndet till mittpunkten och ser om det är mindre eller större än radien. Men tänk på en kurva som Helge von Kochs snöflingekurva eller någon annan, oregelbunden fraktal. Om man ligger i närheten av den så syns bara en mycket taggig kurva och det är omöjligt att avgöra genom att bara titta på punkter i närheten huruvida man ligger utanför eller inuti kurvan. Detsamma gäller för en labyrint. (Se figur 3.) Egenskapen hos en punkt att ligga innanför kurvan är inte en lokal egenskap, det vill säga det går inte att avgöra om punkten har denna egenskap genom att bara undersöka en liten del av planet nära punkten; man måste se hela kurvan.



Figur 3. När är vi inne i kurvan?

5. Topologi

Jordans kurvsats gäller inte bara för cirklar utan även för kurvor som liknar cirklar. Vi behöver därför förklara vad som menas med att en kurva är tillräckligt lik en cirkel. Det matematiska ordet för detta är *homeomorf.* Syftet med att ställa villkor på kurvan är att vi vill undvika de två fallen till vänster i figur 2, alltså en kurva som inte är sluten, liksom en sluten kurva som inte är enkel, dvs. har en dubbelpunkt. Vi kräver alltså att kurvan skall vara homeomorf med en cirkel. Vad betyder nu det? Kalla kurvan för K och en cirkel för C. Vi kräver först att det skall finnas en avbildning av C in i K. I matematiken används ett kort skrivsätt för detta: man skriver avbildningens namn följt av kolon och en pil från C till K, så här:

$$f\colon C \to K.$$

Det innebär att man för varje punkt i C anger en punkt i K som den förstnämnda punkten avbildas på. Ordet *avbildning* betyder detsamma som *funktion*, men det förstnämnda är vanligare i geometriska sammanhang.

Vi kräver dessutom att denna avbildning skall avbilda två olika punkter i C på två olika punkter i K (därmed har vi uteslutit åttan i figur 2, ty för den gäller att två olika punkter på cirkeln avbildas på samma punkt). Vi kräver också att varje punkt i K skall förekomma som bild under f av någon punkt i C. Om dessa krav är uppfyllda, så säger vi att vi har en *bijektion av C på K.*

Vi har dock inte uteslutit en kurva som är ett linjestycke, ty den uppstår genom att man klipper upp cirkeln så att man får en bijektion mellan cirkeln och linjestycket. Eftersom komplementet till en sträcka bara består av en bit, gäller ju inte satsen för sådana kurvor. Tydligen fordras det ytterligare någon egenskap hos avbildningen. Denna egenskap har att göra med närhet, och kan uttryckas med avstånd, men, som vi skall se, även på annat sätt.

Det vi skall kräva är att f är en bijektion och att både f och dess invers är *kontinuerliga*, vad det nu betyder. Vi säger då att f är en *homeomorfism.*

Kontinuitet hos en avbildning f innebär att små ändringar hos ursprungspunkten (argumentet) ger upphov till blott små ändringar hos bildpunkten. Några språng får inte förekomma. Alltså: om punkten q ligger nära punkten p, så skall bilden av q ligga nära bilden av p.

Detta är ett intuitivt talesätt, som måste göras precist och hållbart, dvs. så att det tål olika generaliseringar. Detta kan ske genom att man mäter närhet med hjälp av avstånd, men allmännare genom att man studerar så kallade öppna mängder.

Om man har en mängd X och en mängd Y och en avbildning f av X in i Y, alltså $f\colon X \to Y$ med det beteckningssätt vi infört, så kan man också definiera en avbildning av delmängderna av Y in i delmängderna av X. Om nämligen B är en delmängd av Y så kan vi titta på mängden av alla punkter x som avbildas in i B. Vi kallar denna mängd för *urbilden av B* och betecknar den $f^{-1}(B)$; det är alltså mängden av alla x i X sådana att $f(x)$ tillhör B. Denna avbildning av mängder går alltså åt motsatt håll jämfört med f. Vi kan också avbilda mängder åt samma håll som f: vi kan definiera $f(A)$ som mängden av alla punkter $f(x)$ där x är en godtycklig punkt i A, en given delmängd av definitionsmängden X.

Vi skapar nu familjer av särskilt fina mängder i X och Y och kallar dem för *öppna mängder,* och undrar om urbilden av en öppen mängd alltid är öppen. Om det är fallet, så säger man att avbildningen är *kontinuerlig*. En kontinuerlig avbildning är alltså en som levererar en öppen urbild i X för varje öppen delmängd av Y.

En delmängd A av tallinjen $\mathbf{R}$ kallas *öppen* om det är så att det för varje punkt a som tillhör A finns ett litet intervall $[b, c]$ som ligger helt inne i A, där $b < a < c$. Man kan alltså röra sig fritt litet grand inne i A; därav namnet *öppen*. I planet $\mathbf{R}^2$ har vi en punkt $a = (a_1, a_2)$ i A och det skall då finnas en liten rektangel som innehåller a i sitt inre och som får rum i A. Det visar sig nu, kanske överraskande, att en avbildning $f\colon \mathbf{R} \to \mathbf{R}^2$ är kontinuerlig om och endast om den har den egenskap som vi talade om tidigare, att små ändringar i x leder till små ändringar i läget hos bilden $f(x)$.

Vi har alltså definierat öppna mängder på tallinjen $\mathbf{R}$, men vi kan göra detsamma också på linjen av hela tal $\mathbf{Z}$ (se nästa avsnitt). Därmed kan vi tala om kontinuerliga avbildningar $f\colon \mathbf{Z} \to \mathbf{Z}$ lika väl som kontinuerliga avbildningar $f\colon \mathbf{R} \to \mathbf{R}$. Det finns många analogier mellan de två, och detta är en stor fördel: om man har lärt sig något om de kontinuerliga avbildningarna $\mathbf{R} \to \mathbf{R}$ (det reella fallet) så kan man använda den kunskapen utan att göra misstag också när man sysslar med avbildningar $\mathbf{Z} \to \mathbf{Z}$ (det digitala fallet). Det är just denna analogi som gör att man vill utveckla den digitala geometrin så att den liknar den reella och som gör att vi kan känna oss lika hemma i båda. Den generation som har lärt sig euklidisk geometri får genom analogin lättare att förstå den digitala; i framtiden kanske det kommer att finnas barn som lärt sig digital geometri i skolan och som genom analogin kan lära sig den euklidiska.

Om vi har definierat en familj av öppna delmängder av en given mängd, och om familjen har vissa egenskaper, som vi inte skall beskriva här, så säger vi att vi har en *topologi* (se appendix). Ordet betyder alltså två saker, dels en familj av öppna mängder, dels studiet av sådana strukturer – jämför med ordet *algebra*, som har en liknande dubbelbetydelse.

Med hjälp av öppna mängder kan vi nu definiera begreppet sammanhängande mängd. En mängd A kallas *sammanhängande* om den inte kan delas upp i två mängder i en viss mening, nämligen på det sättet att om vi har två öppna mängder U och V sådana att varje punkt i A ligger i en av dem och sådana att ingen punkt i A ligger i båda, så gäller antingen att A ingår i U eller att A ingår i V. (Annars kan A delas upp i de punkter som ligger i U och de punkter som ligger i V, och det får inte ske om A skall vara sammanhängande.)

En *komponent* är en maximal sammanhängande mängd, dvs. en sammanhängande mängd som inte är äkta delmängd av någon sammanhängande mängd. Vi har nu alla begrepp som behövs för att förstå Jordans kurvsats, både den klassiska och den digitala.

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

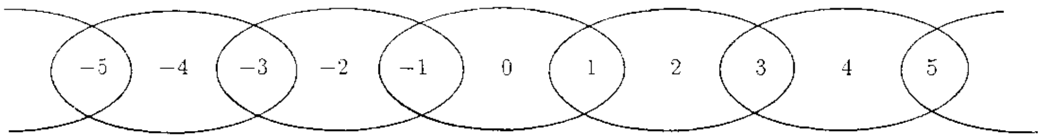
Sats (Jordans kurvsats). *Antag att $f\colon C \to \mathbf{R}^2$ är en homeomorfism av en cirkel C in i $\mathbf{R}^2$. Dess bild $K = f(C)$ delar planet $\mathbf{R}^2$ i precis två delar: $\mathbf{R}^2 \smallsetminus K$ har två komponenter, varken fler eller färre.*

De två orden *homeomorfism* och *komponent* har nu fått sin förklaring. Med $\mathbf{R}^2 \smallsetminus K$ betecknas komplementet av mängden K med avseende på planet $\mathbf{R}^2$. En *Jordankurva* är en homeomorf bild av en cirkel.

6. Khalimskys topologi

Efim Khalimsky hittade på en topologi för de hela talen **Z** som lyder som följer. Man säger att en delmängd A av heltalen **Z** är *öppen* om det är så att för varje jämnt tal $2n$ i A också dess två udda grannar $2n-1$ och $2n+1$ ligger i A. Mängden $\{1\}$ är en öppen mängd, medan $\{0\}$ inte är det. I det första fallet finns det inga udda grannar (grannarna 0 och 2 är jämna); i det andra fallet finns det udda grannar (-1 och 1), men de är inte med i mängden. Den minsta öppna mängd som innehåller $\{0\}$ är $\{-1, 0, 1\}$. Man brukar kalla en mängd vars komplement är öppet för en *sluten* mängd.[2] Mängden $\{0\}$ är sluten.

Mängden av alla jämna tal är inte öppen, medan mängden av alla udda tal är öppen, eftersom den inte innehåller några jämna tal där man kan ställa kravet.



Figur 4. Khalimskylinjen.

De jämna och udda talen spelar alltså helt olika roller i Khalimskys topologi. Vad innebär detta för kontinuiteten hos en avbildning $f\colon \mathbf{Z} \to \mathbf{Z}$? Om vi tar en godtycklig öppen delmängd B av **Z**, så skall tydligen mängden $A = f^{-1}(B)$ av alla punkter n som avbildas in i B vara öppen. Detta innebär att om ett jämnt tal $2n$ tillhör A, så skall även $2n-1$ och $2n+1$ tillhöra A, eftersom A är öppen. Med andra ord, om $f(2n) \in B$, så skall även $f(2n-1)$ och $f(2n+1)$ tillhöra B.

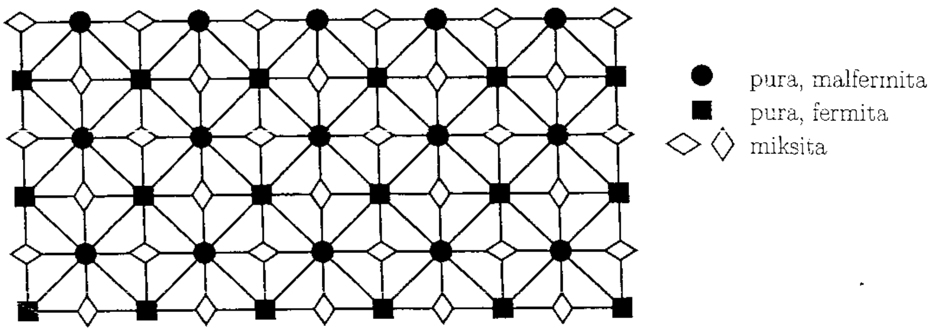
Om $f(2n)$ är udda, kan vi ta $B = \{f(2n)\}$ – det är ju en öppen mängd. Och då måste $f(2n-1) = f(2n) = f(2n+1)$, d.v.s. funktionen måste vara konstant i mängden $\{2n-1, 2n, 2n+1\}$. Om däremot $f(2n)$ är ett jämnt tal, kan vi ta $B = \{f(2n)-1, f(2n), f(2n)+1\}$ och då måste $f(2n-1)$ och $f(2n+1)$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

Vi bildar nu planet $\mathbf{Z}^2$, som består av alla par av heltal, och vi kan ge även det en topologi. En delmängd A av $\mathbf{Z}^2$ kallas *öppen* om den för varje pixel $(x_1, x_2) \in A$ också innehåller pixlarna $(x_1 \pm 1, x_2)$ om x_1 är jämnt och $(x_1, x_2 \pm 1)$ om x_2 är jämnt. Det följer av detta att om A är öppen och $(x_1, x_2) \in A$ med både x_1 och x_2 jämna, så måste A innehålla även de fyra punkterna $(x_1 \pm 1, x_2 \pm 1)$ liksom de fyra punkterna

[2]Adjektivet *sluten* används alltså i två olika betydelser i uttrycken *sluten mängd* och *sluten kurva.*

$(x_1, x_2 \pm 1)$ och $(x_1 \pm 1, x_2)$. Därmed har vi en topologi i planet $\mathbf{Z}^2$. Vi kallar även den för *Khalimskys topologi.*

Vi kan nu tala om en kontinuerlig avbildning $f\colon \mathbf{Z} \to \mathbf{Z}^2$, där både $\mathbf{Z}$ och $\mathbf{Z}^2$ ges Khalimskys topologi. Men vi behöver också en motsvarighet till cirkeln som vi använde i Jordans sats. En cirkel får man av $\mathbf{Z}$ genom att bestämma att något jämnt tal m är lika med noll. Man säger att man *identifierar* m med noll. Det betyder att man för varje tal j också identifierar talet $j + km$ med j för alla $k \in \mathbf{Z}$. Man säger att man räknar *modulo* m. Vi är ju vana att räkna klockslag modulo 12 eller 24. En resa som startar klockan 22 och tar 9 timmar slutar ju klockan 7; additionen lyder $22 + 9 = 7$ modulo 24. Låt oss med $\mathbf{Z}_m$ beteckna heltalen $\mathbf{Z}$ när man räknar modulo m. Dessa tal kan representeras av talen $0, 1, 2, ..., m - 1$. Om vi lägger 1 till $m - 1$ så får vi 0, dvs. vi har gått urtavlan runt. (Att räkna modulo ett udda tal är inte bra i detta sammanhang, eftersom de udda och jämna talen spelar olika roller, och om man räknar modulo ett udda tal, så kan distinktionen inte upprätthållas; m identiferas ju med det jämna talet 0.) Låt oss säga att $\mathbf{Z}_m$ är en *Khalimskycirkel* om m är ett jämt tal och $m \geqslant 4$.



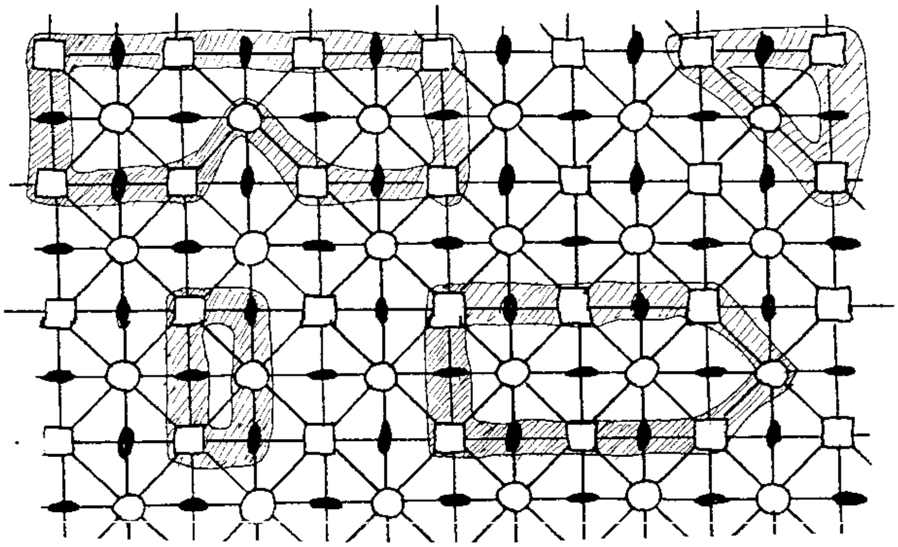
Figur 5. Khalimskyplanet. Svarta cirkelskivor representerar öppna punkter; svarta kvadrater slutna; romberna punkter som varken är öppna eller slutna.

Precis som en Jordankurva i det euklidiska planet $\mathbf{R}^2$ är en homeomorf bild av en cirkel, så är en *digital Jordankurva* en homeomorf bild av en Khalimskycirkel $\mathbf{Z}_m$ in i Khalimskyplanet $\mathbf{Z}^2$. Vi har alltså en homeomorfism $f\colon \mathbf{Z}_m \to f(\mathbf{Z}_m)$ där $f(\mathbf{Z}_m)$ är en delmängd av $\mathbf{Z}^2$. Definitionen är densamma som för de klassiska Jordankurvorna. Kurvan består av m punkter i planet $\mathbf{Z}^2$.

Man kan visa att en Jordankurva kan svänga i en punkt endast om punktens båda koordinater är jämna eller udda, och den kan svänga 45 eller 90 grader där, aldrig 135 grader.

En kurva som ges av en kontinuerlig avbildning $f\colon \mathbf{Z}_m \to \mathbf{Z}^2$ kan faktiskt svänga på det förbjudna sättet, men då på bekostnad av injektiviteten, d.v.s. den måste stanna ett tag i punkten och sedan starta på nytt, precis som en bil måste sakta in i en skarp kurva. De tre kraven injektivitet, kontinuitet hos avbildningen och kontinuitet hos dess invers kan inte förenas med en sväng i en punkt med en jämn och en udda koordinat, och inte heller med en 135-graders sväng.

Ett annat sätt att se detta är att titta på en punkts grannar. På en Khalimskycirkel $\mathbf{Z}_m$ har varje punkt exakt två grannar, en föregångare och en efterföljare. Dessa grannrelationer måste bevaras under en homeomorfism $f\colon \mathbf{Z}_m \to f(\mathbf{Z}_m)$. Men man ser att om kurvan gör en av de förbjudna svängarna, så får minst två punkter minst tre grannar. (Se figur 6.)



Figur 6. Några kurvor. Den längst ned till höger är en homeomorf bild av en Khalimskycirkel, medan de övriga tre inte är det. Och man ser att Jordans kurvsats gäller för den förstnämnda, men inte för de tre andra. Den längst upp till vänster delar in planet i tre delar. De andra två har sammanhängande komplement (de omsluter inga punkter).

7. Jordans kurvsats i det digitala planet

Vi formulerar nu Khalimskys sats.

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

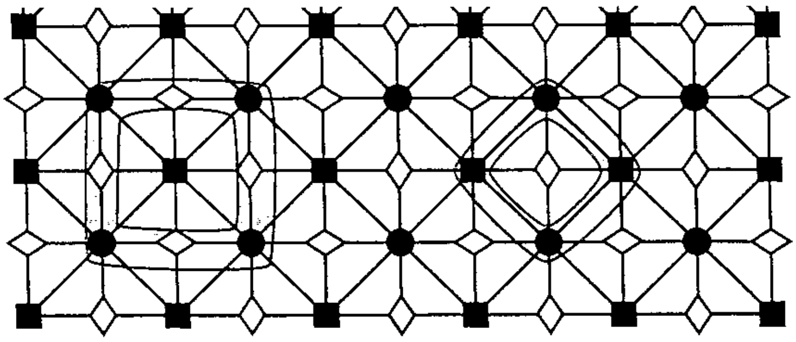
Satsen ser precis likadan ut som den klassiska Jordans kurvsats i avsnitt 4; det är topologierna som är olika.

Jordans klassiska sats är som sagt svår att bevisa. Beviset för den digitala satsen är mycket lättare. Det beror på att en digital Jordankurva bara kan innehålla ändligt många punkter.

Ett bevis består i att man visar att alla Jordankurvor utom de allra kortaste kan förkortas på ett sådant sätt att den nya kortare kurvan också är en Jordankurva. Denna kortning görs så att man kan verifiera att påståendet i satsen gäller för den längre kurvan om det gäller för den nya kortare kurvan. Efter ändligt många steg har man fått en kurva som är så kort att den inte kan kortas mera. (Se figur 7.) Men då är den av en mycket enkel typ, så enkel och kort att man direkt kan verifiera att slutsatsen gäller.

Beviset är alltså ett induktionsbevis som går över kurvans längd. Denna längd skall mätas euklidiskt, d.v.s. horisontala och vertikala segment skall tilldelas längden 1 och diagonaler längden $\sqrt{2}$. (Det går inte att ge diagonalerna längden 1 eller 2.)

Något liknande är förstås inte möjligt i det klassiska fallet; där kan ju en kurva ha en längd som är vilket positivt tal som helst, eller till och med ha oändlig längd.



Figur 7. De minsta Jordankurvorna. Det syns direkt att kurvsatsen gäller.

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

8. Appendix: axiomatisk topologi

För den läsare som förstår mängdsymbolerna $\cap$ och $\cup$ (snitt och union) skall vi i detta appendix beskriva hur man kan bygga upp studiet av öppna mängder axiomatiskt. De familjer av öppna mängder som vi definierat på den reella linjen $\mathbf{R}$ eller det reella planet $\mathbf{R}^2$ har vissa egenskaper, nämligen att unionen av en godtycklig familj av öppna mängder är öppen och snittet av en ändlig familj av öppna mängder är öppet. Vi tar nu dessa egenskaper som utgångspunkter, som axiom. Låt alltså X vara en godtycklig mängd, och låt oss ha en familj $\mathscr{U}$ av delmängder av X som vi kallar öppna.

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

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

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

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

Axiom 3. *Den tomma mängden $\varnothing$ och hela rummet X är öppna.*

Om vi har en familj $\mathscr{U}$ av delmängder av X som uppfyller axiomen, så säger vi att vi har en *topologi* på X, och att X med denna topologi är ett *topologiskt rum.* (Något avstånd behöver inte finnas.) Och om vi har två topologiska rum X och Y och en avbildning f av X in i Y, så säger vi att den är *kontinuerlig* om det är så att mängden $\{x \in X; f(x) \in B\}$, kallad *urbilden* av B, är öppen för varje öppen mängd B i Y.

Man kan visa att de öppna mängderna som vi definierat på den reella tallinjen $\mathbf{R}$ (alltså alla unioner av öppna intervall) uppfyller axiomen, lika väl som de som vi definierat på heltalslinjen $\mathbf{Z}$ (alltså alla mängder som består av udda tal och tripplar $\{2n-1, 2n, 2n+1\}$). De senare har till och med den starkare egenskapen att snittet av en godtycklig familj av öppna mängder är öppet. När det gäller de reella talen är ju snittet av alla öppna intervall $]-1/n, 1/n[$, $n = 1, 2, 3, ...$, lika med punktmängden $\{0\}$, som inte är öppen, men för Khalimskys topologi är snittet av alla öppna mängder som innehåller en given punkt också öppet.

Vi har sett väsentligen två exempel på topologier (på $\mathbf{R}$ och på $\mathbf{Z}$). Men det finns många olika topologier. Fördelen med den axiomatiska uppbyggnaden är att man kan få fram egenskaper som gäller för kontinuerliga avbildningar och sammanhängande mängder i många olika situationer.

Till slut en kommentar om axiom 3. Dess roll är att garantera att det finns minst en öppen mängd, eller två ifall rummet X inte är tomt. Men detta är onödigt. Det är nämligen tillåtet att låta J vara tom, och då följer av axiom 1 att den tomma mängden är öppen, och av axiom 2 att hela rummet X är öppet, ty $\bigcup_{j\in\varnothing} A_j = \varnothing$ och $\bigcap_{j\in\varnothing} A_j = X$. Det är inte konstigare än att den tomma summan är noll och den tomma produkten ett: $\sum_{j\in\varnothing} x_j = 0$ och $\prod_{j\in\varnothing} x_j = 1$.

9. Litteratur

Populära framställningar på svenska

Borgefors, Gunilla (2008?). Räta linjer på datorskärmen – en illustration av rekursivitet. *Nämnaren*, volym ??, sidorna ??–??. (Denna uppsats förklarar de digitala räta linjernas egenskaper.)

Kiselman, Christer (2008). Datorskärmens geometri. I: *Människor och matematik – läsebok för nyfikna.* Nationellt centrum för matematikutbildning. (Denna uppsats innehåller en intressant referens till en populär beskrivning av digital geometri och topologins användning inom denna.)

En populär framställning på annat språk

Christer Kiselman (2003). La geometrio de la komputila ekrano. I: *Internacia Kongresa Universitato*, 1–12. (Ed. Michela Lipari.) Rotterdam: Universala Esperanto-Asocio, 2003. 83 ss. (Denna uppsats är en föregångare till den ovan nämnda.)

Vetenskapliga texter

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

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

Kiselman, Christer O. (2000). Digital Jordan curve theorems. *Discrete Geometry for Digital Imagery,* 9th International Conference, DGCI 2000, Uppsala Sweden, December 13–15, 2000. (Eds. Gunilla Borgefors, Ingela Nyström, Gabriella Sanniti di Baja.) Lecture Notes in Computer Science **1953**, ss. 46–55. Springer.

Kiselman, Christer O. (2004). *Digital Geometry and Mathematical Morphology.* Föreläsningsanteckningar. Uppsala universitet, Matematiska institutionen. Kan hämtas hos `www.math.uu.se/~kiselman`

Melin, Erik (2005). Digital straight lines in the Khalimsky plane. *Mathematica Scandinavica*, **96**, 49–64.

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

Författarens adress: Uppsala universitet, Matematiska institutionen, Box 480, 751 06 Uppsala
Telefon: 018-4713216 (till universitetet); 018-300708 (hem); 0708-870708 (mobil)
Datoradress: `kiselman@math.uu.se` URL: `www.math.uu.se/~kiselman`