Förord

Föreliggande text är resultatet av alltmer bearbetade föreläsningsanteckningar till en fördjupningskurs vid Göteborgs universitet som jag har hållit ett antal (otal) gånger de senaste tio åren. Kursen, och därmed även denna text, vänder sig matematikstudenter som vill se en del av det man gärna hoppar över i de ofta kanske väl strömlinjeformade kurserna på grundnivå. Särskilt blivande ämneslärare ...

De formella förkunskapskraven är grundläggande kurser i analys och algebra, men för större delen av texten räcker det faktiskt med (goda) gymnasiekunskaper; främst är det en del exempel som man får hoppa över ... Speciellt tror jag att texten kan vara läsbar, och förhoppningsvis även läsvärd, för arbetande skollärare som vill fortbilda sig. Vidare skulle nog valda delar kunna användas för (specialarbeten på gymnasiet ???).

Trots att så vitt skilda saker som blblab tas upp i denna kurs så menar jag att man skönja en röd tråd i framställningen. I kapitel 1

diskuterar jag hur logik används inom matematik och jämför vardagligt och matematiskt språk. Vidare försöker jag understryka att det faktum att matematiskt resonerande i princip är (ska vara) helt logiskt stringent, å andra sidan i både muntlig och skriftlig kommunikation man ofta av praktiska skäl men även förståelseskäl har inkonsekvenser balblabla.

balbla logik grundvalen för mateamtik som deduktiv vetenskap, axiomsystem blalbla. Vi presenterar en rad axiomsystem (axiomatiska teorier) varav somliga behandlas mer utförligt i senare kapitel. (ZF, Peano, grupp ring, kropp, absolut geometri)

Nästan all existerande matematik kan formuleras inom Zermelo-Fraenckels mändteori som beskrivs axiomatiskt i kapitel 2. Blablba mängder balbla träning i praktiskt logiskt handlag. blabla

Utifrån de naturliga talen Peano definierar vi heltal, rationella tal etc och reella blabla komplexa. och visar att $\mathbb{C}$ är algebraiskt slutet.???

I kapitel 4 går vi igenom axiom för absolut geometri A blabla Euklides fast uttryckt i moderna termer som reella tal, grupp av avbildningar, etc.

Ett av huvudmålen här är att visa att parallellaxiomet PA är oavhängigt teorin A. I den euklidiska modellen är sant. I det avslutande kapitel 5 går vi igenom Poincare-modellen för ickeeuklidisk hyperbolisk geometri.

balblaba bra träning i modelltänkande, träning på komplexa tal, räkning, vidare utmärkt att ha den modell i tanken om man fortsatta studier i geometri, enär standardexemplet på en mfald med konstant negativ krökning.

Några ord. När jag talar om att fylla igen luckor i de traditionella kurserna så menar jag inte att dessa luckor är brister i de kurserna. Tvärtom talar både pedagogiska, begreppsmässiga och historiska skäl för att man gör som man gör. Många anser att ett avsnitt om logik bör inleda alla universotetsstudier i matematik. blbala OK men min erfarenhet är att många studenter inte kan tillgodogöra sig balab innan man sett det hela i aktion. Pss sätt som att man inför reella talen som en utvidgnin av den ordnade kroppen $\mathbb{Q}$ sådan att supremumuegenskapen gäller. Detta är med säkerhet en klok policy. De flesta studenter skulle nog omedelbart hängas av om man inledde med att blablab. För övrigt blbal rationella tal i skolan utan att bekymra sig om existens och entydighet. Vidare tycker åtminstone jag att man skulle kunna tillåta sig, mer än man gör idag, att räkna analys utan att vara så petig med ϵ och δ-exercisen. Man lär sig mer

För övrigt är detta naturligt sett ur ett historiskt perspektiv

Jag vågar nog sticka ut hakan och säga att all matamtisk inlärning även för professionella matamaiker]balbla komma in i det och se hur det funkar parallellt med att man sätter sig in i de teoretiska grunderna. Man “förstår “ en formel parallellt med att man börjar känna igen den, och man vill gärna se konkreta enkla exempel på vad den underliggande teorin betyder balbla

I grundläggande analyskurs inför man reella tal balba supremumuegenskapen utan att diskutera varför en sådan utvidning av $\mathbb{Q}$ finns och ännu mindre entydghet. I själva verket så lär man sig större delen av skolmatematiken genom erfarenhet utan att i alltför hög grad reflektera över varför det ena och det andra funkar. T ex inför man rationella tal utan att oroa sig för balbabla. Tur är nog det för balbalbla

Man vänjs in i det matematiska språket.

Jag minns min förtvivlan över inkonsekvensen i beteckning av minustecknet dels står för subtraktion, dels självständigt för att beteckna ett negativt tal ????? Ett annat exempel är ett uttryck som $f(x) = 4$, som kan beteckna både att f är 4 för alla x, för det fixa talet x eller ekvationen blaba, dvs mängden av x sådana att $f(x) = 4$. blablablabla

. Om man började diskutera alla dessa tänkbara tolkningar innan man införde funktionsbegreppet så skulle det nog balbla

Ett av huvudsyftena är att framställa matematiken som en dynamisk vetenskap stadd i utveckling; särskilt för blivande skollärare kan betydelsen av denna insikt inte nog betonas. Alla lärare vid högskolan känner igen frågor som “Hur kan man forska i matematik?”, “Var det inte känt redan för flera hundra år sedan hur man räknar ?” etc etc. Matematik som en både empirisk och deduktiv vetenskap.

Delar av min framställning är kraftigt inspirerad av ett kompendium “Något om logik ....” av Olof Hanner.

Tacksam

Hasse Carlsson, Ulf Persson, Jan Smith, Thomas Weibull, ??? ??? samt diverse studenter

Götebirg 99 01 10
Mats Andersson

Innehåll

Förord i

Inledning 1

Kapitel 1. Något om logik 5
1. Matematiskt och vardagligt språk 5
2. Satslogik 9
3. Kvantifikatorer 19
4. Matematiska teorier 23
5. Predikatlogik 28
6. Ytterligare övningar till kapitel 1 37
Några kommentarer 40

Kapitel 2. Mängdteori 41
1. Zermelo-Fraenkels axiom för mängdteorin 41
2. Relation och funktion 47
3. Ekvipotens 52
4. Ordningsrelationer och tal 56
5. Uppräkneliga mängder 58
6. Urvalsaxiomet 61
7. Transfinit induktion 64
8. Ytterligare övningar till kapitel 2 69
Några kommentarer 70

Kapitel 3. Talsystem 71
1. De naturliga talen 71
2. Heltal 75
3. Rationella tal 77
4. Reella tal 78
5. Komplexa tal 85
6. Ytterligare övningar till kapitel 3 87
Några kommentarer 87

Kapitel 4. Geometri 89
1. Axiom för absolut geometri 89
2. Den euklidiska modellen för absolut geometri 94

3. Grundläggande satser i absolut geometri 98
4. Euklidisk geometri, likformighetsteori 107
5. Längd och area i euklidisk geometri 113
6. Några klassiska satser i euklidisk geometri 117
7. Konstruktioner med passare och linjal 123
8. Sfärisk geometri 126
9. Ytterligare övningar till kapitel 4 128
Några kommentarer 129

Kapitel 5. Hyperbolisk geometri 131
1. Möbiusavbildningar 131
2. Poincaré-modellen för hyperbolisk geometri 136
3. Längd i Poincaré-modellen 140
4. Area i Poincaré-modellen 145
5. Ytterligare övningar till kapitel 5 149
Några kommentarer 151

Lösningar och kommentarer till några av övningarna 153

Litteraturförteckning 161

Inledning

Att tala om matematik som en självständig vetenskap är ganska nytt. Historiskt har m. varit intimt kopplad till annan kulturell verksamhet som musik, ingenjörsvetenskap, fysik etc.

Behovet av geometri, att förstå samband mellan längd, area och volym dyker naturligt upp i jordbrukssamhällen, både för lantmäteri, byggnationer samt astronomiska studier, vilka är viktiga för att t.ex. upprätta kalendrar etc. Geomtriska kunskaper uppstod parallellt i de stora flodkulturerna i olika världsdelar ett par tusen år före vår tideräkning och framåt. Man kunde t.ex. beräkna omkrets och area av en cirkel med goda närmevärden till π, och man kände till det vi idag kallar Pytagoras sats. I dessa kulturer utvecklades även aritmetik och algebra; det var nödvändigt att kunna addera och multiplicera tal och att lösa ekvationer.

Vid denna tid var dock matematiken empirisk. På 600-talet började man i Grekland (Thales från Miletos) att försöka att med logiska argument härleda vissa påståenden från andra. Detta utvecklades under de kommande århundradena av en rad tänkare, och man fann det naturligt att försöka härleda alla resultat från att begränsat antal grunsantaganden, axiom. Denna utveckling samanfattades i Euklides verk Elementa som fullbordades ca år 300. Detta verk kom att bli normgivande i över 2000 år, och den axiomatiska metod som Euklides använder är i princip den samma som man använder i matematik idag. Detta innebär ingalunda att matematiken inte har utvecklats sedan Elementa. För en grundlig historisk översikt hänvisas till t.ex. [**?**] och [**?**]. För geometri balba se särskilt [**LÅL**]. Vi ska bara här ge några enkla exempel.

Den matematik man lär sig i skolan, inte bara geometri, är resultatet av tusentals år av mänsklig kulturell utveckling.

Antag att vi har infört de positiva heltalen och förstått vad addition innebär. Då är det lätt att säga vad man man menar med $m{\cdot}n$, nämligen $n{+}n{+}\ldots{+}n$ (m stycken termer). Däremot är det absolut inte självklart att $m \cdot n = n \cdot m$.

UPPGIFT 0.1. Försök bevisa att multiplikation blir kommutativ.

Man bör ha respekt för den grad av abstrakt tänkande det innebär att gå från positiva heltal, till att införa talet 0, och negativa tal, och sedan komma fram till sådana saker som att produkten av två negativa tal blir ett prositivt tal.

An annan sak man exercerar i i de högre årskurserna i skolan är symbolalgebra. För att förstå vidden av kan man titta på ett enkelt exempel.

Exempel 0.1. Antag att sju apelsiner kostar sex kronor mer än tre bananer, samt att tre apelsiner och sex bananer tillsammans kostar 39 kronor. Vad kostar då en apelsin och en banan?

För en person som inte har fätt lära sig symbolalgebra är nog detta en hopplös uppgift. Om man dessutom får veta att priset är i hela kronor, kan man förstås lösa det hela genom att pröva sig fram; det finns ju då bara ändligt många fall att gå igenom.

Med symbolalgebra skriver man upp ekvationssystemet

$$\begin{aligned} 7a - 3b &= \quad 6 \\ 3a + 6b &= \quad 39 \end{aligned}$$

och genom sedvanligt manipulerande finner man lätt att $a = 3$ och $b = 9$. $\square$

På gymnasiet studerar man grundläggande analys. balbla Descartes 1600 talet analytisk geometri ablblaa syntetisk geometri

studerar man inte längre syntetisk geometri,

detta naturligt men man går miste om mateamtik utan tal !!!!

Analys, Leibnitz, Newton

Obs återigen empirisk i meningen att man kände sig fram till resultaten.

Först på 1800-talet genom Cauchy Weierstrass axiomatisk grnd för analys ϵ och δ

Sent 1800 tal Cantor Frega mängdlära Russels paradox !!

Vad är bevis, oändlighet ??

Detta får däremot inte tolkas som att matematiken är fullbordad idag. Detta kanske kan tyckas självklart, men varje mateamtiklärare på högskolan har säkert fått frågan från nybörjande studenter om man verkligen kan forska i matematik, och om inte allt redan är känt. I själva verket har matematiken haft en allt snabbare utveckling. Fram till sekelskiftet menade man att en duktig matematiker kunde behärska strängt taget all känd matematik; idag har man enligt AMS lkssificering ca 100 ?? underavdelningar, och en riktig duktig matematiker kanske kan behärska en handfull av dessa.

1. Matematik handlar inte bara om tal,

2. matematik är en vetenskap stadd i ständig och kraftig utveckling.

3. Forskning balbal

4. Mateamtikens utveckling beror av samhälleliga behov, och av tekniska framsteg.

5. Mycket av det man lär sig i skolan är resultatet av tusentals år av utveckling.

6. M är både en empirisk och deduktiv vetenskap

Påstås att lantmätare kände till formeln för vinkelsummans excess i stora trianglar (kolla Hörmander) långt före Gauss-Bonnets sats

5: Tal har vi redan nämnt. Ekvationer med obekanta. blablablbala. Ex

4. Redan nämnt jordbrukets behov av geometri. Kartritning etc

Mekanik analys newton.

Numerisk analys, datorer. Meningslöst att utveckla teorier för hur man numeriskt löser differentialekvationer om det skulle ta en miljon år att utföra beräkningarna för hand.

3. Forskningen i matte utvecklats explosition?? Finns tillgänglig tid, gott om papper, ljus etc. Många av de problem som sysselsatt forskare sedan antiken, som vinkelns tredelning, cirkelns kvadratur, parallellaximets oavhängighet m.m. löstes på 1800-talet.

balblaa problem Hilberts problem ???

Fermats stora sats bevisades av Andrew Wiles 199?.

Varje år bevisas större och mindre satser som varit kända blabla utvecklas nya teorier, t ex inom numeriska analys etc. Vidare finslipas teorierna. T ex Galois teori kan numera läsas som fördjupningskurs balbla

GÖDEL

1: Varje dominobricka täcker två rutor på ett schackbräde. Kan man täcka över alla rutor utom två diagonalt motstående hörn med 31 dominobrickor? Ge bevis eller motbevisa!

2: Königsbergs broar.

3: Stympade schackbräden. Ett SS är ett schakbräde med sidan 2^n sådant att

KAPITEL 1

Något om logik

Ordet logik har som så många begrepp en rad olika innebörder. Dels är det en beskrivning av hur man resonerar och argumenterar i vetenskapliga och andra formella sammanhang, t ex i matematik, dels är det benämning på en vetenskap, som handlar om att formalisera logiken och studera den för dess egen skull. Vidare används ordet ofta i dagligt tal, t ex som "det är ju ingen logik i detta" med vilket man kanske bara menar "det finns ju inget vett i detta" eller något liknande.

I det här kapitlet ska vi diskutera hur matematiskt språk och matematisk argumentation (bevisföring) fungerar. Vi diskuterar *satslogik* och *kvantifikatorer*. Vidare tar vi upp centrala begrepp som *teori*, *axiom(system)*, *sats*, *modell* och *logisk sanning*. Vi ger även exempel på ett antal viktiga axiomatiska teorier. Några av dessa kommer att behandlas närmare i kommande kapitel. Vi avslutar med ett avsnitt om *(formell, första ordningens) predikatlogik*, där vi mycket kortfattat diskuterar *formella härledningar*, *fullständighetssatsen* samt *kompakthetssatsen*. Från den sistnämnda följer det att inte alla teorier kan uttryckas i predikatlogik.

1. Matematiskt och vardagligt språk

Till skillnad från vad somliga kanske tror är matematikernas språkbruk inte fullständigt stringent och konsekvent, utan precis som i vardagsspråket vimlar det av inkonsekvenser i uttryckssätt, beteckningar etc. Skälet är att man i första hand strävar efter ett smidigt språk där de grundläggande idéerna i det man vill förmedla framgår så tydligt som möjligt. Av samma skäl finns det ofta logiska luckor i bevisen (som läsaren förutsätts att själv fylla i) och det förekommer ofta underförstådda outtalade antaganden. Det senare är särskilt vanligt vid muntlig kommunikation när mottagaren, dvs den som lyssnar, har möjlighet att vid behov fråga. Det händer t ex ofta att man pratar om "en funktion" och underförstår att det rör sig om en reellvärd funktion, definierad på

$\mathbb{R}$, eller t.o.m. att den är deriverbar. Det är dock viktigt att understryka att det i princip är något välbestämt man pratar om. I alla formella sammanhang däremot, som när man definierar nya begrepp, så brukar man vara mycket strikt.

Att införa en *definition* av ett nytt begrepp innebär i princip bara att man inför ett förkortat skrivsätt. T ex säger man "en *positiv funktion*" i stället för "en funktion som bara antar ickenegativa (sic!) värden", och π istället för "halva omkretsen av en cirkel med radien 1". Vidare säger man "f är kontinuerlig på $\mathbb{R}$" istället för " till varje x och varje $\epsilon > 0$ så finns ett $\delta > 0$ sådant att $|f(x) - f(y)| < \epsilon$ om $|x - y| < \delta$". Att hitta de rätta begreppen innebär ofta betydande matematiska framsteg. Exempelvis hade man under lång tid haft en intuitiv känsla för vad en kontinuerlig funktion skulle vara, men den strikta definitionen dröjde till mitten på 1800-talet. Först när man har en definition är det möjligt att ge strikta bevis för påståenden om kontinuerliga funktioner.

Ett utslag av den stora vikten man tillmäter "känslan" för begreppen kan vara att många benämningar på nya matematiska begrepp hämtas ur vardagslivet, såsom grupp, ring, kärve, grodd, stjälk, fiber, överlagring, (vektor-)bunt, etc.

Strikta definitioner krävs förstås även i andra än matematiska sammanhang, t ex i juridiska. Om en lagtext handlar om "fordon" så måste det vara möjligt att avgöra om en sparkstötting är ett fordon eller inte. I vardagsspråket är man normalt inte lika petig, eller åtminstone så skiljer sig ofta definitionerna åt mellan olika talare.

UPPGIFT 1.1. Försök att ge definitioner av några vanliga begrepp, t ex "karbad", "motorväg" och "bil".

De flesta vanliga meningar har inte bara ett lexikaliskt innehåll utan även ett värderande, en aspekt. Det är ju skillnad på "våra barn" och "andras ungar". Jämför också meningarna "I lördags kom paketet" och "I lördags kom äntligen paketet". Vidare har t ex orden "och" och "men" exakt samma logiska funktion men inte bara i dagligt tal utan även när man genomför en matematisk argumentation kan det ha stor betydelse för förståelsen om man väljer det ena eller det andra, se vidare i avsnitt 2. Även en mening som saknar värdeladdade ord (som "ungar" eller "äntligen") kan innebära en värdering. Om det står "Idag var kapten nykter" i loggboken så tolkas det naturligen som att kaptenen ofta är onykter. Uttalandet "På varenda sida jag läst i den här boken har jag hittat fel" är ju sant om man inte läst någon sida alls, eller möjligen bara någon enstaka och råkat hitta ett fel där. Inte desto mindre är ju uttalandet under sådana omständigheter vilseledande.

Det är skillnad på en benämning av (eller ett uttryck för, eller ett namn på) en företeelse och själva företeelsen, dvs det den benämner. T ex är $1/2$, $\frac{1}{2}$, $\int_0^1 x dx$ och $2 - 3/2$ olika uttryck, men de betecknar samma tal. På samma sätt är "Rom", "Italiens huvudstad" och "den eviga staden" olika uttryck för samma sak. I såväl vardagligt som matematiskt språk är det vanligt att man inte skiljer på benämningen och det den benämner. Jämför t ex utsagorna "Maria är en vacker flicka", "Maria är ett vackert namn" och "Maria har fem bokstäver".

Uppgift 1.2. Fundera på vad som skiljer dessa åt!

Den här typen av tvetydighet utnyttjas gärna av korsordsmakare, som till korsordsnyckeln "Stad i Tjeckoslovakien" kan ha lösningen "Prag" såväl som "Oslo".

Vi ska nu titta på några exempel inom matematik.

Exempel 1.1. Betrakta formeln

$$3x = 1 + 1/2 + 1/3 + 1/4. \tag{1.1}$$

I vanligt matematiskt språk säger vi gladeligen "Högerledet i formeln ovan är större än 2." eller "Högerledet i (1.1) är större än 2." trots att det man egentligen menar är att högerledet i den formel som betecknas med (1.1) är en beteckning på ett tal som är större än det tal som 2 betecknar. □

Exempel 1.2. Här är ett annat exempel på tvetydighet i matematiskt språk. Om p och q är polynom så kan uttrycket (formeln)

$$p(x) = q(x) \tag{1.2}$$

betyda en rad olika saker beroende på sammanhanget:

- (i) Att p och q är lika som polynom, dvs att de har precis samma koefficienter.
- (ii) Att x betecknar en fix punkt och polynomen har samma värde i just denna punkt.
- (iii) Att de båda polynomen är lika i varje punkt x.
- (iv) Ekvationen $p(x) = q(x)$, vilken är en beteckning för mängden av alla reella x (eller komplexa tal x eller något annat, vilket får framgå av sammanhanget) sådana att $p(x) = q(x)$, dvs
$$\{x \in \mathbb{R};\ p(x) = q(x)\}.$$

Om man vill understryka att det är det första fallet som avses kan man ibland välja att skriva $p = q$ istället, men detta är inte alltid möjligt. Vill man uttrycka att p är polynomet $x^2 + 1$ så får man skriva $p(x) = x^2 + 1$; vidare kan även $p = q$ tolkas som en ekvation, dvs fall (iv). (I och för sig är fallen (iii) och (i) ekvivalenta om det är fråga om polynom över de reella eller komplexa talen, men detta är då en sats. Dessutom är det inte sant för alla kroppar; t ex är $x^2 = x$ för alla x över kroppen $\mathbb{Z}_2$ men likväl är x och x^2 olika polynom.) □

Uppgift 1.3. Fundera på skillnaden mellan ekvationerna $x^2 = 4$ och $2x^2 = 8$.

Exempel 1.3. I matematisk text skriver man ofta "...på cirkeln $x^2 + y^2 = 2$ ..." istället för "...på cirkeln $\{(x, y) \in \mathbb{R}^2;\ x^2 + y^2 = 2\}$ ..." □

Exempel 1.4. I envariabelanalysen inför man ofta en ny symbol för en sammansatt funktion, dvs om $x = x(t)$ och $f(x)$ är en funktion av x så sätter man t ex $g(t) = f(x(t))$ för den sammansatta funktionen. Kedjeregeln kan då skrivas $g'(t) = f'(x(t))x'(t)$. Ibland blir dock detta så otympligt (i synnerhet om man har att göra med funktioner av flera variabler) att man istället låter f beteckna såväl funktionen av x som av t. Man kan då skriva kedjeregeln

$$\frac{df}{dt} = \frac{df}{dx}\frac{dx}{dt}.$$

I själva verket är inte detta en inkonsekvens om man betraktar f som en funktion inte av x utan som en funktion av punkter, där t och $x = x(t)$ är olika beteckningar (koordinater) för samma punkt. Hur som helst blir en framställning ofta tungrodd om man inte tillåter sig att hoppa lite mellan dessa olika betraktelsesätt. □

Låt oss fundera lite på vad det innebär att lösa en ekvation eller ett system av ekvationer. Att *lösa* ekvationen $x^2 = 4$ innebär att uttrycka mängden $\{x;\ x^2 = 4\}$ på något enklare sätt än det givna. Eftersom $x^2 = 4$ om och endast om $x^2 - 4 = 0$, kan vi dra slutsatsen att $\{x;\ x^2 = 4\} = \{x;\ x^2 - 4 = 0\}$. Genom ytterligare omskrivningar får vi $\{x;\ x^2 = 4\} = \{x;\ (x-2)(x+2) = 0\} = \{x;\ x-2 = 0 \text{ eller } x+2 = 0\} = \{2, -2\}$. I det här fallet är det ganska klart att uttrycket $\{2, -2\}$ är en enklare beskrivning av lösningsmängden än den ursprungliga. Ofta finns det oändligt många lösningar och då får det aktuella behovet avgöra vilket framställningssätt av lösningsmängden som är att föredra.

Exempel 1.5. Om vi vill "lösa ekvationen $x+y+z = 1$" i $\mathbb{R}^3$ så får vi en lösningsskara med två parametrar, t ex $x = t,\ y = s,\ z = 1-s-t$ där s och t löper över $\mathbb{R}$. □

Uppgift 1.4. Antag att vi vill lösa ekvationen $x^2 = 4$, och betrakta följande argumentation: Om $x^2 = 4$ så är $Dx^2 = D4$ och alltså $2x = 0$ dvs $x = 0$. Alltså; om x är en lösning (rot) till ekvationen så måste

$x = 0$. Men 0 löser inte ekvationen, medan däremot 2 och -2 gör det? Förklara detta!

På samma sätt som vi sett att "lösa ett ekvationssystem" inte alltid har en självklar innebörd, kan man fundera över vad det betyder att "beräkna en integral". Det finns ofta en underförstådd innebörd, som "uttryck värdet av integralen med hjälp av ändligt många elementära funktioner" eller "ge ett närmevärde till integralen".

Som nämnts tidigare så innehåller matematiska bevis ofta luckor eller utelämnade steg. Sådana luckor kan markeras med uttryck som "man inser nu (lätt) att", "det är nu klart att", " det följer nu att", "man kan nu övertyga sig om att" etc. Vilka och hur stora steg i argumentationen man kan hoppa över beror förstås på vem man förväntar sig ska läsa texten. Ofta utelämnas relativt långa argument om man kan anta att läsaren är så van vid teorin att hon själv kan (eller vet att hon kan om hon vill) fylla i luckorna. Detta gör man delvis för att spara plats men framför allt för att öka överskådligheten. Om man skulle skriva vetenskapliga artiklar med lika mycket detaljer som i en elementär lärobok, så skulle artiklarna bli oerhört långa och mycket svårlästa. Denna anpassning till vem mottagaren är gäller förstås inte bara för matematisk text. Även om hon är obekant med den speciella företeelsen man ska beskriva, kan man i allmänhet ändå uttnyttja en betydande gemensam kunskapsbas.

Uppgift 1.5. Försök att beskriva så enkelt som möjligt hur man åker kollektivt från matematiska institutionen till Saltholmen, om den förväntade mottagaren är en normal tjugoårig stockholmare. Fundera sedan på hur mycket extra förklaringar och detaljer man måste ta med om mottagaren istället (hypotetiskt) är en person från en helt annan miljö, t ex från medeltiden.

2. Satslogik

Satslogik (*propositional calculus* på engelska) handlar om hur man kan sätta samman olika s.k. satser (eller *utsagor* eller *påståenden*) med *konnektiver* till nya satser, och regler för hur sanningsvärdet av den sammansatta satsen beror av dess delar. Man utgår från vad man intuitivt menar med "tänka logiskt", och satslogiken är därmed ett sätt att formalisera (beskriva och hitta regler för) detta. Som vi ska se så följer man i det vardagliga språket inte alltid dessa regler.

Med *(satslogisk) sats* menas här någonting som har sanningsvärdet *sant* eller *falskt*. Satser betecknar vi med bokstäver (satsvariabler) $A, B, C, \ldots$ eller med bokstäver med index $A_1, A_2, \ldots$ etc. De fem vanliga konnektiven är $\neg$, $\wedge$, $\vee$, $\rightarrow$ och $\leftrightarrow$. De två sista skrivs ibland även $\Rightarrow$ respektive $\Leftrightarrow$. Det är lätt att beskriva satslogikens syntax, dvs hur man får lov att bilda nya satser från tidigare, och jämsides med att vi gör detta så diskuterar vi hur man uttrycker motsvarande bildningar på svenska i matematiska sammanhang. Vi gör även några jämförelser med vardagsspråket.

Anmärkning 2.1. Ska man vara riktigt formell så ska man också införa strikta regler för hur man ska sätta parenteser; dessa behövs t ex för att skilja på $(\neg A) \rightarrow B$ och $\neg(A \rightarrow B)$. Man kan också införa regler som säger att $\neg A \rightarrow B$ ska tolkas som $(\neg A) \rightarrow B$ osv., men vi hoppar över detta och sätter ut parenteser bara så att det ska bli klart vad som avses. □

Anmärkning 2.2. I vanlig matematisk text använder man logiska symboler bara i undantagsfall. Däremot kan de vara praktiska i mer informella sammanhang, t ex när man ska förklara något vid svarta tavlan. Ibland är det praktiskt att belysa vissa satser och definitioner med hjälp av logiska symboler, t ex för att förklara skillnaden mellan kontinuitet och likformig kontinuitet, se exempel 3.6 nedan. □

Negation. Om A är en sats så är $\neg A$ en ny sats, *negationen* av A, som har egenskapen att $\neg A$ är sann om A är falsk och $\neg A$ är falsk om A är sann. Vi kan uttrycka detta med en sanningstabell som

A	$\neg A$
S	F
F	S

Notera att $\neg(\neg A)$ har samma sanningsvärde som A.

En vanlig svårighet i de begynnande universitetsstudierna i matematik är att korrekt negera utsagor. Detta är viktigt bl.a. för att kunna genomföra s.k. motsägelsebevis, se nedan. Om A är ett språkligt uttryck så kan man alltid uttrycka $\neg A$ genom att sätta frasen “Det är inte så att ...” före A. I allmänhet är dock detta till föga hjälp och man måste göra någon språklig omskrivning för förstå innebörden av negationen. Om A står för “Lasse är längre än Lisa” så kan man uttrycka negationen som “Det är inte så att Lasse är längre än Lisa” men det är förstås naturligare att säga “Lasse är jämnlång med eller kortare än Lisa”.

Exempel 2.3. Antag att a_n är en reell talföljd och att vi vill negera utsagan $\lim_{n\to\infty} a_n = 0$. Återigen kan vi förstås köra med "Det är inte så att ...", men detta hjäper oss inte förstå vad negationen innebär. Ett vanligt misstag är att säga att negationen är $\lim_{n\to\infty} a_n \neq 0$. Felet består i att det även kan vara så att gränsvärdet inte existerar alls. Ett korrekt uttryck för negationen är t ex "a_n saknar gränsvärde då $n \to \infty$ eller så är $\lim_{n\to\infty} a_n = B$ där $B \neq 0$." □

Anmärkning 2.4. Betydelsen av en mening kan variera beroende på var man sätter ordet "inte". Jämför meningarna "Hon bad mig inte att göra det" och "Hon bad mig att inte göra det". Här finns dessutom en fälla vid översättning till (eller från) engelska. På engelska får vi "She didn't ask me to do it" respektive "She asked me not to do it". □

En ofta använd bevismetod är att man först antar motsatsen (negationen) till det man vill bevisa, och utifrån detta härleder en motsägelse. Ett bevis av detta slag kallas ett *motsägelsebevis*. Mer formellt går det alltså till på följande vis. Antag att man vill visa påståendet A. Man antar då att negationen $\neg A$ gäller. Utifrån $\neg A$ försöker man sedan härleda en motsägelse, dvs något som är uppenbart falskt. Detta innebär att antagandet $\neg A$ inte kan vara riktigt, och alltså gäller A. Vi ger nu ett vanligt exempel på ett motsägelsebevis.

Proposition 2.5. *Talet $\sqrt{2}$ är inte rationellt.*

Bevis. Vi antar påståendet är falskt, dvs att det verkligen finns positiva heltal p och q sådana att $\sqrt{2} = p/q$. Genom att förkorta bort alla eventuella gemensamma faktorer 2 kan vi anta att åtminstone ett av talen p och q är udda. Efter kvadrering får vi att $p^2/q^2 = 2$ dvs $p^2 = 2q^2$, men då är ju p^2 jämnt och alltså p jämnt. Härur följer att p^2 är delbart med 4 och alltså q^2 delbart med 2, men då måste även q vara jämnt och vi får en motsägelse. Alltså är ursprungsantagandet falskt och därmed är propositionen bevisad. □

Uppgift 2.1. Förklara påståendet i meningen "Genom att ..." i beviset ovan. Förklara också i detalj argumenten i andra och tredje meningarna från slutet.

Uppgift 2.2. Negera utsagan "Om du inte kommer hit ikväll så går jag på bio."

Uppgift 2.3. Negera utsagan "Det finns två positiva irrationella tal a och b sådana att a^b är rationellt."

Uppgift 2.4. Negera utsagan "Blåögda föräldrar får alltid blåögda barn".

Implikation. Om A och B är två satser så kan vi bilda satsen $A \to B$ som har följande sanningstabell:

A	B	$A \to B$
S	S	S
S	F	F
F	S	S
F	F	S

Alltså är $A \to B$ sann utom då B är falsk och A är sann. Speciellt är $A \to B$ sann om A är falsk. Notera att $A \to B$ har samma sanningsvärde som $\neg B \to \neg A$ (kontrollera med sanningstabell om du är osäker).

I matematisk text förekommer en rad olika sätt att uttrycka en implikation språkligt, såsom

- om A (gäller) så (gäller) B
- Antag A. Då gäller B.
- A medför B
- A implicerar B
- B (gäller) om A (gäller)
- B följer av/från A
- A är ett tillräckligt villkor för B
- B är ett nödvändigt villkor för A
- B följer av A
- B om A
- A endast om B

Exempelvis kan vi säga "om $x < -2$ så är $x^2 > 4$" eller "Antag att $x < -2$. Då är $x^2 > 4$." eller "att x är mindre än -2 medför att $x^2 > 4$" osv.

Ett annat exempel är den välkända frasen i gränsvärdesdefinitionen som ibland uttrycks som "$|x - a| < \delta$ medför att $|f(x) - f(a)| < \epsilon$" och ibland som "$|f(x) - f(a)| < \epsilon$ om $|x - a| < \delta$."

Vilket man väljer är en stilistisk fråga, men det betyder inte alls att den saknar betydelse i matematisk text.

Det finns också en hel del "dolda" implikationer i matematisk text. En känd sats i analys är "En deriverbar funktion är kontinuerlig." Notera att vad man menar är "Om en funktion är deriverbar så är den kontinuerlig". Dessutom finns här en "dold" kvantifikator; för att uttrycka det ännu tydligare kan man säga "För varje funktion gäller att om den är deriverbar så är den också kontinuerlig."

Istället för att uttrycka implikation med "om" använder man ibland omvänd ordföljd. Exempelvis "Är f deriverbar är den kontinuerlig" eller

"Bråkar du (så) åker du i säng". I vardagsspråket kan man t.o.m. säga "Du bråkar och du åker i säng".

Notera att en implikation alltid är sann om förledet är falskt. T ex är det matematiska påståendet "om $2 > 4$ så är alla reella tal större än 12" sant även om det ter sig lite underligt. När det gäller vanligt språk ska man vara lite försiktig. Normalt förväntar man sig någon form av (subjektivt) samband mellan leden i implikationen. Ett påstående av typ "Om månen är en ost så ligger Göteborg i England" är av samma skäl sant, men uppfattas väl av de flesta som nonsens. Även när man har "vettiga" meningar så kan det vara så att den vardagliga tolkningen skiljer sig från den strikt logiska. Om pappa säger "Om du bråkar vid matbordet så åker du bums i säng" till sin lille Niklas, så tolkar väl ingen det på annat sätt än att Niklas får vara uppe en stund efter maten om han sitter snällt och äter. Av samma skäl anser man det vara vilseledande (lögn?) att säga "Om du inte skyndar dig nu så missar vi tåget" om man vet att tåget redan har gått.

Det är välkänt att man från $A \to B$ inte kan dra slutsatsen att omvändningen $B \to A$ gäller. Exempelvis är det sant att $x < -2$ medför att $x^2 > 4$, men omvändningen, att $x^2 > 4$ medför $x < -2$, är falsk.

I den politiska retoriken är det vanligt att man (avsiktligt eller oavsiktligt) vänder på implikationer. Om man tror att arbetslösheten ökar om riksbanken höjer räntan, så följer ju inte av detta att arbetslösheten är oförändrad eller minskar bara för att man inte höjer räntan.

Uppgift 2.5. Antag att en viss politiker tror fullt och fast att arbetslösheten kommer att öka nästa år vare sig man fattar beslutet A eller inte. Ljuger han då om han säger "Om vi inte fattar beslutet A så kommer arbetslösheten att öka nästa år"?

I vardagligt språk har man ofta en tidsaspekt, t ex "Om det regnar så *blir* det vått", dvs det måste först regna och sedan blir det vått. Ska man läsa ut $\neg B \to \neg A$ så får den då formuleras som något i stil med "Om det inte är vått så har det inte regnat".

Om man vill bevisa ett påstående av typen $A \to B$ är det ofta enklare att bevisa det ekvivalenta påståendet $\neg B \to \neg A$. Det betyder alltså att man antar att $\neg B$ gäller och utifrån detta försöker härleda $\neg A$. För att exemplifiera detta visar vi följande.

Proposition 2.6. *Om $f'(x) = 0$ för alla x på intervallet I så är f konstant på I.*

Här är A alltså påståendet "$f'(x) = 0$ för alla x på intervallet I" och B är "f är konstant på I". Negationen till B är "f är icke-konstant på I" vilket kan omformuleras som "Det finns två punkter $x_1, x_2 \in I$ sådana att $f(x_1) \neq f(x_2)$." Beviset i sin helhet kan se ut så här.

BEVIS. Antag att f inte vore konstant. Då skulle det finnas två punkter $x_1, x_2 \in I$ sådana att $f(x_1) \neq f(x_2)$. Eftersom nödvändigtvis då $x_1 \neq x_2$ så finns enligt medelvärdessatsen ett tal ξ mellan x_1 och x_2 sådant att

$$f(x_2) - f(x_1) = (x_2 - x_1)f'(\xi).$$

Eftersom $f(x_2) - f(x_1) \neq 0$ följer det att $f'(\xi) \neq 0$, vilket motsäger antagandet att $f'(x) = 0$ på hela intervallet. Alltså är propositionen visad. □

Man får inte förväxla implikation med det ofta förekommande uttrycket "Eftersom A (gäller) så (följer) B." Detta innebär nämligen att A är sant, och att man därur kan sluta sig till att B är sant.

Konjunktion. Om A och B är satser kan vi bilda *konjunktionen* $A \wedge B$, och sanningstabellen är

A	B	$A \wedge B$
S	S	S
S	F	F
F	S	F
F	F	F

Vi kan uttrycka detta språkligt bl.a. som

- A och B
- både A och B
- såväl A som B
- dels A och dels B
- A men B

EXEMPEL 2.7. Notera att ekvationssystemet

$$\begin{cases} x^2 = & 4 \\ x+3 = & 5 \end{cases}$$

innebär

$$x^2 = 4 \wedge x + 3 = 5,$$

och att lösningsmängden är snittmängden av lösningsmängderna till var och en av ekvationerna, dvs

$$\{x;\ x^2 = 4\} \cap \{x;\ x + 3 = 5\}.$$

Likheten mellan symbolerna $\wedge$ och $\cap$ är ingen tillfällighet. □

Ofta kan man förkorta det språkliga uttrycket. Om man tar konjunktionen av påståendena "π är större än noll" och "π är irrationellt" så behöver man inte skriva "π är större än noll och π är irrationellt" utan man förkortar lämpligen till "π är irrationellt och större än noll".

Notera att "och" och "men" har samma logiska innebörd, men att "men" uttrycker ett slags (subjektivt) motsatsförhållande mellan utsagorna. Fundera på skillnaden mellan "jag är sjuk idag men jag har ingen feber" och "jag är sjuk idag och jag har ingen feber", eller paret "jag är sjuk idag och jag ska gå till jobbet" och "jag är sjuk idag men jag ska gå till jobbet (ändå)". Ibland finner man att det kan gå lika bra med vilket som. Det kan nämnas att det finns språk, som ryska, där man vid sidan av orden för "och" och "men" har ett tredje ord vars betydelse ligger emellan dessa. Även i matematiska sammanhang har det betydelse för förståelsen vilket ord man väljer.

Exempel 2.8. Funktionen $f(x) = |x|$ är kontinuerlig *men* inte deriverbar. $\square$

Notera även användningen av "men" i beviset av proposition 2.5, och i följande exempel.

Exempel 2.9. Vi vill finna alla x och y sådana att $x \geq 1$, $xy = -2$ och $y^2 = 4$.

Lösning: Eftersom $x \geq 1$ och/men $xy < 0$ så följer att $y < 0$. Eftersom $y^2 = 4$ och/men $y < 0$ så följer att $y = -2$, och alltså är $x = 1$. $\square$

Disjunktion. Av satser A och B kan vi bilda *disjunktionen* $A \vee B$ som har sanningstabellen

A	B	$A \vee B$
S	S	S
S	F	S
F	S	S
F	F	F

Alltså är $A \vee B$ sann om minst en av A och B är sann. I språket uttrycks disjunktionen som "A eller B". Som i fallet med konjunktion så kan man ofta förkorta formuleringen. Utsagan "x är rationellt eller x är transcendent" kan förkortas till "x är rationellt eller transcendent".

Speciellt är alltså en disjunktion sann om både A och B är sanna. I dagligt språk gör man inte alltid denna tolkning. Om läraren säger "Ni får skriftligt förhör på onsdag eller på torsdag" så är det nog underförstått att det inte blir förhör båda dessa dagar. Vidare om pappa säger

"Antingen sitter du snällt vid bordet, eller så åker du bums i säng" till lille Niklas så har han förstås lurats om Niklas trots exemplariskt uppträdande vid matbordet körs i säng direkt efter maten.

ANMÄRKNING 2.10. Det är värt att notera att många (fler än lille Niklas ovan) uppfattar "eller" i fraser som "antingen A eller B" som ett s.k. *exklusivt eller*; detta innebär att man anser att satsen är sann om och endast om *exakt* en av A och B är sann. Vill man att det ska stå klart att det är ett exklusivt eller man avser, bör man dock inte nöja sig med att sätta "antingen" framför, utan använda en formulering som "A eller B men inte båda" eller något i den stilen. □

EXEMPEL 2.11. Notera att $a \leq 1$ är det samma som $a < 1 \ \vee \ a = 1$. □

EXEMPEL 2.12. Om $x^2 = 4$ är $x = 2 \ \vee \ x = -2$, dvs lösningsmängden är $\{2\} \cup \{-2\} = \{2, -2\}$. □

Om X och Y är mängder så är

$$X \cap Y = \{x;\ x \in X \ \wedge \ x \in Y\}$$

och

$$X \cup Y = \{x;\ x \in X \ \vee \ x \in Y\}.$$

Notera att $A \vee B$ har samma sanningsvärde som $\neg B \to A$. Båda är nämligen falska om och endast om både A och B är falska.

Ekvivalens. Två satser är *ekvivalenta* om de har samma sanningsvärde, dvs vi har sanningstabellen

A	B	$A \leftrightarrow B$
S	S	S
S	F	F
F	S	F
F	F	S

Språkligt uttrycker vi ekvivalens bl.a. som

- A är ekvivalent med B
- om A så B och om B så A
- A om och endast om B
- A är ett nödvändigt och tillräckligt villkor för B

Ibland använder man förkortningen "omm" för "om och endast om".

Notera att $A \leftrightarrow B$ har samma sanningsvärde som $(A \to B) \wedge (B \to A)$. För att visa att A är ekvivalent med B så räcker det alltså att visa att A medför B och att B medför A; detta är för övrigt en mycket vanlig metod.

Det är värt att påpeka att man i matematik har konventionen att man i definitioner använder implikation trots att man noga taget menar ekvivalens. Definitionen av begreppet *positiv funktion* kan t ex se ut på följande sätt.

DEFINITION 2.13. En funktion $f(x)$ på ett intervall I sägs vara positiv, om $f(x) \geq 0$ för alla $x \in I$.

Om man inte skriver ut det explicit så måste man se till att det ändå framgår av sammanhanget att det rör sig om en definition. Efter att definitionen väl är gjord gäller sedan att en funktion $f(x)$ är positiv om och endast om $f(x) \geq 0$ för alla x.

Tautologier. En satslogisk utsaga är en *tautologi* om den är sann för alla tänkbara sanningsvärden på de ingående satsvariablerna. En sats som alltid är falsk, dvs sådan att dess negation är en tautologi, kallas en *kontradiktion*.

EXEMPEL 2.14. Vi ska visa att satsen

$$(A \vee B) \leftrightarrow (\neg B \to A) \tag{2.1}$$

är en tautologi, dvs sann för alla sanningsvärden på A och B. Man kan göra detta genom sätta upp en sanningstabell:

A	B	$(A \vee B)$	$\leftrightarrow$	$(\neg B$	$\to$	$A)$
S	S	S	S	F	S	S
S	F	S	S	S	S	S
F	S	S	S	F	S	F
F	F	F	S	S	F	F

I princip kan man alltid avgöra om en sats är en tautologi med hjälp av sanningstabell, men om satsen innehåller n satsvariabler så blir antalet rader i tabellen 2^n, vilket ju växer mycket snabbt med n. För stora satser är det därför bäst att först försöka reducera antalet fall man måste kolla. Man kan t ex utnyttja att en implikation är falsk omm förledet är sant och efterledet falskt. I vårt exempel kan man därför direkt notera att $\neg B \to A$ är falsk omm $\neg B$ är sann och A är falsk, men detta gäller ju omm både A och B är falska, vilket i sin tur gäller omm $A \vee B$ är falsk. Alltså är (2.1) en tautologi. □

Det kan vara bra att känna till några andra vanliga tautologier. Från $A \to B$ och A kan man dra slutsatsen B, dvs

$$((A \to B) \wedge A) \to B \tag{2.2}$$

är en tautologi. Detta brukar kallas *modus ponens*. En annan tautologi är

$$(A \wedge \neg A) \to B, \tag{2.3}$$

vilken är sann eftersom förledet alltid är falskt. Vi har även de distributiva lagarna

$$A \wedge (B \vee C) \leftrightarrow (A \wedge B) \vee (A \wedge C) \tag{2.4}$$

och

$$A \vee (B \wedge C) \leftrightarrow (A \vee B) \wedge (A \vee C). \tag{2.5}$$

Jämför med motsvarande påstående för mängder,

$$X \cap (Y \cup Z) = (X \cap Y) \cup (X \cap Z)$$

och

$$X \cup (Y \cap Z) = (X \cup Y) \cap (X \cup Z),$$

om X, Y och Z är mängder.

Uppgift 2.6. Visa att (2.4) och (2.5) är tautologier.

Vidare kollar man lätt att $\neg(A \wedge B)$ är ekvivalent med $\neg A \vee \neg B$ och att $A \wedge (B \wedge C)$ är ekvivalent med $(A \wedge B) \wedge C$ så att vi utan tvetydighet kan skriva $A \wedge B \wedge C$. Samma sak gäller för $\vee$ istället för $\wedge$.

Ibland kan man visa att en implikation $\phi \to \psi$ är sann genom en s.k. *härledning*. Man antar att ϕ är sann och utnyttjar s.k. *härledningsregler* för att sluta sig till att då även ψ måste vara sann. Exempel på sådana härledningsregler är modus ponens, se ovan, samt (jmf (2.3)) att man från satserna A och $\neg A$ får dra slutsatsen B.

Exempel 2.15. Vi påstår att

$$((A \to \neg B) \to \neg A) \to (A \to B). \tag{2.6}$$

I stället för att sätta upp en sanningstabell kan vi avgöra detta genom att använda härledningsreglerna ovan. I det följande är varje rad en konsekvens av de föregående.

$$\begin{array}{ccc}
(A \to \neg B) & \to & \neg A \\
(\neg A \vee \neg B) & \to & \neg A \\
\neg(A \wedge B) & \to & \neg A \\
(A \wedge B) & \vee & \neg A \\
(A \vee \neg A) & \wedge & (B \vee \neg A) \\
B \vee \neg A & & \\
A \to B & &
\end{array}$$

I tredje raden från slutet har vi använt (2.4). □

UPPGIFT 2.7. Visa att även omvändningen av (2.6) är sann.

EXEMPEL 2.16. Ett år x är ett skottår om och endast om det är delbart med fyra och det dessutom gäller att det antingen inte är delbart med 100 eller är delbart med 400. Om vi inför satsvariablerna
A: x är ett skottår
B: x är delbart med 4
C: x är delbart med 100
D: x är delbart med 400,
så gäller alltså att

$$B \wedge (\neg C \vee D) \leftrightarrow A,$$

dvs

$$(B \wedge \neg C) \vee (B \wedge D) \leftrightarrow A.$$

Men $B \wedge D \leftrightarrow D$ så vi får att

$$(B \wedge \neg C) \vee D \leftrightarrow A.$$

Exempelvis var inte år 1900 ett skottår, men år 2000 är det. □

UPPGIFT 2.8. Visa att $(\neg A \vee B) \wedge C$ medför $(C \to A) \to B$ men att omvändningen inte är sann.

UPPGIFT 2.9. Använd sanningstabell eller avgör på annat sätt vilka av följande satser som är tautologier.

- (i) $\neg(A \to (B \vee \neg C)) \to (\neg(A \vee B) \vee C)$.
- (ii) $(A \wedge B \wedge C) \vee (A \wedge \neg B \wedge C) \vee (\neg A \wedge \neg B \wedge C) \vee (\neg A \wedge \neg B \wedge \neg C)$.
- (iii) $\neg(A \to (B \vee \neg C)) \to ((A \vee B) \vee \neg C)$.

3. Kvantifikatorer

För att kunna göra en närmare beskrivning av hur man resonerar i matematiska (och andra) sammanhang räcker det inte med bara satslogik utan man behöver även införa s.k. *kvantifikatorer* eller *kvantorer*. Den (enklaste) formaliserade varianten man får kallas *(första ordningens) predikatlogik* (*elementary logic* på engelska). En viss kännedom om predikatlogik tillhör matematisk allmänbildning och en orientering ges i avsnitt 5.

Många matematiska påståenden är s.k. *öppna utsagor*, dvs de innehåller en eller flera variabler, och sanningsvärdet på utsagan kan bero på vilka fixa värden man ger dessa variabler. En utsaga (påstående) som "$x^2 > 4$" är uppenbarligen sann för vissa värden på x men falsk för andra. Låt oss beteckna utsagan med ϕ. Ibland skriver man då $\phi(x)$

för att understryka att variabeln x är *fri* i ϕ. Man säger att en (öppen) utsaga är sann om den är sann för alla fixa värden på variablerna. Till att börja med kan det verka lite lustigt att tillskriva en öppen utsaga ett sanningsvärde, men tänker man efter finner man att detta är praxis i matematisk text. Man kan uppfatta detta som att allkvantifiering underförstås. Exempelvis är utsagan "om $x < -2$ så är $x^2 > 4$" sann i kraft av att den är sann för alla (reella) x.

Vi inför nu symbolerna $\forall$ och $\exists$, *allkvantorn* och *existenskvantorn*. Om ϕ är en utsaga så kan man bilda $\forall x\phi$, och denna är sann omm ϕ själv är sann (för alla x). Man kan också bilda $\exists x\phi$ vilken är sann omm det finns ett fixt val av x så att ϕ är sann. Om ψ är utsagan "$x^2 = 2$", så är alltså $\exists x\psi$ ett sant påstående, men ψ själv och $\forall x\psi$ är falska.

ANMÄRKNING 3.1. Eftersom $\forall x\phi$ är sann omm ϕ är sann kan man fråga sig varför $\forall x$ behövs. Så fort man sätter samman utsagor med konnektiver har det emellertid betydelse om de är slutna med allkvantorer eller inte. Exempelvis kan $\phi(x) \vee \psi(x)$ vara sann utan att för den skull $\phi(x)$ eller $\psi(x)$ är sann. Tag t ex ϕ som $x \geq 0$ och ψ som $-x \geq 0$. $\square$

Liksom är fallet med konnektiverna, finns det många olika sätt att språkligt uttrycka $\forall x\phi$. Här är några exempel;

- för alla x gäller (att) ϕ
- för varje x gäller (att) ϕ
- för ett godtyckligt x gäller (att) ϕ

Vidare påminner vi om exemplet "en deriverbar funktion är kontinuerlig" som ju innehåller en "dold" allkvantor. Läsaren uppmanas att tänka ut fler exempel.

I vanligt språk uttrycker man $\exists x\phi$ exempelvis som

- det finns (ett) x sådant att ϕ (gäller)
- det existerar ett x sådant att ϕ (gäller)
- vi kan välja (ta, hitta,...) x sådant att ϕ (gäller)

Mer konkret förekommer det tredje fallet t ex i en mening som "...och eftersom x är udda kan vi sätta $x = 2m + 1$ där m är ett heltal. ..."

EXEMPEL 3.2. Ibland kan en mening, taget ur sitt sammanhang, vara tvetydig. Meningen "En reell funktion på mängden M är begränsad" kan tolkas, beroende på sammanhanget (och tonvikten på ordet "En" när meningen läses), både som "Det finns en funktion på M som är begränsad" och som "Varje funktion på M är begränsad". Är det det första fallet som avses, kan man förtydliga genom att lägga till "åtminstone", "i alla fall" eller dylikt. $\square$

Om man har flera kvantifikatorer är det viktigt att hålla reda på ordningen.

Exempel 3.3. Låt ϕ vara $x \geq y$, där x och y står för reella tal. Då är $\forall x \exists y \phi$ sann men $\exists y \forall x \phi$ är falsk. Till varje (fixt) x finns det ju något y sådant att $x \geq y$. Däremot finns det ju inte något fixt y som är mindre än eller lika med alla x.

Gillar man inte det exemplet kan man istället välja att kvantifiera över mängden av alla människor som någonsin levat, och låta ϕ stå för "y är mor till x". Också i detta fall är det så att $\forall x \exists y \phi$ är sann men $\exists y \forall x \phi$ är falsk. Varje människa har nämligen en mor, men det finns inte någon som är mor till alla människor.

Det allra enklaste är att välja en mängd med två element och en lämplig relation på denna som i exempel 5.6. □

Alltså är inte alltid

$$\forall x \exists y \phi \to \exists y \forall x \phi$$

sann. Däremot kan man alltid byta plats på två kvantorer av samma slag, dvs

$$\forall x \forall y \phi \leftrightarrow \forall y \forall x \phi$$

är sann, och motsvarande gäller för existenskvantorn.

Uppgift 3.1. Visa (dvs övertyga dig om) att

$$\exists y \forall x \phi \to \forall x \exists y \phi \tag{3.1}$$

är sann för varje utsaga ϕ.

Vi ska nu studera hur kvantorerna är relaterade till negation, och vi börjar med ett exempel.

Exempel 3.4. Låt ϕ stå för "x klarade tentan". Då betyder $\forall x \phi$ att alla (t ex i klassen) klarade tentan, medan $\neg \forall x \phi$ betyder att det inte var så att alla klarade tentan. Det senare innebär att det fanns åtminstone någon i klassen som inte klarade tentan, dvs $\exists x(\neg \phi)$. □

Man kan ganska lätt övertyga sig om att fenomenet i det förra exemplet gäller allmänt, dvs att vi har

$$\forall x \neg \phi \leftrightarrow \neg \exists x \phi \tag{3.2}$$

och (ekvivalent)

$$\exists x \neg \phi \leftrightarrow \neg \forall x \phi. \tag{3.3}$$

Negationen av utsagan "Varje reell funktion på mängden M är begränsad" är alltså "Det finns en reell funktion på mängden M som inte är begränsad”. Den senare meningen kan också uttryckas som “Det finns en obegränsad reell funktion på mängden M”.

Uppgift 3.2. Fundera igenom innebörden av var och en av följande meningar. Vilka uttrycker samma sak?

- (i) Alla reella funktioner på mängden M är obegränsade.
- (ii) Alla reella funktioner på mängden M är inte begränsade.
- (iii) Inte alla reella funktioner på mängden M är begränsade.
- (iv) En reell funktion på M är begränsad.
- (v) Alla reella funktioner på mängden M är inte obegränsade.
- (vi) Alla reella funktioner på mängden M är begränsade.

Hittills har vi underförstått över vilken mängd man kvantifierar. Man kan också införa villkor på kvantifikatorerna. Om ψ är en utsaga så kan vi bilda $(\forall x)_\psi \phi$, vilket definitionsmässigt innebär $\forall x(\psi \to \phi)$, dvs “ϕ gäller för alla x sådana att ψ gäller”. På samma sätt kan man bilda $(\exists x)_\psi \phi$ vilket betyder $\exists x(\psi \wedge \phi)$, dvs “det finns ett x sådant att både ψ och ϕ gäller”.

Proposition 3.5. *Vi har att*

$$\neg(\exists x)_\psi \phi \leftrightarrow (\forall x)_\psi \neg\phi \tag{3.4}$$

och

$$\neg(\forall x)_\psi \phi \leftrightarrow (\exists x)_\psi \neg\phi \tag{3.5}$$

Bevis. Vi börjar med att visa (3.4). Genom att använda definitionen, (3.2) och enkel satslogik så får vi

$$\neg(\exists x)_\psi \phi \leftrightarrow \neg\exists x(\psi \wedge \phi) \leftrightarrow \forall x \neg(\psi \wedge \phi) \leftrightarrow \forall x(\psi \to \neg\phi) \leftrightarrow (\forall x)_\psi \neg\phi.$$

Detta visar (3.4). Man får nu (3.5) genom att ta negationer i båda leden av (3.4). □

Uppgift 3.3. Visa hur man får (3.5).

En vanlig situation är att ψ står för $x > 0$. I detta fall skriver man ofta $\forall x > 0$ och $\exists x > 0$ hellre än $(\forall x)_{x>0}$ respektive $(\exists x)_{x>0}$.

Exempel 3.6. En funktion $f(x)$ på $\mathbb{R}$ är kontinuerlig i punkten y om det till varje $\epsilon > 0$ finns ett $\delta > 0$ sådant att det för varje x gäller att

$$|x - y| < \delta \text{ medför att } |f(x) - f(y)| < \epsilon. \tag{3.6}$$

Om utsagan (3.6) betecknas med ϕ så är alltså f kontinuerlig i y omm $\forall \epsilon > 0 \exists \delta > 0 \forall x \phi$. Man säger att f är kontinuerlig, om den är kontinuerlig i varje punkt y. Vi kan skriva detta som

$$\forall y(\forall \epsilon > 0)(\exists \delta > 0)\forall x \phi. \tag{3.7}$$

Det finns ett annat viktigt besläktat begrepp. Man säger att f är *likformigt kontinuerlig* om det till varje $\epsilon > 0$ finns ett $\delta > 0$ sådant att ϕ gäller för alla y och x; alltså

$$(\forall \epsilon > 0)(\exists \delta > 0)\forall y \forall x \phi. \tag{3.8}$$

Notera skillnaden mellan (3.7) och (3.8). □

Exempel 3.7. Antag att vi vill negera påståendet att f är kontinuerlig på $\mathbb{R}$. Genom att sätta $\neg$ framför (3.7) och använda (3.4) och (3.5) får vi

$$\exists y(\exists \epsilon > 0)(\forall \delta > 0)\exists x \neg \phi.$$

Notera att $\neg \phi$ innebär (kolla!)

$$|x - y| < \delta \text{ och } |f(x) - f(y)| \geq \epsilon.$$

Med ord kan vi därför uttrycka negationen av att f är kontinuerlig på $\mathbb{R}$ som "Det finns ett y och ett positivt tal ϵ sådana att det till varje godtyckligt litet positivt δ går att hitta en punkt x sådan att $|x-y| < \delta$ men likväl $|f(x) - f(y)| \geq \epsilon$. □

I normalfallet använder man inte symboler när man negerar matematiska påståenden utan man försöker känna sig fram till vad negationen innebär, men ibland kan det vara en fördel att kunna stödja sig på en kalkyl som i förra exemplet. Hur man gör är till viss del en fråga om personlig läggning, tycke och smak.

Uppgift 3.4. Negera påståendet att f är likformigt kontinuerlig.

Anmärkning 3.8. Vi har i detta avsnitt diskuterat kvantifiering av öppna utsagor. I informella sammanhang används ibland kvantorerna även som en direkt ersättning av ett språkligt uttryck, som i meningen "En funktion $f(x)$ är positiv på intervallet I om $f(x) \geq 0 \ \forall x \in I$". Fler exempel dyker upp i avsnitt 5. □

4. Matematiska teorier

En *sats* i matematik är en utsaga som man strikt logiskt kan härleda från axiom, andra redan kända satser och kända och eventuellt nya definitioner. När vi väl har lyckats definiera rationella och reella tal så får utsagan "$\sqrt{2}$ är ej rationellt" en innebörd, och som vi sett tidigare är den i själva verket en sats.

Man brukar i praktiken förbehålla benämningen sats till sådana bevisade utsagor som man tycker är särskilt viktiga. En sats som i sammanhanget är lite mindre viktig kallar man ofta *proposition*, och en sats som bara har intresse medan man genomför beviset av en viktigare sats brukar man kalla för ett *lemma*.

ANMÄRKNING 4.1. Det finns ett formellt synsätt att alla sanna matematiska påståenden är satser, oberoende av huruvida de är bevisade eller inte. Detta verkar dock inte stämma med hur matematiker i allmänhet använder ordet sats. Det är vanligt med formuleringar som "den satsen är från 1972" eller "det är en sats av Hörmander", och vad som avses är då att påståendet visades, eller eventuellt både formulerades och visades, 1972 respektive av Hörmander. Det förfaller då vara underförstått att satsen inte fanns tidigare. Å andra sidan förekommer även uttalanden som "..., men det visade sig långt senare att satsen i själva verket även gäller i alla konvexa områden", vilket antyder en bakomliggande uppfattning om att satsen har funnits där hela tiden, men att det dröjde innan någon upptäckte den. Det är även värt att påpeka att inte varje bevisbar utsaga kallas sats av matematiker. Det krävs att utsagan på något sätt är intressant. Vad som är intressant varierar förstås från person till person och från tid till annan. Hur som helst är detta bara en fråga om bruket av ordet sats bland matematiker.

En annan sak är vad det innebär att en matematisk utsaga är *sann*. Vissa menar att det måste finnas ett känt bevis (vissa kräver dessutom ett s.k. konstruktivt bevis, se nedan), andra menar att en utsaga är sann om det finns ett bevis (som eventuellt ännu är oupptäckt). Det finns emellertid påståenden om reella tal såsom kontinuumhypotesen, se avsnitt 5, som man vet inte kan bevisas eller motbevisas utifrån mängdteorins axiom. Om man tillskriver de reella talen en absolut, av människor oberoende, existens, är det rimligt att tänka sig att varje påstående om dessa tal, såsom kontinuumhypotesen, är sant eller falskt. □

De logiska principer som vi har gått igenom så långt är de som man normalt använder i matematisk bevisföring. Om man t ex kan visa att det leder till motsägelse att anta att det inte finns ett objekt med en viss egenskap så anser man att man därmed har bevisat att ett sådant objekt existerar. Detta sätt att resonera är dock inte helt oomstritt. Det finns de som hävdar att man för att bevisa ett existenspåstående ska kunna peka ut eller konstruera objektet i fråga; man talar då om ett *konstruktivistiskt* synsätt. I alla händelser anser de flesta att det är att föredra om man verkligen kan hitta objektet i fråga. Låt oss se på ett enkelt exempel.

EXEMPEL 4.2. Vi vill bevisa följande påstående: Det finns ett par av positiva irrationella tal a och b sådana att a^b är rationellt.

Vi vet redan att $\sqrt{2}$ är irrationellt. Vi påstår att antingen paret $a = \sqrt{2}$, $b = \sqrt{2}$ eller paret $a = \sqrt{2}^{\sqrt{2}}$, $b = \sqrt{2}$ har önskad egenskap. Om nämligen $\sqrt{2}^{\sqrt{2}}$ är rationellt så duger det första paret. Om å andra sidan $\sqrt{2}^{\sqrt{2}}$ är irrationellt så duger det andra paret; då är ju a och b båda irrationella och

$$a^b = (\sqrt{2}^{\sqrt{2}})^{\sqrt{2}} = \sqrt{2}^{\sqrt{2}\sqrt{2}} = \sqrt{2}^2 = 2$$

som är rationellt. Alltså är beviset klart, utan att vi för den skull vet vilket av paren som verkligen har den önskade egenskapen. □

Den vanliga uppfattningen är att de naturliga talen finns, och från dessa kan man sedan definiera heltal, rationella tal, reella tal och komplexa tal, se kapitel 3. Om man nu ska bevisa satser om naturliga tal, vad får man då utgå från? Om man tror på att talen finns, så kan man försöka hitta några påståenden, *axiom* eller postulat, som är självklart sanna och sedan försöka bevisa allt annat utifrån dessa. Följande axiomsystem är från slutet av 1800-talet.

Exempel 4.3 (Peanos axiom). Man inför en symbol 0 och följande axiom:

- P1 0 är ett naturligt tal.
- P2 till varje naturligt tal n finns ett entydigt bestämt naturligt tal n', som kallas efterföljaren till n.
- P3 $n' \neq 0$ för alla n.
- P4 Om $m' = n'$ så är $m = n$.
- P5 Om X är en mängd av naturliga tal sådan att $0 \in X$ och sådan att $n' \in X$ om $n \in X$, så innehåller X alla naturliga tal.

Man känner igen P5 som induktionsaxiomet. Utifrån dessa axiom kan man definiera de vanliga operationerna $\cdot$ och $+$ och bevisa allt man vet är sant om naturliga tal. Vi ska titta närmare på detta i kapitel 3. □

Exempel 4.4. På samma sätt införde Euklides axiom för geometrin. Han menade att linjer och punkter existerade och att de uppfyllde vissa fundamentala egenskaper (axiom) och ur dessa lyckades han bevisa alla då kända satser om geometri. Detta behandlar vi i kapitel 4. □

Man möter ibland i olika sammanhang inom matematiken resonemang som liknar varandra även om de ytligt sett handlar om olika saker. Man kan då försöka abstrahera ut grundläggande gemensamma drag och införa en *axiomatisk teori*. Denna består precis som i de två föregående exemplen av några grundläggande begrepp samt ett antal axiom som reglerar förhållandet mellan dessa begrepp. En *modell* till teorin är en icketom mängd M där varje begrepp i teorin har en *tolkning*, på ett sådant sätt att axiomen är sanna påståenden om M. Av praktiska skäl låter vi M även beteckna hela modellen (dvs mängden tillsammans med tolkningarna av begreppen). Alla utsagor som man kan bevisa utgående från axiomen blir då sanna påståenden om varje modell. En teori sägs vara *konsistent* om det inte går att bevisa en motsägelse i den. Om teorin har en modell så kommer varje påstående i teorin att svara mot ett påstående i modellen, och därför måste teorin vara konsistent.

ANMÄRKNING 4.5. Man säger att en utsaga inom en teori är *sann*, om den är sann för varje modell till teorin. Huruvida det är möjligt att finna ett bevis för satsen är en annan sak. I fallet med predikatlogiska teorier, så finns det alltid ett bevis av en sann utsaga, se nästa avsnitt. □

EXEMPEL 4.6 (Gruppteori). Man har ett element e samt en operation, kallad multiplikation, som till varje ordnat par av element x och y ger ett nytt element xy så att följande är uppfyllt:

- $(xy)z = x(yz)$ för alla x, y, z
- $xe = ex = x$ för alla x
- till varje x finns ett element y sådant att $xy = yx = e$.

Modellerna till denna teori kallas *grupper*. Teorin är konsistent eftersom exempelvis heltalen $\mathbb{Z}$ med operationen $+$ är en modell. En annan modell är mängden av reella inverterbara 2×2-matriser med operationen matrismultiplikation. □

Som illustration bevisar vi en enkel sats i gruppteori.

SATS 4.7. *Antag att G är en grupp och att $xy = y$. Då är $x = e$.*

BEVIS. Antag att $xy = y$. Enligt tredje axiomet så finns ett z sådant att $yz = e$. Enligt första axiomet är nu $e = yz = (xy)z = x(yz) = xe$, men från andra axiomet har vi att $xe = x$, så sammantaget får vi att $e = x$, vilket skulle visas. □

Påståendet ovan är alltså sant för varje grupp. Däremot är påståendet

$$xy = yx \tag{4.1}$$

inte en sats i gruppteori. För att inse detta måste vi hitta en modell för axiomen, dvs en grupp, där (4.1) inte är sann. Man kan exempelvis ta mängden av alla inverterbara 2×2-matriser, där xy svarar mot matrismultiplikation.

UPPGIFT 4.1. Kontrollera detta!

Andra vanliga axiomatiska teorier är teorin för *lineära rum*, *topologiska rum*, *Banachrum*, *mångfalder*, *vektorbuntar*, *kärvar* etc.

Ibland lägger man till extra axiom. Om man till axiomen i gruppteori lägger till påståendet (4.1) som ett extra axiom så får man teorin för *abelska grupper*. Att även denna är konsistent inses av att t ex heltalen $\mathbb{Z}$ med operationen $+$ är en modell.

Man säger att två modeller M_1 och M_2 till en teori är *isomorfa* om det finns en bijektion (enentydig korrespondens) mellan mängderna sådan att tolkningen av ett givet begrepp i de båda modellerna stämmer överens under denna bijektion. Enkelt uttryckt innebär detta att modellerna ser likadana ut, förutom att man har bytt namn på objekten.

EXEMPEL 4.8. Om M_1 är mängden av de reella talen, så är M_1 en modell till gruppteorin, dvs en grupp, om e tolkas som talet 0 och xy tolkas som addition av tal. Mängden M_2 av positiva reella tal är också en modell om e tolkas som 1 och xy tolkas som multiplikation av tal. Man ser lätt att dessa modeller är isomorfa genom den bijektiva avbildningen $\exp\colon M_1 \to M_2$. $\square$

Det är klart att det finns icke-isomorfa modeller till gruppteorin, dvs icke-isomorfa grupper, eftersom vi har sett att (4.1) är sann för somliga grupper men inte för andra. Det är däremot ganska lätt att övertyga sig om att varje modell till Peanos axiom är isomorf med de naturliga talen, se vidare i kapitel 3.1.

Om man tror på att de naturliga talen existerar och uppfyller Peanos axiom så är det självklart att man aldrig kan bevisa någon motsägelse från dessa axiom. Kring sekelskiftet var man intresserad av att på något sätt *bevisa* att det aldrig går att härleda någon motsägelse. För att göra detta fick man anlägga ett formellt synsätt på axiom och bevis utifrån givna axiom; ett sätt är att utnyttja *predikatlogiken* som vi ska titta på lite i nästa avsnitt.

ANMÄRKNING 4.9. De befintliga matematiska teorierna har inte uppkommit bara av en slump utan som ett led i en historisk utveckling för att söka svar på tidigare inom- eller utommatematiska frågor. I princip kan man förstås hitta på definitioner på måfå och ett antal axiom, och på så sätt bilda en ny matematisk teori (som förhoppningsvis är konsistent åtminstone). Det finns naturligtvis olika uppfattningar om vilka matematiska teorier som särskilt är värda att ägna sig åt, precis som åsikterna om annan kulturell verksamhet varierar kraftigt. $\square$

ANMÄRKNING 4.10. Det faktum att matematiken innehåller en massa satser som bevisas från andra satser betyder emellertid inte att matematisk förståelse bara går ut på att överblicka långa räckor av logiska argument. Förståelsen bygger i hög grad på idéer och förklaringar som ofta beskrivs med bildspråk och som analogier till något mer lättbegripligt. En forskare i matematik sitter därför normalt inte heller och laborerar med logiska argument utan försöker, genom att betrakta olika mer konkreta exempel och genom att försöka hitta analogier till redan kända teorier eller sätt att resonera, komma fram till nya teorier eller nya resultat (satser), synsätt och metoder inom någon befintlig teori. □

5. Predikatlogik

I det här avsnittet ska vi kortfattat beskriva *predikatlogik*, eller mer exakt *första ordningens predikatlogik*.

I (en) formell predikatlogik har man till att börja med ett *alfabet* bestående av *(individ)variabler* $x, y, z, \ldots x_1, x_2, \ldots$, *individkonstanter* $a, b, c, \ldots, a_1, a_2, \ldots$, en uppsättning *predikat(symboler)* t ex

$$A(x), B(x_1, x_2) \ldots A_1(x, y, z)$$

etc samt predikatet $=$. Antalet variabler efter själva predikatet markerar om det är *ettställigt*, *tvåställigt* osv. Predikatet $=$ är tvåställigt och man skriver variablerna på var sin sida om det, dvs $x = y$. Istället för $\neg(x = y)$ skriver vi ofta $x \neq y$. Man kan också ha funktionssymboler som $f(x)$, $g(x, y)$ etc. Slutligen ingår i alfabetet de vanliga konnektiverna samt kvantorsymbolerna $\forall$ och $\exists$.

Vi ska nu beskriva predikatlogikens syntax. Ett uttryck t är en *term* om antingen t är en individvariabel eller en individkonstant eller $f(t_1, t_2, \ldots)$ där f är en funktionssymbol och t_j är andra termer. Detta är en s.k. *induktiv* definition. De enklaste termerna är alltså variabler och individkonstanter och man kan sedan bygga upp mer och mer komplexa termer genom funktionssymboler. (Notera att vi i satslogiken har en rekursiv definition av vad som är (satslogiska) satser.)

Begreppet *formel* definieras också induktivt. Ett uttryck ϕ är en formel om antingen ϕ är $A(t_1, t_2, \ldots)$, där A är ett predikat och t_j är termer (antalet ska svara mot ställigheten hos A förstås), eller ϕ är $\forall x \psi$ eller $\exists x \psi$, där x är en variabel och ψ är en annan formel, eller ϕ är en satslogisk kombination av andra formler. Definitionen ger alltså principerna för hur man kan bygga upp mer och mer komplexa formler.

Om ϕ bara består av ett predikat $A(t_1, t_2, \ldots, t_m)$, så säger man att alla variabler som förekommer i termerna $t_1, \ldots, t_m$ är *fria* i formeln ϕ.

När man för en godtycklig formel ϕ bildar $\forall x\phi$ eller $\exists x\phi$ så *binds* alla x i ϕ som förekommer fria, dvs de x som inte redan tidigare är bundna av någon annan kvantor.

EXEMPEL 5.1. I formeln $\forall x(A(x,y) \to \exists y B(x,y,f(x,z)))$ är alla x bundna, vänstra förekomsten av y är fri medan den högra är bunden, och z är fri. □

DEFINITION 5.2. En *tolkning* av (en) predikatlogik är en icketom mängd M, kallad tolkningens *domän*, i vilken man till varje individkonstant har tillordnat ett fixt element i M, till varje predikat har tillordnat en motsvarande *relation* (med rätt ställighet) i mängden M, där $=$ svarar mot likhetsrelationen på M, och man till varje funktionssymbol har tillordnat en motsvarande funktion på M. En formel ϕ är *sann* i denna tolkning om motsvarande påstående i M är sant.

ANMÄRKNING 5.3. En enställig relation R på en mängd M är helt enkelt en delmängd till M. För varje $x \in M$ gäller alltså antingen relationen R eller så gäller den inte, dvs $x \in R$ eller $x \notin R$. En tvåställig relation R är en delmängd till $M \times M$ (se kapitel 2), dvs för varje ordnat par av element $\langle x, y\rangle$ så gäller antingen att $\langle x, y\rangle \in R$ eller att $\langle x, y\rangle \notin R$. På liknande sätt definieras m-ställig relation, se vidare i kapitel 2. □

ANMÄRKNING 5.4. Sista meningen i definition 5.2 kräver en förklaring. Man kan visa med induktion att varje formel svarar mot ett påstående om modellen och följaktligen har ett sanningsvärde. Detta är väl intuitivt självklart och hur som helst så hoppar vi över det formella argumentet och hänvisar till någon bok i logik. Kom ihåg att en öppen utsaga U är sann om och endast om motsvarande utsaga $\forall x U$ är sann. Alltså gäller att en formel ϕ är sann i en tolkning om och endast om formeln $\forall x\phi$ är sann. □

EXEMPEL 5.5. Låt alfabetet bestå av de logiska symbolerna, $=$, lämpliga variabler $x, y, z, \ldots$ och det tvåställiga predikatet $A(x,y)$. Låt ϕ vara formeln $\exists y\forall x A(x,y)$. I tolkningen $M = \mathbb{R}$, och $A(x,y)$ svarande mot $x \geq y$, är ϕ falsk, men i tolkningen $M = \mathbb{N}$ (de naturliga talen), och $A(x,y)$ svarande mot $x \geq y$, är ϕ sann. □

UPPGIFT 5.1. Låt ϕ vara formeln från föregående exempel och den tolkning där M är mängden av studenter som är registrerade på kursen i Logik och geometri i år, och $A(x,y)$ tolkas som "x är längre än eller jämnlång med y". Avgör om ϕ är sann i denna tolkning.

EXEMPEL 5.6. Ett enkelt sätt att hitta en modell i vilken $\exists y \forall x A(x,y)$ är falsk men $\forall x \exists y A(x,y)$ är sann är att ta en mängd med två element, t ex $M = \{0,1\}$, och relationen $\{\langle 0,0\rangle, \langle 0,1\rangle\}$. $\square$

EXEMPEL 5.7. Låt alfabetet bestå av de logiska symbolerna, $=$, lämpliga variabler $x, y, w, \ldots$ och det treställiga predikatet $A(x,y,z)$. En tolkning kan vara att $M = \mathbb{R}$ och $A(x,y,z)$ tolkas som $x - y = z$. En annan tolkning är att $M = \{\text{delmängder till } \mathbb{N}\}$ och A tolkas som $x \cap y = z$. Formeln $\forall x \forall y (A(x,y,z) \to A(y,x,z))$ är falsk i första tolkningen men sann i den andra. $\square$

EXEMPEL 5.8. Låt alfabetet bestå av de logiska symbolerna, $=$, lämpliga variabler $x, y, z, \ldots$, individkonstanten 0 och funktionssymbolen $f(x,y)$. Låt vidare ϕ vara formeln

$$\forall x \exists y (f(x,y) = 0).$$

I tolkningen där $M = \mathbb{Z}$, 0 tolkas som noll, och $f(x,y)$ som $x + y$, är formeln ϕ sann, men i tolkningen där man byter $\mathbb{Z}$ mot $\mathbb{N}$ är formeln ϕ falsk. $\square$

DEFINITION 5.9. En formel ϕ är *logiskt sann* om den är sann i varje tolkning. Låt Ψ vara en (eventuellt oändlig) mängd av formler. En formel ϕ är en *logisk konsekvens* av Ψ om ϕ är sann i varje tolkning där alla formlerna i Ψ är sanna. En *(predikatlogisk) teori* är en mängd av formler Ψ. Formlerna i Ψ kallas då *axiom* för teorin. En *modell* till teorin Ψ är en tolkning där alla formlerna i Ψ är sanna. En formel ϕ som är sann i varje modell till teorin Ψ, dvs en logisk konsekvens av Ψ, kallas en *sats* i teorin Ψ.

Notera att formeln ϕ är sann i en tolkning omm $\tilde{\phi}$ är sann i denna tolkning, där $\tilde{\phi}$ betecknar den slutna formel man får genom att binda alla fria variabler med allkvantorer. Det följer att ϕ är en logisk konsekvens av formeln ψ (noga taget av den mängd Ψ av formler som bara består av formeln ψ) om och endast om $\tilde{\psi} \to \tilde{\phi}$ är en logisk sanning.

UPPGIFT 5.2. Tänk igenom detta!

EXEMPEL 5.10. Låt ϕ vara $x = a$ och ψ vara $y = a$. Vi påstår att ψ en logisk konsekvens av ϕ och $\forall x\phi \to \forall y\psi$ är en logisk sanning, men att däremot inte $\phi \to \psi$ är en logisk sanning.

Att ϕ är sann i en modell innebär att $\forall x\phi$ är sann, vilket i sin tur betyder att i denna modell varje element är lika med (tolkningen av) individkonstanten a. Detta innebär förstås att modellen bara har ett enda element, och därför är även ψ sann. Det följer nu att ψ är en logisk konsekvens av ϕ. Av samma skäl är $\forall x\phi \to \forall y\psi$ är en logisk sanning.

Formeln $\phi \to \psi$ är sann i varje modell med bara ett element, men falsk i varje modell med fler än ett element, och därför är formeln inte en logisk sanning. $\square$

EXEMPEL 5.11. Antag att $\phi(x)$ och $\psi(x)$ är formler med endast en fri variabel x, t ex enställiga predikat. Vi vill visa att

$$\forall x(\phi(x) \to \psi(x)) \to (\forall x\phi(x) \to \forall x\psi(x))$$

är en logisk sanning. Om M är en tolkning så kommer $\phi(x)$ och $\psi(x)$ att svara mot enställiga relationer, dvs delmängder, M_1 respektive M_2 av M. Alltså innebär $\forall x(\phi(x) \to \psi(x))$ att M_1 är en delmängd till M_2. Vidare innebär $\forall x\phi(x)$ att $M_1 = M$ och på samma sätt innebär $\forall x\psi(x)$ att $M_2 = M$. Nu följer det med satslogik att hela påståendet är sant i tolkningen M. Eftersom M var en godtycklig tolkning så är alltså utsagan logiskt sann.

Däremot är inte omvändningen

$$(\forall x\phi(x) \to \forall x\psi(x)) \to \forall x(\phi(x) \to \psi(x)) \tag{5.1}$$

en logisk sanning för alla formler ϕ och ψ. För att se detta kan vi låta ϕ och ψ vara de enställiga predikaten $A(x)$ respektive $B(x)$. Om $M = \{0, 1\}$ och vi tolkar $A(x)$ som $\{0\}$ och $B(x)$ som $\emptyset$ så kommer formeln (5.1) inte att vara sann. $\square$

UPPGIFT 5.3. Visa att

$$\exists y\forall x\phi \to \forall x\exists y\phi \tag{5.2}$$

är logiskt sann för varje formel ϕ.

Hur kan man då i allmänhet avgöra om en given formel är logiskt sann eller inte? Det finns ingen algoritm som i fallet med satslogik där man kan använda en sanningstabell. Man kan t.o.m. bevisa att det inte kan finnas någon sådan algoritm. Vad man kan göra är att försöka hitta ett bevis. Allt som följer med satslogik är förstås sant. Om man t ex redan vet att ϕ och $\phi \to \psi$ är logiskt sanna, så följer att ψ är logiskt sann. Vidare är formeln ϕ en logisk konsekvens av $\forall x\phi$ (och tvärtom). Man kan också lätt övertyga sig om att om man i en formel ϕ byter alla förekomster av variabeln x mot variabeln y, där y inte alls förekommer i ϕ, och kallar den nya formeln ψ, så är ψ en logisk konsekvens av ϕ (och vice versa).

EXEMPEL 5.12. Vi vill visa att $\forall x\exists yA(x, y) \leftrightarrow \forall y\exists xA(y, x)$. Man kan förstås inse detta direkt, eftersom båda leden "säger samma sak". Om man vill utnyttja regeln ovan så kan man resonera så här: Låt oss börja med $\forall x\exists yA(x, y)$. Ur denna kan vi dra slutsatsen $\forall x\exists zA(x, z)$

och denna implicerar i sin tur $\forall w \exists z A(w,z)$ som ger $\forall y \exists z A(y,z)$ som ger $\forall y \exists x A(y,x)$. På samma sätt åt andra hållet. $\square$

Vidare är det klart att formeln $\forall x \phi$ logiskt implicerar ψ som erhållits genom att ersätta varje fri förekomst av x i ϕ med individkonstanten a. För fullständighetens skull ger vi följande mer allmänna variant av det sistnämnda:

Anmärkning 5.13. Man säger att en term t är *fri för* variabeln x i formeln ϕ om ingen fri förekomst av x i ϕ påverkas av någon kvantifikator med avseende på någon variabel som ingår i t. Antag att t är fri för x i ϕ, och låt ψ vara den formel man får då man byter ut alla fria förekomster av x i ϕ mot t. Man kan då övertyga sig om att ψ är en logisk konsekvens av $\forall x \phi$. T ex är y fri för x i $A(x,y) \wedge \forall y \forall x B(x,y)$, men däremot inte i $\exists y A(x,y)$. Speciellt är alltså t fri för x om t inte innehåller några variabler alls, t ex om t är en individkonstant a. Vidare är t fri för x i ϕ om t är en variabel y som inte alls förekommer i ϕ. I båda dessa fall är alltså ψ en logisk konsekvens av ϕ, där ψ är den formel man får från ϕ genom att byta ut alla fria förekomster av x mot t. $\square$

Uppgift 5.4. Visa påståendet i anmärkning 5.13. Visa också med exempel att påståendet inte behöver vara sant om man gör avkall på villkoret att t ska vara fri för x.

Uppgift 5.5. Antag att t är fri för x i ϕ och antag att ψ uppkommer från ϕ som ovan. Visa att $\exists x \phi$ är en logisk konsekvens av ψ.

Några exempel på predikatlogiska teorier. Vi ska nu titta på några exempel på teorier som kan formuleras i predikatlogik. Man bör understryka att detta inte är möjligt för alla matematiska teorier. Som vi ska se i nästa avsnitt finns det exempelvis ingen predikatlogisk teori vars modeller är precis alla *ändliga* grupper. Peanos axiomatisering av de naturliga talen kan inte heller uttryckas i predikatlogik.

Exempel 5.14 (Teorin för mängder med två element). Vi låter denna teori ha som enda axiom

$$\exists x \exists y \forall z (\neg(x = y) \wedge (z = x \vee z = y)).$$

Modellerna till denna teori är precis de mängder som har exakt två element. $\square$

Exempel 5.15 (En teori som saknar modeller). Betrakta teorin med det enda axiomet

$$\forall x \forall y (\neg(x = y)).$$

Eftersom $=$ svarar mot likhetsrelationen i varje tolkning så finner vi att denna teori saknar modell. $\square$

Det är lätt att beskriva gruppteori med predikatlogik, jmf. exempel 4.6.

EXEMPEL 5.16 (Gruppteori). Låt alfabetet (utöver de logiska symbolerna, $=$ och variabler) bestå av en individkonstant e samt en funktionssymbol $f(x,y)$ (som vi skriver xy). Som axiom har vi

$$\begin{aligned} &(xy)z = x(yz), \\ &\quad xe = x \quad \wedge \quad ex = x \\ &\forall x \exists y (xy = e \quad \wedge \quad yx = e). \end{aligned}$$

En modell till denna teori kallas en *grupp*. $\square$

Vi vet redan från avsnitt 4 att formeln

$$\exists y(xy = y) \to x = e$$

är en sats i gruppteori. Även för så här enkla satser är det dock rätt knepigt att genomföra beviset med symboler. Den som vill kan försöka att skriva ned ett sådant bevis med utgångspunkt från beviset i avsnitt 4.

Det är lätt att se att inte bara gruppteori utan även de andra vanliga algebraiska strukturerna, som teorin för *ringar, kroppar, linjära rum, moduler* etc, kan formuleras i predikatlogik.

Det vi talar om i detta avsnitt är *första ordningens* predikatlogik. Den utmärks av att man tillåter kvantifiering över elementen i en modell men inte över delmängder till modellen. Om detta är möjligt talar man om andra ordningens predikatlogik.

Det är därför inte möjligt att uttrycka Peanos axiom med predikatlogik. Problemet är axiom P5 (induktionsaxiomet) (jmf exempel 4.3) som kvantifierar över mängder; observera att det kan läsas som "För varje delmängd X gäller att ...". Istället får man hålla till godo med den svagare formuleringen (a_7) nedan. Den uttalar sig bara om sådana delmängder av en modell som kan beskrivas med en predikatlogisk formel.

EXEMPEL 5.17 (Teorin T). Utöver de vanliga symbolerna innehåller alfabetet en individkonstant 0 samt tre funktionssymboler $S(x)$ (som vi också skriver x'), $f(x,y)$ (som vi oftast skriver $x+y$) och $g(x,y)$ (som vi skriver som xy eller $x \cdot y$). Vi har följande axiom

(a_1) $$x' \neq 0$$

(a_2) $$x' = y' \to x = y$$

$$(a_3) \qquad x + 0 = x$$

$$(a_4) \qquad x + y' = (x + y)'$$

$$(a_5) \qquad x \cdot 0 = 0$$

$$(a_6) \qquad x \cdot y' = x \cdot y + x$$

$$(a_7) \qquad (\phi(0) \wedge \forall x(\phi(x) \to \phi(x'))) \longrightarrow \forall x \phi(x),$$

där $\phi(x)$ är en formel med fri variabel x.

Noga taget är alltså (a_7) inte ett axiom utan oändligt många, nämligen ett för varje formel $\phi(x)$. Man känner igen (a_7) som en beskrivning av induktionsprincipen, dvs om ett påstående gäller för 0, och det vidare är så att det gäller för x' om det gäller för x, så följer att det gäller för alla x. Standardmodellen för denna teori är de naturliga talen $\mathbb{N}$, där 0 tolkas som 0, + som + och $\cdot$ som $\cdot$, och $'$ som efterföljarfunktionen, dvs $x' = x + 1$. Vi ska strax se att denna teori även har andra modeller. □

Några andra viktiga exempel på predikatlogiska teorier som vi kommer att stöta på i kommande kapitel är *Axiomatisk mängdteori* och *Absolut geometri.*

Kompakthetssatsen. Följande resultat som vi formulerar utan bevis är grundläggande i vad som kallas *modellteori*, och har flera intressanta konsekvenser.

SATS 5.18 (Kompakthetssatsen). *Antag att Φ är en mängd av (predikatlogiska) formler. Om varje ändlig delmängd av Φ har en modell, så finns det en modell för hela Φ.*

Med hjälp av denna sats kan man visa att teorin för ändliga grupper inte går att formulera i predikatlogik.

PROPOSITION 5.19. *Det finns ingen predikatlogisk teori vars modeller är precis alla ändliga grupper.*

Med samma bevis får man att det inte finns någon predikatlogisk teori vars modeller är alla ändliga mängder, eller alla ändliga ringar etc.

BEVIS. Antag att Φ har varje ändlig grupp som modell. Lägg till de nya individkonstanterna $a_1, a_2, a_3, \ldots$ samt de nya formlerna (axiomen) $a_1 \neq a_2,\ a_1 \neq a_3,\ a_2 \neq a_3, \ldots$. Eftersom det finns godtyckligt stora ändliga grupper, så har varje ändlig delmängd av $\Phi \cup \{a_1 \neq a_2,\ a_1 \neq$

$a_3,\ a_2 \neq a_3, \ldots\}$ en modell. Enligt kompakthetssatsen finns alltså en modell G för hela formelmängden $\Phi \cup \{a_1 \neq a_2,\ a_1 \neq a_3,\ a_2 \neq a_3 \ldots\}$. Eftersom G speciellt är en modell för formelmängden $\{a_1 \neq a_2,\ a_1 \neq a_3,\ a_2 \neq a_3 \ldots\}$ så följer att G inte är ändlig. Eftersom G är en modell till Φ har vi visat att Φ har även oändliga modeller. □

Exempel 5.20. Betrakta igen teorin T i exempel 5.17. Man kan fråga sig om det följer från axiomen att varje y i varje modell är något av elementen $0,\ 0',\ 0'',\ 0''',\ 0^{(4)}, \ldots$? Svaret är nej, se övning 5.6, och häri ligger lite av predikatlogikens svaghet. I vårt vanliga språk kan vi ju lätt uttrycka att "varje x är $0^{(m)}$ för något heltal m" vilket säger samma sak som axiom P5 i exempel 4.3 men denna typ av axiom går inte att formulera i predikatlogik. □

Uppgift 5.6. Visa att det finns en modell till teorin i exempel 5.17 där alla formlerna $a \neq 0,\ a \neq 0',\ a \neq 0''$, etc är sanna.

En modell av detta slag brukar kallas en *icke-standardmodell*.

Formella härledningar. För att formalisera bevisföring (i predikatlogik) så inför man ett system av *härledningsregler*. En härledningsregel är en princip enligt vilken man får en formel ur ändligt många andra givna formler. Den ska vara sådan att den den erhållna formeln är en logisk konsekvens av de andra formlerna. Ett exempel på ett sådant system av härledningsregler är följande:
1. Från formlerna $\phi_1, \ldots, \phi_k$ får man dra slutsatsen ϕ, om ϕ är en satslogisk konsekvens av $\phi_1, \ldots, \phi_k$.
2. Från ϕ får man dra slutsatsen $\forall x \phi$.
3. Antag att ϕ är en formel och t är en term som är fri för x i ϕ (jmf anmärkning 5.13 ovan) och att ψ är den formel man får om varje förekomst av x ersätts med t. Då får man dra slutsatsen $\forall x \phi \to \psi$. Speciellt kan man ta $t = x$. Om ϕ saknar fria förekomster av x så får man dra slutsatsen $\forall x(\phi \to \psi) \to (\phi \to \forall x \psi)$.

Vi ska inte här fördjupa oss i hur det kommer sig att just detta är ett "tillräckligt" system av härledningsregler, utan hänvisar till en bok om predikatlogik.

En *formell härledning* (eller ett *formellt bevis*) i predikatlogik av en sats ϕ i en predikatlogisk teori Ψ, är en ändlig rad av formler $\phi_1, \phi_2, \ldots, \phi_m$, där varje formel antingen är ett axiom, dvs en formel ur Ψ, eller följer ur de föregående ϕ_j genom någon av härledningsreglerna, och där $\phi = \phi_m$. Om det finns en härledning av ϕ från axiomen Ψ så skriver vi

$$\Psi \vdash \phi.$$

Notera att om $\Psi \vdash \phi$ så är ϕ en logisk konsekvens av Ψ, eller annorlunda uttryckt, ϕ är en sats i teorin Ψ. Detta följer av att man med induktion inser att var och en av formlerna i en härledning av ϕ från Ψ är en sats i teorin Ψ.

Att även omvändningen är sann, är innehållet i följande grundläggande resultat som bevisades av Gödel 1930.

Sats 5.21 (Gödels fullständighetssats). *Om Φ är en teori och ϕ är en logisk konsekvens av Φ, dvs en sats i Φ, så finns en härledning av ϕ från Φ, dvs $\Phi \vdash \phi$.*

Satsen säger alltså att varje sant påstående i en predikatlogisk teori kan härledas (bevisas). Det bör dock understrykas att beviset av fullständighetssatsen är ett rent existensbevis, dvs det anger inte alls någon metod för hur man kan hitta en härledning av en given sats, eller metod att avgöra om en given formel är en sats i teorin Φ. Dock har fullständighetssatsen många viktiga konsekvenser; t ex medför den omedelbart kompakthetssatsen.

Bevis av kompakthetssatsen från fullständighetssatsen. Om Φ saknar modell så är $\phi \wedge \neg\phi$ sann i varje modell till Φ, så $\phi \wedge \neg\phi$ är en logisk konsekvens av Ψ. Enligt fullständighetssatsen finns därför en härledning av formeln $\phi \wedge \neg\phi$ från Φ, dvs $\Phi \vdash \phi \wedge \neg\phi$. Men en härledning involverar bara ett ändligt antal av formlerna i Φ och därför finns en ändlig delmängd Φ' av Φ sådan att $\Phi' \vdash \phi \wedge \neg\phi$. Detta strider dock mot antagandet att Φ' har en modell, eftersom $\phi \wedge \neg\phi$ är en motsägelse och därför inte kan vara sann i någon modell. Alltså följer att Φ måste ha någon modell. □

Uppgift 5.7. Visa att fullständighetssatsen kan omformuleras på följande vis:

Sats 5.22. *Om Φ är en satsmängd ur vilken man inte kan härleda en motsägelse, så har Φ en modell.*

En satsmängd Φ kallas *konsistent* om det inte går att härleda en motsägelse ur Φ. Det följer att Φ är konsistent om och endast om Φ har en modell.

Uppgift 5.8. Visa detta!

6. Ytterligare övningar till kapitel 1

Uppgift 6.1. Fundera på vad som skiljer utsagorna "Susan har fem bokstäver" och "Susanne har fem bokstäver".

Uppgift 6.2. Vilken eller vilka av följande formler är logiskt sanna? Ange till varje icke-sann utsaga en tolkning i vilken den är falsk.

$$(i) \qquad \forall x \exists y A(x,y) \to \exists y \forall x A(x,y)$$

$$(ii) \qquad \exists x \forall y A(x,y) \to \forall y \exists x A(x,y)$$

$$(iii) \qquad \forall y \exists x A(x,y) \to \exists y \forall x A(x,y)$$

$$(iv) \qquad \exists x \forall y A(x,y) \to \forall z \exists x A(x,z)$$

$$(v) \qquad \exists x \forall y A(x,y) \to \forall z \exists x A(z,x)$$

$$(vi) \qquad \exists x \forall y A(x,y) \to \exists x \forall y A(y,x)$$

$$(vii) \qquad \forall y \exists x A(x,y) \to \forall y \exists x A(y,x)$$

$$(viii) \qquad \forall x(A(x) \to B(x)) \to (\forall x A(x) \to \forall x B(x))$$

$$(ix) \qquad (\forall x A(x) \to \forall x B(x)) \to \forall x(A(x) \to B(x))$$

$$(x) \qquad \forall x(A(x) \vee B(x)) \to (\forall x A(x) \vee \forall x B(x)$$

$$(xi) \qquad (\forall x A(x) \vee \forall x B(x)) \to \forall x(A(x) \vee B(x))$$

$$(xii) \qquad \forall x(A(x) \wedge B(x)) \to (\forall x A(x) \wedge \forall x B(x))$$

$$(xiii) \qquad (\forall x A(x) \wedge \forall x B(x)) \to \forall x(A(x) \wedge B(x))$$

Uppgift 6.3. Visa att det för varje $m > 0$ finns en satslogisk utsaga ϕ_m innehållande m satsvariabler $A_1, \ldots, A_m$ sådan att ϕ är sann om och endast om ett jämnt antal av A_j är sanna.

Uppgift 6.4. Visa att det till varje satslogisk utsaga finns en ekvivalent utsaga som bara består av konnektiverna $\neg$ och $\vee$ samt satsvariabler.

Uppgift 6.5. Antag att man till varje kombination av sanningsvärden på satsvariablerna $A_1, \ldots, A_m$ har angett antingen S eller F. Visa att det finns en sats ϕ som är sann precis för de kombinationer av sanningsvärden på $A_1, \ldots, A_m$ för vilka man har angett S.

UPPGIFT 6.6. Vilken eller vilka (om någon) av följande utsagor uttrycker att det finns precis två element så att $A(x)$ gäller?

$$(i) \qquad \exists x \exists y (A(x) \wedge A(y) \wedge x \neq y \wedge \forall z (A(z) \rightarrow (z = x \vee z = y)))$$

$$(ii) \qquad \exists x \exists y \forall z (A(x) \wedge A(y) \wedge x \neq y \wedge (A(z) \rightarrow (z = x \vee z = y)))$$

$$(iii) \qquad \exists x \exists y (A(x) \wedge A(y) \wedge x \neq y \wedge \forall z (z = x \vee z = y \vee \neg A(z)))$$

UPPGIFT 6.7. Ange en predikatlogisk formel vars modeller är alla mängder med minst två element.

UPPGIFT 6.8. Ange en predikatlogisk formel vars modeller är alla mängder med högst två element.

UPPGIFT 6.9. Ange en predikatlogisk formel vars modeller är precis alla mängder med exakt tre element.

UPPGIFT 6.10. Ange en predikatlogisk formel vars modeller är alla mängder med två eller tre element.

UPPGIFT 6.11. Låt f vara en funktion på $\mathbb{R}$. Avgör vilken eller vilka av följande utsagor uttrycker negationen av att $f(x) \to 1$ då $x \to 0$.

- (i) Det finns $B \neq 1$ och $\epsilon > 0$ sådana att det till varje $\delta > 0$ finns x så att $|f(x) - B| < \epsilon$ men $|x| \geq \delta$.
- (ii) Det finns $\epsilon > 0$ sådant att det till varje $\delta > 0$ gäller att $|f(x) - 1| \geq \epsilon$ om $|x| < \delta$.
- (iii) Det finns B sådant att $f(x) \to B$ då $x \to 0$ och $B \neq 1$.
- (iv) Till varje $\epsilon > 0$ finns $\delta > 0$ och x sådant att $|f(x) - 1| \geq \epsilon$ men $|x| < \delta$.

UPPGIFT 6.12. Låt f vara en funktion på $\mathbb{R}$. Vilken eller vilka av följande utsagor uttrycker precis definitionen av att f är kontinuerlig på $\mathbb{R}$?

- (i) Till varje $\epsilon > 0$ finns ett $\delta > 0$ så att det till varje x och y gäller att $|f(x) - f(y)| < \epsilon$ om $|x - y| < \delta$.
- (ii) Till varje x och varje $\epsilon > 0$ finns ett $\delta > 0$ så att det för varje y gäller att $|f(x) - f(y)| < \epsilon$ om $|x - y| < \delta$.
- (iii) Till varje $\epsilon > 0$ och varje x finns ett $\delta > 0$ så att det för varje y gäller att $|f(x) - f(y)| < \epsilon$ om $|x - y| < \delta$.
- (iv) Till varje x och y och varje $\epsilon > 0$ finns $\delta > 0$ så att $|f(x) - f(y)| < \epsilon$ om $|x - y| < \delta$.

UPPGIFT 6.13. Låt f vara en funktion på $\mathbb{R}$. Vilken eller vilka av följande utsagor uttrycker negationen av att f är kontinuerlig på $\mathbb{R}$?

- (i) Det finns $\epsilon > 0$ och y sådant att det till varje $\delta > 0$ finns x så att $|f(x) - f(y)| < \epsilon$ men $|x - y| \geq \delta$.
- (ii) Det finns $\epsilon > 0$ sådant att det till varje $\delta > 0$ finns f och $x, y \in \mathbb{R}$ sådana att $|x - y| < \delta$ och $|f(x) - f(y)| \geq \epsilon$.
- (iii) För varje y finns $\epsilon > 0$ och x sådana att $|x - y| < \delta$ för alla $\delta > 0$ och $|f(x) - f(y)| \geq \epsilon$.
- (iv) Det finns $\epsilon > 0$ och y sådant att det till varje $\delta > 0$ finns x så att $|f(x) - f(y)| \geq \epsilon$ men $|x - y| < \delta$.

Uppgift 6.14. Visa direkt med definitionen att $f(x) = x^2$ är kontinuerlig. Visa direkt med definitionen att $f(x) = x^2$ inte är likformigt kontinuerlig.

Uppgift 6.15. Låt a_n vara en reell talföljd. Vilken eller vilka av följande utsagor uttrycker precis negationen av att $a_n \to 0$ då $n \to \infty$?

- (i) Det finns ett $\epsilon > 0$ sådant att det till varje N gäller att $|a_n| \geq \epsilon$ för alla n sådana att $n \geq N$.
- (ii) Det finns ett $\epsilon > 0$ sådant att det till varje N finns ett $n \geq N$ sådant att $|a_n| \geq \epsilon$.
- (iii) Det finns ett $\epsilon > 0$ och ett N sådant att att $|a_n| \geq \epsilon$ om $n \geq N$.
- (iv) Till varje N finns ett $\epsilon > 0$ så att $|a_n| \geq \epsilon$ för all n sådana att $n \geq N$.

Uppgift 6.16. Låt f vara en funktion. Vilket eller vilka av följande påståenden är negationen av att $f(a_n) \to 0$ för alla följder a_n sådana att $a_n \to 0$ då $n \to \infty$?

- (i) Det finns $\epsilon > 0$ och N s.a. $|f(a_n)| \geq \epsilon$ för alla $n \geq N$ och varje följd a_n s.a. $a_n \to 0$ då $n \to \infty$.
- (ii) Det finns en följd a_n och ett tal N s.a. $|f(a_n)| \geq \epsilon$ för alla $\epsilon > 0$ om $n \geq N$.
- (iii) Det finns en följd a_n och $\epsilon > 0$ s.a. $a_n \to 0$ då $n \to \infty$ och så att det för varje N finns $n \geq N$ s.a. $|f(a_n)| \geq \epsilon$.
- (iv) Det finns en följd a_n och $\epsilon > 0$ s.a. det för varje N finns $n \geq N$ s.a. $|f(a_n)| \geq \epsilon$ och s.a. det för varje $\epsilon' > 0$ finns ett M s.a. $|a_n| < \epsilon'$ om $n \geq M$.

Uppgift 6.17. Vilken eller vilka av följande utsagor är negationen av att funktionen $f(x)$ är begränsad på $\mathbb{R}$? (Att f är begränsad betyder att det finns M sådant att $|f(x)| \leq M$ för alla x.)

- (i) Till varje $\omega > 0$ finns reella tal y och x sådana att $f(x) > \omega$ och $f(y) < -\omega$.
- (ii) Det finns ett $\omega > 0$ sådant att det för alla x gäller att $f(x) > \omega$ eller $f(x) < -\omega$.

- (iii) Till varje $\omega > 0$ finns $x \in \mathbb{R}$ sådant att $f(x) > \omega$ eller $f(x) < -\omega$.
- (iv) För varje $\omega > 0$ och $x \in \mathbb{R}$ så är $f(x) > \omega$ eller $f(x) < -\omega$.

UPPGIFT 6.18. En mängd M av kontinuerliga funktioner på $(0,1)$ är *ekvikontinuerlig*, om det till varje $\epsilon > 0$ finns ett $\delta > 0$ s.a. $|f(x) - f(y)| < \epsilon$ om $|x - y| < \delta$ och $f \in M$. Vilken eller vilka av följande utsagor uttrycker negationen av att M är ekvikontinuerlig?

- (i) Det finns $f \in M$ och $\epsilon > 0$ sådana att det till varje $\delta > 0$ finns $x, y \in (0,1)$ så att $|f(x) - f(y)| < \epsilon$ men $|x - y| \geq \delta$.
- (ii) Det finns $\epsilon > 0$ sådant att det till varje $\delta > 0$ finns $f \in M$ och $x, y \in (0,1)$ sådana att $|x - y| < \delta$ och $|f(x) - f(y)| \geq \epsilon$.
- (iii) För varje $f \in M$ finns $\epsilon > 0$ och $x, y \in (0,1)$ sådana att $|x - y| < \delta$ för alla $\delta > 0$ och $|f(x) - f(y)| \geq \epsilon$.
- (iv) Det finns $\epsilon > 0$ sådant att det till varje $\delta > 0$ finns $f \in M$ och $x, y \in (0,1)$ sådana att $|f(x) - f(y)| \geq \epsilon$ men $|x - y| < \delta$.

Några kommentarer

Grundläggande framställningar av logik finner man i [**DD**] eller [**EM**]. Här finns noggranna definitioner och och bevis av de resultat, som vi bara har berört ytligt i detta kapitel.

Härledningsregler och formella härledningar med s.k. *naturlig deduktion* kan man läsa om i [**DD**]. Detta har stor betydelse inom datalogi.

I [**DD**] finns även en framställning av grundläggande modellteori.

KAPITEL 2

Mängdteori

I det här kapitlet ska vi diskutera mängdbegreppet i matematik och beskriva hur det s.k. *mängduniversum* är uppbyggt. Vi går först igenom Zermelo-Fraenkels axiomatisering av mängdteori. Vi ska även se hur man kan definiera naturliga tal (och i sin förlängning heltal, rationella och reella tal, funktioner etc, kort sagt all matematik man stöter på i tidigare kurser) utgående från mängdteori. Vi ska även definiera vad det innebär att två mängder har samma antal element, samma *kardinalitet*, och visa att olika oändliga mängder kan ha olika kardinalitet.

1. Zermelo-Fraenkels axiom för mängdteorin

Mängdteorin introducerades av Cantor på 1880-talet. Till en början möttes den av stor misstro och motstånd men betraktas idag som grundläggande för all matematik, och så gott som all existerande matematik kan i princip formuleras inom mängdteorin. Mängdteorin handlar förstås om mängder. Vad är då en mängd? Följande försök till definition kanske inte gör en så mycket klokare: En mängd är en samling av (matematiska) objekt betraktade som en enhet. Detta leder ju genast till frågan vad ett objekt är. Är t ex $\mathbb{N}$, mängden av alla naturliga tal, ett objekt i sig? En poäng med mängdteori är just att även oändliga mängder ska kunna betraktas som objekt i sig, och att dessa ska kunna vara element i nya mängder. Vi ska se senare att ett reellt tal naturligen definieras som en viss oändlig mängd, och naturligtvis vill vi kunna tala om mängder av reella tal.

I mängdteorin gjorde man det första systematiska studiet av (hur man kan handskas med) oändliga mängder. Eftersom en del av denna verksamhet stred mot somligas sunda förnuft försökte man att formalisera det hela och ange ett antal axiom och/eller regler som bestämde och begränsade hur man skulle få handskas med mängder. Detta visade sig inte vara alldeles lätt. I de första försöken att bilda en formaliserad teori upptäckte man motsägelser. Dessa berodde löst talat på att man "bildade" nya mängder alltför ohämmat; man tillät sig att bilda alla mängder av typen $\{x;\ P(x)\}$, där $P(x)$ är ett villkor, t ex en predikatlogisk formel. Den mest kända motsägelsen är Russels paradox.

Exempel 1.1 (Russels paradox). Låt M vara mängden av alla mängder som inte är ett element i sig själv. Man ser genast då att $M \in M$ om $M \notin M$ men även att $M \notin M$ om $M \in M$. Alltså får man en motsägelse. □

Uppgift 1.1. Ett mer populärt sätt att formulera samma "paradox" är: "Barberaren i Sevilla rakar precis de invånare som inte rakar sig själva. Vem rakas då barberaren av?" Är detta verkligen en paradox? Försök att förklara!

För att undvika detta fick man vara lite mer försiktig med hur man bildar nya mängder. Vi ska nu gå igenom en uppsättning axiom för mängdteori ur vilken man, såvitt man vet, inte kan härleda några motsägelser. Det visar sig enklast att bara tala om mängder, så att varje element i en mängd själv är en mängd, istället för att ha något slags *urelement*. Axiomen beskriver en intuitiv idé om hur mängduniversum är uppbyggt med utgångspunkt från den tomma mängden och ett antal mängdbildningsprinciper.

Anmärkning 1.2. På grund av den skepsis som mängdteorin möttes av i början, satte man upp som mål att *bevisa* att den aldrig skulle kunna leda till några motsägelser. Handskandet med oändliga mängder skulle alltså legitimeras genom att visa att det aldrig kunde leda till något felaktigt om "riktig" dvs "ändlig matematik". Detta mål fick överges på 30-talet, då Gödel *bevisade* att det inte går att bevisa motsägelsefrihet för mängdteorin (utan att gå utanför mängdteorin själv). En vanlig ståndpunkt bland matematiker nu för tiden är att mängduniversum existerar oberoende av oss själva och våra axiom. Då är varje påstående om mängduniversum sant eller falskt och om bara axiomen för mängdteori är sanna påståenden om detta mängduniversum så följer det att mängdteorin är motsägelsefri eftersom mängduniversum då är en modell. En helt annan sak är vilka satser om mängduniversum som går att *bevisa* från de givna axiomen (eller någon delmängd av dem); se anmärkning 6.8. □

Vad vi presenterar här är en predikatlogisk teori, som brukar kallas Zermelo–Fraenkels axiomatisering av mängdteorin ZF, som beskrivs av axiomen ZFI t.o.m. ZFVIII nedan. Utöver de logiska symbolerna och $=$ har vi ett tvåställigt predikat $\in$. För att det ska bli någorlunda hanterbart så formulerar vi dock de flesta axiomen på svenska, och lämnar som övning att fundera ut hur de kan formuleras i predikatlogik. Vi kommer också att införa nya predikatsymboler, som man kan se som förkortade skrivsätt, t ex $x \notin y$ för $\neg(x \in y)$. Att $x \in y$ gäller utläser vi som att "x *tillhör* y" eller "x är ett *element* i y".

ZF I (extensionalitetsaxiomet) Om två mängder innehåller samma element så är de lika.

Uttryckt med symboler är detta

$$\forall y \forall z (\forall x (x \in y \leftrightarrow x \in z) \rightarrow y = z).$$

Axiomet uttrycker det för vår intuition självklara att en mängd är entydigt bestämd av sina element.

Om a och b är mängder sådana att varje element i b också är ett element i a säger man att b är en *delmängd* till a. Vi skriver detta $b \subset a$; obs att detta även inbegriper fallet att $b = a$. Nästa axiom uttrycker möjligheten att bilda delmängder.

ZF II (delmängdsaxiomet) Givet ett villkor (en formel med en fri variabel) och en mängd a så existerar den mängd vars element är de element i a som uppfyller villkoret.

Givet en mängd a och formel ϕ så är alltså

$$b = \{x \in a;\ \phi\}$$

en mängd. I praktiken uttrycker man ofta villkoret på svenska men man måste då förvissa sig om att det åtminstone i princip är möjligt att formulera det predikatlogiskt, jmf exemplet nedan. (Noga taget är inte ZF II *ett* axiom utan en oändlig mängd axiom; ett för varje utsaga ϕ med en fri variabel x.)

I nästa avsnitt inför vi ett axiom, replacementaxiomet, som är en förstärkning av delmängdsaxiomet.

Exempel 1.3. Vi ska senare se att alla de naturliga talen tillsammans utgör en mängd $\mathbb{N}$. Enligt delmängdsaxiomet är då även till exempel alla jämna kvadrater en mängd för de är " alla heltal x sådana att det finns ett heltal y sådant att $y^2 = x$" eller mer formellt

$$M = \{x \in \mathbb{N};\ (\exists y \in \mathbb{N})(y^2 = x)\}.$$

□

Exempel 1.4. Betrakta nu "alla naturliga tal som kan uttryckas med färre än 90 bokstäver på svenska". Skenbart är detta en mängd precis som i förra exemplet, men vi ska se att så inte är fallet. Antag därför att det finns en mängd a som består av precis de tal som kan uttryckas med färre än 90 bokstäver på svenska. Eftersom det bara finns ändligt många kombinationer av 90 bokstäver så måste a vara ändlig, och alltså måste den innehålla ett största element. Betrakta nu talet m som är "ett plus det största naturliga tal som kan uttryckas på svenska med färre än nittio bokstäver". Man ser att å ena sidan $m \in a$ fast å andra sidan att $m \notin a$. $\square$

Uppgift 1.2. Försök förstå vari "felet" består!

Om a betecknar en godtycklig mängd så kan vi definiera den tomma mängden som

$$\emptyset = \{x \in a;\ x \neq x\}. \tag{1.1}$$

Eftersom denna mängd saknar element följer det från ZF I att den inte beror på valet av a. Noga taget kräver detta att man vet att det finns åtminstone någon mängd. man kan säkerställa detta genom att införa en godtycklig individkonstant. Alternativt kan vi lägga till en individkonstant $\emptyset$ till vårt alfabet och ett axiom i stil med $\forall x(x \notin \emptyset)$. I så fall inser man lätt att (1.1) blir en sats i teorin.

Vi kan använda idén i Russels paradox till att visa att att det inte finns någon universell mängd; dvs en i vilken alla mängder är element. Låt m vara en godtycklig mängd och låt $a = \{x \in m;\ x \notin x\}$. Då är a en mängd enligt delmängdsaxiomet och vi påstår att $a \notin m$. Om nämligen $a \in m$ så har vi att $a \in a$ medför $a \notin a$ och att $a \notin a$ medför $a \in a$, vilket är en motsägelse. Alltså innehåller inte m alla mängder.

ZF III (paraxiomet) Givet mängder a och b. Då existerar den mängd $\{a, b\}$ vars element är a och b; dvs

$$\forall y \forall z \exists u \forall x (x \in u \leftrightarrow x = y \vee x = z)$$

Låt a vara en icketom mängd (av mängder). Om b är ett element i a så är

$$\{x \in b;\ \forall y(y \in a \to x \in y)\},$$

en mängd enligt delmängdsaxiomet. Denna kallas *snittet* av mängderna i a. Alltså består snittet av precis de element x som är element i var och en av de y som är element i a. Speciellt gäller alltså att snittet inte beror på valet av b. Vi betecknar snittet

$$\bigcap_{y \in a} y$$

(ibland används även beteckningen $\cap a$). Om a bara har två element, säg $a = \{b, c\}$, så skriver vi snittet som

$$b \cap c.$$

Om b och c är två mängder vilka som helst så är $\{b, c\}$ en mängd enligt paraxiomet och alltså är snittet av b och c en mängd. Detta följer även utan paraxiomet eftersom $a \cap b = \{x \in b;\ x \in a\}$. Två mängder är *disjunkta* om de har tomt snitt. Vi inför beteckningen $a \setminus b$ för mängden $\{x \in a;\ x \notin b\}$.

Anmärkning 1.5. Om i en given situation alla mängder man arbetar med är delmängder till någon fix grundmängd M, så kallas $M \setminus a$ för komplementet av a (med avseende på M) och betecknas $\complement a$ eller a^c. Notera dock att begreppet "komplement av en mängd " saknar mening om man inte specificerar en grundmängd. □

ZF IV (unionsaxiomet) Givet en mängd a (av mängder). Då existerar den mängd vars element är precis de mängder x sådana att det finns ett $y \in a$ sådant att $x \in y$. Den kallas *unionen* av mängderna i a och vi betecknar den

$$\bigcup_{y \in a} y$$

(ibland används beteckningen $\cup a$), eller $b \cup c$ om a bara består av de två elementen b och c.

Exempel 1.6. Om $a = \{\{b, c\}, \{c, d\}, \{b\}\}$ så är unionen $\{b, c, d\}$. □

Uppgift 1.3. Visa att om a och b är godtyckliga mängder så är

$$b \setminus \bigcup_{y \in a} y = \bigcap_{y \in a} (b \setminus y).$$

Vi kan nu definiera de naturliga talen i mängdteorin. Låt 0 definieras som den tomma mängden $\emptyset$. Vi låter sedan $1 = \{0\} = \{\emptyset\}$, $2 = \{0, 1\}$ dvs $2 = \{\emptyset, \{\emptyset\}\}$ etc. Allmänt är då

$$n = n - 1 \cup \{n - 1\} = \{0, 1, 2, \ldots, n - 1\}.$$

Speciellt ser vi att n är en mängd med n element.

ZF V (oändlighetsaxiomet) Det finns en mängd som har precis de naturliga talen $0, 1, 2, \ldots$ som element.

Vi betecknar denna mängd med $\mathbb{N}$, mängden av de naturliga talen. Speciellt medför det här axiomet att det finns oändliga mängder (se nedan för exakt definition).

ZF VI (potensmängdsaxiomet) Givet en mängd a så existerar den mängd vars element är alla delmängder till a. Denna mängd betecknas $P(a)$.

Exempel 1.7. Om a har två element, säg $a = \{b, c\}$, så är

$$P(a) = \big\{\{b, c\}, \{b\}, \{c\}, \emptyset\big\}.$$

$\square$

Nu kommer ett axiom som bl a säkerställer att ingen mängd är ett element i sig själv.

ZF VII (regularitetsaxiomet) Varje icketom mängd a innehåller ett element b sådant att $a \cap b = \emptyset$.

Man tänker sig att mängduniversum är uppbyggt i steg i enlighet med ZF II till ZF VI. I första steget har man bara den tomma mängden. I nästa steg kan man bilda mängden av den tomma mängden etc. I varje steg sedan kan man bilda mängder vars element är redan tidigare bildade mängder. Enligt det här synsättet är regularitetsaxiomet rimligen sant eftersom om b är ett element i a som har bildats vid ett så tidigt steg som möjligt så är elementen i b från något ännu tidigare steg och kan därför inte vara element i a.

Exempel 1.8. Om $a = \{\{\emptyset\}, \{\{\emptyset\}, \emptyset\}\}$ så duger det första elementet som b, dvs $a \cap \{\emptyset\} = \emptyset$, men inte det andra eftersom $a \cap \{\{\emptyset\}, \emptyset\} = \{\{\emptyset\}\} \neq \emptyset$. $\square$

Proposition 1.9. *Det finns ingen mängd som är ett element i sig själv, dvs $c \notin c$ för alla mängder c.*

Bevis. Tag en godtycklig mängd c och betrakta mängden $a = \{c\}$. Enligt regularitetsaxiomet finns $b \in a$ sådant att $b \cap a = \emptyset$. Men eftersom a bara har ett element så är $b = c$ och alltså $c \cap \{c\} = \emptyset$, dvs $c \notin c$. $\square$

Uppgift 1.4. Visa att det inte finns mängder a och c sådana att $a \in c \in a$.

Vi kan nu visa att de naturliga talen $\mathbb{N}$ så som de är definierade ovan, uppfyller Peanos axiom, jmf avsnitt 4 i kapitel 1. De två första axiomen, P1 och P2, är uppenbara. Eftersom $n' = n \cup \{n\}$ så är den icketom och alltså skild från $\emptyset = 0$, vilket visar P3. Om $n' = m'$ så är $m \cup \{m\} = n \cup \{n\}$, och speciellt har vi då att $m \in n \cup \{n\}$ och $n \in m \cup \{m\}$. Om nu $m = n$ är vi klara; om inte så har vi att $m \in n$ och att $n \in m$, dvs $n \in m \in n$, och detta strider mot regularitetsaxiomet, se övning 1.4. Alltså gäller P4. Axiom P5, slutligen, följer direkt av oändlighetsaxiomet.

2. Relation och funktion

Vi ska nu definiera begreppet *ordnat par* av två mängder. Givet mängder a och b, sätt

$$(a, b) = \{\{a\}, \{a, b\}\}$$

som ju är en mängd enligt paraxiomet. Det väsentliga här är inte den exakta definitionen utan att ordnade par har följande karakteristiska egenskap:

PROPOSITION 2.1. *Om a, b, c, d är mängder så är $(a, b) = (c, d)$ om och endast om $a = c$ och $b = d$.*

BEVIS. Om $a = c$ och $b = d$ är det självklart att $(a, b) = (c, d)$, så vi har att visa omvändningen. Vi antar alltså att $(a, b) = (c, d)$, dvs $\{\{a\}, \{a, b\}\} = \{\{c\}, \{c, d\}\}$. Antag först dessutom att $a \neq b$. Då är $\{a\} \neq \{a, b\}$ och alltså har mängden $\{\{a\}, \{a, b\}\}$ två element; ett av dem består i sin tur av ett element och det andra består av två element. Därför har även $\{\{c\}, \{c, d\}\}$ två element och alltså måste $c \neq d$, ty annars skulle $\{c, d\} = \{c\}$ och därför $\{\{c\}, \{c, d\}\}$ vara lika med $\{\{c\}\}$ och alltså bara ha ett element. Nu följer det att $\{a\} = \{c\}$ och $\{a, b\} = \{c, d\}$, vilket ger att $a = c$ och $b = d$. Om $a = b$ måste vi ha att $c = d$, så antagandet blir att $\{\{a\}\} = \{\{c\}\}$ vilket medför att att $a = c$ och följaktligen $b = d$. $\square$

Givet två mängder A och B kan vi nu definiera den *kartesiska produkten*

$$A \times B = \{(a, b);\ a \in A \text{ och } b \in B\}.$$

Vid första anblicken kan detta tyckas lite äventyrligt eftersom vi inte har skrivit vad den är en delmängd av (jmf delmängdsaxiomet). Man kan dock ganska lätt övertyga sig om att den blir en delmängd till $P(P(A \cup B))$. Eftersom $a, b \in A \cup B$ så har vi att $\{a\}$ och $\{a, b\}$ är element i $P(A \cup B)$, och därför är $\{\{a\}, \{a, b\}\}$ en delmängd till $P(A \cup B)$ och alltså ett element i $P(P(A \cup B))$, dvs $\langle a, b\rangle \in P(P(A \cup B))$.

Låt X och Y vara två mängder. En *relation* R mellan X och Y är en delmängd till produktmängden $X \times Y$. Det betyder att för varje val av $x \in X$ och $y \in Y$ så gäller antingen att $(x, y) \in R$ (detta skrivs ofta som xRy eller $R(x, y)$) eller att $(x, y) \notin R$.

EXEMPEL 2.2. Om a är en mängd så är $\in$ en relation mellan a och dess potensmängd $P(a)$. Formellt alltså

$$R = \{(x, y) \in a \times P(a);\ x \in y\}.$$

$\square$

EXEMPEL 2.3. Låt X vara mängden av alla människor och låt Y vara mängden av alla förnamn. Vi kan då definiera en relation R sådan att xRy om och endast om x heter y. I allmänhet kommer det att svara mer än ett y till varje x och mer än ett x till varje y. $\square$

ANMÄRKNING 2.4. Noga taget är väl inte en mängd av människor en mängd i mängduniversum, men den som bekymras av detta kan ju i så fall låta varje människa svara mot ett unikt naturligt tal; gäller det svenskar kan man med fördel ta personnumret. På så sätt svarar den mängd av människor man betraktar mot en mängd i mängduniversum. På liknande sätt kan man identifiera mängden av alla namn med en mängd i mängduniversum. $\square$

Om $X = Y$ säger man att R är en *(tvåställig) relation* på X.

EXEMPEL 2.5. Vi har relationen "y är kvadraten på x" på $\mathbb{N}$. Alltså

$$R = \{(x, y) \in \mathbb{N} \times \mathbb{N};\ x^2 = y\}.$$

$\square$

Ett specialfall av en relation från X till Y är en funktion. En relation f från X till Y är en *funktion* om till varje $x \in X$ finns ett entydigt $y \in Y$ sådant att $(x, y) \in f$. Att f är en funktion från X till Y skrivs ofta $f\colon X \to Y$ och att $(x, y) \in f$ skrivs ofta(st) $y = f(x)$ eller $x \mapsto y$. I stället för "funktion" används termerna "avbildning" eller "transformation".

Om $f\colon X \to Y$ och $M \subset X$ så kan man definiera en funktion $f|_M\colon M \to Y$, kallad *restriktionen av f till delmängden M* och som definieras genom att $f|_M(x) = f(x)$ för $x \in M$.

EXEMPEL 2.6. Låt X vara mängden av alla människor och låt Y vara mängden av alla förnamn. Om vi då säger att xfy om och endast om y är tilltalsnamnet på x, så är f en funktion. $\square$

EXEMPEL 2.7. Om X är en mängd så finns det en funktion $I_X\colon X \to X$, identitetsavbildningen på X, som definieras genom att $I_X(x) = x$ för alla $x \in X$. $\square$

En funktion $f\colon X \to Y$ är *injektiv (en injektion)* om $f(x) = f(x')$ medför att $x = x'$. Den är *surjektiv (en surjektion)* om det till varje

$y \in Y$ finns något $x \in X$ sådant att $y = f(x)$. Om den är både injektiv och surjektiv kallas den *bijektiv (en bijektion).*

EXEMPEL 2.8. Definiera en funktion $f\colon \mathbb{N} \to \mathbb{N}$ genom att $f(x) = (x-4)^2 + 5$. Då är inte f vare sig injektiv eller surjektiv. Däremot är $f|_A$ injektiv om $A = \{x \in \mathbb{N};\ x \geq 4\}$. $\square$

Om $f\colon X \to Y$ och $g\colon Y \to Z$ så kan man bilda *sammansättningen* $g \circ f\colon X \to Z$, som definieras genom att $g \circ f(x) = g(f(x))$.

UPPGIFT 2.1. Antag att $g \circ f$ är injektiv. Visa att f måste vara injektiv. Måste även g vara injektiv? Antag att $g \circ f$ är surjektiv. Följer det då att f eller g är surjektiv?

Om $f\colon X \to Y$ är en bijektion så finns det en entydigt bestämd funktion $f^{-1}\colon Y \to X$ sådan att $f \circ f^{-1} = I_X$ och $f^{-1} \circ f = I_Y$. Man säger att f är *inverterbar* och att f^{-1} är *inversen* till f.

UPPGIFT 2.2. Visa att en bijektiv funktion är inverterbar, och omvänt, att f är bijektiv om f är inverterbar.

För en godtycklig funktion $f\colon A \to B$ och $E \subset B$ inför vi beteckningen

$$f^{-1}(E) = \{x \in A;\ f(x) \in E\}.$$

Om $F \subset A$ sätter vi

$$f(F) = \{b \in B;\ b = f(x) \text{ för något } x \in F\}.$$

EXEMPEL 2.9. Låt $\mathbb{R}$ beteckna de reella talen. En formell definition kommer i kapitel 3. Låt vidare $f(x) = x^2$. Sätt $A^+ = \{x \in \mathbb{R};\ x > 2\}$, $A^- = \{x \in \mathbb{R};\ x < -2\}$, $B^+ = \{y \in \mathbb{R};\ y > 4\}$ och $B^- = \{y \in \mathbb{R};\ y < -4\}$. Då är $f(A^+) = B^+$ och $f(A^-) = B^+$. Vidare är $f^{-1}(B^+) = A^+ \cup A^-$ och $f^{-1}(B^-) = \emptyset$. $\square$

EXEMPEL 2.10. Vi ska visa att om $f\colon X \to Y$ och $A, B \subset X$ så är

$$f(A \cup B) = f(A) \cup f(B) \quad \text{och} \quad f(A \cap B) \subset f(A) \cap f(B). \tag{2.1}$$

Nämligen; $y \in f(A \cup B)$ omm $y = f(x)$ för något $x \in A \cup B$. Detta gäller omm $y = f(x)$ för $x \in A$ eller $x \in B$ vilket gäller omm $y \in f(A)$ eller $y \in f(B)$ dvs $y \in f(A) \cup f(B)$. Detta visar likheten i (2.1).

Inklusionen kan visas på liknande sätt eller genom följande: Eftersom $A \cap B \subset A$ så $f(A \cap B) \subset f(A)$, och på samma sätt $f(A \cap B) \subset f(B)$, så $f(A \cap B) \subset f(A) \cap f(B)$. Inklusionen i (2.1) kan vara strikt; t ex är, med beteckningarna från exempel 2.9, $f(A^+ \cap A^-) = \emptyset$ men $f(A^+) \cap f(A^-) = B^+ \neq \emptyset$. $\square$

UPPGIFT 2.3. Antag att $f\colon X \to Y$ och $A, B \subset Y$. Visa att $f^{-1}(A \cap B) = f^{-1}(A) \cap f^{-1}(B)$ och $f^{-1}(A \cup B) = f^{-1}(A) \cup f^{-1}(B)$.

Uppgift 2.4. Visa att $F \subset f^{-1}(f(F))$ men att likhet inte behöver gälla. Vad kan man säga om förhållandet mellan E och $f(f^{-1}(E))$?

Exempel 2.11. Givet A och B så finns det en funktion

$$\pi_A \colon A \times B \to A,$$

kallad *projektionen* av $A \times B$ på A, som är definierad genom att $\pi_A(a,b) = a$. På samma sätt definieras $\pi_B \colon A \times B \to B$. □

Anmärkning 2.12. Det skulle vara mer konsekvent att skriva $\pi_A(\langle a,b\rangle)$ istället för $\pi_A(a,b)$. Man brukar dock välja det enklare skrivsättet eftersom det ändå är klart att det rör sig om det ordnade paret $\langle a,b\rangle$. □

Om A och B är två mängder och $B \neq \emptyset$ så definierar man B^A som mängden av alla funktioner från A till B. För en motivering av denna beteckning, se övning 3.6.

Exempel 2.13. Låt A vara en icketom mängd. Det finns en naturlig bijektion mellan 2^A, dvs mängden av funktioner från A till $2 = \{0,1\}$, och potensmängden $P(A)$. Man tar helt enkelt och identifierar funktionen $f \colon A \to \{0,1\}$ med delmängden $\{x \in A;\ f(x) = 1\}$ av A, eller mer formellt uttryckt, vi definierar en avbildning $\Phi_A \colon 2^A \to P(A)$ genom att sätta $\Phi_A(f) = \{x \in A;\ f(x) = 1\}$. Man kollar lätt att Φ_A är surjektiv och injektiv. Om $\Phi_A(f) = \Phi_A(g)$ så måste $f(x) = g(x)$ för alla x så f och g är samma funktion, och alltså är Φ_A är injektiv. Om B är en godtycklig delmängd till A så är $B = \Phi_A(f)$, där $f(x)$ är funktionen som är 1 för $x \in B$ och 0 för $x \in A \setminus B$. Alltså är Φ_A surjektiv. □

Antag att A, B och C är tre mängder. Det finns en naturlig bijektion mellan $(A\times B)\times C$ och $A\times(B\times C)$, definierad genom att $((a,b),c) \mapsto (a,(b,c))$. Ofta skriver man därför helt enkelt bara $A \times B \times C$.

Anmärkning 2.14. Antag att A är en mängd (av mängder). Man kan definiera produkten av mängderna i A, $\Pi_{y\in A} y$, som mängden av alla funktioner $f \colon A \to \cup_{y\in A} y$ sådana att $f(a) \in a$ för alla $a \in A$. Att denna mängd är icketom är en (om)formulering av urvalsaxiomet, se vidare avsnitt 6. □

Om vi har en bijektion $g \colon B \to A$ och sätter $a_b = g(b)$ så är A mängden av alla a_b sådana att $b \in B$. Vi kan därför skriva $\Pi_{b\in B} a_b$ istället för $\Pi_{y\in A} y$ och t ex $\cup_{b\in B} a_b$ istället för $\cup_{y\in A} y$. Man säger att B är en *indexmängd*.

Exempel 2.15. Antag att vi har $A = \{A_0, A_1\}$. Vi använder alltså $2 = \{0,1\}$ som indexmängd här. Vi ska visa att definitionen i anmärkning 2.14 av $\Pi_{j\in 2} A_j$ är ekvivalent med den tidigare definitionen av

$A_0 \times A_1$; dvs att det finns en (naturlig) bijektion $\Psi\colon A_0 \times A_1 \to \Pi_{j\in 2} A_j$. Enkelt uttryckt så identifierar vi helt enkelt $(a_0, a_1) \in A_0 \times A_1$ med den funktion från $2 = \{0, 1\}$ vars funktionsvärde är a_0 i punkten 0 och a_1 i punkten 1. Om f är en funktion på 2 vars värde i punkten 0 ligger i A_0 och vars värde i punkten 1 ligger i A_1 så svarar den då mot elementet $(f(0), f(1)) \in A_0 \times A_1$. Alltså $\Psi(a_0, a_1)(0) = a_0$ och $\Psi(a_0, a_1)(1) = a_1$, och omvänt $\Psi^{-1}(f) = (f(0), f(1))$ om $f \in \Pi_{j\in 2} A_j$. $\square$

DEFINITION 2.16. En (tvåställig) relation xRy på en mängd X kallas en *ekvivalensrelation* om följande gäller:
(r) aRa för alla a.
(s) aRb medför bRa.
(t) aRb och bRc medför aRc.

En relation som uppfyller (r) kallas *reflexiv*, om den uppfyller (s) kallas den *symmetrisk* och om den uppfyller (t) kallas den *transitiv*.

UPPGIFT 2.5. Låt X vara mängden av alla människor och låt xRy betyda " x är (hel)syskon till y". Avgör om R är reflexiv, symmetrisk och/eller transitiv. Hur blir det om man ändrar "(hel)syskon" till "hel- eller halvsyskon" ?

UPPGIFT 2.6. Låt X vara mängden av alla elever i en given skolklass och låt xRy vara relationen "x är kär i y". Diskutera rimligheten i att R är symmetrisk, reflexiv eller transitiv.

Om R är en ekvivalensrelation så skriver man ofta $\sim$ istället för R.

En *partition* av en mängd X är en mängd av parvis disjunkta delmängder vars union är hela X.

PROPOSITION 2.17. *Om $\sim$ är en ekvivalensrelation på X så finns en unik partition av X sådan att $a, b \in X$ ligger i samma delmängd om och endast om $a \sim b$.*

Mängderna i partitionen kallas *ekvivalensklasser*, och a och b ligger alltså i samma ekvivalensklass om och endast om $a \sim b$.

Omvänt, om vi har en partition av X given, så svarar den mot ekvivalensrelationen $\sim$, där $a \sim b$ om och endast om a och b hör till samma mängd i den givna partitionen. Det finns alltså en bijektion mellan mängden av alla partitioner av X och mängden av alla ekvivalensrelationer, jmf övning 2.7.

BEVIS. För varje $a \in X$ bildar vi mängden $X_a = \{x \in X;\ x \sim a\}$. Till att börja med är varje X_a icketom eftersom $\sim$ är reflexiv. Antag att $X_a \cap X_b \neq \emptyset$. Då finns något z sådant att $z \sim a$ och $z \sim b$. Eftersom

$\sim$ är symmetrisk så har vi att $a \sim z$. Enligt transitiviteten är då $a \sim b$. Om $w \in X_a$ så $w \sim a$, men eftersom $a \sim b$ så har vi därför att $w \sim b$, dvs $w \in X_b$. Alltså $X_a \subset X_b$. På samma sätt ser man att $X_b \subset X_a$. Vi har alltså visat att X_a och X_b antingen är disjunkta eller identiska. Eftersom varje a ligger i en av dessa mängder, nämligen X_a, innebär detta att $\{X_a\}_{a \in X}$ är en mängd av parvis disjunkta delmängder till X vars union är hela X, dvs en partition av X. $\square$

Exempel 2.18. Låt X vara mängden av alla elever i någon fix grundskola, och säg att $x \sim y$ om och endast om x och y går i samma årskurs. Då är $\sim$ en ekvivalensrelation på X och ekvivalensklasserna är precis de olika årskurserna. $\square$

Uppgift 2.7. Visa att mängden av ekvivalensrelationer på en mängd X är en mängd. Mer noggrant uttryckt: Visa att det, givet en mängd X, finns en mängd Y vars element är precis alla ekvivalensrelationer på X.

Vi avslutar detta avsnitt med att införa det sista axiomet i teorin ZF. Antag att X och Y är mängder. En relation $\phi(x, y)$ mellan X och Y sägs vara *funktionell* om $\phi(x, y) = \phi(x, y')$ medför att $y = y'$.

ZF VIII (replacementaxiomet) Antag att ϕ är en funktionell relation mellan mängderna X och Y. Då finns en delmängd Z till Y som består av precis de $z \in Z$ sådana att det finns $x \in X$ så att $\phi(x, z)$.

Enkelt uttryckt så är innebär detta att bilden av en funktionell relation är en mängd.

Uppgift 2.8. Visa att replacementaxiomet medför delmängdsaxiomet.

3. Ekvipotens

Vi ska nu definiera vad det betyder att två mängder har lika många element. Hur avgör vi i vardagslag huruvida två mängder vi stöter på har lika många element? Det vanligaste svaret är förstås att vi helt enkelt räknar var och en av dem och sedan jämför. Att *räkna* en (ändlig) mängd betyder precis att hitta en bijektion till något naturligt tal n. Dock finns det ibland möjlighet att fastställa att två mängder har samma antal element utan att räkna dem; genom att direkt finna en bijektion mellan dem.

Exempel 3.1. Låt A vara mängden av alla gifta män i Sverige och B mängden av alla gifta kvinnor (änkemän och änkor undantagna). Då kan vi omedelbart avgöra att det finns en bijektion mellan A och B och följaktligen kan vi dra slutsatsen att de har lika många element, utan att för den skull känns till antalet. □

Definition 3.2. Vi säger att två mängder A och B är *ekvipotenta* (att de har samma *kardinalitet* eller att de är *likmäktiga*) om det finns en bijektion mellan A och B.

Om A och B är ekvipotenta skriver vi $A \simeq B$.

Exempel 3.3. Vi ska visa att att $A^C \times B^C$ är ekvipotent med $(A \times B)^C$ genom att ange en bijektion $\Psi\colon A^C \times B^C \to (A \times B)^C$. Observera att $A^C \times B^C$ består av alla (f, g) där $f\colon C \to A$ och $g\colon C \to B$. Varje sådan definierar en funktion $\Psi(f, g)\colon C \to A \times B$ genom att $\Psi(f, g)(c) = (f(c), g(c))$ (som ju är ett element i $A \times B$). Om $\langle f, g\rangle \neq \langle f', g'\rangle$ så finns $c \in C$ sådant att $f(c) \neq f'(c)$ eller $g(c) \neq g'(c)$. Alltså är $\langle f(c), g(c)\rangle \neq \langle f'(c), g'(c)\rangle$ vilket medför att $\Psi(f, g) \neq \Psi(f', g')$, och därför är Ψ injektiv. Om $h\colon C \to A \times B$ så är $h = \Psi(f, g)$ där $f = \pi_A \circ h$ och $g = \pi_B \circ h$ och sålunda är Φ surjektiv. Alltså har vi en bijektion $\Psi\colon A^C \times B^C \to (A \times B)^C$ och alltså är $A^C \times B^C$ och $(A \times B)^C$ ekvipotenta. □

Uppgift 3.1. Visa att $(A^B)^C$ är ekvipotent med $A^{B \times C}$.

Uppgift 3.2. Antag att B och C är disjunkta. Visa att $A^B \times A^C$ är ekvipotent med $A^{B \cup C}$.

Uppgift 3.3. Antag att $A \simeq B$. Visa att $2^A \simeq 2^B$.

En mängd A är *ändlig* om det finns en bijektion från A till n för något $n \in \mathbb{N}$. Om A inte är ändlig kallas den *oändlig*. Exempelvis är $\mathbb{N}$ oändlig.

Det är välkänt att om man räknar en ändlig mängd på två olika sätt så får man samma resultat. Man kan alltså tala om *antalet* element i mängden. Formellt innebär detta

Proposition 3.4. *Om A är en mängd sådan att $A \simeq n$ och $A \simeq m$ så är $n = m$.*

Bevis. Om $A \simeq n$ och $A \simeq m$ så finns en bijektion $\phi\colon n \to m$. Vi påstår nu att om $\phi\colon n \to m$ är en bijektion så är $n = m$. Det räcker att visa att detta är sant för $n = 0$ samt att det är sant för n' givet att man vet att det är sant för n. Detta kallas (svag) induktion. Om $n = 0 = \emptyset$ är det klart att även $m = 0$. Antag nu att påståendet är sant för n och antag att $\phi\colon n' \to m$ är en bijektion. Då är $m \neq \emptyset$ och alltså är $m = \ell'$

för något ℓ. Kom ihåg att $n' = n \cup \{n\}$ så n är ett element i n'. Antag att $\phi(n) = k \in \ell'$. Låt ψ vara den bijektion på ℓ' sådan att $\ell \mapsto k$, $k \mapsto \ell$ och alla andra element i ℓ' hålls fixa. Då kommer $\psi \circ \phi\colon n' \to \ell'$ att vara en bijektion sådan att $n \mapsto \ell$. Följaktligen är restriktionen av $\psi \circ \phi$ till n en bijektion $n \to \ell$ och enligt induktionsantagandet är därför $n = \ell$. Det följer nu att $n' = \ell' = m$. $\square$

Uppgift 3.4. Visa att en ändlig mängd inte kan vara ekvipotent med en äkta delmängd till sig själv.

Uppgift 3.5. Visa att varje delmängd till en ändlig mängd är ändlig.

Uppgift 3.6. Låt A och B vara ändliga mängder och beteckna antalet element med $|A|$ respektive $|B|$. Visa att B^A har $|B|^{|A|}$ element.

Exempel 3.5. Antag att $A \simeq B$ och $A' \simeq B'$, och att $A \cap A' = \emptyset$ och $B \cap B' = \emptyset$. Då följer att $A \cup A' \simeq B \cup B'$. Om nämligen $f\colon A \to B$ och $f'\colon A' \to B'$ är bijektioner så får man en bijektion $F\colon A \cup A' \to B \cup B'$ om man sätter $F(x) = f(x)$ för $x \in A$ och $F(x) = f'(x)$ för $x \in A'$. Rita en figur! $\square$

I praktiken när man ska visa att två mängder är ekvipotenta är det i allmänhet inte så lätt att direkt hitta en bijektion mellan sina båda mängder. Då har man ofta nytta av följande sats.

Sats 3.6 (Schröder–Bernsteins sats). *Antag att A och B är två mängder och att det finns injektioner $f\colon A \to B$ och $g\colon B \to A$. Då är A och B ekvipotenta.*

Bevis. Vi ska dela upp A i tre disjunkta mängder A_0, A_1 och A_∞. Tag $a \in A$. Om a ligger i bilden av g så finns entydigt $b_1 \in B$ sådant att $g(b_1) = a$. Om nu b_1 ligger i bilden av f så finns entydigt $a_1 \in A$ sådant att $f(a_1) = b_1$. Eventuellt ligger nu detta a_1 i bilden av g och då finns entydigt $b_2 \in B$ sådant att $g(b_2) = a_1$. Om nu detta b_2 ligger i bilden av f så finns entydigt $a_2 \in A$ sådant att $f(a_2) = b_2$ etc. Vi kallar alla dessa a_j och b_j föregångare till a. Givet ett $a \in A$ finns nu tre möjligheter. Eventuellt tar detta slut efter ett jämnt antal steg, dvs man hamnar på ett element a_j som inte ligger i bilden av g. En annan möjlighet är att det hela tar slut efter ett udda antal steg. Sista möjligheten är att att man kan hitta oändligt många föregångare. De $a \in A$ för vilka det första alternativet gäller bildar mängden A_0, nästa grupp är A_1 och de med oändligt många föregångare bildar A_∞.

På samma sätt delar man nu upp B i tre disjunkta delmängder B_0, B_1 och B_∞. Enligt exempel 3.5 räcker det nu att visa att $A_0 \simeq B_1$, $A_1 \simeq B_0$ och $A_\infty \simeq B_\infty$.

Vi påstår nu att $f|_{A_0}\colon A_0 \to B_1$ är en bijektion. Om nämligen $a \in A_0$ så kan man gå bakåt som ovan ett jämnt antal steg innan det blir stopp, men då kan man ju gå bakåt ett steg till från $f(a)$, dvs totalt ett udda antal steg, så $f(a) \in B_1$. Alltså avbildar $f|_{A_0}$ mängden A_0 in i B_1. Vidare är $f|_{A_0}$ injektiv eftersom redan f är det. Man finner lätt på samma sätt att varje $b \in B_1$ ligger i bilden så $f|_{A_0}\colon A_0 \to B_1$ är surjektiv. Alltså är $A_0 \simeq B_1$, och på samma sätt visas att $A_1 \simeq B_0$. Slutligen finner man på liknande sätt att $f|_{A_\infty}\colon A_\infty \to B_\infty$ är en bijektion och beviset är därmed klart. $\square$

Ett annat sätt att formulera Schröder–Bernsteins sats är "om A är ekvipotent med en delmängd till B och B är ekvipotent med en delmängd till A så är A och B ekvipotenta". Däremot har vi inte (ännu) bevisat att det för varje par av mängder A och B är så att A är ekvipotent med en delmängd till A eller tvärtom. Detta är svårare än man kanske omedelbart tror och bygger på urvalsaxiomet UA, se avsnitt 6.

Det är lätt att se att M inte är ekvipotent med $P(M)$ om M är en ändlig mängd, jmf övning 3.6 och exempel 2.13. Det är också lätt att se att $\mathbb{N}$ inte är ekvipotent med $P(\mathbb{N})$. Betrakta nämligen en godtycklig funktion $\psi\colon \mathbb{N} \to P(\mathbb{N})$ och sätt $A_k = \psi(k)$. Då har vi en uppräknelig följd $A_0, A_1, A_2, \ldots$ av delmängder till $\mathbb{N}$. Låt nu A vara den delmängd till $\mathbb{N}$ som innehåller 0 omm $0 \notin A_0$, 1 omm $1 \notin A_1$, etc. Då kommer mängden A att vara skild från var och en av mängderna A_j, så ψ är inte surjektiv och följaktligen inte bijektiv. Med samma argument kan vi visa följande allmänna resultat.

Sats 3.7. *Antag att M är en mängd. Då är inte M ekvipotent med $P(M)$.*

Bevis. För en godtycklig funktion $\psi\colon M \to P(M)$, sätt

$$A = \{x \in M;\ x \notin \psi(x)\}.$$

Då är inte $A = \psi(a)$ för något $a \in M$ eftersom i så fall $a \in A$ medför $a \notin \psi(a) = A$ och vice versa. Alltså finns ingen surjektion från M till $P(M)$ och därför finns ingen bijektion mellan till M och $P(M)$. $\square$

Exempel 3.8. Antag att $f\colon A \to B$ är injektiv. Vi påstår att det då finns en surjektion $g\colon B \to A$. Att f är injektiv betyder nämligen att $f\colon A \to B'$ är bijektiv, där $B' = f(A)$. Tag ett $a \in A$, och definiera g som f^{-1} på B' och låt $g(x) = a$ för $x \in B \setminus B'$. Då blir $g\colon B \to A$ surjektiv.

I beviset av Sats 3.7 fann vi att det inte finns någon surjektion $f\colon M \to P(M)$. Alltså följer det att det inte kan finnas någon injektion

$P(M) \to M$, dvs $P(M)$ är inte ekvipotent med någon delmängd till M. $\square$

Uppgift 3.7. Visa direkt att det inte finns någon injektion $f\colon P(M) \to M$ genom att betrakta $A = \{f(B);\ f(B) \notin B\}$.

4. Ordningsrelationer och tal

I detta avsnitt definierar vi begreppet ordningsrelation och påminner om några grundläggande egenskaper hos tal. Vi har redan sett att de naturliga talen $\mathbb{N}$ så som vi har definierat dem uppfyller Peanos axiom. Utifrån dessa axiom kan man definiera vanlig aritmetik, dvs addition och multiplikation, och utvidga till heltalen $\mathbb{Z}$, de rationellla talen $\mathbb{Q}$ samt de reella talen $\mathbb{R}$ och de komplexa talen $\mathbb{C}$. Vidare har man ordningsrelationer på $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$ och $\mathbb{R}$. (Noga taget har vi redan utnyttjat ordningsrelationerna på $\mathbb{N}$ och $\mathbb{R}$ i några exempel i föregående avsnitt.) Definitioner av dessa talsystem, deras ordningsrelationer samt grundläggande satser finner man i kapitel 3. Den som vill följa en helt logisk linje i framställningen bör läsa om ordningsrelationer i detta avsnitt och sedan hoppa till kapitel 3, läsa detta i sin helhet och sedan fortsätta med avsnitt 5 i detta kapitel.

Definition 4.1. En relation R på en mängd M är en *ordningsrelation* om den uppfyller
(r) aRa
(a) aRb och bRa medför att $a = b$.
(t) aRb och bRc medför att aRc

Man känner igen egenskaperna (r) och (t) från definitionen av ekvivalensrelation; egenskapen (a) kallas *antisymmetri*.

Vi betecknar oftast en ordningsrelation med $\leq$ och vi skriver $x < y$ om $x \leq y$ och $x \neq y$. En ordningsrelation $\leq$ på en mängd M är *total* om för varje par $x, y \in X$, antingen $x \leq y$ eller $y \leq x$. Man uttrycker även detta som att M är en *totalt ordnad* mängd. En mängd M med en ordningsrelation som (man inte vet) är total kallas *partiellt ordnad.* En delmängd X till en partiellt ordnad mängd M blir själv partiellt ordnad genom att ta restriktionen av ordningsrelationen på M till X.

Uppgift 4.1. Kontrollera att restriktionen av en ordningsrelation till en delmängd är en ordningsrelation.

Exempel 4.2. På potensmängden $P(M)$ till en mängd M är $\subset$ en ordningsrelation. Om M har fler än ett element så är ordningsrelationen inte total. □

Exempel 4.3. Om $\leq$ är en ordningsrelation på en mängd M så är även omvändningen $\geq$ en ordningsrelation på M. □

I kapitel 3 inför vi en total ordningsrelation på de naturliga talen $\mathbb{N}$, utgående från Peanos axiom, sådan att $m \leq n$ om och endast om n erhålls från m genom att applicera efterföljarfunktionen ett ändligt antal gånger. Vi inför även totala ordningsrelationer på $\mathbb{Q}$ och på $\mathbb{R}$. För de senare gäller att om $r < s$ så finns t sådant att $r < t < s$. Detta är däremot inte sant för (ordningsrelationen på) $\mathbb{N}$.

Exempel 4.4. På $\mathbb{R}^2$ kan vi definiera en ordningsrelation R genom att säga att $(x, y)R(x', y')$ omm $x \leq x'$ och $y \leq y'$. Denna är inte total. □

Låt M vara en totalt ordnad mängd. Man säger att $m \in M$ är ett *minsta element* om $m \leq x$ för alla $x \in M$. Om M har ett minsta element så är det entydigt på grund av (a) i definitionen av ordningsrelation.

Definition 4.5. Man säger att en ordningsrelation $\leq$ på M är en *välordning* om varje icketom delmängd X av M har ett minsta element.

Det följer genast att en välordnad mängd M är totalt ordnad, ty om $x, y \in M$ så har $\{x, y\}$ ett minsta element, dvs $x \leq y$ eller $y \leq x$.

Exempel 4.6. Varje totalt ordnad ändlig mängd är välordnad. För att se detta räcker det att visa att varje totalt ordnad ändlig mängd A har ett minsta element. Tag ett godtyckligt element $a_0 \in A$. Om inte detta är ett minsta element så finns något mindre säg a_1. Om inte detta är ett minsta så finns något mindre element a_2 etc. Efter ändligt många steg kommer vi att hitta det minsta. Det följer att om A är en totalt ordnad ändlig mängd så är $A = \{b_0, b_1, \ldots, b_m\}$ där $b_0 < b_1 < \ldots < b_m$. □

Exempel 4.7. De reella talen $\mathbb{R}$ är en totalt ordnad mängd som inte är välordnad; exempelvis har det öppna intervallet $(0, 1)$ inte något minsta element eftersom det för varje $x > 0$ finns r sådant att $0 < r < x$. □

Man säger att den *starka induktionsprincipen* gäller för en mängd M om för varje delmängd X gäller: Om för varje $m \in M$ gäller att $m \in X$ givet att $x \in X$ för alla $x < m$, så följer det att $X = M$.

Eftersom en enställig relation på en mängd i princip inte är något annat än en delmängd och vice versa, så kan man formulera den starka induktionspricipen på följande sätt:

Antag att $\phi(x)$ är en enställig relation på M sådan att $\phi(m)$ gäller givet att $\phi(x)$ gäller för alla $x < m$. Då gäller $\phi(x)$ för alla $x \in M$.

Proposition 4.8. *Antag att M är välordnad. Då gäller den starka induktionsprincipen för M.*

Bevis. Antag att X har angiven egenskap. Om $M \setminus X$ inte är tom så har den ett minsta element m eftersom M är välordnad. Detta betyder dock att $x \in X$ för alla $x < m$ men att $m \notin X$, vilket är en motsägelse. Alltså är $M \setminus X$ tom, dvs $X = M$. □

Omvänt är det lätt att se att om M är en totalt ordnad mängd sådan att den starka induktionsprincipen gäller så är M välordnad. Tag nämligen $Z \subset M$ och antag att Z saknar ett minsta element. Detta betyder att $m \in M \setminus Z$ om $x \in M \setminus Z$ för alla $x < m$. Om den starka induktionsprincipen nu gäller så följer det att $M \setminus Z = M$ dvs $Z = \emptyset$. Det följer att M är välordnad.

I kapitel 3 visar vi att det naturliga talen $\mathbb{N}$ är välordnade och att alltså den starka induktionsprincipen är giltig.

5. Uppräkneliga mängder

Definition 5.1. En mängd X som är ekvipotent med $\mathbb{N}$, de naturliga talen, kallas *uppräknelig* (*countable* på engelska).

Om X är uppräknelig har vi alltså en bijektion $f\colon \mathbb{N} \simeq X$, så X utgörs precis av elementen $f(0), f(1), f(2), \ldots$, dvs man kan "räkna upp" elementen i X. Ofta skriver man x_j istället för $f(j)$ så att $X = \{x_0, x_1, x_2, \ldots\}$.

Exempel 5.2. Per definition är $\mathbb{N}$ uppräknelig men även $\mathbb{Z}$, mängden av heltal, är det eftersom vi t ex har uppräkningen

$$0, 1, -1, 2, -2, 3, \ldots$$

av $\mathbb{Z}$, dvs $f\colon \mathbb{N} \simeq \mathbb{Z}$ där $f(0) = 0, f(1) = 1, f(2) = -1, \ldots$. □

Vidare är $\mathbb{N} \times \mathbb{N}$ är uppräknelig eftersom vi kan ta

$$(0,0), (1,0), (0,1), (2,0), (1,1), (0,2), (3,0), (2,1), (1,2), (0,3), (4,0), \ldots.$$

Man kan också visa att $\mathbb{N} \times \mathbb{N}$ är uppräknelig genom att använda Schröder–Bernsteins sats. Eftersom varje positivt naturligt tal har en entydig primtalsfaktorisering så är $f((n,m)) = 2^n 3^m$ en injektion $\mathbb{N} \times \mathbb{N} \to \mathbb{N}$. En injektion $g\colon \mathbb{N} \to \mathbb{N} \times \mathbb{N}$ ges t ex av $g(n) = (n, 0)$.

Antag nu att A och B är uppräkneliga mängder. Då har vi alltså bijektioner $f\colon \mathbb{N} \to A$ och $g\colon \mathbb{N} \to B$. Om vi sätter $F(n,k) = \langle f(n), g(k) \rangle$ får vi en bijektion $F\colon \mathbb{N} \times \mathbb{N} \to A \times B$. Om nu G är en bijektion $G\colon \mathbb{N} \to \mathbb{N} \times \mathbb{N}$ så blir $F \circ G$ en bijektion $\mathbb{N} \to A \times B$, och alltså följer att $A \times B$ är uppräknelig. Vi kan skriva detta argument på följande vis: Eftersom $\mathbb{N} \simeq A$ och $\mathbb{N} \simeq B$ så följer att $\mathbb{N} \times \mathbb{N} \simeq A \times B$. Eftersom vidare $\mathbb{N} \simeq \mathbb{N} \times \mathbb{N}$ så får vi att $\mathbb{N} \simeq A \times B$, jmf övning **??**.

Uppgift 5.1. Visa direkt att om A och B är uppräkneliga så är $A \times B$ uppräknelig.

Exempel 5.3. Mängden av de rationella talen, $\mathbb{Q}$, är uppräknelig. För att se detta, observera först att varje rationellt tal kan skrivas entydigt som a/b för heltal a och b om vi kräver att man har förkortat så långt som möjligt, $b > 0$, och att $0 = 0/1$. Alltså får vi en injektiv avbildning $\mathbb{Q} \to \mathbb{Z} \times \mathbb{N}$ genom att $a/b \mapsto (a,b) \in \mathbb{Z} \times \mathbb{N}$. Eftersom vi har $\mathbb{Z} \times \mathbb{N} \simeq \mathbb{N}$, och sammansättningen av en bijektion och en injektion blir en injektion, så får vi en injektion $\mathbb{Q} \to \mathbb{N}$. Trivialt har vi en injektion $\mathbb{N} \to \mathbb{Q}$, så det följer av Schröder–Bernsteins sats att $\mathbb{Q} \simeq \mathbb{N}$. $\square$

Uppgift 5.2. Försök att hitta en direkt uppräkning av $\mathbb{Q}$.

Exempel 5.4. Vi ska nu visa att varje delmängd X till $\mathbb{N}$ är ändlig eller uppräknelig. Vi utnyttjar att $\mathbb{N}$ är välordnad. Om inte X är tom så finns ett minsta element x_0. Om nu inte $X \setminus \{x_0\}$ är tom så finns ett minsta element x_1 i denna mängd . Om nu inte $X \setminus \{x_0, x_1\}$ är tom så finns ett minsta element x_2 etc. Antingen dyker den tomma mängden upp efter ett ändligt antal steg och då är ju $X = \{x_0, \ldots, x_m\}$ och alltså ändlig. Om inte så får vi en oändlig växande följd $x_0 < x_1 < x_2 < \ldots$. Vi påstår nu att i så fall $X = \{x_0, x_1, \ldots\}$. Om inte så finns ett minsta element m i $X \setminus \{x_0, x_1, \ldots\}$. Men eftersom $x_0, x_1..$ är växande och oändlig måste då $x_k > m$ för något k. Detta strider mot att x_k var det minsta i $X \setminus \{x_0, \ldots, x_{k-1}\}$ i vilken ju m är ett element. $\square$

Lemma 5.5. *En mängd B är ändlig eller uppräknelig om och endast om det finns en surjektion $f\colon \mathbb{N} \to B$.*

Bevis. Om $f\colon \mathbb{N} \to B$ är surjektiv, så är $N_b = \{x \in \mathbb{N};\ f(x) = b\}$ icketom för varje $b \in B$. Definiera $g\colon B \to \mathbb{N}$ genom att $g(b) = \min N_b$. Då är g injektiv, så B är ekvipotent med en delmängd till $\mathbb{N}$ och alltså ändlig eller uppräknelig enligt exempel 5.4. Omvänt, om B är uppräknelig så finns per definition en bijektion $f\colon \mathbb{N} \to B$. Om B är ändlig, säg $B = \{b_0, \ldots, b_m\}$, så sätt $f(k) = b_k$ för $k \le m$ och t ex $f(k) = b_0$ för $k > m$. Då blir $f\colon \mathbb{N} \to B$ surjektiv. $\square$

PROPOSITION 5.6. *Antag att var och en av A_0, $A_1, \dots$ är ändlig eller uppräknelig. Då är unionen $\cup_{k\in\mathbb{N}} A_k$ ändlig eller uppräknelig.*

BEVIS. Eftersom varje A_k är ändlig eller uppräknelig så finns enligt lemma 5.5 surjektioner $g_k : \mathbb{N} \to A_k$, $k = 0, 1, 2, \dots$. Definiera nu en avbildning $F : \mathbb{N} \times \mathbb{N} \to \cup_{k\in\mathbb{N}} A_k$, genom att $F(n, k) = f_k(n)$. Då blir F surjektiv, och eftersom $\mathbb{N} \simeq \mathbb{N} \times \mathbb{N}$ så finns det en surjektion $\mathbb{N} \to \cup_{k\in\mathbb{N}} A_k$. Enligt lemma 5.5 igen är därför $\cup_{k\in\mathbb{N}} A_k$ ändlig eller uppräknelig. □

EXEMPEL 5.7. Vi ska nu visa att $\mathbb{N}^{\mathbb{N}} \simeq 2^{\mathbb{N}}$. Observera först att vi har en naturlig injektion $2^{\mathbb{N}} \to \mathbb{N}^{\mathbb{N}}$, genom att funktionen $f : \mathbb{N} \to 2$ betraktas som en funktion $f : \mathbb{N} \to \mathbb{N}$. För den omvända injektionen noterar vi att varje funktion $f : \mathbb{N} \to \mathbb{N}$ är en delmängd till $\mathbb{N} \times \mathbb{N}$, så vi får en injektion $\mathbb{N}^{\mathbb{N}} \to 2^{\mathbb{N}\times\mathbb{N}} \simeq 2^{\mathbb{N}}$, jmf övning 3.3. Det följer nu från Schröder–Bernsteins sats att $\mathbb{N}^{\mathbb{N}} \simeq 2^{\mathbb{N}}$. □

UPPGIFT 5.3. Antag att $m \geq 2$. Visa att $2^{\mathbb{N}} \simeq m^{\mathbb{N}}$.

SATS 5.8. *Mängden $\mathbb{R}$ av de reella talen är ekvipotent med $2^{\mathbb{N}}$.*

Vi ska bara utnyttja ett par enkla egenskaper hos reella tal. Till att börja med är $\mathbb{R}$ ekvipotent med intervallet $(0, 1)$ eftersom vi har bijektionen $f : \mathbb{R} \to (0, 1)$ där $f(x) = \frac{1}{\pi}(\frac{\pi}{2} + \arctan x)$. Vidare ska vi utnyttja att varje tal $x \in (0, 1)$ svarar mot en oändlig decimalutveckling $0.d_0d_1d_2 \dots$, där varje $d_j \in \{0, 1, \dots, 9\}$, och det inte finns något N sådant att $d_j = 9$ för alla $j > N$.

BEVIS. Vi har en injektion från $(0, 1)$ till $10^{\mathbb{N}}$, genom att $0.d_1d_2 \dots$ avbildas på $f : \mathbb{N} \to 10$, definierad genom $f(k) = d_k$. Den är nästan bijektiv; man missar de $f : \mathbb{N} \to 10$ som är sådana att $f(k) = 9$ för alla k större än något N. Eftersom $10^{\mathbb{N}} \simeq 2^{\mathbb{N}}$ enligt uppgift 5.3 så får vi i alla fall en injektion $(0, 1) \to 2^{\mathbb{N}}$. (Man kan hitta en sådan injektion direkt genom att utnyttja en binär decimalutveckling.) Vi hittar lätt en injektion $2^{\mathbb{N}} \to (0, 1)$ genom att låta $g : \mathbb{N} \to 2$ avbildas på $0.g(0)g(1)g(2) \dots$ om g inte är identiskt noll och låta nollfunktionen t ex avbildas på talet 0.2. Det följer nu från Schröder–Bernsteins sats att $(0, 1) \simeq 2^{\mathbb{N}}$, och alltså $\mathbb{R} \simeq 2^{\mathbb{N}}$. □

UPPGIFT 5.4. Visa att $2^{\mathbb{N}} \times 2^{\mathbb{N}} \simeq 2^{\mathbb{N}}$.

UPPGIFT 5.5. Visa att $\mathbb{R} \simeq \mathbb{R}^2$.

Eftersom $\mathbb{R} \simeq 2^{\mathbb{N}}$, $\mathbb{Q} \simeq \mathbb{N}$ och $2^{\mathbb{N}}$ inte är ekvipotent med $\mathbb{N}$, så följer det att $\mathbb{R}$ inte är ekvipotent med $\mathbb{Q}$. Eftersom $\mathbb{Q} \subset \mathbb{R}$ så innebär detta att det finns "fler" reella tal än vad det finns rationella (eller heltal). Man kan då fråga sig om varje delmängd av $\mathbb{R}$ är ekvipotent med antingen $\mathbb{Q}$ eller $\mathbb{R}$. Påståendet att så är fallet är den s.k. *kontinuumhypotesen*, om vilken man vet att den är oavhängig axiomen i mängdteorin inklusive urvalsaxiomet UA (se avsnitt 6), dvs den går inte att bevisa eller motbevisa, se anmärkning 6.8.

6. Urvalsaxiomet

Om man har en ändlig mängd a av disjunkta mängder är det lätt att bilda en ny mängd b som består av precis ett element ur var och en av dem. T ex kan man utan vidare (?) bilda ett elevråd på en skola bestående av en representant från varje skolklass. Om däremot a är oändlig så krävs det ytterligare ett axiom för att valet ska kunna bli utfört i alla mängderna.

UA (urvalsaxiomet) Låt a vara en mängd av disjunkta icketomma mängder. Då finns en mängd b som består av precis ett element ur var och en av mängderna i a.

Detta axiom ser kanske oskyldigt ut, men har många viktiga konsekvenser, varav en del är mycket användbara, t ex för kardinaltalsaritmetik, i funktionalanalys m m, medan andra konsekvenser strider mot en del människors intuition, såsom t ex Banach–Tarskis paradox, se anmärkning 6.7. Somliga matematiker anser därför att UA i möjligaste mån bör undvikas, och man försöker ofta att bevisa så mycket som möjligt inom en matematisk teori utan att använda UA.

Exempel 6.1. Antag att vi har en partition av X. Enligt UA finns då en mängd med precis ett element ur var och en av ekvivalensklasserna. $\square$

Proposition 6.2. *Antag att $f\colon A \to B$ är surjektiv. Då finns en injektion $g\colon B \to A$ sådan att $f \circ g = I_B$.*

Alltså: Om $f\colon A \to B$ är surjektiv så är B ekvipotent med en delmängd till A.

Bevis. Antag $f\colon A \to B$ är surjektiv. Om $A_b = f^{-1}(b) = \{a;\ f(a) = b\}$ så är $\{A_b;\ b \in B\}$ en mängd av icketomma disjunkta mängder. Enligt UA finns en mängd E som består av precis ett element $g(b)$ ur var och en av dem. Detta definierar en injektion $g\colon B \to A$ och självklart är $f \circ g = I_B$. $\square$

Uppgift 6.1. Visa att UA följer från proposition 6.2.

Vi har alltså att proposition 6.2 är ekvivalent med UA. Det finns många andra vanliga satser som följer från UA och som visar sig vara ekvivalenta, såsom Tychonoffs produktsats, välordningsatsen och Zorns lemma. De två sistnämnda ska vi titta lite närmare på i nästa avsnitt.

Exempel 6.3. Vi påstår att varje oändlig mängd M har en uppräknelig delmängd.

Eftersom M är oändlig är den icketom. Därför kan vi ta ett element a_0. Eftersom M är oändlig är $M \setminus \{a_0\}$ icketom. Tag ett element a_1 ur denna mängd. Igen eftersom M är oändlig så är $M \setminus \{a_0, a_1\}$ icketom så vi kan ta ett element a_2 etc. Vi får på detta sätt en uppräknelig delmängd $A = \{a_0, a_1, a_2, \ldots\}$.

Det är värt att notera att man här bara använder det s.k. uppräkneliga urvalsaxiomet, se anmärkning 6.9. $\square$

Proposition 6.4. *En mängd är oändlig om och endast om den är ekvipotent med en äkta delmängd av sig själv.*

Uppgift 6.2. Visa proposition 6.4.

Exempel 6.5. Vi ska nu utnyttja UA till att visa att man dela upp enhetscirkeln $T = \{z \in \mathbb{C};\ |z| = 1\}$ i uppräkneligt många *kongruenta* (dvs den ena kan erhållas från den andra genom en rotation) parvis disjunkta delmängder.

Fixera ett irrationellt tal r, och definiera för varje $m \in \mathbb{Z}$ avbildningen $\theta^m\colon T \to T$ som ges av $\theta(z) = e^{mr2\pi i}z$, dvs så att θ^m innebär "vridning vinkeln $mr2\pi$". Vi har att $\theta^m \circ \theta^n = \theta^{m+n}$, och speciellt är alltså θ^m inverterbar med inversen θ^{-m}.

För $z, w \in T$ säger vi att $z \sim w$ om det finns $m \in \mathbb{Z}$ så att $w = \theta^m(z)$. Man kollar lätt att $\sim$ är en ekvivalensrelation. Eftersom θ^0 är identitetsavbildningen så gäller $z \sim z$. Om $w = \theta^m(z)$ så är $z = \theta^{-m}(w)$, dvs $z \sim w$ medför att $w \sim z$. Beviset för transitiviteten lämnar vi som övning. Sålunda ger $\sim$ en uppdelning av T i disjunkta ekvivalensklasser. Enligt UA finns nu en mängd A som består av precis ett element från var och en av dessa ekvivalensklasser. Det är klart att mängderna $A_k = \theta^k(A)$, $k = 0, 1, -1, 2, -2, \ldots$ är kongruenta, och vi påstår vidare att de är parvis disjunkta och att deras union är hela T.

Vi visar först att om $A_k \cap A_m \neq \emptyset$ så är $A_k = A_m$. Om $A_k \cap A_m \neq \emptyset$ så finns $z, w \in A$ sådana att $\theta^k(z) = \theta^m(w)$. Men då är $w = \theta^{k-m}(z)$ och alltså $z \sim w$. Eftersom A bara innehåller ett enda element ur varje ekvivalensklass följer att $z = w$. Men då är alltså $\exp((k-m)r2\pi i)z = z$, vilket medför att $\exp((k-m)r2\pi i) = 1$ (eftersom $z \neq 0$) och alltså $(k-m)r2\pi = n2\pi$ för något heltal n. Eftersom r är irrationellt följer därför att $k - m = 0$ och alltså att $A_k = A_m$.

Vi visar nu att unionen av alla A_k är hela T. Varje punkt z ligger i någon ekvivalensklass och är alltså ekvivalent med någon punkt $a \in A$. Detta betyder att $z = \theta^m(a)$ för något m och alltså $z \in A_m$. $\square$

ANMÄRKNING 6.6. Den som har läst en kurs i integrationsteori inser att dessa delmängder A_k inte kan vara *mätbara* (detta betyder att man inte i någon naturlig mening kan tillskriva dem en längd, eller ett s.k. mått). Eftersom de är kongruenta måste de alla ha samma mått c, och eftersom det får plats oändligt många disjunkta sådana mängder på cirkeln, som ju har ändlig längd 2π, så måste i så fall $c = 0$. Å andra sidan finns bara uppräkneligt många A_k, och om var och en har mått noll så måste enligt integrationsteorin deras union också ha mått noll, vilket strider mot att unionen är hela cirkeln. $\square$

ANMÄRKNING 6.7. Genom att tillämpa ett besläktat (men avsevärt mer raffinerat) resonemang på den slutna enhetsbollen $\{(x, y, z) \in \mathbb{R}^3;\ x^2 + y^2 + z^2 \leq 1\}$ i $\mathbb{R}^3$ kan man visa att den går att dela upp i *ändligt* många (5 stycken!) delar som man efter lämpliga kongruenstransformationer (stela förflyttningar i rummet) kan sätta ihop till två enhetsbollar. Detta brukar kallas Banach–Tarskis paradox. För ett bevis se [**OH**]. $\square$

ANMÄRKNING 6.8. Man har *bevisat* att UA är oberoende av de övriga axiomen i ZF, dvs vare sig UA eller dess negation är en logisk konsekvens av de andra axiomen, givet att teorin ZF själv är motsägelsefri. Inte desto mindre kan man ju fråga sig om UA är sant eller inte. Om man anser att mängduniversum existerar, så måste ju rimligen UA vara ett sant eller ett falskt påstående om detta mängduniversum. Eftersom UA medför resultat som strider mot somligas intuition, så kan man detta som intäkt för att UA är falskt, och en del matematiker anser också att UA helt enkelt inte är sant. En vanligare uppfattning är nog att UA är sann och att vissa konsekvenser ter sig märkliga mer antyder att vår intuition om oändliga mängder är svag. En hel del matematiker intar den mer pragmatiska ståndpunkten att utan att tillskriva UA ett bestämt sanningsvärde, t ex på grund av att man inte anser att mängduniversum existerar i någon djupare mening, kan man ändå tillåta sig att använda det vid behov eftersom man i alla fall aldrig kan bevisa någon motsägelse.

Man har också bevisat att kontinuumhypotesen KH är oberoende av ZF samt UA. Den vanligaste uppfattningen om KH är nog att vi inte har tillräcklig förståelse för att kunna bedöma dess sanningsvärde. Man har därför, utan större framgång, försökt hitta andra påståenden som man tror är sanna och som skulle medföra antingen kontinuumhypotesen eller dess negation. Till skillnad från UA åberopas därför inte KH (eller dess negation) i vanlig matematisk bevisföring. □

ANMÄRKNING 6.9. Vi har här talat om det allmänna UA. Om man bara behöver göra ett uppräkneligt antal val så räcker det med det s.k. *uppräkneliga urvalsaxiomet* UUA. Man kan visa att UUA är strikt svagare än UA, dvs att UA inte följer från ZF och UUA, men att UUA likväl är oberoende av ZF.

Det är värt att notera att man behöver UUA, men inte UA, för att visa påståendena i exempel 6.3 och övning 6.2.

När man använder UA i vanliga matematiska sammanhang så nämner man det ofta, men UUA används i allmänhet utan kommentar. Här är ett enkelt exempel från elementär analys. En punkt p är en *hopningspunkt* till en mängd E av reella tal, om varje omgivning till p innehåller något element ur E. Man säger att p är en *sekventiell hopningspunkt* om det finns en följd a_k i E som konvergerar mot a. Det är klart att varje sekventiell hopningspunkt är en hopningspunkt, menför att bevisa omvändningen måste man anlita UUA.

Om man formulerar Bolzano-Weierstrass sats som att varje begränsad mängd av reella tal har en sekventiell hopningspunkt behöver man därför UUA. Däremot kan man visa att varje begränsad talföljd a_k har en konvergent delföljd utan UUA. Detta beror på att man då redan har en välordning av $\{a_k\}$ vilken gör alla nödvändiga val för oss. För en vidare diskussion av detta, se t ex [**GM**]. □

7. Transfinit induktion

SATS 7.1 (Välordningssatsen). *Varje mängd kan välordnas, dvs på varje mängd kan man hitta en ordningsrelation som är en välordning.*

Välordningssatsen är ekvivalent med UA, dvs man kan bevisa den från UA och omvänt följer UA från välordningssatsen. Beviset av välordningssatsen är ganska involverat så vi hoppar över det. Däremot är det lätt att få tillbaka UA från välordningssatsen. Tag nämligen en mängd $M = \{a_\alpha\}_{\alpha\in A}$. Vi kan anta att alla a_α är disjunkta, ersätt annars varje a_α med t ex $a'_\alpha = \{(x,\alpha);\ x \in a_\alpha\}$. Låt nu X vara unionen av alla a_α. Enligt välordningssatsen kan vi välordna X. Varje a_α är nu en delmängd till X och har alltså ett minsta element $f(\alpha)$. Bilden av denna funktion f är nu den sökta mängden.

Ett *avsnitt* X' till en välordnad mängd X är en delmängd sådan att om $x \in X'$ och $x' \leq x$ så $x' \in X'$.

Uppgift 7.1. Visa att X' är ett avsnitt till X om och endast om antingen $X' = X$ eller $X' = \{x \in X;\ x < a\}$ för något $a \in X$.

Uppgift 7.2. Antag att A är en mängd av välordnade mängder så att givet två mängder ur A, den ena är ett avsnitt av den andra. Visa att $\cup A$ har en total ordningsrelation vars restriktion till varje mängd a i A är den ursprungliga ordningen på a. Visa att ordningen på $\cup A$ är en välordning.

Anmärkning 7.2. Det går att visa att varje välordnad mängd är isomorf (se nedan) med ett bestämt *ordinaltal.* Ordinaltalen är bildade på följande vis. Till att börja med tar man $0 < 1 < 2 < 3 < \ldots$, dvs de naturliga talen. Kom ihåg att $n = n{-}1 \cup \{n{-}1\}$. Man bildar nu unionen av alla dessa naturliga talen och kallar den ω (istället för $\mathbb{N}$ i detta sammanhang), jmf övning 7.2. Sedan kan man bilda $\omega + 1 = \omega \cup \{\omega\}$ etc, och vi får

$$
\begin{aligned}
0 < 1 < 2 < 3 < \ldots < \omega < \omega + 1 < \\
< \omega + 2 < \ldots < 2\omega < 2\omega + 1 < 2\omega + 2 < \ldots < 3\omega < \\
< 3\omega + 1 < \ldots < \ldots < \omega^2 < \omega^2 + 1 < \ldots
\end{aligned}
$$

där varje ordinaltal är unionen av de föregående, och varje ordinaltal på detta sättet är en välordnad mängd, jmf övning 7.2 igen. Allmänt definierar man ett ordinaltal till att vara en välordnad mängd α sådan att det för varje $x \in \alpha$ gäller att $x = \{y \in \alpha;\ y < x\}$. Man kan visa att det för varje par av ordinaltal gäller att det ena är en delmängd till det andra, jmf Sats 7.8. □

Som bekant är induktionsbevis en vanlig metod när man vill visa ett påstående gäller för alla naturliga tal $\mathbb{N}$. Som vi har sett i avsnitt 4 så är (den starka) induktionsprincipen för $\mathbb{N}$ bara en omformulering av att $\mathbb{N}$ är välordnad, och eftersom varje mängd kan välordnas är det därför rimligt att tänka sig att man kan utvidga tekniken med induktionsbevis till överuppräkneliga mängder. Detta brukar kallas *transfinit induktion.*

Exempel 7.3. Låt oss använda transfinit induktion för att visa att varje vektorrum har en bas. En delmängd B till ett vektorrum X är en bas om dels B är lineärt oberoende, dels varje $x \in X$ går att skriva som en (ändlig) lineärkombination av element ur B.

För att få en idé om hur beviset ska gå till, låt oss först anta att vi bara har uppräkneligt många (eller rent av ändligt många) vektorer $v_0, v_1, \ldots$ givna. Vi önskar då hitta en lineärt oberoende delmängd B

av dessa sådan att varje v_j är en (ändlig) lineärkombination av element ur B.

Vi betraktar nu till att börja med den första vektorn v_0. Om den är lineärt oberoende (vilket den är såvida den inte råkar vara nollvektorn) lägger vi den i mängden B som ska bli vår bas. Sedan tittar vi på nästa vektor, dvs den minsta i $V \setminus \{v_0\}$. Om den är lineärt oberoende med v_0 så lägger vi in den i B annars kastar vi den. Detta förfarande upprepas så att givet att vi redan har testat alla alla v_k upp till och med $k = n$ på detta sätt, så betraktar vi v_{n+1}. Om den är lineärt oberoende av dem man hittills har lagt in i B så lägger man in även v_{n+1} i B, annars kastar man den och betraktar istället v_{n+2} etc.

Uppgift 7.3. Visa att mängden B så konstruerad blir lineärt oberoende och att varje v_j ligger i dess lineära hölje.

Vi ska nu göra om detta argument för hela vektorrummet V med transfinit induktion. Vi börjar med att välordna V. Låt sedan Q vara mängden av alla delmängder $A \subset V$ sådana att A är lineärt oberoende och att varje $v \in V$ som är mindre än eller lika med något $a \in A$ spänns av $A \cap \{a';\ a' \leq v\}$. Detta betyder intuitivt att varje A i Q är en mängd av den typen man har erhållit som i det ändliga fallet genom att successivt ha kollat alla v upp till en viss storlek.

Vi påstår nu att om $A, A' \in Q$ så är $A \subset A'$ eller $A' \subset A$. Om de inte är lika så finns ett minsta a_0 som ligger i ena men inte i andra, säg $a_0 \in A \setminus A'$. Detta betyder att $A \cap \{v;\ v < a_0\} = A' \cap \{v;\ v < a_0\}$. Om nu det finns något element $v \in A'$ sådant att $a_0 \leq v$ så ligger a_0 i lineära höljet av $A' \cap \{v;\ v \leq a_0\}$, men alltså i lineära höljet av $A' \cap \{v;\ v < a_0\} = A \cap \{v;\ v < a_0\}$ vilket strider mot att $a_0 \in A$. Alltså kan inte ett sådant $v \in A'$ finnas och alltså är $A' = A \cap \{v;\ v < a_0\}$.

Låt nu $B = \cup Q$ dvs B är unionen av alla A i Q. Vi påstår nu att $B \in Q$. Till att börja med är det självklart att B är lineärt oberoende; detta beror på att varje ändlig delmängd ur B ligger i en och samma A. Vore de lineärt beroende så vore de det redan i A vilket strider mot att A är en lineärt oberoende mängd. Antag nu att v' är mindre än något element i B. Då är det även mindre än något element i något $A \in Q$ och alltså spänns v' av $A \cap \{v;\ v \leq v'\} \subset B \cap \{v;\ v \leq v'\}$. Alltså $B \in Q$.

Det återstår att visa att varje $v \in V$ ligger i lineära höljet av B. Antag motsatsen; och låt v_0 vara det minsta som *inte* ligger i höljet. Till att börja med måste då v_0 vara större än varje element i B. Betrakta nu $A = B \cup \{v_0\}$. Till att börja med är den lineärt oberoende. Om v' är mindre än eller lika med något element i A så är antingen $v' = v_0$ i vilket fall självklart v' spänns av $A \cap \{v;\ v \le v'\}$. Om v' är mindre eller lika med något element i B gäller samma sak eftersom $B \in Q$. Slutligen gäller samma sak om v' är större än alla element i B men mindre än v_0 eftersom ju då v' spänns av B pga definitionen av v_0. Alltså finner vi att $A \in Q$ vilket motsäger definitionen av B. Beviset är klart! $\square$

Trots den synnerligen enkla idén i beviset är själva genomförandet en smula trassligt. När man genomför transfinit induktion brukar man därför istället använda en med välordningssatsen besläktad sats som är känd som Zorns lemma.

Antag att X är en (partiellt) ordnad mängd. En delmängd X' till X har en *majorant* m om $x \le m$ för alla $x \in X'$. En majorant kan, men behöver inte, ligga i X'. Man säger att m är ett *maximalt element* i X om det inte finns något x i X sådant att $m < x$. Notera att inte nödvändigtvis $x \le m$ för alla $x \in X$. Notera att även om X bara är partiellt ordnad så kan mycket väl en delmängd X' till X vara totalt ordnad. Betrakta t ex delmängden $\{(x, x);\ x \in \mathbb{R}\}$ till $\mathbb{R}^2$ i exempel 4.4.

SATS 7.4 (Zorns lemma). *Antag att X är en partiellt ordnad mängd sådan att varje totalt ordnad delmängd har en majorant. Då har X ett maximalt element.*

Som tidigare nämnts så kan Zorns lemma bevisas från UA och omvänt kan man få tillbaka UA från Zorns lemma. Emellertid hoppar vi över dessa bevis. Istället ska vi tillämpa Zorns lemma och ge ett nytt bevis för att varje vektorrum har en bas.

EXEMPEL 7.5. Låt X bestå av alla lineärt oberoende delmängder till vektorrummet V. För $A, B \in X$ sätter vi $A \le B$ om och endast om $A \subset B$. Då är $\le$ en partiell ordning på X. Antag nu att X' är en totalt ordnad delmängd till X (dvs X' består av en massa lineärt oberoende delmängder till V sådana att för varje par av dem så är den ena en delmängd till den andra). Vi påstår nu att X' har en majorant. Tag nämligen unionen M av alla mängderna i X'. Då är M förstås en delmängd till V men den är också lineärt oberoende; detta beror på att varje ändlig delmängd till M redan måste ligga i en fix mängd $A \in X'$ just på grund av att X' är totalt ordnad. Alltså är denna ändliga delmängd lineärt oberoende och alltså är M det.

Förutsättningarna i Zorns lemma är alltså uppfyllda och vi har därför ett maximalt element $B \in X$, dvs en lineärt oberoende mängd B sådan att det inte finns någon annan lineärt oberoende mängd $B' \supset B$. Men då är ju klart att B är en bas för om t ex v inte låg i det lineära höljet av B så skulle ju $B' = B \cup \{v\}$ vara en sådan. $\square$

Exempel 7.6. Notera att man inte bara behöver ha reella vektorrum; man kan lika gärna ha vektorrum över t ex Q. Då är $\mathbb{R}$ ett vektorrum över $\mathbb{Q}$ och alltså finns en bas. $\square$

En bijektion mellan två ordnade mängder sägs vara en *isomorfi* (och mängderna sägs vara isomorfa) om den bevarar ordningsrelationen, dvs om $f(x) \leq f(y)$ om och endast om $x \leq y$.

Som en ytterligare illustration av transfinit induktion kan vi visa

Lemma 7.7. *Om A och B är välordnade mängder så finns högst en isomorfi från A till något avsnitt B_1 till B.*

Bevis. Låt $f_1 \colon A \to B_1$ och $f_2 \colon A \to B_2$ vara två sådana isomorfier. Om inte $f_1(x) = f_2(x)$ för alla $x \in A$ så kan vi ta a_0 som det minsta elementet i mängden

$$\{x \in A;\ f_1(x) \neq f_2(x)\}.$$

Då är

$$f_1(\{x \in A;\ x < a_0\}) = f_2(\{x \in A;\ x < a_0\}) = B'.$$

Eftersom a_0 är det minsta elementet i $A \setminus \{x \in A;\ x < a_0\}$ och f_1 bevarar ordning så är $f_1(a_0)$ det minsta elementet i $B_1 \setminus B'$. Men B_1 är ett avsnitt av B och $B' \subset B_1$, och därför är $f_1(a_0)$ det minsta elementet i $B \setminus B'$. På samma sätt är då $f_2(a_0)$ det minsta elementet i $B \setminus B'$, dvs $f_1(a_0) = f_2(a_0)$, vilket motsäger definitionen av a_0. $\square$

Med liknande teknik kan man visa

Sats 7.8. *Av två välordnade mängder är åtminstone den ena isomorf med ett avsnitt till den andra.*

För ett bevis, se t ex **[JK]**.

Sats 7.9. *Av två mängder är åtminstone den ena ekvipotent med en delmängd av den andra.*

Bevis. Välordna först båda mängderna. Sedan följer påståendet genast från Sats 7.8. $\square$

Om två mängder A och B är ekvipotenta kan vi skriva detta $|A| = |B|$. Om A är ekvipotent med en delmängd till B och B inte är ekvipotent med en delmängd till A skriver vi detta $|A| < |B|$. Från Sats 7.9 och Schröder–Bernsteins sats finner vi då att för två godtycklig mängder A och B gäller precis en av följande; $|A| = |B|$, $|A| < |B|$ eller $|B| < |A|$.

Från Zorns lemma följer även, se t ex **[JK]**, SATS 7.10. *Om A är oändlig så är $A \times A \simeq A$.*

8. Ytterligare övningar till kapitel 2

UPPGIFT 8.1. Uttryck mängdaxiomen IV, V, VII och UA med predikatlogiska formler (axiom VI är lite svårare).

UPPGIFT 8.2. Antag att a är en mängd. Visa att det finns en mängd b sådan att elementen i b är precis alla partitioner av a.

UPPGIFT 8.3. Låt A vara en uppräknelig mängd. Visa att mängden av alla ändliga delmängder till A är uppräknelig.

UPPGIFT 8.4. Antag att A är en ändlig mängd och f en funktion $f: A \to A$. Visa att det finns en icketom delmängd $B \subset A$ sådan att $f(B) = B$.

UPPGIFT 8.5. Låt $A = B = \mathbb{N}$ och låt $f(x) = 2x$ och $g(x) = 3x$. Då är förutsättningarna i Schröder–Bernsteins sats uppfyllda och beviset ger en metod att konstruera en bijektion mellan A och B (dvs från $\mathbb{N}$ till sig själv) utgående från f och g. Beskriv vad den blir!

UPPGIFT 8.6. Gör samma sak med $A = \mathbb{N}$, $B = \mathbb{N} \times \mathbb{N}$, $f(x) = (x, 0)$ och $g(x, y) = 2^x 3^y$.

UPPGIFT 8.7. Visa att om A är uppräknelig och B är ändlig eller uppräknelig så är $A \cup B$ uppräknelig.

UPPGIFT 8.8. Om B är oändlig och A är uppräknelig (eller ändlig) så är $A \cup B$ ekvipotent med B.

UPPGIFT 8.9. Visa att $\mathbb{R} \simeq \mathbb{R}^n$.

UPPGIFT 8.10. Visa att $\mathbb{R} \simeq \mathbb{R}^{\mathbb{N}}$.

UPPGIFT 8.11. Visa att $\mathbb{C}$, mängden av komplexa tal, är ekvipotent med $\mathbb{R}$.

UPPGIFT 8.12. Visa att mängden av algebraiska tal A (dvs mängden av alla komplexa tal som är rötter till någon polynomekvation med rationella koefficienter) är uppräknelig.

UPPGIFT 8.13. Visa att mängden av alla transcendenta tal, $\mathbb{R} \setminus A$, inte är uppräknelig.

UPPGIFT 8.14. Är $\{X \subset \mathbb{N};\ 2X \cap X \simeq \mathbb{N}\}$ uppräknelig?

UPPGIFT 8.15. Är mängden av alla bijektioner från $\mathbb{N}$ till $\mathbb{N}$ uppräknelig?

UPPGIFT 8.16. Låt M vara en mängd med minst två element. Visa att det finns en bijektion $f\colon M \to M$ sådan att $f(x) \neq x$ för alla $x \in M$.

UPPGIFT 8.17. Antag att $A \neq \emptyset$ och att $2^A \simeq A^2$. Vad kan sägas om A?

UPPGIFT 8.18. Antag att $B_0,\ B_1, \dots$ är en följd av *ändliga* mängder sådana att $B_0 \supset B_1 \supset \dots$. Visa att om varje B_k är icketom så är snittet $\cap_1^\infty B_j$ icketomt. Visa med exempel att detta inte behöver vara sant om mängderna inte är ändliga.

UPPGIFT 8.19. Antag att vi har en följd av ändliga icketomma mängder $A_0,\ A_1, \dots$ och avbildningar $f_k\colon A_{k+1} \to A_k$. Visa att det finns en följd av element $a_0,\ a_1, \dots$ sådana att $a_0 \in A_0$, $a_1 \in A_1$ etc sådana att $f_k(a_{k+1}) = a_k$ för alla $k \geq 0$.

UPPGIFT 8.20. Betrakta mängden av alla satslogiska satser som endast innehåller satsvariablerna $A_0, \dots, A_{m-1}$. Sätt $\phi \sim \psi$ omm $\phi \leftrightarrow \psi$ är en tautologi och visa att $\sim$ är en ekvivalensrelation. Visa att det finns 2^{2^m} olika ekvivalensklasser.

UPPGIFT 8.21. Visa att det halvöppna intervallet $[0,1)$ är ekvipotent med det öppna intervallet $(0,1)$ genom att hitta en explicit bijektion.

UPPGIFT 8.22. Låt $\leq$ vara restriktionen av ordningsrelationen i exempel 4.4 på $\mathbb{R}^2$, till $M = \{(x,y) \in \mathbb{R}^2;\ 0 \leq x,\ 0 \leq y,\ x+y \leq 1\}$. Avgör om M har ett största och/eller minsta element, och om M har något maximalt och/eller minimalt element.

UPPGIFT 8.23. Låt M vara delmängden till $\mathbb{R}^2$ från förra uppgiften. Vilka element i $\mathbb{R}^2$ är majoranter till M?

Några kommentarer

En grundläggande framställning av mängdteori finns i [**EM**] och [**JK**]. Den senare sträcker sig längre och innehåller även bevis för att UA och KH är konsistenta med övriga mängdaxiomen (givet att dessa är konsistenta). I [**GM**] finns en omfattande diskussion om olika former av UA.

KAPITEL 3

Talsystem

I förra kapitlet visade vi hur man kan definiera de naturliga talen i mängdteori. Vi ska nu gå igenom grundläggande aritmetik för de naturliga talen. Vidare ska visa hur man kan utvidga de naturliga talen till heltalen $\mathbb{Z}$, de rationella talen $\mathbb{Q}$ samt de reella och komplexa talen $\mathbb{R}$ respektive $\mathbb{C}$.

1. De naturliga talen

I avsnitt 1 i förra kapitlet såg vi att de naturliga talen $\mathbb{N}$, såsom vi definierade dem där, uppfyller Peanos axiom om vi låter n' vara $n \cup \{n\}$. Eftersom det är dessa axiom vi utgår från i detta kapitel så skriver vi upp dem igen.

P1 0 är ett naturligt tal.
P2 Till varje naturligt tal n finns en entydigt bestämd efterföljare n'.
P3 $n' \neq 0$ för alla n
P4 Om $n' = m'$ så är $n = m$.
P5 Antag att $X \subset \mathbb{N}$. Om $0 \in X$ och om $n' \in X$ närhelst $n \in X$ så är $X = \mathbb{N}$.

Elementen i en modell till dessa axiom kallar vi (naturliga) tal. Axiom P5 ger möjlighet att utföra induktionsbevis. Om $\phi(n)$ är en utsaga för varje naturligt tal n sådan att $\phi(0)$ är sann och $\phi(n)$ medför $\phi(n')$ så är mängden $\{n \in \mathbb{N};\ \phi(n) \text{ sann}\}$ lika med hela $\mathbb{N}$ enligt P5, dvs $\phi(n)$ gäller för alla $n \in \mathbb{N}$.

Proposition 1.1. *För alla tal n gäller att $n' \neq n$, och varje tal n utom 0 är efterföljare till ett bestämt tal.*

Bevis. Vi använder induktion och låter $\phi(n)$ stå för $n' \neq n$. Enligt P3 är $0' \neq 0$. Antag $n' \neq n$. Då är $(n')' \neq n'$ enligt P4. Alltså är beviset av första påståendet klart.

Entydigheten i det andra påståendet följer genast av P4 och existensen är en omedelbar konsekvens av P5. $\square$

Vi kan nu definiera $+$ och $\cdot$ rekursivt genom

$$\begin{aligned} m+0&=m\\ m+n'&=(m+n)'\\ m\cdot 0&=0\\ m\cdot n'&=(m\cdot n)+m \end{aligned}$$

Istället för $m\cdot n$ skriver vi ofta mn. De naturliga talen med denna definition av $+$ och $\cdot$ är en modell till den predikatlogiska teori T som är beskriven i exempel 5.17 i kapitel 1.

Med induktion bevisar man de vanliga räknelagarna

$$\begin{array}{lll} m+n=n+m, & mn=nm & \text{kommutativitet}\\ (k+m)+n=k+(m+n), & (km)n=k(mn) & \text{associativitet}\\ k(m+n)=km+kn & & \text{distributivitet} \end{array}$$

Som exempel bevisar vi kommutativiteten för addition.

Bevis av kommutativiteten för addition. Vi visar att $m+n=n+m$ genom induktion över m. Tag först $m=0$. Då är ju

$$0+n=n+0 \tag{1.1}$$

sann om $n=0$. Antag att vi kan (1.1) för fixt n. Då är $0+n'=(0+n)'=(n+0)'=n'=n'+0$, och alltså följer det att (1.1) gäller för alla n. Antag nu att vi har visat

$$m+n=n+m \tag{1.2}$$

för ett givet m och alla n. Vi vill då se att (1.2) även gäller för m' och alla n. Vi vet redan att $m'+0=0+m'$. Antag att vi också vet att $m'+n=n+m'$ för något visst n. Då har vi att

$$\begin{aligned} m'+n'=(m'+n)'=(n+m')'=(n+m)''=(m+n)''=&\\ =(m+n')'=(n'+m)'=n'+m'.& \end{aligned}$$

Detta visar att $m'+n=n+m'$ för alla n. Alltså är induktionen över m klar, och därmed beviset. □

Man kan notera att detta bevis liksom alla liknande grundläggande verifikationer fungerar lika bra i predikatlogik, dvs man utnyttjar bara det predikatlogiska axiomet (a_7) i teorin T som ersätter det allmänna induktionsaxiomet P5.

Uppgift 1.1. Visa någon av de övriga räknelagarna t ex associativiteten för addition.

Uppgift 1.2. Visa att om $n + k = m + k$ för några naturliga tal n,m,k, så är $n = m$.

Det betyder att ekvationen $n+x = m$ har entydig lösning om den är lösbar. Vi vill nu införa en den vanliga ordningsrelationen på $\mathbb{N}$ genom att säga att $n \leq m$ om och endast om $n + x = m$ är lösbar. Vi har då några saker att kontrollera.

Uppgift 1.3. Visa att för givna tal n och m så är $n + x = m$ eller $m + x = n$ lösbar.

Om båda dessa ekvationer är lösbara, säg $n+k = m$ och $m+\ell = n$, så är enligt associativiteten och kommutativiteten $n+m+k+\ell = n+m$. Alltså är $k + \ell = 0$ och vilket medför att $\ell = 0$ och alltså är $m = n$. Om båda ekvationerna $n + x = m$ och $m + x = k$ är lösbara så följer det lätt att även $n + x = k$ är lösbar. Sammantaget innebär detta att $\leq$ är en total ordningsrelation på $\mathbb{N}$.

Med hjälp av ordningsrelationen kan vi formulera den starka induktionsprincipen, jmf avsnitt 4.

Sats 1.2. *Antag att $\phi(k)$ är en relation på $\mathbb{N}$ sådan att $\phi(n')$ gäller om $\phi(k)$ gäller för alla $k \leq n$. Då gäller $\phi(n)$ för alla n.*

Bevis. Definiera en ny relation $\psi(k)$ sådan att $\psi(k)$ omm $\phi(\ell)$ för alla $\ell < k$. Då gäller till att börja med $\psi(0)$ trivialt eftersom det inte finns något $\ell < 0$. Antag nu att $\psi(k)$ gäller, dvs att $\phi(\ell)$ gäller för alla $\ell < k$. Då gäller enligt antagandet i satsen även $\phi(k)$ och alltså $\psi(k')$. Med den enkla induktionsprincipen följer nu att $\psi(k)$ gäller för alla k och detta betyder att även $\phi(k)$ gäller för alla k. □

Enligt resonemanget i avsnitt 4 i kapitel 2 följer det att $\mathbb{N}$ är välordnad.

Uppgift 1.4. Visa direkt att $\mathbb{N}$ är välordnad, utgående från (den svaga) induktionsprincipen.

Som vi nämnt tidigare så finns, så när som på isomorfi, endast en modell till Peanos axiom, dvs om M_1 och M_2 båda är modeller så finns en bijektion $\phi\colon M_1 \to M_2$ sådan att $\phi(0) = 0$ (snarare tolkningen av symbolen 0) och $\phi(x') = (\phi(x))'$. Om nämligen ϕ definieras på detta sätt, så följer av P5 att den därmed är definierad för alla $x \in M_1$, och av samma skäl följer att ϕ är surjektiv. Injektiviteten är klar eftersom (tolkningarna av) $0, 0', 0'', \ldots$ alla är olika. Om vi tittar på modellerna till den predikatlogiska teorin T som beskrivs i exempel 5.17 i kapitel 1 så är saken annorlunda.

Exempel 1.3. Tag en ny individkonstant a och bilda den nya teorin $T \cup \{a \neq 0, a \neq 0', a \neq 0'', a \neq 0''', \dots\}$. Varje ändlig delmängd av denna formelmängd har en modell, nämligen $\mathbb{N}$, och enligt kompakthetssatsen, se kapitel 1.5, så har T' en modell M. Men denna är alltså en modell till T, och dessutom sådan att det finns ett element (tolkningen av) a som inte är innehållet i mängden $\{0, 0', 0'', \dots\}$. Alltså är inte denna modell isomorf med $\mathbb{N}$. En modell av detta slag brukar kallas en *icke-standardmodell* till T. □

Gödel visade 1931 att det finns en formel ϕ i teorin T (dvs som är uppbyggd bara med alfabetet för T) som varken kan bevisas eller motbevisas i teorin T. Denna formel innebär alltså ett påstående om naturliga tal för vilken det inte finns ett bevis (utifrån axiomen i T). Konstruktionen av denna formel ϕ är mycket raffinerad. Till att börja med så tilldelar man varje predikatlogisk formel ett naturligt tal, dess *Gödeltal*; vidare tilldelar man varje predikatlogisk härledning (bevis) ett tal, så att existensen av ett (predikatlogiskt) bevis för en viss formel ϕ svarar mot existensen av ett naturligt tal n sådant att ett visst påstående $\Psi(n)$ är sant. Slutligen lyckades Gödel konstruera en sluten formel ϕ som säger "det finns inte något (predikatlogiskt) bevis av mig". Denna formel ϕ är ett påstående i teorin T som varken kan bevisas eller motbevisas i T. Om det finns ett bevis i T så leder detta till motsägelse eftersom den då säger att det inte finns ett sådant bevis. Omvänt, om det finns ett bevis för $\neg\phi$ i T så är det alltså sant att det finns ett bevis för ϕ vilket är omöjligt såvida inte T är inkonsistent. Notera dock att det påstående om naturliga tal som ϕ svarar mot är sant, dvs att ϕ är sann i tolkningen $\mathbb{N}$ av teorin T. Detta beror på att om ϕ vore falsk så skulle det finnas ett bevis för ϕ vilket ju leder till motsägelse. Existensen av denna formel ϕ kallas *Gödels ofullständighetssats* och är ett av de absolut mest remarkabla resultaten inom logiken.

Vad händer då om man utvidgar teorin T genom att lägga till formeln ϕ eller dess negation som ett nytt axiom? Gödels konstruktion ger då en ny formel som inte kan bevisas eller motbevisas osv. Man kan förstås lägga till som axiom en *maximal* konsistent mängd av formler; dvs en mängd av formler som är konsistent och sådan att det inte går att lägga till en enda ytterligare formel utan att mängden blir inkonsistent. Man kan som axiom t ex ta alla formler som är sanna påståenden om de naturliga talen. Då följer det att varje formel eller dess negation kan härledas, men man har inte vunnit något eftersom det inte längre finns någon möjlighet att mekaniskt avgöra om en given formel är ett axiom eller inte, och följaktligen finns inte längre någon möjlighet att avgöra vad som är en härledning.

Ett *formellt system* innebär att man har en uppsättning härledningsregler sådana att man på något mekaniskt sätt kan avgöra vad som är en härledning. Man kan använda Gödels konstruktion i varje formellt system som innehåller tillräckligt mycket aritmetik, för att konstruera utsagor som är oavgörbara inom systemet. Exempelvis kan man ta en godtycklig första ordningens teori som innehåller T, där man har något sätt att avgöra vilka utsagor som är axiom. Man kan också ta en andra ordningens teori som innehåller T, t ex Peanos axiom och en uppsättning formella härledningsregler.

Det följer av ofullständighetssatsen bl a att man inte kan bevisa motsägelsefriheten hos teorin T utan att redan från början anta att $\mathbb{N}$ är motsägelsefri. För närmare diskussion hänvisar vi till en bok i logik, t ex [**EM**].

ANMÄRKNING 1.4. Gödels *fullständighetssats* (sats 5.21 i kapitel 1) säger att alla logiska konsekvenser av axiomen i en första ordningens teori kan härledas. Ofullständighetssatsen medför alltså att det finns formler i teorin T som inte är logiska konsekvenser av axiomen; dvs axiomsystemet är ofullständigt. För den andra ordningens teorin som ges av Peanos axiom vet vi att alla modeller är isomorfa och att följaktligen varje påstående eller dess negation är en logisk konsekvens av axiomen. Givet att man har bestämt något mekaniskt härledningssystem, så visar ofullständighetsatsen att det finns logiska konsekvenser av axiomen som inte kan härledas med det givna härledningssystemet. I detta fall är det alltså inte axiomen utan härledningsreglerna som är ofullständiga. □

UPPGIFT 1.5. Visa att det omedelbart följer från Gödels ofullständighetssats (och sats 5.22 i kapitel 1) att det finns icke-standardmodeller till T.

2. Heltal

DEFINITION 2.1. En mängd med två operationer $+$ och $\cdot$ och två element 0 och 1 är en *kommutativ ring med etta* om följande är uppfyllt:

$$x + y = y + x, \quad xy = yx \tag{2.1}$$

$$(x + y) + z = x + (y + z), \quad (xy)z = x(yz) \tag{2.2}$$

$$x(y + z) = xy + xz \tag{2.3}$$

$$1 \cdot x = x, \quad 0 + x = x, \tag{2.4}$$

och dessutom ekvationen $\alpha + x = \beta$ alltid är lösbar.

Vi har alltså en (predikatlogisk) teori vars modeller kallas kommutativa ringar (med etta). Lösningen till $\alpha + x = \beta$ är entydig och betecknas $\beta - \alpha$. Antag nämligen att

$$\alpha + x = \alpha + x'. \tag{2.5}$$

Tag z sådant att $\alpha + z = 0$. Genom att addera z på båda sidorna av (2.5) så följer att $x = x'$.

Notera att $\mathbb{N}$ uppfyller (2.1) till (2.4) men att ekvationen $a + x = b$ inte alltid är lösbar. Vi ska nu visa att man kan utvidga $\mathbb{N}$ till en ring, *heltalsringen* $\mathbb{Z}$. Det skenbart enklaste sättet är att lägga till element $-1, -2, \ldots$ och stipulera räkneregler av typ $(-m)+n = x$ om $m+x = n$ är lösbar och $(-m) + n = -x$ om $n + x = m$ är lösbar etc. Detta är fullt möjligt men eftersom man vill att räknereglerna (2.1) till (2.4) fortfarande ska gälla så leder detta till en massa olika fall som måste kollas igenom. Vi ska därför presentera en elegantare konstruktion av $\mathbb{Z}$, som dessutom kan användas i andra sammanhang. Idén är att man istället betraktar alla formella differenser $a - b$ där a och b är naturliga tal och identifierar två olika sådana uttryck $a-b$ och $c-d$ omm $a+d = b + c$. Formellt inför man på $\mathbb{N} \times \mathbb{N}$ relationen $\sim$ genom att

$$\langle a, b\rangle \sim \langle c, d\rangle \qquad \text{omm} \qquad a + d = b + c.$$

PROPOSITION 2.2. *$\sim$ är en ekvivalensrelation.*

BEVIS. Vi lämnar reflexiviteten och symmetrin som övning och visar bara transitiviteten. Antag att $\langle a, b\rangle \sim \langle c, d\rangle$, dvs $a + d = b + c$ och $c + f = d + e$. Det följer då att $a + d + c + f = b + c + d + e$ vilket pga associativiteten och kommutativiteten är detsamma som att $a + f + g = b + e + g$, där $g = c + d$. Men då följer att $a + f = b + e$ dvs $\langle a, b\rangle \sim \langle e, f\rangle$. □

Vi låter nu $\mathbb{Z}$ vara mängden av ekvivalensklasser, och låter $[a, b]$ beteckna den ekvivalensklass som innehåller $\langle a, b\rangle$. Alltså är $[a, b] = [c, d]$ omm $a + d = b + c$. Speciellt är $[a, b] = [a + c, b + c]$ för alla c. Vi definierar operationen $+$ på $\mathbb{Z}$ genom att sätta

$$[a, b] + [c, d] = [a + c, b + d].$$

För att denna definition ska ha mening måste vi kontrollera att den inte beror på valet av representanter $\langle a, b\rangle$ och $\langle c, d\rangle$ för respektive ekvivalensklasser, dvs vi måste kontrollera att om $\langle a, b\rangle \sim \langle a_1, b_1\rangle$ och $\langle c, d\rangle \sim \langle c_1, d_1\rangle$ så är $\langle a + c, b + d\rangle \sim \langle a_1 + c_1, b_1 + d_1\rangle$. Men detta i sin tur betyder att vi antar att $a + b_1 = b + a_1$ och $c + d_1 = d + c_1$ och från detta vill visa att $a + c + b_1 + d_1 = b + d + a_1 + c_1$, vilket dock är självklart.

Vi inför sedan operationen $\cdot$ på $\mathbb{Z}$ genom

$$[a,b]\cdot[c,d]=[ac+bd,ad+bc].$$

Igen måste man verifiera att detta inte beror på valet av representanter, men det lämnar vi som övning. Nästa steg är att kontrollera att räknelagarna (2.1) till (2.4) gäller för $\mathbb{Z}$. Detta görs också helt rutinmässigt, och som exempel verifierar vi distributiviteten:

$$\begin{gathered}[e,f]([a,b]+[c,d])=[e,f][a+c,b+d]=\\ [e(a+c)+f(b+d),e(b+d)+f(a+c)]=\\ [ea+fb,fa+eb]+[ec+fd,ed+fc]=[e,f][a,b]+[e,f][c,d].\end{gathered}$$

Vi noterar nu att $x=[a+d,b+c]$ är en lösning till ekvationen $[c,d]+x=[a,b]$, och att alltså $\mathbb{Z}$ är en kommutativ ring med etta.

Det återstår att se att $\mathbb{N}$ naturligt kan inbäddas i $\mathbb{Z}$, dvs att $\mathbb{N}$ kan uppfattas som en delmängd till $\mathbb{Z}$ sådan att operationerna $+$ och $\cdot$ stämmer överens. Mer formellt betyder detta att det finns en injektiv avbildning $\psi\colon\mathbb{N}\to\mathbb{Z}$ sådan att

$$\begin{multline}\psi(0)=0,\quad \psi(1)=1,\\ \psi(a+b)=\psi(a)+\psi(b),\quad \psi(ab)=\psi(a)\psi(b).\end{multline}\tag{2.6}$$

Man tar helt enkelt och låter det naturliga talet a avbildas på $[a,0]$ och kontrollerar att (2.6) gäller. När detta väl är gjort kan vi uppfatta $\mathbb{N}$ som en delstruktur till $\mathbb{Z}$, eller om vi så vill, $\mathbb{Z}$ som en utvidgning av $\mathbb{N}$.

Slutligen kan vi införa en total ordningsrelation på $\mathbb{Z}$ genom att säga att $\alpha\le\beta$ omm $\beta-\alpha$ är ett naturligt tal. Vi lämnar som övning att kolla att det alltid gäller att $\alpha-\beta$ eller $\beta-\alpha$ är ett naturligt tal, och att båda gäller omm $\alpha=\beta$.

3. Rationella tal

DEFINITION 3.1. En ring med etta är en *kropp* om ekvationen $ax=b$ alltid är lösbar för $a\neq 0$.

Ringen $\mathbb{Z}$ har den extra egenskapen att vara ett *integritetsområde*. Detta innebär att $\alpha\beta=0$ omm α eller β är noll. Man säger också att $\mathbb{Z}$ saknar nolldelare. Varje integritetsområde R kan utvidgas till en kropp $Q(R)$ genom att man betraktar alla utryck av typen α/β där $\beta=0$ och identifierar två sådana utryck α/β och $\alpha'\beta'$ omm $\alpha\beta'=\beta\alpha'$. Formellt inför man en ekvivalensrelation på $R\times R\setminus\{0\}$ genom att

$\langle \alpha, \beta \rangle \sim \langle \alpha', \beta' \rangle$ omm $\alpha\beta' = \beta\alpha'$, och låter $Q(R)$ vara mängden av ekvivalensklasser. På liknande sätt som i förra avsnittet kan man sedan införa operationerna $+$ och $\cdot$ genom att sätta

$$[\alpha, \beta] + [\gamma, \delta] = [\alpha\delta + \beta\gamma, \beta\delta], \qquad [\alpha, \beta][\gamma, \delta] = [\alpha\gamma, \beta\delta].$$

Man har igen att kontrollera att högerleden inte beror på valet av representanter och att räknereglerna (2.1) till (2.4) gäller, och att det blir en kropp som har R som en inbäddad delring, genom att $\alpha \in R$ avbildas på klassen $[\alpha, 1]$. Vi lämnar allt detta som övning eller hänvisar till någon bok i grundläggande algebra. I fallet att man utgår från $\mathbb{Z}$ så kallas den resulterande kroppen för $\mathbb{Q}$, de rationella talen.

Varje rationellt tal α går alltså att skriva som $\alpha = a/b$ där a och b är heltal, och vi säger att $\alpha \geq 0$ omm a och b har samma tecken. Man kollar lätt att detta inte beror på representationen och att $\leq$ är en total ordning på $\mathbb{Q}$ sådan att den för den inbäddade delringen $\mathbb{Z}$ stämmer överens med ordningsrelationen på $\mathbb{Z}$. Vidare kollar man lätt att

$$0 \leq a \text{ och } 0 \leq b \text{ medför att } 0 \leq a + b \text{ och } \ 0 \leq ab. \tag{3.1}$$

En kropp med en total ordningsrelation som uppfyller (3.1) kallas en *ordnad kropp*, och alltså är $\mathbb{Q}$ alltså en ordnad kropp.

4. Reella tal

I förra avsnittet fann vi att $\mathbb{Q}$ är en ordnad kropp. En brist i $\mathbb{Q}$ är att inte varje uppåt begränsad mängd har ett supremum. Vi påminner om att ett tal som är större eller lika med alla talen i en mängd M av tal kallas en majorant till M. Om det finns en minsta majorant till M så kallas denna supremum av M. Det är lätt att se att mängden $\{x \in \mathbb{Q};\ x^2 < 2\}$ inte har någon minsta rationell majorant.

Uppgift 4.1. Kontrollera detta genom att visa att om r vore en minsta rationell majorant till denna mängd, så skulle $r^2 = 2$.

Sats 4.1. *Det finns en ordnad kropp $\mathbb{R}$ som innehåller $\mathbb{Q}$ som ordnad delkropp, sådan att följande är uppfyllt:*
(i) Om $x, y \in \mathbb{R}$ och $x < y$ så finns $r \in \mathbb{Q}$ sådant att $x < r < y$.
(ii) Det finns inget största eller minsta element i $\mathbb{R}$.
(iii) Varje icketom uppåt begränsad delmängd till $\mathbb{R}$ har en minsta majorant, dvs ett supremum.

Notera att $\mathbb{Q}$ själv uppfyller de två första villkoren i sats 4.1 men inte villkor (iii). Detta villkor är vad som i många analysböcker kallas *supremumaxiomet* eller *supremumegenskapen*.

Man kan konstruera utvidgningen på flera olika sätt men det visar sig att alla är isomorfa, se sats 4.15. Ett vanligt sätt är att använda s.k. *Dedekindsnitt*, se beviset av sats 4.15, men här ska vi istället använda Cauchyföljder. Dels är detta naturligt eftersom det stämmer överens med att vi ofta ser på reella tal som oändliga decimalutvecklingar, dels är det en typ av konstruktion som används i många andra sammanhang i matematik.

DEFINITION 4.2. En följd $(a_k) = (a_0, a_1, \ldots)$ av rationella tal är en *rationell Cauchyföljd* (vi skriver då $(a_k) \in RC$) om det till varje (rationellt) $\epsilon > 0$ finns ett (naturligt tal) N sådant att $|a_k - a_j| < \epsilon$ om $j, k \geq N$.

EXEMPEL 4.3. Låt $a_0 = 3$, $a_1 = 3,1$, $a_2 = 3,14$, $a_3 = 3,141$ etc. Då är $(a_k) \in RC$ för om $j, k \geq N$ så är $|a_k - a_j| < 10^{-N}$. På samma sätt ser man att varje decimalutveckling är ett element i RC. $\square$

EXEMPEL 4.4. Antag att (a_k) är en rationell talföljd och att $a_k \to a$ då $k \to \infty$ för något $a \in \mathbb{Q}$. Detta betyder definitionsmässigt att det till varje (rationellt) $\epsilon > 0$ finns ett N sådant att $|a_k - a| < \epsilon$ om $k \geq N$.

Men då följer att $|a_k - a_j| \leq |a_k - a| + |a - a_j| < \epsilon + \epsilon = 2\epsilon$ om $j, k \geq N$. Detta visar att (a_k) en Cauchyföljd. $\square$

Vi säger att en rationell följd (a_k) är en *nollföljd* om $a_k \to 0$ då $k \to \infty$. Speciellt är alltså nollföljder i RC.

EXEMPEL 4.5. Om $a_k = 1/k$ så är (a_k) en nollföljd. $\square$

ANMÄRKNING 4.6. Från analysen vet vi att en (rationell) följd (a_k) har ett (reellt) gränsvärde om och endast om den är en Cauchyföljd. Vår konstruktion går ut på att definiera reella tal som gränsvärden av följder i RC.

LEMMA 4.7. *(i) Om $(a_k) \in RC$ så är (a_k) begränsad, dvs det finns ett tal M sådant att $|a_k| \leq M$ för alla k.*
(ii) Om $(a_k) \in RC$, $(b_k) \in RC$ och $\alpha \in \mathbb{Q}$ så är (αa_k), $(a_k + b_k)$ och $(a_k b_k)$ i RC.
(iii) Om en av (a_k) och (b_k) är en nollföljd så är $(a_k b_k)$ en nollföljd, och om båda är nollföljder så är $(a_k + b_k)$ en nollföljd.

BEVIS. (i): Om $(a_k) \in RC$ så finns ett N sådant att $|a_k - a_j| < \epsilon = 1$ om $k, j \geq N$. Men då är $|a_k| \leq |a_N| + 1$ för alla $k \geq N$. Sätt $m = \max(|a_0|, |a_1|, \ldots, |a_N|)$. Då är $|a_k| \leq m + 1$ för alla k.

(ii): Antag att (a_k) och (b_k) är i RC. Givet $\epsilon > 0$ finns därför N_1 sådant att $|a_k - a_j| < \epsilon/2$ om $j, k \geq N_1$ och N_2 sådant att $|b_k - b_j| < \epsilon/2$ om $j, k \geq N_2$. Alltså har vi att

$$|(a_k + b_k) - (a_j + b_j)| \leq |a_k - a_j| + |b_k - b_j| < \epsilon/2 + \epsilon/2 = \epsilon$$

om $j, k \geq N = \max(N_1, N_2)$. Detta visar att $(a_k + b_k)$ är en Cauchyföljd.

Enligt (i) så finns ett tal M sådant att $|a_k| \leq M$ och $|b_k| \leq M$ för alla k. Till $\epsilon > 0$ finns nu (precis som i förra argumentet) ett N sådant att $|a_k - a_j| < \epsilon/2M$ och $|b_k - b_j| < \epsilon/2M$ om $j, k \geq N$. Men då är

$$\begin{aligned} |a_k b_k - a_j b_j| &= |a_k b_k - a_k b_j + a_k b_j - a_j b_j| \leq \\ &\leq |a_k||b_k - b_j| + |a_k - a_j||b_j| \leq M\epsilon/2M + M\epsilon/2M = \epsilon \end{aligned}$$

om $j, k \geq N$, vilket visar att $(a_k b_k)$ är en Cauchyföljd. Det återstående påståendet i (ii) visas på liknade sätt. Del (iii) lämnas som övning. $\square$

Uppgift 4.2. Bevisa resten av (ii) samt del (iii) av lemmat.

Enligt Lemma 4.7 så är $(a_k - b_k)$ i RC om (a_k) och (b_k) är i RC, och vi inför en relation $\sim$ på RC genom att $(a_k) \sim (b_k)$ om och endast om $(a_k - b_k)$ är en nollföljd. Man kollar lätt att $\sim$ är en ekvivalensrelation. Symmetrin och reflexiviteten är (nästan) triviala. Om $(a_k - b_k)$ och $(b_k - c_k)$ är nollföljder, så är $(a_k - c_k) = (a_k - b_k) + (b_k - c_k)$ en nollföljd enligt lemmat, vilket visar att $\sim$ är reflexiv.

Vi låter nu $\mathbb{R}$, de reella talen, vara mängden av ekvivalensklasser i RC. Ekvivalensklassen för $(a_k) \in RC$ betecknar vi med $[a_k]$. Vi definierar $+$ och $\cdot$ på $\mathbb{R}$ genom

$$[a_k] + [b_k] = [a_k + b_k], \quad [a_k][b_k] = [a_k b_k].$$

Vi måste visa att operationerna är väldefinierade. Antag därför att $(a'_k) \sim (a_k)$ och $(b'_k) \sim (b_k)$. Vi måste då visa att

$$(a_k + b_k) \sim (a'_k + b'_k) \quad \text{och} \quad (a_k b_k) \sim (a'_k b'_k). \tag{4.1}$$

Vi nöjer oss med att verifiera den andra likheten i (4.1) och lämnar den första som övning. Vi vill visa att $(a_k b_k - a'_k b'_k)$ är en nollföljd. Men detta följer genast av lemmat eftersom $a_k b_k - a'_k b'_k = a_k(b_k - b'_k) + (a_k - a'_k)b'_k$.

Antag nu att $a \in \mathbb{Q}$. Följden $(a) = (a, a, a, \ldots)$ är då uppenbarligen i RC och definierar alltså ett element $[a]$ i $\mathbb{R}$. Om $a, b \in \mathbb{Q}$ och $a \neq b$ så är $(a - b)$ inte en nollföljd och alltså är $[a] \neq [b]$. Speciellt har vi element $[0]$ och $[1]$ i $\mathbb{R}$. Observera att (a_k) är en nollföljd om och endast om $[a_k] = [0]$. Det är klart att räknereglerna (2.1) till (2.4) är uppfyllda för $\mathbb{R}$. Detta följer omedelbart av definitionen av $+$ och $\cdot$ samt av att de gäller för $\mathbb{Q}$. Ekvationen $[a_k] + x = [b_k]$ har lösningen $x = [b_k - a_k]$, så $\mathbb{R}$

är en kommutativ ring med nollelement $[0]$. Eftersom $[a+b] = [a]+[b]$ och $[ab] = [a][b]$ så är avbildningen $\mathbb{Q} \to \mathbb{R}$ en injektiv (ring)homorfism, en inbäddning, dvs vi kan betrakta $\mathbb{Q}$ som en delring till $\mathbb{R}$.

PROPOSITION 4.8. *Ekvationen* $[a_k]x = [b_k]$ *har lösning i* $\mathbb{R}$ *om* $[a_k] \neq 0$.

Alltså är $\mathbb{R}$ en kropp som har $\mathbb{Q}$ som delkropp. För beviset av Proposition 4.8 behöver vi följande enkla lemma.

LEMMA 4.9. *Antag att* $(a_k) \in RC$ *inte är en nollföljd. Då finns ett positivt rationellt tal* q *och ett naturligt tal* N *sådana att* $|a_k| \geq q$ *för alla* $k \geq N$.

BEVIS. Att (a_k) inte är en nollföljd betyder att det finns ett tal $\epsilon > 0$ sådant att det till varje M finns $k \geq M$ sådant att $|a_k| \geq \epsilon$. Eftersom (a_k) är en Cauchyföljd så finns ett N sådant att $|a_j - a_k| < \epsilon/2$ för $j, k \geq N$. För ett godtyckligt $j \geq N$ har vi alltså att

$$|a_j| = |a_k + a_j - a_k| \geq |a_k| - |a_j - a_k| \geq \epsilon - \epsilon/2 = \epsilon/2.$$

Alltså följer lemmat om man tar $q = \epsilon/2$. □

BEVIS AV PROPOSITION 4.8. Enligt lemmat finns q och N sådana att $|a_k| \geq q$ om $k \geq N$. Definiera nu en ny talföljd a'_k genom att sätta $a_k = q$ för $k < N$ och $a'_k = a_k$ för $k \geq N$. Då är $(a'_k) \sim (a_k)$ eftersom $a'_k - a_k$ är noll för alla $k \geq N$. Definiera nu (c_k) genom att sätta $c_k = 1/a'_k$. Vi påstår att $(c_k) \in RC$. Tag ett $\epsilon > 0$. Eftersom a'_k är en Cauchyföljd så finns M sådant att $|a'_j - a'_k| < q^2\epsilon$ om $k, j \geq M$. Vi kan anta att $M \geq N$. Men för $k, j \geq M$ har vi då att

$$|c_j - c_k| = \left| \frac{a'_k - a'_j}{a'_j a'_k} \right| < \frac{q^2\epsilon}{q^2} = \epsilon,$$

vilket alltså visar att $(c_k) \in RC$. Nu är $[a_k][c_k] = [a'_k][c_k] = [a'_k c_k] = [1]$ och alltså löser $x = [c_k b_k]$ ekvationen. □

Vi ska nu införa en ordningsrelation på $\mathbb{R}$. Antag att $(a_k) \in RC$. Vi säger att $(a_k) > 0$ om det finns ett rationellt tal $q > 0$ och ett naturligt tal N sådana att $a_k \geq q$ för alla $k \geq N$. Vi säger att $(a_k) < 0$ om $(-a_k) > 0$.

UPPGIFT 4.3. Visa att om $(a_k) > 0$ och $(a'_k) \sim (a_k)$ så $(a'_k) > 0$.

Om $\alpha \in \mathbb{R}$ så säger vi att $\alpha > 0$ om $\alpha = [a_k]$ och $(a_k) > 0$. Enligt övning 4.3 är detta oberoende av valet av representant (a_k) för α.

PROPOSITION 4.10. *För* $(a_k) \in RC$ *gäller exakt en av* $(a_k) \sim (0)$, $(a_k) > 0$ *eller* $(a_k) < 0$.

Bevis. Det är klart att de är uteslutande. Antag att (a_k) inte är en nollföljd. Enligt Lemma 4.9 finns då $q > 0$ och N sådana att $|a_k| \geq q$ för $k \geq N$. Eftersom $(a_k) \in RC$ så finns $M \geq N$ sådant att $|a_M - a_j| < q$ för $j \geq M$. Om $a_M \geq q$ så måste därför $a_j \geq q$ för alla $j \geq M$ och om $a_m < -q$ så är $a_j < -q$ för alla $j \geq M$. □

För varje $\alpha \in R$ gäller alltså precis ett av $\alpha < 0$, $\alpha = 0$ eller $\alpha > 0$.

Definition 4.11. Om $\alpha, \beta \in \mathbb{R}$ så säger vi att $\alpha \leq \beta$ om $\beta - \alpha \geq 0$.

Det följer att $\leq$ är en total ordning(srelation) på $\mathbb{R}$.

Notera att om α är rationellt, dvs $\alpha = [a, a, \ldots]$ där $a \in \mathbb{Q}$, så är $\alpha > 0$ om och endast om $a > 0$, så för rationella tal stämmer vår nya relation $\leq$ överens med den vanliga. Man kollar också lätt att $\alpha + \beta \geq 0$ och $\alpha\beta \geq 0$ om $\alpha, \beta \geq 0$.

Uppgift 4.4. Gör detta!

Alltså har vi att

Proposition 4.12. *De reella talen* $\mathbb{R}$ *är en ordnad kropp med* $\mathbb{Q}$ *som ordnad delkropp.*

Vi kan nu införa konvergensbegreppet för reella tal. Om α och α_k är en följd av reella tal så säger vi att $\alpha_k \to \alpha$ (då $k \to \infty$) om det till varje (rationellt) $\epsilon > 0$ finns ett naturligt tal N sådant att $|\alpha - \alpha_k| < \epsilon$ för alla $k \geq N$.

Från inledningen kommer vi ihåg att själva idén i konstruktionen var att se till att varje $(a_k) \in RC$ skulle få ett gränsvärde genom att definiera detta tal som klassen $[a_k]$ av alla ekvivalenta följder. Nästa proposition säger att vi har lyckats i detta uppsåt.

Proposition 4.13. *Tag en* $(a_k) \in RC$. *Om* $\alpha_k = [a_k, a_k, \ldots]$ *och* $\alpha = [a_0, a_1, a_2, \ldots]$ *så gäller att* $\alpha_k \to \alpha$.

Enkelt uttryckt innebär detta att varje rationell Cauchyföljd (a_k), betraktad som en följd av reella tal, har ett reellt gränsvärde, nämligen det reella tal α som följden (a_k) representerar.

Bevis. Givet ett rationellt $\epsilon > 0$, tag N sådant att $|a_j - a_k| < \epsilon$ för $j, k \geq N$. Tag ett fixt $n \geq N$ och definiera talföljden (b_k) genom

$$b_k = a_n - a_k + 2\epsilon.$$

Låt $\beta = [b_0, b_1, \ldots]$. Eftersom $b_k \geq \epsilon$ för $k \geq N$ så är $\beta > 0$. Men $\beta = \alpha_n - \alpha + 2\epsilon$, (här betecknar alltså ϵ motsvarande reella tal om man ska vara petig) så $\alpha_n - \alpha + 2\epsilon > 0$ dvs $\alpha - \alpha_n < 2\epsilon$. På liknande sätt visar man att $\alpha_n - \alpha < 2\epsilon$. Alltså har vi att

$$|\alpha_n - \alpha| < 2\epsilon \quad \text{om} \quad n \geq N.$$

Eftersom ϵ var godtyckligt betyder detta $\alpha_n \to \alpha$. $\square$

I fortsättningen skiljer vi inte på rationella tal och motsvarande tal i $\mathbb{R}$, utan vi skriver istället $a_n \to \alpha$.

För att slutgiltigt kunna bevisa sats 4.1 behöver vi ett litet lemma till.

LEMMA 4.14. *Om* $\alpha \in \mathbb{R}$ *och* $0 \leq \alpha < 1/m$ *för alla* $m \in \mathbb{N}$ *så är* $\alpha = 0$.

BEVIS. Antag att $\alpha > 0$ och att $\alpha = [a_0, a_1, \ldots]$. Då finns rationellt $q > 0$ och N sådana att $a_k \geq q$ för alla $k \geq N$. Eftersom $1/m - \alpha > 0$ (och $1/m - \alpha$ representeras av följden $(1/m - a_k)$) så finns $q' > 0$ och N' sådant att $1/m - a_k \geq q' > 0$ för alla $k \geq N'$. Om $k \geq \max(N, N')$ så är alltså

$$q \leq a_k \leq 1/m - q' < 1/m,$$

dvs $q < 1/m$ vilket är en motsägelse om m är stort. $\square$

BEVIS AV SATS 4.1. Det återstår att visa att $\mathbb{R}$ uppfyller (i), (ii) och (iii). Antag att x och y är reella tal, $x \leq y$ och att det inte finns något rationellt tal mellan dem. Fixera ett $m \in \mathbb{N}$. Låt c vara det största heltalet som är mindre än eller lika med x och låt j vara det största heltalet sådant att $c + j/m \leq x$. Enligt antagandet har vi då att $c + j/m \leq x \leq y \leq c + (j+1)/m$. Alltså är $0 \leq y - x \leq 1/m$. Eftersom m är godtyckligt så följer det av Lemma 4.14 att $x = y$. Om $x < y$ så finns alltså något rationellt tal mellan dem. Detta visar (i).

Eftersom varje Cauchyföljd är begränsad, så finns till varje reellt tal ett heltal som är större (och något annat som är mindre). Alltså följer (ii).

Antag nu att $M \subset \mathbb{R}$ är uppåt begränsad. Låt för varje n a_n vara det största tal av typen $p2^{-n}$, där p är ett heltal, sådant att a_n inte är en majorant till M. (Ett sådant största tal finns eftersom M är uppåt begränsad.) Då är $a_n + 2^{-n}$ en majorant till M. Observera nu att $a_n \leq a_{n+1}$, för antingen är $a_{n+1} = a_n$ eller så är $a_{n+1} = a_n + 2^{-(n+1)}$. Alltså har vi att

$$a_n \leq a_{n+1} \leq \ldots \leq a_{n+m} < a_n + 2^{-n}.$$

Den sista olikheten beror på att a_{n+m} inte är en majorant, medan $a_n + 2^{-n}$ är det. Alltså är $|a_j - a_k| < 2^{-n}$ om $j, k \geq n$ så $(a_k) \in RC$. Enligt Proposition 4.13 finns $\alpha \in \mathbb{R}$ sådant att $a_k \to \alpha$. Men även $a_n + 2^{-n} \to \alpha$. Eftersom varje $a_n + 2^{-n}$ är en majorant så är även α en majorant. Eftersom å andra sidan ingen a_n är en majorant, och $a_n \to \alpha$ så kan det inte heller finnas någon mindre majorant än α. Alltså har M en minsta majorant. $\square$

De egenskaper vi har visat om vår utvidgning $\mathbb{R}$ av $\mathbb{Q}$ bestämmer utvidgningen så när som på isomorfi.

SATS 4.15. *Antag att H är en ordnad kropp som innehåller $\mathbb{Q}$ som ordnad delkropp och som uppfyller (i), (ii) och (iii). Då finns en ordningsbevarande kroppsisomorfi $\Phi\colon \mathbb{R} \to H$.*

BEVISSKISS. En delmängd M till $\mathbb{Q}$ kallas ett *snitt*, eller ett *Dedekindsnitt*, om

- (1) $M \neq \mathbb{Q}$ och $M \neq \emptyset$
- (2) $a \in M$ och $b < a$ medför att $b \in M$
- (3) M har inget största element.

Man inser lätt att för varje reellt tal α så är mängden $M = \{r \in Q;\ r < \alpha\}$ ett snitt. Beviset går ut på att visa att om R är en ordnad kropp som innehåller $\mathbb{Q}$ och sådan att (i), (ii) och (iii) gäller, så finns en bijektion mellan elementen i R och mängden av snitt av $\mathbb{Q}$. På detta sätt får man en bijektion mellan R och $\mathbb{R}$. Vi hoppar över detaljerna! □

Vi avslutar detta avsnitt med ett par kända resultat om reella tal.

SATS 4.16. *Antag att (a_n) är en reell Cauchyföljd. Då finns ett reellt tal a sådant att $a_n \to a$ då $n \to \infty$.*

Definitionerna av Cauchyföljd och konvergens är bokstavligen desamma som för rationella följder.

BEVIS. Till varje a_n finns ett rationellt tal a'_n sådant att $|a_n - a'_n| < 1/n$. Eftersom (a_n) är en Cauchyföljd och $a_n - a'_n \to 0$ så följer det att även (a'_n) är en Cauchyföljd. (Givet $\epsilon > 0$ så finns N sådant att $|a_k - a_j| < \epsilon/3$ om $j, k \geq N$. Men det finns även N' sådant att $|a_k - a'_k| < \epsilon/3$ om $k \geq N'$. Alltså är $|a'_k - a'_j| < \epsilon$ om $j, k \geq \max(N, N')$.) Enligt proposition 4.13 finns därför $a \in \mathbb{R}$ sådant att $a'_n \to a$. Men då följer att även $a_n \to a$. □

SATS 4.17 (Bolzano-Weierstrass sats). *Antag att (a_n) är en begränsad reell talföljd. Då finns en delföljd (b_j), där $b_j = a_{k_j}$, som är konvergent.*

BEVIS. Antag att alla a_k ligger i intervallet I_0 med längd C. Dela nu detta i två lika stora delintervall. Åtminstone i ett av dessa delintervall måste det finnas oändligt många av talen a_k. Kalla detta I_1 och dela det i två lika stora delintervall. I ett av dessa, kalla det I_2, finns oändligt många a_k etc. Tag till varje j ett tal $b_j = a_{k_j}$ i I_j. Det följer att $|b_j - b_i| \leq 2^{-j}C$ om $i \geq j$, så b_j är en Cauchyföljd och alltså konvergent enligt sats 4.16. □

Uppgift 4.5. Visa att det för varje ickenegativt reellt tal r finns ett entydigt ickenegativt reellt tal $s = \sqrt{r}$ sådant att $s^2 = r$. Ledning: Betrakta $\{x \geq 0;\ x^2 \leq r\}$.

Uppgift 4.6. Antag att $a_j \geq 0$ och att $a_k \to a$. Visa att $\sqrt{a_k} \to \sqrt{a}$.

5. Komplexa tal

På $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$ inför vi operationerna $+$ och $\cdot$ genom att sätta $(a,b)+(b,c) = (a+c, b+d)$ och $(a,b)(c,d) = (ac-bd, ad+bc)$. Man visar lätt att detta blir en kropp med $\mathbb{R}$ som en delkropp om vi identifierar $r \in \mathbb{R}$ med $(r,0) \in \mathbb{R}^2$. Vi kallar $\mathbb{R}^2$ med dessa operationer för $\mathbb{C}$, de *komplexa talen*. Om man inför beteckningen i för $(0,1)$ så finner man att $i^2 = -1$. Varje komplext tal z kan skrivas entydigt som $z = a + ib$ där a och b är reella tal. Dessa kallas *real-* respektive *imaginärdelen* av z. Om $z = a + ib$ så kallas $\bar{z} = a - ib$ för konjugatet till z. Till ett komplext tal $z = a + ib$ definierar vi dess absolutbelopp $|z|$ som det ickenegativa reella tal (jämför övning 4.5) sådant att $|z|^2 = z\bar{z} = a^2+b^2$. Vi lämnar som övning att verifiera triangelolikheten

$$|z + w| \leq |z| + |w|.$$

Ur denna följer även att $|w|-|z| \leq |z-w|$. Vi påminner också om polär framställning av komplexa tal. Med beteckningen $e^{i\theta} = \cos\theta + i\sin\theta$ kan varje komplext tal skrivas som $z = re^{i\theta}$, där $r = |z|$ och θ är bestämt så när som på en heltalsmultipel av 2π. Konsultera i nödfall någon elementär bok i algebra.

Vi definierar konvergens för en följd av komplexa tal z_j på det uppenbara sättet. Om $z_j = a_j + ib_j$ så $z_j \to z = a + ib$ omm $a_j \to a$ och $b_j \to b$. Detta är förstås detsamma som att $|z_j - z| \to 0$ eftersom $|z_j - z|^2 = (a_j - a)^2 + (b_j - b)^2$. Notera också att om $z_j \to z$ så $|z_j| \to |z|$ och om dessutom $w_j \to w$ så $z_j + w_j \to z + w$ och $z_j w_j \to zw$. Detta bevisas precis som för rationella eller reella talföljder.

Vi ska avrunda detta kapitel med att bevisa *algebrans fundamentalsats*. Ett uttryck $p(z) = a_0 + a_1 z + a_2 z^2 + \cdots + a_n z^n$, där $a_k \in \mathbb{C}$, kallas ett komplext polynom. Notera att om $z_j \to z$ så följer att $p(z_j) \to p(z)$. Detta följer genast av att summor och produkter bevarar konvergens. Vidare $|p(z_j)| \to |p(z)|$.

Sats 5.1 (Algebrans fundamentalsats). *Antag att $p(z)$ är ett komplext polynom av grad ≥ 1. Då finns något z sådant att $p(z) = 0$.*

Sats 5.2 (Bolzano–Weierstrass sats). *Antag att* z_k *är en följd av komplexa tal i en begränsad mängd. Då finns en delföljd* z_{k_ℓ} *som är konvergent.*

Bevis. Antag $z_k = a_k + ib_k$. Att z_k är begränsad medför att både a_k och b_k är begränsade reella talföljder, och alltså finns en delföljd a_{k_j} som är konvergent enligt sats 4.17. Men b_{k_j} är en begränsad följd så det finns en delföljd $b_{k_{j_\ell}}$ som är konvergent. Eftersom en delföljd till en konvergent följd är konvergent, så följer att $w_\ell = a_{k_{j_\ell}} + ib_{k_{j_\ell}}$ är en konvergent delföljd till z_k. □

Bevis av sats 5.1. Notera att $\{|p(z)|;\ z \in \mathbb{C}\}$ är en nedåt begränsad mängd av reella tal, så $\beta = \inf\{|p(z)|;\ z \in \mathbb{C}\}$ existerar och $\beta \geq 0$. Om $p(z) = a_n z^n + a_{n-1}z^{n-1} + \dots$ så är

$$
\begin{aligned}
|p(z)| \geq |a_n||z|^n - |a_{n-1}||z|^{n-1} - \dots - |a_0| \geq \\
\geq |z|^n \left(|a_n| - \left(\frac{1}{|z|}|a_{n-1}| + \dots + \frac{1}{|z|^n}|a_0| \right) \right).
\end{aligned}
$$

Om $a_n \neq 0$ så finns därför ett R sådant att $|p(z)| \geq 2\beta$ om $|z| \geq R$. Enligt definitionen av β så finns det en följd z_k sådan att $|f(z_k)| \to \beta$ då $k \to \infty$. Eftersom $|z_k| \leq R$ för alla (stora) k, så finns enligt sats 5.2 en konvergent delföljd $w_j = z_{k_j} \to a$. Men då har vi att $|p(w_j)| \to |p(a)|$ och eftersom $|p(w_j)|$ är en delföljd till $|p(z_k)|$ så följer att $|p(w_j)| \to \beta$, och alltså $|p(a)| = \beta$.

Om $\beta = 0$ så är vi klara. Låt oss därför anta att $\beta > 0$. Genom att ersätta $p(z)$ med $q(z) = (1/\beta)p(z+a)$ kan vi anta att $\beta = 1$ och $a = 0$. Alltså är

$$p(z) = 1 + a_m z^m + a_{m+1}z^{m+1} + \dots + a_{m+k}z^{m+k}.$$

Genom att byta ut z mot αz för lämpligt α (jämför övning 6.3), kan vi anta att $a_m = -1$, dvs

$$p(z) = 1 - z^m + a_{m+1}z^{m+1} + \dots a_{m+k}z^{m+k}.$$

För små positiva reella x har vi nu med triangelolikheten att

$$
\begin{aligned}
|p(x)| \leq 1 - x^m + |a_{m+1}|x^{m+1} + \dots + |a_{m+n}|x^{m+n} = \\
= 1 - x^m(1 - |a_{m+1}|x - \dots - |a_{m+n}|x^{m+n}),
\end{aligned}
$$

och detta är strikt mindre än 1 för små positiva x, eftersom uttrycket i parentesen då är strikt positivt. Detta visar att det finns x sådana att $|p(x)| < 1 = \beta$ vilket strider mot definitionen av β. Alltså måste $\beta = 0$ och vi är klara. □

6. Ytterligare övningar till kapitel 3

UPPGIFT 6.1. Visa att om a_n är en växande uppåt begränsad följd av reella tal så är den konvergent.

UPPGIFT 6.2. Antag att man vet att kroppen $\mathbb{Q}$ finns. Fundera på hur man kan förklara för säg en högstadieelev att

$$\frac{p}{q}\frac{p'}{q'} = \frac{pp'}{qq'}.$$

UPPGIFT 6.3. Visa att det till givet $w \in \mathbb{C}$ finns $z \in \mathbb{C}$ sådant att $z^m = w$.

UPPGIFT 6.4. Antag att $f(x)$ är en reellvärd kontinuerlig funktion på det slutna intervallet $[0,1]$ och att $f(0) = -1$ och $f(1) = 1$. Visa att det finns ett $\xi \in [0,1]$ sådant att $f(\xi) = 0$.

Några kommentarer

I **[EM]** finns en grundlig framställning av Gödels konstruktion av oavgörbara utsagor och hur detta medför att man inte kan bevisa att teorin T är konsistent. Konstruktionen av de reella talen med hjälp av Dedekindsnitt finns beskriven i **[OH]**.

KAPITEL 4

Geometri

Sedan mycket länge har man känt till och utnyttjat geometri. Exempelvis kände redan babylonierna mer än 1600 år f Kr till Pythagoras sats på empiriska grunder. Den deduktiva metoden, dvs att utgående från vissa förhoppningsvis självklara påståenden, postulat eller axiom, genom logisk argumentation dra nya slutledningar, började utvecklas under 600-talet f Kr av grekiska matematiker. Det definitiva genombrottet kom när Euklides omkring 300 f Kr fullbordade sitt epokgörande arbete Elementa, som innehåller 465 satser om geometri, talteori och algebra, alla härledda ur några få axiom.

Euklides införde ett axiomsystem för geometri som i modernt språkbruk brukar kallas *absolut geometri*, vilket vi beteckar A, samt det s.k. parallellaxiomet PA. I den vanliga modellen för A, nämligen geometri i planet, är PA också sann. Under mer än två tusen år försökte man utan framgång bevisa PA från de övriga axiomen, ända tills man på 1800-talet visade att PA är oberoende av de övriga, genom att hitta modeller till A där PA är falskt, s.k. ickeeuklidiska modeller. För att konstruera en sådan modell måste man anlägga ett helt axiomatiskt synsätt, och frigöra sig från den intuitiva idén om vad grundbegreppen *punkt* och *linje* i A ska betyda; de behöver inte tolkas som punkter och linjer i "vanlig" mening, utan bara på ett sådant sätt att axiomen i A är uppfyllda.

1. Axiom för absolut geometri

I beskrivningen av teorin för absolut geometri, A, använder vi modernt språkbruk, med begrepp som reella tal, avbildningar, grupp etc. Den som är intresserad av en framställning som mer direkt ansluter sig till Euklides egen kan t ex läsa i [**LÅL**].

Begreppen i teorin A är:

- En mängd av element kallade *punkter*.
- En mängd av element kallade *linjer*.

- Relationen $\in$ mellan mängden av punkter och mängden av linjer. Vi säger att punkten p *ligger på* linjen ℓ eller att ℓ *går genom* p om $p \in \ell$.
- På mängden av punkter på en linje har vi två totala ordningsrelationer, varav den ena är den motsatta till den andra,
- En mängd av bijektiva avbildningar $\mathcal{K}$ av mängden av punkter på sig själv, kallade *kongruensavbildningar* eller *kongruenstransformationer*.

Dessa ska uppfylla följande axiom:

A1 För två olika punkter finns en och endast en linje genom punkterna.

Linjen genom punkterna a och b kallar vi linjen ab. Det följer att två olika linjer kan ha högst en skärningspunkt. Två linjer som inte skär varandra sägs vara *parallella*.

A2 Mängden av punkter på en linje är, med endera ordningsrelationen, ordningsisomorf med den ordnade mängden $\mathbb{R}$ av reella tal.

Det följer att varje linje går genom oändligt många punkter. Antag att $a \in \ell$. På grund av A2 kan vi tala om att två punkter på ℓ ligger på samma sida om a. Mängden av alla punkter som ligger på en och samma sida om a (inklusive a själv) kallas en *stråle* från (eller med utgångspunkt i) a. De två strålarna från a som ligger i samma linje kallas *motsatta*. Den av dem som innehåller $b \in \ell$ kallas strålen ab.

Av tre olika punkter på en linje så ligger alltid en av dem *mellan* de två andra. Mängden av punkter på linjen ab som ligger mellan a och b (inklusive a och b själva) kallas sträckan ab, och a och b kallas sträckans *ändpunkter*.

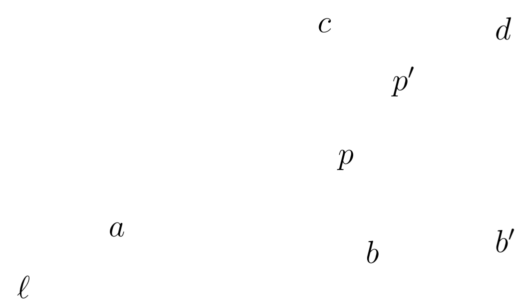
A3 Mängden av de punkter som inte ligger på linjen ℓ kan indelas i två icketomma mängder kallade *sidorna* av ℓ. Två punkter utanför ℓ ligger på samma sida av ℓ om och endast om sträckan mellan dem inte skär ℓ.

Det följer att relationen "ligga på samma sida som" är en ekvivalensrelation. Antag att två olika linjer ℓ och ℓ' skär varandra i a, och antag att p och q är två punkter på ℓ som båda är skilda från a. Då ligger p och q på samma stråle från a om och endast om de ligger på samma sida om ℓ'.

A4 Mängden av kongruenstransformationer $\mathcal{K}$ bildar en grupp under sammansättning; speciellt har varje $\phi \in \mathcal{K}$ sin invers i $\mathcal{K}$. En $\phi \in \mathcal{K}$ avbildar linjer på linjer och bevarar (eller omkastar) ordningsrelationen på linjer. Givet tre punkter a, b, c som inte alla ligger på samma linje, och tre andra punkter a', b', c' som inte heller ligger på samma linje, så finns det en och endast en $\phi \in \mathcal{K}$ sådan att $\phi(a) = a'$, strålen ab avbildas på strålen $a'b'$ samt sidan (som bestäms av) c av linjen ab avbildas på sidan c' av linjen $a'b'$.

Man säger att två punktmängder är *kongruenta* om det finns någon $\phi \in \mathcal{K}$ som tar den ena mängden på den andra. Eftersom $\mathcal{K}$ är en grupp så är kongruens en ekvivalensrelation.

Vi ska nu definiera begreppet *vinkel*. Fixera en punkt a och betrakta två strålar ab och ac från a. Låt vinkeln $\wedge bac$ vara mängden av alla strålar från a som skär sträckan bc. Alltså är $\wedge bac$ samma som $\wedge cab$. Vi påstår att denna mängd bara beror på strålarna ab och ac och inte på valet av punkter b och c på dessa. Låt ad vara en stråle från a som skär sträckan bc i, säg, punkten p. Det räcker att visa att om b' är en annan punkt på strålen ab så skär strålen ad även sträckan $b'c$. Låt ℓ vara linjen ad. Eftersom denna skär linjen ab i a så följer det, jmf kommentaren efter A3 ovan, att b och b' ligger på samma sida om ℓ. Eftersom sträckan bc skär ℓ så ligger b och c på olika sidor om ℓ och följaktligen ligger även b' och c på olika sidor om ℓ, och därför skär sträckan $b'c$ linjen ℓ i en punkt p'. Vi vill nu se att p' ligger på strålen ad. Men p ligger på samma sida om linjen ab som c, vilken i sin tur ligger på samma sida som p'. Alltså ligger p och p' på samma sida om linjen ab och därför på samma stråle från a.



Strålarna ab och ac kallas *vinkelbenen* till vinkeln $\wedge bac$.

Mängden av strålar från a som ligger utanför en vinkel $\wedge bac$ (inklusive vinkelbenen) som ovan kallas också en vinkel vid a.

Av praktiska skäl ska vi lägga till ytterligare ett axiom. Det kan bevisas från de övriga, men är å andra sidan lätt att verifiera direkt i de modeller till teorin A som vi ska studera.

A5 Ingen sträcka är kongruent med en äkta delsträcka till sig själv, och ingen vinkel är kongruent med en äkta delvinkel till sig själv.

Vi kan nu definiera ytterligare några begrepp.

Spegling i linje. Antag att c inte ligger på linjen ab. Då finns exakt en $\phi \in \mathcal{K}$ som tar a på sig själv, strålen ab på sig själv och som avbildar c på en punkt c' på andra sidan av linjen ab. På grund av A5 måste ϕ ta b på sig själv, och eftersom b är en godtycklig punkt ($\neq a$) på strålen så måste ϕ hålla varje punkt på strålen fix. Eftersom den motsatta strålen också avbildas på sig själv följer på samma sätt att varje punkt på den hålls fix, dvs varje punkt på linjen ab hålls fix av ϕ. Denna ϕ kallas *spegling* i linjen ab.

Mittpunkt av sträcka. Betrakta sträckan ab och välj en punkt c som inte ligger på linjen ab. Tag den entydiga $\phi \in \mathcal{K}$ sådan att $a \mapsto b$, strålen ab går på strålen ba och c avbildas på en punkt c' på andra sidan om linjen ab. Enligt A5 kommer då b att avbildas på a. Därför avbildar $\phi \circ \phi$ punkten a på sig själv, sträckan ab på sträckan ab och c på samma sida, och därför är $\phi \circ \phi$ identitetsavbildningen på grund av entydigheten i A4. Låt p vara skärningen mellan sträckan cc' och linjen ab. Eftersom $c = \phi \circ \phi(c)$ följer det att $\phi(c') = c$, och följaktligen avbildas sträckan cc' på sig själv, så $\phi(p) = p$. Det följer att ap och pb är kongruenta, och enligt A5 är därför p en entydigt bestämd punkt på sträckan ab som kallas *mittpunkten* på sträckan.

Längd av sträckor. Vi ska nu införa ett längdbegrepp för sträckor sådant att två sträckor har samma längd om och endast om de är kongruenta, och sådant att

$$|ac| = |ab| + |bc|. \tag{1.1}$$

om b ligger mellan a och c på linjen ac.

Fixera först en sträcka I_0 och säg att den definitionsmässigt har längd 1. Om ab är en godtycklig sträcka, kan vi efter en kongruenstransformation anta att I_0 är sträckan aa_1 på strålen ab. Vi kan avsätta sträckan I_0 från punkten a_1 och få sträckan a_1a_2 på samma stråle etc. Vi påstår att varje punkt på strålen kommer att ligga på någon av dessa sträckor a_ka_{k+1}. Om inte, eftersom strålen ab är ordnad som $\mathbb{R}^+$, så existerar supremum m av alla punkter på strålen ab som ligger i någon av sträckorna a_ka_{k+1}. Avsätt I_0 från m mot a och säg att vi får sträckan md. Då ligger d på någon sträcka $a_{k-1}a_k$ och alltså ligger m på sträckan a_ka_{k+1}, vilket är en motsägelse.

När man stoppat in så många I_0 som får plats på sträckan ab, säg s stycken, så tar man om möjligt och stoppar in sträckan I_1 som är halva sträckan I_0. Man kan få in högst en sådan. Sedan tar man om möjligt och stoppar in I_2, som är halva I_1, etc. Vi definierar sedan längden $|ab|$ av sträckan ab som

$$|ab| = s + s_1 2^{-1} + s_2 2^{-2} + \dots,$$

där s_k är 1 eller 0 beroende på om sträckan I_k fick plats eller inte. Vi lämnar som övning att visa

Uppgift 1.1. Visa att varje (icketrivial) sträcka har positiv längd; att två sträckor är kongruenta om och endast om de har samma längd, och att (1.1) gäller.

En *cirkel* med *medelpunkt* m och *radie* r är mängden av alla punkter p som har avståndet r till punkten m. En linje genom medelpunkten till en cirkel skär cirkeln i exakt två punkter. Sträckan mellan dessa punkter kallas en *diameter* till cirkeln.

Uppgift 1.2. Visa att varje cirkel har en entydigt bestämd radie och medelpunkt.

Storlek på vinklar. Om ab och ac är motsatta strålar och ad är ytterligare en stråle så kallas $\wedge bad$ och $\wedge cad$ *sidovinklar* till varandra. En vinkel som är kongruent med sin sidovinkel kallas *rät*. Vi påstår att alla räta vinklar är kongruenta. Antag motsatsen. Efter en lämplig kongruenstransformation kan vi anta att strålarna ad_1 och ad_2 bildar räta vinklar med linjen ab, och att d_1 och d_2 båda ligger på samma sida om ab. Antag exempelvis att strålen ad_2 ligger i vinkeln $\wedge cad_1$. Vinkeln $\wedge cad_2$ är då en äkta delmängd av $\wedge cad_1$ som är kongruent med $\wedge bad_1$, vilken i sin tur är en del av $\wedge bad_2$. Men denna är kongruent med cad_2, så vi får att $\wedge cad_2$ är kongruent med en äkta delvinkel av sig själv, vilket motsäger A5.

d_2 d_1

c a b

Att det verkligen finns räta vinklar inses genom att spegla i en linje ℓ. En punkt p utanför linjen övergår då i en punkt q på andra sidan, och linjen pq skär linjen ℓ i vinklar som efter spegling övergår i sina sidovinklar och alltså måste vara räta.

Två linjer som skär varandra i räta vinklar sägs vara *normaler* till varandra. Genom en given punkt p utanför linjen ℓ finns som vi sett ovan alltid en linje som är normal till ℓ. Den är dessutom entydig. Om nämligen pa är normal till ℓ, där a ligger på ℓ, så övergår vid spegling pa i qa som också är normal till ℓ. Men då bildar strålarna ap och aq båda rät vinkel mot ℓ och är därför delar av samma linje, och alltså är a skärningen av pq och ℓ.

Betrakta vinkeln $\wedge bac$ vid a, eller dess komplementvinkel. Vi kan anta att sträckorna ab och ac är lika långa. Den $\phi \in \mathcal{K}$ som håller a fix och tar ab på ac och c i en punkt på samma sida om ac som b, måste ta b på c, eftersom ab och ac är lika långa, och c på b, av samma skäl samt av A5. Mittpunkten p på sträckan bc ligger då fix och linjen ap delar vinkeln $\wedge bac$ och dess komplementvinkel i två kongruenta vinklar. Linjen ap kallas *bisektris* till vinkeln bac (och dess komplementvinkel). Som för sträckor kan vi nu införa mått på vinklar, och vi normaliserar så att räta vinklar får måttet $\pi/2$.

Om två vinklar $\wedge abc$ och $\wedge a'b'c'$ är kongruenta, dvs lika stora, så skriver vi $\wedge abc = \wedge a'b'c'$.

Anmärkning 1.1. Den som störs av att vi här använder talet π utan att inom teorin ge en definition, kan tills vidare tänka på det som enbart en beteckning för två räta vinklar. Se vidare avsnitt 4. $\square$

En vinkel som är mindre än en rät, kallas *spetsig* och en som är större kallas *trubbig*.

När två linjer skär varandra i en punkt p uppstår fyra vinklar. Ett par av dessa som inte har något gemensamt vinkelben kallas *vertikalvinklar*. De är lika stora eftersom de har gemensam sidovinkel.

Först efter att ha bevisat så mycket han kunde från (motsvarigheten till) axiomen i A, dvs A1 till A4 (eller A5), så införde Euklides parallellaxiomet, vilket kan formuleras på följande vis:

Parallellaxiomet PA Givet en linje ℓ och en punkt p som inte ligger på ℓ, så finns det exakt en linje genom p som är parallell med ℓ.

Vi låter A+PA beteckna teorin vars axiom är A1 till A5 samt PA.

2. Den euklidiska modellen för absolut geometri

Vi ska nu presentera en modell för A där PA är sann. Senare, i avsnitt 4, ska vi att se att varje modell till A+PA är isomorf med denna modell, vilket berättigar oss att tala om *den* euklidiska modellen.

Som punktmängd tar vi $\mathbb{R}^2 = \{x = (x_1, x_2);\ x_1, x_2 \in \mathbb{R}\}$ och en *linje* låter vi vara en mängd av punkter som löser en ekvation av typ

$$a \cdot x + m = a_1x_1 + a_2x_2 + m = 0,$$

där $(0,0) \neq (a_1, a_2) = a \in \mathbb{R}^2$ och $m \in \mathbb{R}$. Att en punkt x *ligger på* linjen ℓ ska helt enkelt betyda att x är ett element i punktmängden ℓ i $\mathbb{R}^2$. En *kongruensavbildning* $\phi \in \mathcal{K}$ ska vara en avbildning $\phi : \mathbb{R}^2 \to \mathbb{R}^2$ på formen $\phi(x) = Ax + b$, där A är en ortogonal 2×2-matris (dvs $A^t A = I$), och $b \in \mathbb{R}^2$.

Vi ska nu verifiera att axiomen A1 t o m A5 är uppfyllda. Vi börjar med att notera att linjen $a_1 x_1 + a_2 x_2 + m = 0$ (där alltså a_1 eller a_2 är nollskild), är bilden av den injektiva avbildningen

$$\mathbb{R} \ni t \mapsto tv + c, \tag{2.1}$$

där $v = (a_2, -a_1)$ och $(a_1, a_2) \cdot c = -m$. En sådan avbildning kallas en (linjär) parametrisering av linjen. Omvänt är bilden av en avbildning (2.1), där $v \neq 0$, en linje.

A1: Genom två olika punkter a och b i $\mathbb{R}^2$ går linjen som ges av

$$t \mapsto t(b-a) + a = tb + (1-t)a.$$

Vi vill se att denna är entydig. Antag att linjen som ges av $s \mapsto sv + c$ också går genom a och b. Då är $a = s_0 v + c$ och $b = s_1 v + c$ för några tal s_0 och s_1 med $s_0 \neq s_1$, så

$$sv + c = t(b-a) + a$$

om $t(s_1 - s_0) + s_0 = s$. Men $t \mapsto s = t(s_1 - s_0) + s_0$ är en bijektion $\mathbb{R} \to \mathbb{R}$, vilket betyder att de båda parametriseringarna beskriver samma linje. Linjen genom a och b kan även erhållas på ekvationsform som

$$(x_1 - a_1)(b_2 - a_2) - (x_2 - a_2)(b_1 - a_1) = 0.$$

A2: En linje får sin ordningsrelation (och sin omvända ordningsrelation) genom en parametrisering (2.1). Den erhållna ordningsrelationen beror inte på valet av parametrisering för som ovan ser vi att två olika parametriseringar svarar mot ett koordinatbyte $s = \alpha t + \beta$ på $\mathbb{R}$, vilket bevarar (eller kastar om) ordningsrelationen på $\mathbb{R}$.

Om a och b är två olika punkter så parametriseras sträckan ab av $[0,1] \ni t \mapsto tb + (1-t)a = a + t(b-a)$.

A3: Om linjen ℓ ges av ekvationen $a \cdot x + m = 0$, så uppdelas $\mathbb{R}^2 \setminus \ell$ i de disjunkta mängderna $\ell^+ = \{x \in \mathbb{R}^2;\, a \cdot x + m > 0\}$ och $\ell^- = \{x \in \mathbb{R}^2;\, a \cdot x + m < 0\}$. Tag två punkter α och β utanför ℓ och sätt $A = a \cdot \alpha + m$ och $B = a \cdot \beta + m$. Antag först att $A \neq B$. Då är skärningen mellan linjen $t \mapsto t(\beta - \alpha) + \alpha$ och ℓ den punkt som svarar mot $t = A/(A - B)$, och detta tal t ligger i intervallet $(0,1)$ om och endast om A och B har olika tecken. Detta innebär att sträckan $\alpha\beta$ skär ℓ om och endast om α och β ligger på olika sidor om ℓ. Om $A = B$

ligger förstås α och β på samma sida om ℓ och linjen $\alpha\beta$ är parallell med ℓ, så sträckan $\alpha\beta$ skär inte inte ℓ.

A4 (och A5): Vi visar först att $\mathcal{K}$ är en grupp, dvs att varje $\phi \in \mathcal{K}$ har en invers ϕ^{-1} i $\mathcal{K}$ och att $\psi \circ \phi \in \mathcal{K}$ om ϕ och ψ ligger i $\mathcal{K}$. Om $\phi(x) = Ax + \alpha$ och $\psi(y) = By + \beta$ så är

$$\psi \circ \phi(x) = BAx + B\alpha + \beta = Cx + \gamma.$$

Nu är C ortogonal, ty

$$C^tC = (BA)^tBA = A^tB^tBA = A^tIA = A^tA = I,$$

så $\psi \circ \phi \in \mathcal{K}$. Vidare om ϕ är given och ψ väljs så att $B = A^t$ och $\beta = -A^t\alpha$ så blir $\psi \circ \phi(x) = \phi \circ \psi(x) = x$, dvs $\psi \circ \phi = \phi \circ \psi = I$. Det följer nu att $\mathcal{K}$ är en grupp ty A^t är ortogonal om A är det, enligt lemma 2.1 nedan.

En liknande kalkyl visar att linjer avbildas på linjer av $\phi \in K$ och att ordningsrelationen bevaras (eller kastas om).

Uppgift 2.1. Utför denna kalkyl!

Vi tittar nu lite närmare på avbildningarna i $\mathcal{K}$.

Lemma 2.1. *En 2×2-matris A är ortogonal om och endast om för några α, β med $\alpha^2 + \beta^2 = 1$ vi har*

$$A = \begin{pmatrix} \alpha & -\beta \\ \beta & \alpha \end{pmatrix}$$

eller

$$A = \begin{pmatrix} \alpha & \beta \\ \beta & -\alpha \end{pmatrix} = \begin{pmatrix} \alpha & -\beta \\ \beta & \alpha \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Notera att matrisen längst till höger svarar mot en avbildning S som tar övre halvplanet på det undre, och vice versa, samt håller x_1-axeln fix. Eftersom S ska vara en avbildning i $\mathcal{K}$, följer det att S måste vara spegling i x_1-axeln.

Bevis. Att

$$A = \begin{pmatrix} a & c \\ b & d \end{pmatrix}$$

är ortogonal, betyder att $a^2 + b^2 = 1$, $c^2 + d^2 = 1$ samt att $ac + bd = 0$. Den senare ekvationen innebär att $(c, d) = t(b, -a)$ för något t. Alltså är

$$1 = c^2 + d^2 = t^2(b^2 + a^2) = t^2,$$

så $t = \pm 1$ och därför $(c, d) = \pm(b, -a)$ och lemmat följer. $\square$

Vi kan nu verifiera resten av A4 samt A5. Låt $e = (1, 0)$. Då är $Ae = (\alpha, \beta)$ om A är som i lemmat. Om $b = (b_1, b_2)$ är en godtycklig punkt kan vi alltså hitta en avbildning A, som tar positiva x_1-axeln på strålen från origo genom b, genom att välja (α, β) så att $t(\alpha, \beta) = (b_1, b_2)$ för ett positivt tal t. Eftersom A tar linjer på linjer följer det att den tar övre halvplanet på en given sida av linjen ℓ genom origo och b och undre halvplanet på andra sidan av ℓ. Efter att vid behov komponera med avbildningen S, jmf kommentaren efter lemmat, så kan vi få övre halvplanet att avbildas på önskad sida av ℓ.

Antag nu att vi har en given punkt a, en stråle ab och en sida c av linjen ab. Om A tar positiva x_1-axeln på strålen från origo genom $b - a$ så kommer $\psi(x) = Ax + a$ att ta origo på a och positiva x_1-axeln på strålen ab. Genom att vid behov komponera med avbilningen S kan vi få övre halvplanet att avbildas på önskad sida av linjen ab. Om $\psi'(x)$ är en motsvarande avbildning för a', $a'b'$ och c', så blir $\phi = \psi' \circ \psi^{-1}$ en avbildning i $\mathcal{K}$ som tar a på a', strålen ab på strålen $a'b'$ och sidan c på sidan c'. Alltså har vi visat på existensen av en avbildning ϕ med önskad egenskap. Det återstår att visa entydigheten. Antag därför att ϕ' också gör samma sak. Låt σ vara en avbildning som tar a på origo, strålen ab på positiva x-axeln samt sidan c på övre halvplanet. Då är $T = \sigma \circ \phi^{-1} \circ \phi' \circ \sigma^{-1}$ en avbildning i $\mathcal{K}$ som tar origo på sig själv, positiva x_1-axeln på sig själv samt övre halvplanet på sig självt. Eftersom $T(x) = Ax + c$ och $T(0) = 0$ så följer att $c = 0$. Eftersom positiva x_1-axeln går på sig själv så följer, jmf lemmat, att $\alpha = 1$ och $\beta = 0$. Om dessutom T tar övre halvplanet på sig självt så följer det att T är identitetsavbildningen I. Alltså är $\sigma \circ \phi^{-1} \circ \phi' \circ \sigma^{-1} = I$ och därför är $\phi^{-1} \circ \phi' = I$, dvs $\phi' = \phi$. Alltså är entydigheten visad och därmed är axiom A4 verifierat.

Det följer från argumenten ovan även att första villkoret i A5 är uppfyllt. På liknande sätt inser man att ingen vinkel är kongruent med en äkta delvinkel till sig själv. Genom lämpliga sammansättningar kan man anta att ϕ tar en vinkel vid origo med positiva x-axeln som ena vinkelbenet på en delvinkel av denna, som också har positiva x-axeln som ena vinkelbenet. Men då tar ϕ origo på sig själv, positiva x-axeln på sig själv och övre halvplanet på sig själv, så ϕ är identiteten.

Vi noterar att i denna modell även parallellaxiomet är uppfyllt; dvs att givet en linje ℓ och en punkt p utanför ℓ, så finns det precis *en* linje genom p som inte skär ℓ. Efter en lämplig kongruenstransformation kan vi anta att ℓ är x_1-axeln, dvs $x_2 = 0$, och att punkten p är $(0, t)$ för något $t \neq 0$. En linje är parallell med x_1-axeln om och endast om den ges på formen $x_2 = s$ för $s \neq 0$, så enda möjligheten är linjen $x_2 = t$.

Eftersom vi nu sett att $\mathbb{R}^2$ med ovanstående tolkningar av begreppen linjer och kongruenstransformationer är en modell till teorin A, så följer från den allmänna teorin för A att (storlek på) vinklar och (längd av) sträckor har en mening, bara vi fixerar en enhetssträcka. Tag sträckan från origo till $e = (1, 0)$ som enhetssträcka. Eftersom avbildningarna i lemma 2.1 är kongruensavbildningar följer det att varje punkt (α, β) med $\alpha^2 + \beta^2 = 1$ också har avstånd 1 till origo. Vi kan nu definiera $\cos\theta$ och $\sin\theta$ genom att sätta $(\cos\theta, \sin\theta) = (\alpha, \beta)$ om θ är vinkeln mellan positiva x_1-axeln och strålen från origo genom (α, β); här tänker vi oss att $-\pi < \theta \leq \pi$ och att positiva vinklar är de för vilka $\beta > 0$. Man kan sedan utvidga definitionen av $\sin\theta$ och $\cos\theta$ till alla reella θ genom att kräva att de ska vara 2π-periodiska. Sätt nu

$$R_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

Eftersom R_θ är en kongruensavbildning så bevarar den vinklar och det följer att en godtycklig stråle från origo med säg vinkel ϕ gentemot positiva x_1-axeln avbildas på en stråle med vinkel $\phi + \theta$ gentemot positiva x_1-axeln. Alltså innebär R_θ rotation runt origo med vinkeln θ (man kan här tillåta negativa θ). Lemma 2.1 innebär alltså att varje ortogonal avbildning A är $A = R_\theta$ eller $A = R_\theta S$ för något θ.

En godtycklig avbildning i $\mathcal{K}$ kan alltså beskrivas som att man först eventuellt speglar i x_1-axeln, sedan roterar en viss vinkel θ runt origo och slutligen translaterar med β.

Vi lämnar som övning att verifiera den vanliga formeln för avståndet mellan två punkter.

Uppgift 2.2. Visa att avståndet mellan punkterna $a = (a_1, a_2)$ och $b = (b_1, b_2)$ ges av $\sqrt{(b_1 - a_1)^2 + (b_2 - a_2)^2}$.

Uppgift 2.3. Ange den kongruenstransformation $\phi \in \mathcal{K}$ som tar $(0, 0)$ på $(2, 1)$, strålen $(0, 0), (0, 1)$ på strålen strålen $(2, 1), (-3, 2)$ och som tar övre halvplanet på den sidan om linjen genom $(2, 1)$ och $(-3, 2)$ som innehåller $(4, 0)$.

3. Grundläggande satser i absolut geometri

Huvudresultatet i det här avsnittet är följande.

Sats 3.1. *I en modell till teorin A så gäller precis ett av följande fall: I det ena fallet finns till varje linje ℓ och punkt p utanför ℓ exakt* *en linje ℓ' genom p som är parallell med ℓ. I detta fall är vinkelsumman i alla trianglar π.*

I det andra fallet finns för varje linje ℓ och punkt p utanför ℓ oändligt många linjer genom p som är parallella med ℓ. I detta fall är vinkelsumman mindre än π i alla trianglar.

Om PA är uppfyllt är alltså vinkelsumman π i alla trianglar, och om PA inte gäller så finns alltid flera olika parallella linjer genom p, och vinkelsumman är mindre än π i alla trianglar. Teorin A+PA kallas *euklidisk geometri*. Vi ska se i avsnitt 4 att alla modeller till denna teori är isomorfa, dvs att det väsentligen bara finns en sådan modell. Teorin A plus negationen av PA kallas *hyperbolisk geometri*. Även för denna teori är alla modeller isomorfa, vilket dock är svårare att visa. Vi studerar en modell till denna teori i kapitel 5.

För att komma fram till sats 3.1 måste vi först studera de grundläggande satserna om kongruenta trianglar.

Kongruensfallen. Tre punkter a, b, och c som inte ligger på samma linje definierar en *triangel* abc med dessa tre punkter som *hörn*.

Givet en triangel abc så säger vi att en punkt *ligger i* triangeln om den ligger på den sida om var och en av linjerna ab, ac respektive bc som bestäms av den tredje punkten. Det följer genast att om två punkter ligger i en triangel så ligger även sträckan mellan dem i triangeln, se även övning 9.2.

Vinklarna vid punkterna a, b och c, dvs vinklarna $\wedge bac$, $\wedge abc$ och $\wedge acb$, kallas för triangelns *(inner)vinklar*. När det inte finns risk för missförstånd skriver vi gärna $\wedge a$, $\wedge b$ och $\wedge c$ istället. Deras sidovinklar kallas *yttervinklar* till triangeln. Sträckorna ab, bc, och ca kallas triangelns *sidor*.

När vi säger att trianglarna abc och $a'b'c'$ är kongruenta så menar vi att det finns en kongruenstransformation sådan att $a \mapsto a'$, $b \mapsto b'$ och $c \mapsto c'$. Därmed har vi också att $|ab| = |a'b'|$, $|ac| = |a'c'|$ och $|bc| = |b'c'|$ samt att $\wedge a = \wedge a'$, $\wedge b = \wedge b'$ och $\wedge c = \wedge c'$. Euklides bevisade tre satser om kongruenta trianglar, de s.k. kongruensfallen.

SATS 3.2 (Första kongruensfallet). *Om vi har två trianglar abc och $a'b'c'$ sådana att $|ab| = |a'b'|$, $|ac| = |a'c'|$ och $\wedge a = \wedge a'$, så är trianglarna kongruenta.*

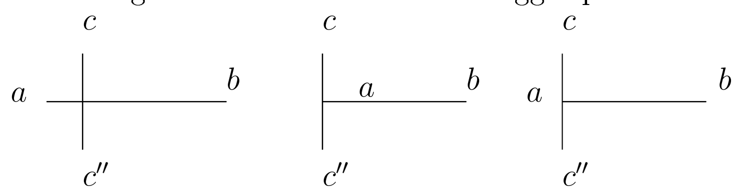
BEVIS. Tag en en kongruensavbildning sådan att ab avbildas på $a'b'$ och så att c hamnar på en punkt c'' på samma sida om $a'b'$ som c'. Eftersom $\wedge a = \wedge a'$ kommer då $a'c'$ och $a'c''$ att ligga på samma stråle från a'. Eftersom de är lika långa måste därför $c' = c''$. $\square$

Korollarium 3.3 (Basvinkelsatsen). *I en likbent triangel är basvinklarna lika stora.*

Bevis. Antag att vi i abc har att $|ab| = |ac|$. Då är enligt sats 3.2 abc kongruent med acb och alltså är $\wedge c = \wedge b$. □

Sats 3.4 (Andra kongruensfallet). *Om vi har två trianglar abc och $a'b'c'$ sådana att $|ab| = |a'b'|$, $|ac| = |a'c'|$ och $|bc| = |b'c'|$, så är trianglarna kongruenta.*

Bevis. Flytta först $a'b'c'$ så att $a'b'$ sammanfaller med ab och c' hamnar på en punkt c'' så att denna och c ligger på motsatta sidor om ab. Det gäller nu att se att abc och abc'' är kongruenta. Dra sträckan cc'' och antag först att varken a eller b ligger på denna.



Då är acc'' och bcc'' likbenta, så i var och en av dessa är basvinklarna lika stora. Genom att ta summan av dessa vinklar vid c respektive c'' (eller skillnaden, om sträckan cc'' inte skär sträckan ab) så får vi att $\wedge acb = \wedge ac''b$, och alltså är trianglarna kongruenta enligt första kongruensfallet. Fallet när a eller b ligger på sträckan cc'' lämnas som övning. □

Sats 3.5 (Tredje kongruensfallet). *Om abc och $a'b'c'$ är två trianglar sådana att $|ab| = |a'b'|$, $\wedge a = \wedge a'$ och $\wedge b = \wedge b'$, så är trianglarna kongruenta.*

Bevis. Flytta ab så att den sammanfaller med $a'b'$ och så att c hamnar på en punkt c'' på samma sida om $a'b'$ som c'. Eftersom $\wedge a = \wedge a'$ följer att $a'c'$ och $a'c''$ ligger på samma stråle från a'. På motsvarande sätt följer att $b'c'$ och $b'c''$ ligger på samma stråle från b'. Det följer att såväl c' som c'' ligger på båda linjerna $a'c'$ och $b'c'$, och alltså är $c' = c''$. □

Korollarium 3.6 (Basvinkelsatsens omvändning). *Om två vinklar i en triangel är lika så är motstående sidor lika.*

Uppgift 3.1. Bevisa korollariet!

Sats 3.7. *I en triangel är varje yttervinkel större än var och en av de motstående innervinklarna.*

Bevis. Låt triangeln vara abc och låt d ligga på förlängningen av ab över b (dvs d är en punkt på strålen ab så att b ligger på sträckan ad). Vi vill visa att $\wedge cbd > \wedge c$. Låt p vara mittpunkten på sträckan bc.

c q

p

a b d

Förläng ap över p till en punkt q sådan att $|ap| = |pq|$. Eftersom $\wedge apc = \wedge bpq$ (vertikalvinklar) så är trianglarna pca och pbq kongruenta enligt det första kongruensfallet, och alltså är $\wedge c = \wedge pbq < \wedge pbd$. □

Uppgift 3.2. Visa att minst två av vinklarna i en triangel måste vara spetsiga.

Proposition 3.8. *En triangel kan delas upp i två rätvinkliga trianglar.*

Bevis. Antag att $\wedge b$ och $\wedge c$ är spetsiga i den givna triangeln abc. Dra normalen från a till linjen bc och kalla skärningspunkten p. Det gäller nu att visa att p ligger mellan b och c. Eftersom vinklarna vid b och c inte är räta så kan inte p sammanfalla med någon av dessa punkter. Om p ligger utanför sträckan bc, säg på förlängningen över c,

a a

b p c b c p

så blir $\wedge bca$ yttervinkel till triangeln cpa med motstående vinkel rät, vilket i ljuset av sats 3.7 strider mot antagandet att $\wedge bca$ är spetsig. □

Sats 3.9. *I en triangel är vinkeln stående mot en större sida större än vinkeln stående mot en mindre sida.*

Bevis. Antag att vi i abc har att $|ab| > |ac|$. Då kan vi ta p på ab sådan att $|ap| = |ac|$. Dra nu pc. Då har vi att

a

p

b

c

$\wedge acb > \wedge acp = \wedge apc > \wedge abc$, där likheten beror på att acp är likbent, och sista olikheten följer av sats 3.7. Alltså är $\wedge acb > \wedge abc$. □

Sats 3.10 (Triangelolikheten). *Summan av (längderna av) två sidor i en triangel är större än (längden av) den tredje.*

Bevis. Tag i triangeln abc och förläng ab över b till en punkt d sådan att $|bd| = |bc|$, samt dra cd. Då är $\wedge adc = \wedge bcd$ enligt basvinkelsatsen

d

b

a c

och alltså är $\wedge adc < \wedge acd$. Det följer nu av basvinkelsatsen och sats 3.9 att $|ac| < |ad|$. □

Parallellaxiomet och vinkelsumma i trianglar. En linje som skär två övriga linjer ℓ och ℓ' kallas en *transversal* till dessa, och ett par av vinklar som uppstår, en vid vardera av de två linjerna, och som ligger på var sin sida om transversalen, kallas *alternatvinklar.*

ℓ'

ℓ

Vi påminner om att två linjer är parallella om de inte skär varandra. Det följer av sats 3.7 att om ett par av alternatvinklar till en given transversal är lika så är ℓ och ℓ' parallella.

Sats 3.11. *Givet en linje ℓ och en punkt p utanför ℓ så finns det åtminstone en linje ℓ' genom p som är parallell med ℓ.*

Bevis. Dra en linje ℓ'' genom p som skär ℓ och dra sedan en linje ℓ' genom p så att ℓ'' blir transversal till ℓ och ℓ' med lika alternatvinklar. □

Summan av två vinklar i en triangel är alltid mindre än π, eftersom sats 3.7 medför att summan är mindre än summan av ena vinkeln och dess yttervinkel. Detta kan skärpas.

Sats 3.12 (Saccheri). *Summan av vinklarna i en triangel är alltid mindre än eller lika med π.*

Bevis. Vi påstår först att det till en given triangel med en vinkel α finns en annan triangel med samma vinkelsumma och vars ena vinkel är mindre än eller lika med $\alpha/2$. För att se detta upprepar vi beviset av sats 3.7. Med beteckningarna där, och säg $\wedge a = \alpha$, följer det att abc har samma vinkelsumma som abq. Men nu är $\wedge qab + \wedge bqa = \wedge qab + \wedge cap = \wedge a$ så en av vinklarna i abq är mindre än eller lika med $\alpha/2$.

Antag nu att triangeln T har vinkelsumma $\pi + \epsilon$ där $\epsilon > 0$, och låt ena vinkeln vara α. Enligt ovan finns nu en triangel T_1 med samma vinkelsumma och ena vinkeln mindre än eller lika med $\alpha/2$. Med induktion kan vi nu hitta en triangel T_k med samma vinkelsumma och vars ena vinkel är mindre än eller lika med $\alpha/2^k$. För tillräckligt stort k får vi alltså en triangel med vinkelsumma $\pi + \epsilon$ som har en vinkel som är mindre än ϵ. Detta leder till motsägelse eftersom summan av de övriga två vinklarna är mindre än π. □

För en triangel T definierar vi *defekten* $d(T)$ som π minus summan av vinklarna i T. Enligt föregående sats är alltså $d(T) \geq 0$ för alla T. Om triangeln T delas upp i två deltrianglar T_1 och T_2 genom att man från ena hörnet i T drar en sträcka till motstående sida, så kontrollerar man lätt att

$$d(T) = d(T_1) + d(T_2). \tag{3.1}$$

Om speciellt $d(T) = 0$ så följer att $d(T_1) = d(T_2) = 0$.

SATS 3.13. *Om $d(T) = 0$ för någon triangel T så gäller detta för alla trianglar T.*

BEVISSKISS. Antag $d(T) = 0$. Genom upprepad användning av (3.1) följer att $d(T') = 0$ för alla trianglar T' som är (kongruenta med) en deltriangel till T; speciellt gäller detta för alla små T'. Men en godtycklig triangel T'' kan delas in i små trianglar, och med (3.1) igen följer att $d(T'') = 0$. □

PROPOSITION 3.14. *Om det finns mer än en linje ℓ' genom p som inte skär linjen ℓ så finns också en triangel med vinkelsumma mindre än π.*

BEVIS. Dra en normal till ℓ genom p och låt skärningspunkten vara a. Om en stråle från p som ej är vinkelrät mot normalen likväl är parallell med ℓ, så kommer alla strålar från p med större vinkel mot normalen också att vara parallella med ℓ. Låt α vara infimum av alla sådana vinklar, $\epsilon = \pi/2 - \alpha$, och låt ℓ' vara motsvarande linje genom p. Idén är nu att bilda en rätvinklig triangel med hörn i a, p samt en punkt c långt bort på linjen ℓ. Summan av vinklarna vid a och p i denna triangel kommer att vara mindre än $\pi/2 + \alpha$ så det räcker att visa att man kan välja c så att vinkeln vid c blir mindre än ϵ.

p ℓ'

$\alpha - \epsilon$

a b ℓ c

En stråle från p med vinkel $\alpha - \epsilon$ mot normalen kommer att skära ℓ i en punkt b. Tag nu en punkt c på ℓ bortom b sådan att $|bc| > |bp|$. Då är $\wedge bcp < \wedge bpc < \epsilon$ enligt sats 3.9. Därför är vinkelsumman i acp mindre än $\pi/2 + \alpha + \epsilon = \pi$. □

Av det ovanstående resonemanget leds man till att talet $\alpha = \alpha(\ell, p)$ blir mindre när avståndet mellan ℓ och p ökar. Detta tal kallades *parallellvinkeln* av Lobachevski. Det ger en möjlighet att definiera avstånd på ett intrikat sätt i hyperbolisk geometri. Vi ska titta närmare på detta i kapitel 5.

PROPOSITION 3.15. *Om det genom p bara finns en linje som är parallell med ℓ så finns en triangel med vinkelsumma π.*

BEVIS. Tag två punkter a och b på ℓ och dra sträckorna pa och pb. Till var och en av dessa kan vi dra en stråle från p sådan att sträckan blir en transversal med lika alternatvinklar.

p

a b

Båda dessa strålar är då delar av linjer som är parallella med ℓ och enligt antagandet är de alltså delar av en och samma linje. Det följer nu att vinkelsumman i pab är π. □

BEVIS AV SATS 3.1. Satsen följer från satserna 3.13 till 3.15. □

Ytterligare några resultat i absolut geometri. Vi avrundar detta avsnitt med några välkända resultat som gäller i absolut geometri. Antag att C är en cirkel med medelpunkt m och låt a och b vara två punkter på cirkeln. Då ligger sträckan ab helt inne i cirkeln, dvs varje punkt p på sträckan (förutom ändpunkterna) har mindre än eller samma avstånd till cirkelns medelpunkt än radien som är $|ma| = |mb|$. Detta följer genast från sats 3.9, se figur,

m

a p b

eftersom en av vinklarna vid p måste vara trubbig eller rät. Det följer att alla andra punkter på ℓ ligger utanför cirkeln. En sådan sträcka ab kallas en *korda* till cirkeln C. Speciellt ser vi att en linje skär en cirkel i högst två punkter.

En linje som skär en cirkel i endast en punkt p kallas en *tangent* till cirkeln i p.

Uppgift 3.3. Visa att det i varje punkt på en cirkel finns en och endast en tangent.

Om för två trianglar abc och $a'b'c'$ vi har att $|ab| = |a'b'|$ och $|ac| = |a'c'|$ så följer det av det det första och andra kongruensfallet att $\wedge a = \wedge a'$ om och endast om $|bc| = |b'c'|$. Men vi har också:

Proposition 3.16. *Om för två trianglar abc och $a'b'c'$ vi har att $|ab| = |a'b'|$ och $|ac| = |a'c'|$ så är $\wedge a' < \wedge a$ om och endast om $|b'c'| < |bc|$.*

Bevis. Vi kan anta att $|ac| \geq |ab|$ och att sträckan ab sammanfaller med $a'b'$. Om vi vidare antar att $\wedge a' < \wedge a$ så kommer strålen ac' att skära sträckan bc i en punkt p. Denna ligger nu på sträckan ac' eftersom $|ap|$ är mindre än den största av $|ab|$ och $|ac|$, jmf resonemanget kring förra figuren.

Enligt basvinkelsatsen är $\wedge ac'c = \wedge acc'$. I triangeln $bc'c$ har vi alltså att vinkeln vid c' är större än vinkeln vid c, ty

$$\wedge bc'c > \wedge ac'c = \wedge acc' > \wedge bcc',$$

och alltså är $|bc| > |bc'|$ enligt sats 3.9. Från detta följer propositionen. □

Uppgift 3.4. Visa att två olika cirklar kan skära varandra i högst två punkter.

Fixera två positiva tal r, R med säg $0 < r < R$. För varje tal α mellan 0 och π kan vi bilda en triangel abc med $|ab| = r$, $|ac| = R$ och

$\wedge a = \alpha$. Låt $f(\alpha)$ vara $|bc|$. Enligt proposition 3.16 har vi att $f(\alpha)$ är en växande funktion av α. Vi påstår att den även är kontinuerlig.

Låt a, b vara fixa och antag α och α' är två olika vinklar och c och c' motsvarande punkter. Då är enligt triangelolikheten $|ac| \leq |ac'| + |cc'|$ och $|ac'| \leq |ac| + |cc'|$ så vi får att

$$\big||ac| - |ac'|\big| \leq |cc'|.$$

Det räcker alltså att visa att $|cc'| \to 0$ då $\alpha' \to \alpha$.

Från proposition 3.16 följer att $|cc'|$ avtar när $|\alpha - \alpha'|$ avtar. Det räcker därför att se att det till varje $\epsilon > 0$ finns en stråle ac' med $|ac'| = |ac|$ sådan att $|cc'| < \epsilon$. Fixera c och avsätt en sträcka cc'' med längd ϵ från c vinkelrätt ut från strålen ac. Av sats 3.9 följer att $|ac''| > |ac|$. Låt c' vara den punkt på ac'' sådan att $|ac'| = |ac|$. Eftersom nu $\wedge acc' = \wedge ac'c$ följer att $\wedge cc'c''$ är trubbig och av sats 3.9 följer nu att $|cc'| < \epsilon$. Vi har därmed visat att $f(\alpha)$ är kontinuerlig.

Uppgift 3.5. Visa att proposition 3.16 följer från kontinuiteten av $f(\alpha)$, satsen om mellanliggande värde samt det första kongruensfallet.

Uppgift 3.6. Antag att vi har tre positiva tal A, B, C sådana att $A < B + C$, $B < C + A$ samt $C < A + B$. Visa att det finns en triangel med dessa tal som sidlängder.

Per definition så bevarar kongruensavbildningar linjer, deras ordningsrelationer samt (längder av) sträckor. Vi avsluter med att notera att detta i själva verket karakteriserar kongruensavbildningar.

Proposition 3.17. *Antag att ϕ är en bijektiv avbildning som avbildar linjer på linjer, bevarar (eller kastar om) deras ordningsrelationer, samt bevarar längder av sträckor. Då är ϕ en kongruenstransformation.*

Bevis. Notera först att ϕ avbildar sträckor på sträckor och strålar på strålar eftersom den avbildar linjer på linjer och bevarar (eller kastar om) ordningsrelationen. Notera att varje sammansättning $\psi \circ \phi$, där $\psi \in \mathcal{K}$, har samma egenskaper som ϕ antas ha. Eftersom $\mathcal{K}$ är en grupp så räcker det att visa att en viss sådan sammansättning ligger i $\mathcal{K}$. Fixera två punkter a och b och låt ℓ vara linjen genom dem. Efter en lämplig sammansättning kan vi anta att $a \mapsto a$ och att strålen ab avbildas på sig själv. Eftersom avstånd bevaras följer det att varje punkt på ℓ avbildas på sig själv. Tag nu en punkt c på ena sidan av ℓ. Genom att eventuellt komponera med en spegling i ℓ kan vi anta att c avbildas på en punkt c' på samma sida om ℓ som c. Eftersom sträckor avbildas på sträckor följer det att alla punkter på denna sida av ℓ avbildas på denna sida, och följaktligen att alla punkter på andra sidan avbildas på andra sidan. Sträckorna ac och bc avbildas nu på sträckorna ac' respektive bc'. Eftersom deras längder dessutom bevaras är trianglarna abc och abc' är kongruenta enligt det andra kongruensfallet. Det följer att $c = c'$. Eftersom samma argument gäller för en godtycklig punkt c utanför ℓ så följer det att avbildningen är identitetsavbildningen och alltså ett element i $\mathcal{K}$. □

4. Euklidisk geometri, likformighetsteori

I det här avsnittet ska vi anta att PA gäller och utveckla den s.k. likformighetsteorin. En konsekvens av denna är att alla modeller till A+PA (dvs teorin vars axiom är axiomen för A samt PA) är isomorfa med modellen vi betraktade i avsnitt 2.

Proposition 4.1. *En transversal till två parallella linjer måste ha lika alternatvinklar.*

Detta följer genast från beviset av sats 3.11 och PA. Det följer också omedelbart från PA att ℓ är parallell med ℓ'' om ℓ är parallell med ℓ' och ℓ' är parallell med ℓ''.

Om två sinsemellan parallella linjer skär två andra sinsemellan parallella linjer så uppstår en fyrhörning som kallas en *parallellogram.*

Proposition 4.2. *I en parallellogram är motstående sidor lika långa.*

Bevis. Låt ab vara parallell med cd, ac parallell med bd och dra diagonalen bc.

c d

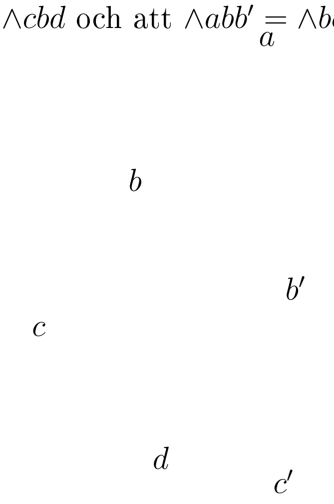
a b

Eftersom alternatvinklar vid parallella linjer är lika följer att $\wedge abc = \wedge bcd$ och $\wedge acb = \wedge cbd$. Enligt tredje kongruensfallet är därför abc kongruent med dcb och från detta följer satsen. □

Sats 4.3 (Transversalsatsen). *Antag vi har två strålar från a, att b och c ligger på ena strålen och att b' och c' ligger på den andra, så att bb' och cc' är parallella. Då är*

$$\frac{|ab|}{|ac|} = \frac{|ab'|}{|ac'|} \quad \textit{och} \quad \frac{|bc|}{|ac|} = \frac{|b'c'|}{|ac'|}.$$

Bevis. Det räcker att visa den första likheten, för den andra är en enkel konsekvens (övning 4.1). Låt oss till att börja med anta att $|ac| = 2|ab|$. Tag en punkt d på cc' sådan att bd blir parallell med $b'c'$. Från proposition 4.1 följer (kom ihåg att vertikalvinklar är lika) att $\wedge a = \wedge cbd$ och att $\wedge abb' = \wedge bcd$.



Eftersom $|ab| = |bc|$ följer det att abb' och bcd är kongruenta (tredje kongruensfallet) och alltså att $|ab'| = |bd|$. Från proposition 4.2 följer att $|bd| = |b'c'|$. Detta medför sammantaget att $|ac'| = 2|ab'|$.

Man kan lätt utvidga detta resonemang till fallet när $|ac| = m|ab|$ för något naturligt tal m som figuren antyder, men det lämnar vi som övning.

Från detta får man satsen i fallet att $|ab|$ och $|ac|$ båda är en heltalsmultipel av en fix sträcka, t ex 2^{-m} för något m. Det allmänna fallet följer sedan genom en enkel gränsövergång; man måste förstås då gå tillbaka till definitionen av längd och vi lämnar detaljerna till läsaren. $\square$

Uppgift 4.1. Visa att andra likheten i transversalsatsen följer från den första,

Uppgift 4.2. Fyll i detaljerna i beviset för sats 4.3.

Man säger att två trianglar abc och $a'b'c'$ är *likformiga* om $\wedge a = \wedge a'$, $\wedge b = \wedge b'$, $\wedge c = \wedge c'$ samt

$$\frac{|ab|}{|a'b'|} = \frac{|ac|}{|a'c'|} = \frac{|bc|}{|b'c'|}.$$

SATS 4.4. *Antag att abc och $a'b'c'$ är två trianglar. Följande tre påståenden brukar kallas första, andra respektive tredje likformighetsfallet.*

(1) Om $\wedge a = \wedge a'$ och

$$\frac{|ab|}{|a'b'|} = \frac{|ac|}{|a'c'|}$$

så är trianglarna likformiga.

(2) Om

$$\frac{|ab|}{|a'b'|} = \frac{|ac|}{|a'c'|} = \frac{|bc|}{|b'c'|}, \tag{4.1}$$

så är trianglarna likformiga.

(3) Om $\wedge a = \wedge a'$ och $\wedge b = \wedge b'$ så är trianglarna likformiga.

BEVIS. (1) Efter en förflyttning kan vi anta att $a' = a$, att b' ligger på strålen ab och att c' ligger på strålen ac. Tag nu c'' på denna stråle så att $b'c''$ blir parallell med bc.

a

b' c''

c

b d

Från antagandet samt transversalsatsen har vi att

$$\frac{|ac|}{|ac'|} = \frac{|ab|}{|ab'|} = \frac{|ac|}{|ac''|}$$

från vilket det följer att $c' = c''$. Enligt sats 4.1 är vinklarna vid b' och c' samma som motsvarande vinklar vid b och c. Det återstår att visa att

$$\frac{|ac|}{|ac'|} = \frac{|bc|}{|b'c'|}. \tag{4.2}$$

Tag d på ac sådan att $c'd$ blir parallell med $b'b$. Då är $|b'c'| = |bd|$ enligt proposition 4.2 och med transversalsatsen tillämpad på strålarna ca

och cb får vi att

$$\frac{|bc|}{|bd|} = \frac{|ac|}{|ac'|}.$$

Sammantaget får vi (4.2), och alltså är första likformighetsfallet visat.

(2) Vi kan anta att $a' = a$ och att b' ligger på ab. Tag c'' som förut. Enligt förra beviset är då abc och $ab'c''$ likformiga och speciellt gäller (4.1) med c'' istället för c'. Sätter vi samman detta med antagandet får vi att $ab'c''$ och $ab'c'$ har lika långa sidor, och följaktligen är kongruenta enligt det andra kongruensfallet. Alltså är abc och $ab'c'$ likformiga.

(3) Beviset av det tredje likformighetsfallet lämnar vi till läsaren. □

Uppgift 4.3. Visa det tredje likformighetsfallet.

I en rätvinklig triangel kallas de två sidorna till den räta vinkeln *kateter* och den tredje sidan kallas *hypotenusan.*

Sats 4.5 (Pythagoras sats). *I en rätvinklig triangel är kvadraten av (längden av) hypotenusan lika med summan av kvadraterna (av längderna) av kateterna.*

Bevis. Antag att triangeln är abc och att $\wedge c$ är rät. Normalen från c till ab kommer att skära i en punkt d mellan a och b eftersom vinklarna vid a och b är spetsiga (jmf beviset av sats 3.8).

c

a d b

Nu är cdb och acb likformiga enligt det tredje likformighetsfallet (en gemensam vinkel och en rät). Därför är

$$\frac{|db|}{|cb|} = \frac{|cb|}{|ab|}$$

och alltså $|db||ab| = |cb|^2$. På samma sätt får man att $|ad||ab| = |ca|^2$ och om man lägger samman dessa får man $|ab|^2 = |cb|^2 + |ca|^2$. □

Koordinatsystem och den euklidiska modellens entydighet. Låt M' vara en godtycklig modell för euklidisk geometri. Vi ska nu visa hur man kan införa ett koordinatsystem i modellen M'. Detta har intresse i sig, men kommer också att ge oss ett bevis för att M' är isomorf med modellen M för euklidisk geomtri som är beskriven i avsnitt 2.

Vi inför ett koordinatsystem i M' på följande vis: Fixera en enhetssträcka och välj två ortogonala linjer som vi kallar x_1-axeln och x_2-axeln, och kalla deras skärningspunkt origo. Notera att en linje är normal till den ena av dessa axlar om och endast om den är en parallell med den andra. Välj en stråle från origo av x_1-axeln och kalla den den positiva halvaxeln. Varje punkt på x_1-axeln får nu en *koordinat* x_1 som är avståndet till origo om punkten ligger på den positiva halvaxeln och minus avståndet till origo om man är på den andra strålen. Notera att avståndet mellan två punkter x_1 och y_1 på x_1-axeln är $|y_1 - x_1|$. Gör likadant med x_2-axeln. Låt en godtycklig punkt få *koordinaterna* (x_1, x_2) om normalerna från p till x_1-axeln respektive x_2-axeln skär i punkter med koordinaterna x_1 respektive x_2. Notera att två normaler från två givna punkter x_1 och x_2 på de respektive axlarna inte är parallella och därför måste skära varandra i en fix punkt. Vi får alltså en bijektion mellan punkterna i modellen M' och $\mathbb{R}^2$, dvs punkterna i modellen M.

Antag att vi har två punkter med koordinaterna (x_1, x_2) och (y_1, y_2). Enligt proposition 4.2 och Pythagoras sats följer det att avståndet mellan dessa punkter är $\sqrt{(y_1 - x_1)^2 + (y_2 - x_2)^2}$.

(y_1, y_2)

y_2

x_2 (x_1, x_2)

y_1 x_1

Låt nu ℓ vara en linje i modellen M' och välj två olika punkter a och b på ℓ med koordinaterna (a_1, a_2) respektive (b_1, b_2). Om $x = (x_1, x_2)$ är en godtycklig punkt på linjen ℓ så följer det av likformighetsteorin att

$$(b_1 - a_1)(x_2 - a_2) = (x_1 - a_1)(b_2 - a_2),$$

dvs punkternas koordinater uppfyller ett linjärt samband.

(b_1, b_2)

(a_1, a_2)

(x_1, x_2)

Omvänt kan man lätt se att varje sådant linjärt samband definierar en linje.

Ordningsrelationen på linjen ℓ svarar mot ordningsrelationen på x_1-axeln, eller eventuellt x_2-axeln om ℓ är parallell med denna. Detta stämmer överens med den ordningsrelation vi har tilldelat linjerna i vår modell M. Alltså har vi en enentydig korrespondens (bijektion) mellan mängden av linjer i M' och mängden av linjer i vår modell M, och ordningsrelationerna stämmer överens. Låt $\mathcal{K}'$ beteckna gruppen av kongruensavbildningar i modellen M', och låt $\mathcal{K}$ beteckna motsvarande grupp som hör till M. Det återstår nu att visa att vi har en enentydig korrespondens (gruppisomorfi) mellan $\mathcal{K}'$ och $\mathcal{K}$. Vi har sett att längd av sträckor i M' stämmer överens med längd i M, jmf övning 2.2. Eftersom varje $\phi \in \mathcal{K}$ inducerar en bijektiv avbildning i modellen M' som tar linjer på linjer så att ordningsrelationen bevaras (eller kastas om) och som bevarar avstånd, så följer det av proposition 3.17 att den inducerade avbildningen är en kongruensavbildning i M', dvs ett element i $\mathcal{K}'$. Vi får på detta sätt en injektiv grupphomomorfism $\mathcal{K} \to \mathcal{K}'$ (eller enklare uttryckt, $\mathcal{K}$ är en delgrupp till $\mathcal{K}'$). Men vi vet att varje förflyttning som ska gå att göra enligt axiom A4 kan göras med något element i $\mathcal{K}$ och på grund av entydigheten i A4 följer det då att $\mathcal{K} = \mathcal{K}'$ och alltså är vi klara.

Anmärkning 4.6. Istället för att använda proposition 3.17 så kan man ganska lätt visa direkt att varje avbildning i $\mathcal{K}$ svarar mot en avbildning i $\mathcal{K}'$ och sedan som ovan dra slutsatsen att de är lika. Eftersom vi vet att $\mathcal{K}$ genereras av translationer, rotationer runt origo samt spegling i x_1-axeln, så räcker det att kontrollera att var och en av dessa avbildningar finns i $\mathcal{K}'$.

Låt t ex $\phi(x) = x - a$. Notera att $-a$ ligger på linjen ℓ genom a och origo. I $\mathcal{K}'$ finns alltså en avbildning ϕ' som tar a på 0, 0 på $-a$ och som bevarar sidorna av linjen ℓ. Det är nu lätt att kontrollera att $\phi' = \phi$. Vi lämnar detta som övning liksom de övriga verifikationerna. $\square$

UPPGIFT 4.4. Visa att $\phi' = \phi$. Visa också att det finns en avbildning i $\mathcal{K}'$ som svarar mot avbildningen R_θ, jmf avsnitt 2, samt mot avbildningen S.

Eftersom vi nu har infört koordinatsystem, och sett att alla modeller till euklidisk geometri är isomorfa med modellen M från avsnitt 2, kan vi i fortsättningen, när vi utvecklar teorin för den euklidiska geometrin, arbeta direkt i modellen M, där man alltså representerar punkter med koordinater, linjer med ekvationer etc. Man talar om *analytisk geometri* eller att man använder *analytiska* metoder. Dessa har otvivelaktigt skördat stora framgångar alltsedan koordinatsystem infördes på 1600-talet. T ex när vi ska studera längd- och areabegreppet i nästa avsnitt är dessa metoder fördelaktiga. Likaså i avsnitt 7 där vi studerar (o)möjligheten att utföra vissa geometriska konstruktioner med passare och linjal. I ljuset av detta är det intressant att notera att många klassiska satser, av vilka vi ska titta på några i avsnitt 6, har korta eleganta *syntetiska*, dvs koordinatfria (icke-analytiska), bevis, medan bevis med analytiska metoder ofta leder till trassliga kalkyler.

5. Längd och area i euklidisk geometri

Längd av sträckor har vi redan behandlat. Med elementär analysteknik kan man inse att cirkeln har en väldefinierad längd, och om vi definierar talet π som längden av en halvcirkel med radie 1 så följer det att cirkelbågen som svarar mot vinkeln θ får precis längden θ, om $0 < \theta < 2\pi$.

UPPGIFT 5.1. Försök visa att enhetscirkeln har en väldefinierad längd genom att approximera inifrån och utifrån med polygoner.

UPPGIFT 5.2. Visa att en cirkel med radie r har längd $2\pi r$.

På samma sätt vet vi att om $[\alpha, \beta] \ni t \mapsto \gamma(t)$ är en någorlunda reguljär kurva i planet, t ex $\gamma(t)$ är styckvis kontinuerligt deriverbar, så har den en väldefinierad längd

$$\int_\alpha^\beta |\gamma'(t)| dt. \tag{5.1}$$

Vi övergår nu till att studera areor. Areabegreppet i euklidisk geometri visar sig vara kopplat till likformighet och till Pythagoras sats. Vi vill definiera *area* (eller *mått*) för begränsade områden. Det är inte möjligt att definiera area av vilken delmängd D som helst; det krävs att D är (Lebesgue-)mätbar. Vi ska nöja oss med att titta på en mindre klass, klassen av Riemann-mätbara mängder, vilken omfattar alla områden vars rand är styckvis kontinuerligt deriverbar, t ex cirklar, ellipser och polygonområden, speciellt trianglar och rektanglar.

Om måttet (arean) av området D betecknas $m(D)$ så vill vi att följande ska gälla:

- (i) $m(D) \geq 0$ för alla Riemann-mätbara D.
- (ii) Om D är en disjunkt union av Riemann-mätbara delområden $D_1, \ldots, D_m$ så är D Riemann-mätbart och $m(D) = \sum_{j=1}^{m} m(D_j)$.
- (iii) Om D är Riemann-mätbart och D' är kongruent med D så är D' Riemann-mätbart och $m(D) = m(D')$.
- (iv) Om D är en kvadrat med sidan 1 så är $m(D) = 1$.

I (ii) kan man tillåta att områdena överlappar på mängder med måttet noll. I (iv) kan man tillåta både slutna och öppna kvadrater.

Vi bara skissar definitionen här och hänvisar till en lärobok i flervariabelanalys för mer detaljer. Låt oss anta att det existerar en funktion m, definierad för säg alla polygonområden, som uppfyller (i) till (iv). Då följer att $m(Q) = 2^{-2k}$ om Q är en kvadrat med sidan 2^{-k}. Vi fixerar ett koordinatsystem och studerar axelparallella kvadrater Q som har sidlängd 2^{-k} för något heltal k och som har sitt nedre vänstra hörn i en punkt $(\ell 2^{-k}, m2^{-k})$ för några heltal ℓ och m. Detta är s.k. *dyadiska kvadrater*. Vi säger att ett område E ligger i klassen $\mathcal{Q}$ om det är en ändlig union av parvis disjunkta (så när som på att deras kanter kan överlappa) dyadiska kvadrater. Enligt (ii) så ska $m(E)$ vara summan av areorna av de ingående kvadraterna, och eftersom representationen av E som en disjunkt union av dyadiska kvadrater väsentligen är entydig, följer det att talet $m(E)$ är väldefinierat. För en godtycklig mängd D i planet sätter vi nu

$$\underline{m}(D) = \sup\{m(E);\ E \in \mathcal{Q},\ E \subset D\} \quad \text{och}$$
$$\overline{m}(D) = \inf\{m(E);\ E \in \mathcal{Q},\ E \supset D\}.$$

Man ser lätt att $\underline{m}(D) = m(D) = \overline{m}(D)$ om D är ett område i klassen $\mathcal{Q}$. För ett godtyckligt område D är det klart att $\underline{m}(D) \leq \overline{m}(D)$ och vi säger att D är Riemann-mätbart med *måttet* (eller *arean*) $m(D)$ om båda dessa tal är lika. Detta är helt enkelt definitionen av Riemann-integralen

$$\int\int_D dx_1 dx_2.$$

UPPGIFT 5.3. Antag att D är ett område sådant att det till varje $\epsilon > 0$ finns mätbara D_1 och D_2 sådana att $D_1 \subset D \subset D_2$ och

$$m(D_2) - m(D_1) < \epsilon.$$

Visa att D är mätbart och att

$$m(D) = \inf\{m(D_2);\ D_2 \text{ mätbart},\ D_2 \supset D\}.$$

Det är självklart att (i) gäller med vår definition av mått.

UPPGIFT 5.4. Kontrollera att (ii) gäller!

Antag nu att vi kan visa att varje translat av en dyadisk kvadrat är mätbar och har samma mått som den ursprungliga. Då gäller detta för godtyckliga mängder ur $\mathcal{Q}$ enligt (ii). Om nu D är en mätbar mängd så finns till $\epsilon > 0$ mängder $E_1, E_2 \in \mathcal{Q}$ sådana att $E_1 \subset D \subset E_2$ och $m(E_2) - m(E_1) < \epsilon$. Om E_1', E_2' och D' betecknar motsvarande translaterade mängder så har vi att $E_1' \subset D' \subset E_2'$ och $m(E_2') - m(E_1') < \epsilon$. Det följer nu att D' är mätbart och att $m(D') = m(D)$, jmf övning 5.3.

Eftersom varje translation är en sammansättning av en translation i x_1-led och en i x_2-led så kan vi nöja oss med att visa fallet med x_1-led. Vidare kan vi anta att sidan är 1. På detta sätt har vi reducerat det hela till ett väsentligen envariabelt problem och detaljerna lämnar vi som övning.

UPPGIFT 5.5. Genomför dessa!

UPPGIFT 5.6. Visa att varje triangel med en axelparallell sida är Riemann-mätbar.

Vi har alltså klarat av (iii) och (iv) för translationer. Det återstår att behandla fallen spegling i x_1-axeln samt rotation runt origo. Som tidigare räcker det att klara fallet att området D är en dyadisk kvadrat Q. Fallet med spegling är trivialt så låt oss betrakta en rotation runt origo. Genom att sätta samman med lämpliga translationer kan vi anta att Q har ett hörn i origo och att Q' är bilden av Q efter en vridning runt origo. Som figuren visar får man den axelparallella rektangeln från Q' genom att translatera lämpliga trianglar. Det följer från övning 5.6 och (ii) att Q' har samma mått som rektangeln. Å andra sidan inser man lätt att rektangeln har samma mått som Q, genom att translatera lämpliga delrektanglar, och alltså har Q och Q' samma mått.

Exempel 5.1. Eftersom den s.k. *dilatationen* $x \mapsto rx$, $r > 0$, avbildar en kvadrat med sidan α på en kvadrat med sidan $r\alpha$, följer det att en mätbar mängd med måttet m avbildas på en mätbar mängd med måttet $r^2 m$, jmf övning 5.3. □

Exempel 5.2. Låt $A(r)$ vara arean av en cirkelskiva med radie r. Från det förra exemplet följer att $A(r) = r^2 A(1)$. Vi ska nu visa att $A(1)$, dvs arean av Δ, är π.

Om C är cirkelringen $C = \{x \in \mathbb{R}^2;\ r \le |x| < R\}$ så är det rimligt att tänka sig att

$$2\pi r(R-r) < m(C) < 2\pi R(R-r) \tag{5.2}$$

i ljuset av övning 5.2. Givet ett naturligt tal n så kan vi dela upp enhetskivan Δ i cirkelringarna $C_j = \{x \in \mathbb{R}^2;\ (j-1)/n \le |x| < j/n\}$, $1 \le j \le n$. Då får vi att $2\pi(j-1)/n^2 < m(C_j) < 2\pi j/n^2$ och alltså att

$$2\pi \sum_{j=1}^{n} (j-1)/n^2 < m(\Delta) < \sum_{j=1}^{n} j/n^2$$

dvs

$$2\pi \frac{n(n-1)}{2n^2} < m(\Delta) < \frac{(n+1)n}{2n^2}.$$

Om $n \to \infty$ så får vi att $m(\Delta) = \pi$.

Argumentet ovan kan uttryckas lite elegantare: Från (5.2) följer att $A(r)$ är deriverbar och att $A'(r) = 2\pi r$. Eftersom $A(0) = 0$ följer det att $A(r) = r^2\pi$. □

En *höjd* i en triangel är en normal till en av sidorna (eventuellt utgående från förlängningen av sidan) som går genom det tredje hörnet. Om man väljer en höjd så kallas sidan ifråga för *basen*.

h

b

Ur figuren följer det med ett enkelt kongruensresonemang att arean av triangeln är (längden av) basen gånger (längden av) höjden genom två.

Med areabegreppet kan man få ett nytt bevis för Pythagoras sats.

β γ

α

Från figuren och formeln för triangelns area är det nämligen klart att $\gamma^2 + 4\alpha\beta/2 = (\alpha+\beta)^2$, vilket efter förenkling blir $\gamma^2 = \alpha^2 + \beta^2$.

6. Några klassiska satser i euklidisk geometri

Vi ska nu titta på ytterligare några klassiska satser i euklidisk geometri. De flesta av dessa har enkla eleganta syntetiska bevis, medan bevis med analyska metoder är trassligare.

PROPOSITION 6.1 (Yttervinkelsatsen). *En yttervinkel till en triangel är summan av de motstående innervinklarna.*

BEVIS. Detta följer genast av att vinkelsumman är π och att summan av en innervinkel och dess yttervinkel är π. $\square$

Två punkter a och b på en cirkel delar denna i två *cirkelbågar*. Låt ab beteckna en av dessa och låt c vara en punkt på den andra. Då kallas $\wedge acb$ en *randvinkel* (eller *periferivinkel*) stående på bågen ab. Om m är cirkelns medelpunkt kallas vinkeln $\wedge amb$ (som innehåller bågen ab; den kan alltså få vara större än π) bågens *medelpunktsvinkel*.

SATS 6.2 (Satsen om medelpunktsvinkel). *Medelpunktvinkeln är dubbelt så stor som randvinkeln stående på samma båge.*

Speciellt gäller alltså att randvinkeln är rät om den står på en diameter. Vi har även andra omedelbara konsekvenser.

KOROLLARIUM 6.3 (Randvinkelsatsen). *I en cirkel är alla randvinklar stående på samma båge lika stora.*

KOROLLARIUM 6.4. *I en fyrhörning som är inskriven i en cirkel är summan av motstående vinklar π.*

Detta beror på att motsvarande medelpunktsvinklar utgör ett helt varv, dvs 2π.

För att till fullo uppskatta tjusningen med följande syntetiska bevis, så uppmunas läsaren att först försöka knåpa ihop ett analytiskt.

BEVIS AV SATS 6.2. Betrakta först fallet då ena vinkelbenet, säg, ac går genom m. Då är $\wedge amb$ yttervinkel till triangeln mbc och alltså lika med summan av vinklarna vid b och c. Men dessa vinklar är lika, eftersom triangeln är likbent, $|mb| = |mc|$, och alltså är $\wedge amb$ dubbelt så stor som $\wedge acb$, vilket visar satsen i detta specialfall.

c

c c

m m

b d

a b a d a b

Om diametern cmd skär bågen ab i en punkt får vi från det redan bevisade fallet att $\wedge amd = 2\wedge acd$ och $\wedge bmd = 2\wedge bcd$. Genom addition får vi att $\wedge amb = 2 \wedge acb$. Tredje fallet får man på liknande sätt fast genom att subtrahera istället för att addera. $\square$

SATS 6.5. *Vinkeln mellan en tangent till en cirkel och en korda genom tangeringspunkten är lika med en randvinkel stående på kordan.*

Bevis. Eftersom alla randvinklar är lika, kan vi ta den sådan att vinkelbenet vid tangeringspunkten blir en diameter. Då blir vinkeln vid kordans andra ändpunkt rät, och satsen följer. □

Sats 6.6 (Kordasatsen). *Om två kordor ab och cd i en cirkel skär varandra i p (eventuellt efter förlängning) så är*

$$|pa| \cdot |pb| = |pc| \cdot |pd|.$$

Bevis. Dra sträckorna ac och bd. Antingen står $\wedge bac$ och $\wedge bdc$ på samma båge och är därför lika, eller så står de på motsatta bågar och då är summan π. (Man kan undvika det senare fallet genom att vid behov låta c och d byta plats.)

a b p

c

d

I båda fallen följer det att trianglarna pac och pdb har två motsvarande vinklar lika alltså är likformiga. Detta ger den önskade relationen. Beviset går igenom vare sig p ligger utanför eller innanför cirkeln. □

Låt p vara mittpunkten på sträckan ab och låt ℓ vara normalen genom p till ab. Denna kallas *mittpunktsnormalen* och kännetecknas av att varje punkt $p \in \ell$ har samma avstånd till a som till b. Detta beror på att b och a är varandras bilder under spegling i linjen ℓ.

Antag nu att vi har tre punkter a, b, c givna som inte ligger på samma linje. Då är mittpunktsnormalerna till ab och bc inte parallella och skär varandra sålunda i en punkt p. Punkten p har nu samma avstånd till a som till b och till b som till c så p ligger på mittpunktsnormalen till ac. Alltså har vi att mittpunktsnormalerna till alla tre punktparen skär varandra i en punkt. Vi har visat:

Givet en triangel så finns det exakt en cirkel som som går genom hörnen på triangeln.

Denna cirkel kallas den *omskrivna cirkeln* till triangeln.

Det är värt att notera att argumentet fungerar lika bra i absolut geometri, under förutsättning att mittpunktsnormalerna inte är parallella, vilket de mycket väl kan vara, se exempel 3.3 i kapitel 5.

På liknande sätt kan man inse att bisektriserna till en triangel alltid skär varandra i en och samma punkt, vars avstånd till triangelns tre sidor är detsamma. Detta leder till existensen av en *inskriven cirkel.* Vi lämnar detaljerna som övning. Detta argument fungerar även i absolut geometri.

Uppgift 6.1. Visa existensen av en inskriven cirkel till en given triangel.

En *median* är en sträcka som sammanbinder ett hörn i en triangel med mittpunkten på motstående sida.

Sats 6.7 (Mediansatsen). *Medianerna till en triangel skär varandra i en punkt. Dess avstånd till ett hörn är två tredjedelar av längden av medianen till det hörnet.*

Bevis. Kalla triangeln T. Dra de tre sträckorna som sammanbinder mittpunkterna på sidorna i T. Den nya triangel T' som uppstår har enligt det första likformighetsfallet sidor som är hälften så långa som de i T, och speciellt är de båda trianglarna likformiga, enligt det andra likformighetsfallet. Det följer nu av det tredje likformighetsfallet att medianerna till T är medianer till T'.

Om nu inte medianerna skär varandra i en punkt så uppstår en ny triangel T''. På grund av likformigheten måste förhållandet mellan sidorna i T'' och sidorna i T, vara detsamma som förhållandet mellan sidorna i T'' och sidorna i T', vilket är en motsägelse. Alltså skär medianerna varandra i en punkt. (Man kan också upprepa konstruktionen och på så sätt finna att T'' inte kan ha någon utsträckning.) Vi vet nu att medianerna skär varandra i en punkt p. På grund av likformigheten igen följer att avståndet till ett hörn i T är dubbelt så stort som avståndet till motsvarande hörn i T',

d

p a

men det senare avståndet är precis längden av den del av medianen som går från p till motstående sida i T. Beviset är klart. □

Skärningspunkten kallas triangelns *tyngdpunkt.*

Uppgift 6.2. Visa att tyngdpunkten är lika med triangelns tyngdpunkt i fysikalisk mening.

Det är även sant att de tre höjderna till en triangel skär varandra i en punkt. För att se detta, bilda till en given triangel T' en dubbelt så stor triangel T så att de förhåller sig till varandra som i beviset av mediansatsen. Det följer att höjderna till T' är precis mittpunktsnormalerna till T, vilka vi vet skär varandra i en punkt.

Ur Pythagoras sats följer dess omvändning:

Om i en triangel summan av kvadraterna på två av sidorna är lika med kvadraten på den tredje sidan, så är triangeln rätvinklig.

För att se detta, kalla sidlängderna α, β och γ, så att $\alpha^2+\beta^2=\gamma^2$, och bilda en rätvinklig triangel med kateterna α och β. Enligt Pythagoras sats är då hypotenusan γ och alltså är denna rätvinkliga triangel kongruent med den ursprungliga.

En annan enkel konsekvens av Pythagoras sats är *parallellogramlagen*:

I en parallellogram är summan av kvadraterna av diagonalerna lika med summan av kvadraterna av all fyra sidorna.

Uppgift 6.3. Visa parallellogramlagen!

Av likformighetsteorin följer det att $\cos\theta$ och $\sin\theta$ är förhållandena mellan kateten vid θ och hypotenusan, respektive motstående katet och hypotenusan, i en rätvinklig triangel. Följande formel kallas *cosinussatsen.*

Om α, β och γ är sidlängderna i en triangel där γ är motstående sida till vinkeln θ, så är

$$\gamma^2=\alpha^2+\beta^2-2\alpha\beta\cos\theta.$$

Uppgift 6.4. Bevisa cosinussatsen!

Avbildningen $x\mapsto rx$, dvs $(x_1,x_2)\mapsto(rx_1,rx_2)$, där r är ett positivt tal, kallas en *dilatation.* Det är klart att en dilatation är bijektiv.

Uppgift 6.5. Visa att en dilatation avbildar linjer på linjer, bevarar (eller kastar om) ordningsrelationen, att den bevarar vinklar, samt att en sträcka ab avbildas på en sträcka $a'b'$ sådan att $|a'b'|=r|ab|$.

Gruppen av avbildningar som genereras av translationer, speglingar, rotationer och dilatationer, kallas gruppen av *likformighetsavbildningar.*

Uppgift 6.6. Visa att två trianglar är likformiga om och endast om det finns en likformighetsavbildning ϕ som tar den ena på den andra.

Vi säger att en avbildning $T\colon\mathbb{R}^2\to\mathbb{R}^2$ är *affin* om den kan skrivas

$$T(x)=Ax+b,$$

där A är en inverterbar 2×2-matris och $b\in\mathbb{R}^2$. Man kontrollerar lätt att mängden $\mathcal{A}$ av affina transformationer är en grupp, som innehåller gruppen av likformighetsavbildningar.

UPPGIFT 6.7. Visa att en $\phi \in \mathcal{A}$ avbildar linjer på linjer och bevarar (eller kastar om) ordningsrelationerna, samt att mittpunkter på sträckor bevaras.

UPPGIFT 6.8. Visa att det givet två trianglar finns en och endast en $\phi \in \mathcal{A}$ som tar den ena triangeln på den andra.

Transformationerna i $\mathcal{A}$ bevarar inte vinklar, men relativa avstånd på linjer, t ex mittpunkter. Därför avbildas medianerna i en triangel på medianerna i bildtriangeln.

UPPGIFT 6.9. Använd detta för att ge ett nytt bevis för mediansatsen.

För fler klassiska resultat hänvisar vi till [**LÅL**] och [**OH**].

7. Konstruktioner med passare och linjal

De enda perfekta geometriska figurerna var cirklar och linjer ansåg man i antiken. Därför ville man när man skulle göra geometriska konstruktioner begränsa sig till att använda *passare* och *linjal*.

En *konstruktion* innebär att man utifrån två punkter (vars avstånd kan tas som enhetssträcka) successivt skaffar sig en större och större punktmängd genom följande två operationer:

(1) Genom ett godtyckligt par av de givna punkterna får man dra en rät linje (linjal).

(2) Man får dra en cirkel med medelpunkt i en av de givna punkterna och radie lika med avståndet mellan ett godtyckligt par av de givna punkterna (passare).

Nya punkter är de man får som skärningar mellan cirklar och linjer som man erhållit genom operationerna (1) och (2).

ANMÄRKNING 7.1. Ibland har man restriktionen i operation (2) att cirkeln måste gå genom en av de tidigare punkterna (man tänker sig att passaren "slår ihop" så fort den lämnar pappret). Man kan dock visa att vår operation (2) kan uppnås genom ett antal av dessa mer restriktiva operationer, så att det i slutändan blir samma sak. □

Man ser lätt att man kan avsätta en given sträcka från en given annan punkt på en given stråle från denna punkt. Vidare bildar man lätt mittpunkten av sträckan ab. Dra nämligen en cirkel genom a med centrum i b och tvärtom. Linjen genom cirklarnas skärningspunkter blir normal till linjen ab och delar sträckan ab i två lika delar. Man kan också föreskriva en punkt p genom vilken en normal till en given linje ℓ ska gå. Genom att dra en cirkel med medelpunkt i p och tillräckligt stor radie, så får man två punkter a och b på ℓ med lika avstånd till p, och man får den önskade normalen som ovan. Det är nu lätt att dra en linje genom p som är parallell med ℓ. Dra först normalen genom p till ℓ och dra normalen till denna. Om vi inför ett koordinatsystem på ett sådant sätt att enhetssträckan är konstruerbar, följer det nu att en punkt är konstruerbar om och endast om dess koordinater är konstruerbara. I fortsättningen kan vi alltså tala om huruvida vissa tal är konstruerbara istället för vissa punkter.

Man ser lätt att om a och b är konstruerbara så är även $a + b$ och $a - b$ det. Antag nu att $b \neq 0$. Bilda triangeln med hörn i $(0, 0)$, $(0, a)$ och $(b, 0)$. Linjen genom $(1, 0)$ som är parallell med hypotenusan kommer att skära y-axeln i a/b. Alltså kan man konstruera a/b och på liknande sätt ab.

Uppgift 7.1. Visa att om a kan konstrueras så kan även $\sqrt{a}$ konstrueras. Ledning: Visa att om talen α och β är konstruerbara så kan man konstruera en triangel med hypotenusan β och ena kateten α. Tag $\beta = (1 + a)/2$ och $\beta = (1 - a)/2$ om $a < 1$.

Uppgift 7.2. Antag man har en vinkel given samt en stråle ab från a. Visa att man kan konstruera en vinkel vid a som är lika stor som den givna vinkeln och som har ab som ena vinkelbenet.

Uppgift 7.3. Visa att en given vinkel (en vinkel definieras av tre punkter) kan delas i två lika vinklar.

Några till synes enkla konstruktioner klarade man dock inte, som t ex att givet en cirkel hitta en kvadrat med samma area, eller att dela en given vinkel i tre lika delar. Detta förklaras av att det, som det senare har visat sig, helt enkelt är omöjligt. Vi ska här ge ett bevis för att vinkeln $\pi/3$ inte kan tredelas.

Uppgift 7.4. Visa att vinkeln $\pi/3$ verkligen kan bildas.

Kvadratroten $\sqrt{a}$ ur ett positivt tal a är som bekant det entydigt bestämda positiva tal vars kvadrat är a. Antag att c är ett rationellt tal sådant att $\sqrt{c}$ inte är rationell, och betrakta mängden $\mathbb{Q}(\sqrt{c})$ av alla tal $a + b\sqrt{c}$ där $a, b \in \mathbb{Q}$. Det är klart att $\mathbb{Q}(\sqrt{c})$ är en ring. Om $a + b\sqrt{c} \neq 0$ så är $a^2 - b^2c \neq 0$, (ty annars vore ju $\sqrt{c}$ rationellt) och alltså ligger

$$\frac{1}{a + b\sqrt{c}} = \frac{a - b\sqrt{c}}{a^2 - b^2c}$$

i $\mathbb{Q}(\sqrt{c})$. Det följer att $\mathbb{Q}(\sqrt{c})$ i själva verket är en kropp. Antag nu att $c_2 \in \mathbb{Q}(\sqrt{c_1})$ men saknar kvadratrot i $\mathbb{Q}(\sqrt{c_1})$. På samma sätt kan vi då adjungera $\sqrt{c_2}$ och bilda en ny kropp $\mathbb{Q}(\sqrt{c_1}, \sqrt{c_2})$, som består av alla tal som kan skrivas $a + b\sqrt{c_2}$ där $a, b \in \mathbb{Q}(\sqrt{c_1})$ etc. Vi säger att ett tal α kan uttryckas genom ändligt många (kvadrat)rotsutdragningar om α ligger i någon kropp $\mathbb{Q}(\sqrt{c_1}, \sqrt{c_2}, \ldots, \sqrt{c_m})$. (Enkelt uttryckt, betyder detta att talet α går att skriva upp med rationella tal och ändligt många (kvadrat)rottecken.) Det följer från ovan att varje sådant tal α kan konstrueras med passare och linjal om alla talen $c_1, \ldots, c_m$ är positiva. Omvänt har vi

Proposition 7.2. *Om ett tal α är konstruerbart så kan man uttrycka det genom ett ändligt antal (kvadrat)rotutdragningar.*

Bevis. Vid skärning av linje eller cirkel med linje eller cirkel ges skärningspunkterns koordinater som lösningar till ekvationer av första eller andra graden (med koefficienter som är rationella uttryck i tidigare konstruerade tal), och dessa lösningar kan alltid uttryckas med kvadratrötter. Propositionen följer nu med induktion. □

Sats 7.3. *Vinkeln $\pi/3$ kan inte tredelas med passare och linjal.*

Bevis. Om detta var möjligt så skulle vi kunna konstruera talet $\alpha = 2\cos(\pi/9)$. Nu är

$$\cos(3\theta) = 4\cos^3\theta - 3\cos\theta,$$

och eftersom $\cos(3\pi/9) = 1/2$ så betyder detta att α är ett nollställe till polynomet

$$f(x) = x^3 - 3x - 1.$$

Man ser lätt att $f(x)$ saknar rationella rötter; om $f(p/q) = 0$ där p/q är relativt prima, så är $q^3 = p(p^2 - 3q^2)$ vilket medför att p delar q^3, och detta är en motsägelse.

Antag nu att man kan uttrycka en rot α till $f(x)$ genom ändligt många (kvadrat)rotutdragningar. Det finns då en rot $\alpha = a + b\sqrt{c_m} \in \mathbb{Q}(\sqrt{c_1}, \ldots, \sqrt{c_m})$, dvs $a, b, c_m \in \mathbb{Q}(\sqrt{c_1}, \ldots, \sqrt{c_{m-1}})$ men $\alpha \notin \mathbb{Q}(\sqrt{c_1}, \ldots, \sqrt{c_{m-1}})$. Vi kan också anta att $f(x)$ saknar rot i $\mathbb{Q}(\sqrt{c_1}, \ldots, \sqrt{c_{m-1}})$. Att $f(\alpha) = 0$ innebär att

$$0 = a' + b'\sqrt{c_m},$$

där a' och b' är rationella uttryck i a, b och c_m. Därför måste $b' = 0$ ty annars skulle $\sqrt{c_m}$, och följaktligen α, ligga i $\mathbb{Q}(\sqrt{c_1}, \ldots, \sqrt{c_{m-1}})$. Men detta betyder att även $\alpha' = a - b\sqrt{c_m}$ är en rot. Låt nu r vara den tredje roten till $f(x)$. Eftersom produkten av samtliga rötter är 1 så följer att $r = (a^2 - b^2 c_m)^{-1}$ (man kan även uppnå detta genom att dividera ut

$(x-\alpha)(x-\alpha')$ från $f(x)$) och alltså ligger roten r i $\mathbb{Q}(\sqrt{c_1},\dots,\sqrt{c_{m-1}})$ vilket är en motsägelse. □

Omöjligheten att med passare och linjal konstruera en kvadrat med arean π följer av det faktum att π är transcendent. Ett bevis för transcendensen av π kan man hitta i [**IS**].

8. Sfärisk geometri

I detta avsnitt ska vi mycket kort nämna något om geometri på sfären

$$S^2 = \{(x_1, x_2, x_3) \in \mathbb{R}^3;\ x_1^2 + x_2^2 + x_3^2 = 1\}$$

i $\mathbb{R}^3$. För att hålla framställningen kort ska vi, utan att ge bevis, utnyttja några välkända fakta om geometri i $\mathbb{R}^3$. Avståndet $d(a, b)$ mellan två punkter a och b på S^2 är lika med vinkeln mellan motsvarande vektorer i $\mathbb{R}^3$ och

$$\cos d(a, b) = a \cdot b = a_1 b_1 + a_2 b_2 + a_3 b_3. \tag{8.1}$$

Vidare ska vi ta för givet att rimliga delmängder till S^2 har en välbestämd area och att totalarean är 4π.

Linjerna på S^2 ska vara *storcirklarna*, och en storcirkel är skärningen med S^2 och ett plan genom origo i $\mathbb{R}^3$. Man ser genast att denna geometri inte uppfyller axiomen A1 till A4. T ex skär varje par av linjer varandra i exakt två punkter, vilka är antipodiska till varandra (två punkter a och b är *antipodiska* om $a = -b$). Vidare är inte en storcirkel (på något naturligt sätt) ordningsisomorf med $\mathbb{R}$. Inte desto mindre finns det flera likheter mellan geometrin på S^2 och den hyperboliska geometrin.

Varje linje (storcirkel) ℓ upp sfären i två halvsfärer, *sidorna* av ℓ. På varje linje har vi en slags ordningsrelation som gör linjen ordningsisomorf med enhetscirkeln i $\mathbb{R}^2$. Två olika punkter a och b på en linje på S^2 bestämmer via denna ordningsrelation två delmängder vilka båda kallas *sträckor* med ändpunkter a och b. Givet två olika punkter a och b på S^2 som inte är antipodiska till varandra så finns det exakt en linje genom dessa punkter.

Det finns en grupp av avbildningar, gruppen av rotationer av $\mathbb{R}^3$, som håller S^2 fixt. Dessa avbildningar bevarar vinklar, avstånd och areor. Givet två punkter a och a' så finns det en rotation ϕ av $\mathbb{R}^3$ som tar a på a'. Man kan också kräva att en given linje genom a avbildas på en given linje genom a'. Man kan även kräva att riktningen bevaras, dvs att små sträckor i en given riktning från a på linjen ifråga avbildas på små sträckor i en given riktning från a'. Om man dessutom kräver att en given sida av linjen genom a ska avbildas på en given sida av linjen genom a' så blir avbildningen ϕ entydigt bestämd, jmf axiom A4 i teorin A.

En *(sfärisk) triangel* består av tre punkter (hörnen) och de tre sträckorna (sidorna) som sammanbinder hörnen parvis så att sidorna är mindre än π.

Tre plan i $\mathbb{R}^3$ som bara har origo som gemensam skärningspunkt delar upp $\mathbb{R}^3$ i åtta delar. Alltså kommer tre linjer på S^2 som inte alla tre går genom en och samma punkt, att dela upp S^2 i åtta trianglar som är parvis kongruenta eftersom de ligger antipodiskt. Låt T vara en av dessa, med hörn a, b och c och vinklar α, β respektive γ. Linjerna ab och ac ger upphov till fyra områden. Det av dem som innehåller T kallar vi O_a. Eftersom O_a upptar $\alpha/2\pi$ av hela sfärens area som är 4π så är arean av O_a lika med 2α. Triangeln $O_a \setminus T$ kallar vi T_a. Vi får alltså att

$$m(T) + m(T_a) = 2\alpha.$$

På samma sätt definerar vi T_b och T_c och får relationerna $m(T) + m(T_b) = 2\beta$ och $m(T) + m(T_c) = 2\gamma$. Men samtidigt måste vi ha att trianglarna T, T_a, T_b och T_c tillsammans har arean 2π eftersom de tillsammans utgör precis halva sfären, och alltså har vi att

$$m(T) + m(T_a) + m(T_b) + m(T_c) = 2\pi.$$

Sätter vi samman detta får vi

$$m(T) = \alpha + \beta + \gamma - \pi. \tag{8.2}$$

Alltså är vinkelsumman i en triangel alltid *större* än π och avvikelsen, som kallas triangelns *excess*, är lika med arean, vid lämplig normalisering. Detta ska jämföras med defekt av en triangel i hyperbolisk geometri som, vilket vi ska se i kapitel 5, också är lika med arean.

Antag att sidorna i en rätvinklig triangel har längder A, B respektive C. Genom att använda en rotation i $\mathbb{R}^3$ kan vi anta att hörnet med rät vinkel ligger i $(1, 0, 0)$ och att de båda andra hörnen ligger i $(\cos A, \sin A, 0)$ respektive $(\cos B, 0, \sin B)$. Vi får då att

$$\cos C = \cos A \cos B. \tag{8.3}$$

Notera att eftersom $\cos x \approx 1 - x^2/2$ så innebär (8.3) att $C^2 \approx A^2 + B^2$ om sidlängderna A, B och C är små, vilket ju stämmer med Pythagoras sats i det euklidiska fallet.

Anmärkning 8.1. Detta otillfredställande faktum att två linjer alltid skär varandra i ett par av antipodiska punkter kan man bli kvitt genom att identifiera varje par av antipodiska punkter med en punkt. Mängden av alla par av antipodiska punkter betecknas $\mathbb{P}^2$ och kallas det tvådimensionella *projektiva rummet*. Det är klart att begreppet linje får mening även i $\mathbb{P}^2$, och det följer att varje par av olika linjer skär varandra i exakt en punkt. Gruppen av rotationer i $\mathbb{R}^3$ inducerar nu en grupp av, vad man kallar, kongruensavbildningar på $\mathbb{P}^2$. Man kan visa att denna geometri innehåller den euklidiska geometrin, dvs att punkterna och linjerna och kongruensavbildningarna i planet kan uppfattas som delmängder av motsvarande objekt för $\mathbb{P}^2$. □

9. Ytterligare övningar till kapitel 4

Uppgift 9.1. Fixera en punkt 0. Visa att man kan välja en orientering så att man kan definiera en avbildning $a \mapsto \psi(a)$ genom att man från a avsätter sträckan $0a$ vinkelrätt mot sig själv (i orienteringens riktning) och låter $\psi(a)$ vara punkten där man då hamnar. Är ψ en kongruenstransformation (absolut geometri)?

En mängd kallas *konvex* om för varje par av punkter i mängden sträckan mellan dem också ligger i mängden.

Uppgift 9.2. Visa att en fylld triangel (dvs dess hörn och sidor samt alla punkterna i triangeln) är konvex (absolut geometri).

Uppgift 9.3. Visa att en cirkelskiva, dvs mängden av alla punkter som har högst avstånd r till en fix punkt m, är konvex.

Uppgift 9.4. Antag ℓ och ℓ' är linjer och $a \in \ell$ och $a' \in \ell'$ sådana att linjen mellan a och a' skär ℓ och ℓ' rätvinkligt. Visa att om $b \in \ell$ och $b' \in \ell'$ så är $|bb'| \geq |aa'|$. Visa också att det i det ickeeuklidiska fallet är likhet om och endast om $b = a$ och $b' = a'$ (absolut geometri).

Uppgift 9.5. Ge en innebörd åt begreppet avstånd mellan två linjer. Låt ℓ vara en linje och a en punkt utanför ℓ. Visa att det går exakt en linje ℓ' genom a sådan att avståndet mellan ℓ' och ℓ är lika stort som avståndet mellan a och ℓ (absolut geometri).

Uppgift 9.6. Antag att två cirklar tangerar varandra i A (antag att den ena ligger i den andra). Dra genom A två vinkelräta kordor AB och AC, en i vardera cirkeln. Visa att radierna till B och C är parallella (euklidisk geometri).

Uppgift 9.7. Två cirklar tangerar varandra i A. En gemensam tangent tangerar ena cirkeln i B och den andra i C. Visa att vinkeln BAC är rät (euklidisk geometri).

Uppgift 9.8. Dra normalerna från en punkt P inuti en liksidig triangel mot triangelns sidor. Visa att summan av dessa normalers längder är lika med längden av triangelns höjd (euklidisk geometri).

Uppgift 9.9. Två cirklar skär varandra i punkterna A och B, och P är en punkt på ena cirkeln. Linjerna PA och PB skär den andra cirkeln i C och D. Visa att kordan CD är parallell med tangenten i P (euklidisk geometri).

Uppgift 9.10. Antag att C och D är två punkter på samma sida om linjen AB. Antag att i trianglarna ACB och ADB vinklarna vid C respektive D är lika stora. Visa att punkterna A, B, C och D ligger på en cirkel (euklidisk geometri).

Uppgift 9.11. Antag att i en fyrhörning båda summorna av motstående vinkelpar är π. Visa att hörnen ligger på en cirkel (euklidisk geometri).

Uppgift 9.12. Två cirklar skär varandra i A och B. Tag två punkter C och D på ena cirkeln och punkter E och F på den andra cirkeln sådana att C, A och E ligger på en rät linje och D, B, F ligger på en parallell linje. Visa att $|CE| = |DF|$ (euklidisk geometri).

Uppgift 9.13. Två cirklar med radier r_1 och r_2 skär varandra med räta vinklar. Vad är avståndet mellan cirklarnas medelpunkter (euklidisk geometri)?

Uppgift 9.14. Två cirklar med radie 1 skär varandra med vinklar θ. Hur långa bli cirkelbågarna mellan skärningspunkterna (euklidisk geometri)?

Några kommentarer

För en mer omfattande framställning av absolut och euklidisk geometri som mer direkt ansluter sig till Euklides framställning i Elementa, se [**LÅL**]. Denna innehåller även en hel del historik. Se även [**OH**]. En mer utförlig diskussion om konstruktioner och kroppsutvidgningar finns i [**IS**]. En trevlig översiktlig framställning av olika geometrier, med flera historiska notiser, finns i [**JL**].

KAPITEL 5

Hyperbolisk geometri

Vi ska nu studera en modell, Poincaré-modellen, för absolut geometri i vilken parallellaxiomet inte gäller; en modell för *hyperbolisk geometri*. I likhet med det euklidiska fallet är det sant att alla modeller för hyperbolisk geometri är är isomorfa, men detta kommer vi inte att bevisa. Poincaré-modellen är konstruerad inom den euklidiska modellen för absolut geometri vilket betyder att A+¬PA är konsistent givet att A+PA är konsistent. Närmare bestämt konstrueras Poincarémodellen i enhetsskivan i $\mathbb{R}^2$, men för att kunna beskriva gruppen av kongruensavbildningar måste vi först studera de geometriska egenskaperna hos en större klass av avbildningar i planet som kallas Möbiusavbildningar och som speciellt innehåller alla likformighetsavbildningar i planet.

1. Möbiusavbildningar

Det visar sig praktiskt att använda komplexa tal så låt oss titta på hur likformighetsavbildningarna uttrycks med komplexa tal. Om vi identifierar $\mathbb{R}^2$ med de komplexa talen, genom att en punkt (a, b) i planet svarar mot det komplexa talet $z = a + ib$, så är translationer avbildningar av typen $z \mapsto z + \beta$ för $\beta \in \mathbb{C}$, spegling i realaxeln (x-axeln) är avbildningen $z \mapsto S(z) = \bar{z}$, och rotation med vinkel θ är avbildningen $z \mapsto e^{i\theta} z$. Varje euklidisk kongruensavbildning ϕ ges alltså som $\phi(z) = \lambda z + \beta$ där $|\lambda| = 1$ eller som $\phi(z) = \lambda \bar{z} + \beta$. En dilatation med $r > 0$ ges av $z \mapsto rz$.

Givet en linje så finns en euklidisk kongruenstransformation som avbildar linjen på y-axeln. Eftersom denna kan uttryckas som $z + \bar{z} = 2 \operatorname{Re} z = 0$ så ges en godtycklig linje av en ekvation av typ $\bar{a}z + a\bar{z} = r$ där $|a| = 1$ och $r \geq 0$.

Vi ska nu titta på en större klass av avbildningar, nämligen alla som går att skriva som

$$\psi(z) = \frac{az + b}{cz + d} \tag{1.1}$$

för några komplexa tal a, b, c, d, där man utesluter det urartade fallet när $ad = bc$. Observera att ψ är definierad för varje z i planet utom $-d/c$. Sådana här avbildningar brukar kallas *Möbiusavbildningar.*

Om vi låter *M-cirkel* betyda (euklidisk) cirkel eller (rät) linje i planet så har vi

Proposition 1.1. *En Möbiusavbildning avbildar varje M-cirkel på en annan M-cirkel, och skärningsvinkeln mellan två M-cirklar bevaras.*

Bevis. Eftersom

$$\psi(z) = \frac{az+b}{cz+d} = \frac{a}{c} + \frac{a}{c}\Big(\frac{b}{a} - \frac{d}{c}\Big)\frac{1}{z+d/c},$$

ser man att varje Möbiustransformation är en sammansättning av avbildningar av typerna
(i) $z \mapsto z + \beta$,
(ii) $z \mapsto \alpha z$, där $\alpha \neq 0$, och
(iii) $z \mapsto 1/z$,

så det räcker att visa propositionen för var och en av dessa.

En avbildning av typ (i) är en translation. Om $\alpha = re^{i\theta}$ på polär form så betyder $z \mapsto \alpha z$ att man dels roterar z runt origo med vinkeln θ och dels multiplicerar z med faktorn $r > 0$. Att translationer och rotationer avbildar linjer på linjer och bevarar vinklar vet vi redan. Vi påstår nu att även en dilatation $z \mapsto rz$ bevarar vinklar och avbildar linjer på linjer (medan däremot en sträcka avbildas på en ny sträcka som är r gånger den ursprungliga). Detta är självklart för linjer genom origo samt vinklar och sträckor vid origo, men eftersom $rz = r(z-a)+ra$ så följer det även för vinklar och sträckor vid a, eftersom vi redan vet att translationer bevarar längder och vinklar. Det följer att avbildningar av typ (i) och (ii) även avbildar cirklar på nya cirklar.

Det återstår att titta på avbildningen $z \mapsto w = 1/z$. Vi studerar bilden av cirkeln $|z-a|^2 = r^2$ under denna avbildning. Eftersom

$$|z-a|^2 = |z|^2 - a\bar{z} - \bar{a}z + |a|^2$$

kan cirkelns ekvation skrivas

$$|z|^2 - a\bar{z} - \bar{a}z + |a|^2 - r^2 = 0,$$

och insättning av $w = 1/z$ och förlängningning med $|w|^2$ ger

$$(|a|^2 - r^2)|w|^2 - aw - \bar{a}\bar{w} + 1 = 0. \tag{1.2}$$

Om nu $|a| \neq r$, dvs att den ursprungliga cirkeln inte går genom origo, så är (1.2) ekvivalent med

$$|w - b|^2 = t^2,$$

där

$$t^2 = \frac{r^2}{(|a|^2 - r^2)^2}$$

och

$$b = \frac{\bar{a}}{|a|^2 - r^2}.$$

Alltså blir bilden i detta fall en cirkel med medelpunkt b och radie t. I fallet att den ursprungliga cirkeln går genom origo får vi som bild linjen

$$1 - aw - \bar{a}\bar{w} = 0.$$

Varje linje som inte går genom origo kan ges på denna form, och eftersom $z \mapsto 1/z$ är sin egen invers har vi även visat att en sådan linje avbildas på en cirkel genom origo. En linje genom origo ges av parametriseringen $t \mapsto te^{i\theta}$, och bilden blir då $t \mapsto te^{-i\theta}$ som också är en linje genom origo.

Det gäller nu att visa att skärningsvinkeln mellan två M-cirklar (dvs mellan deras tangenter i skärningspunkten) bevaras under $z \mapsto 1/z$. Observera att vi först kan multiplicera med ett nollskilt tal a, eftersom vi kan kompensera detta, efter att ha applicerat avbildningen, genom att ånyo multiplicera med a. Därför kan vi anta att M-cirklarna skär varandra i punkten 1, och eftersom det räcker att jämföra med skärningsvinkeln gentemot x-axeln kan vi anta att ena M-cirkeln, eller snarare dess tangent i skärningspunkten, är x-axeln och att den andra är en annan linje ℓ genom 1, se figur.

Eftersom 1 avbildas på 1 och bilden av ℓ är en cirkel (eftersom ℓ inte går genom origo) så är bilden C av ℓ en cirkel genom punkten 1. Vi påstår nu att C tangerar ℓ i 1. Om inte så skär den nämligen ℓ i exakt en punkt till, säg i a, och då måste a ligga antingen över eller under

x-axeln. Eftersom $z \mapsto 1/z$ är sin egen invers och x-axeln avbildas på sig själv, följer att unionen av ℓ och C avbildas på sig själv under avbildningen, men om säg a ligger i övre halvplanet så ligger bilden av a i undre halvplanet vilket leder till motsägelse. Alltså är ℓ tangent till C och följaktligen har de samma vinkel gentemot x-axeln. $\square$

ANMÄRKNING 1.2. Avbildningen $z \mapsto 1/\bar{z}$, dvs $re^{i\theta} \mapsto (1/r)e^{i\theta}$, kallas *inversion* i enhetscirkeln. Avbildningen $z \mapsto 1/z$ är alltså en inversion sammansatt med en spegling i x-axeln. Man säger att punkterna z och $1/\bar{z}$ är *konjugerade* till varandra med avseende på enhetscirkeln. $\square$

Betrakta Möbiusavbildningen (1.1). Observera att $|\phi(z)| \to \infty$ då $z \to -d/c$, och att $\phi(z) \to a/b$ då $|z| \to \infty$. Det är därför naturligt att sätta $\phi(-d/c) = \infty$ och $\phi(\infty) = a/c$, och om vi låter $\widehat{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$ så finner vi lätt att ϕ blir en bijektion på $\widehat{\mathbb{C}}$, och inversen är en ny Möbiusavbildning. Enligt utredningen ovan räcker det nämligen att kontrollera detta för avbildningar på formen $z \mapsto az + b$ samt inversionen $z \mapsto 1/\bar{z}$. Den första av dessa är en bijektion av $\mathbb{C}$ och tar ∞ på ∞, medan inversionen är en bijektion på $\mathbb{C} \setminus \{0\}$ och tar 0 på ∞ och vice versa. Man kallar $\widehat{\mathbb{C}}$ för *det utvidgade talplanet*.

Antag att a, b, c är tre olika punkter i $\widehat{\mathbb{C}}$. Då är

$$\phi(z) = \frac{z-a}{b-a}\frac{b-c}{z-c} \tag{1.3}$$

i en Möbiusavbildning sådan att $a \mapsto 0$, $b \mapsto 1$ och $c \mapsto \infty$. Antag nu att

$$\phi = \frac{\alpha z + \beta}{\gamma z + \delta}$$

är sådan att $0 \mapsto 0$, $1 \mapsto 1$ och $\infty \mapsto \infty$. Eftersom $\infty \mapsto \infty$ så måste till att börja med $\gamma = 0$ och $\delta \neq 0$. Efter förkortning kan vi anta att $\delta = 1$ och alltså $\phi(z) = \alpha z + \beta$. Eftersom $0 \mapsto 0$ så är nu $\beta = 0$ och $1 \mapsto 1$ medför att $\alpha = 1$. Eftersom mängden av Möbiusavbildningar är en grupp så följer det nu:

Givet tre olika punkter a, b, c i $\widehat{\mathbb{C}}$ och tre andra olika punkter a', b', c' i $\widehat{\mathbb{C}}$ så finns en och endast en Möbiusavbildning ϕ sådan att $a \mapsto a'$, $b \mapsto b'$ och $c \mapsto c'$.

Vi avslutar med att bestämma alla Möbiusavbildningar som avbildar $\widehat{\mathbb{R}} = \mathbb{R} \cup \{\infty\}$ på sig själv. Till att börja med är det klart att varje Möbiusavbildning med reella koefficienter gör detta. Omvänt, om en Möbiusavbildning ϕ tar $\widehat{\mathbb{R}}$ på sig själv, så måste speciellt 0, 1 och ∞ vara bilderna av tre punkter a, b, c i $\widehat{\mathbb{R}}$. Men den entydiga Möbiusavbildningen ϕ som gör detta ges av formel (1.3), och får därför reella koefficienter. Alltså har vi visat att en Möbiusavbildning avbildar $\widehat{\mathbb{R}}$ på sig själv om och endast om den kan skrivas på formen (1.1) med a, b, c, d reella.

Anmärkning 1.3. Man kan identifiera punkterna i det utvidgade talplanet $\widehat{\mathbb{C}}$ med punkterna på sfären $S^2 \subset \mathbb{R}^3$ genom avbildningen

$$\begin{aligned} z = re^{i\theta} \mapsto \psi(z) &= \frac{1}{1+|z|^2}(2z, |z|^2 - 1) = \\ &= \frac{1}{1+r^2}(2r\cos\theta, 2r\sin\theta, r^2 - 1), \quad \infty \mapsto (0, 0, 1), \end{aligned}$$

där vi har identifierat det komlpexa z-planet med x_1x_2-planet. Geometriskt innebär denna avbildning att man låter z-planet vara x_1x_2-planet i $\mathbb{R}^3$, sedan tar linjen från z till "nordpolen" $(0, 0, 1)$ på S^2, och låter $\psi(z)$ vara skärningen med denna linje och S^2.

$\psi(z)$

z

Motsvarande avbildning från S^2 till (det utvidgade) planet kallas *stereografisk projektion.* Det är sant men inte alldeles självklart att M-cirklar i planet svarar mot cirklar på S^2. Vi lämnar åt läsaren att fundera över detta. Vidare kan man visa att rotationerna av S^2 svarar precis mot gruppen som genereras av speglingen $z \mapsto \bar{z}$ samt de Möbiusavbildningar på formen (1.1) sådana att $AA^* = I$, om A är matrisen

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

och A^* är konjugatet av motsvarande transponerade matris. En komplex matris sådan att $AA^* = I$ kallas *unitär.* □

2. Poincaré-modellen för hyperbolisk geometri

Låt Δ beteckna den öppna enhetsskivan i (det komplexa tal-)planet. Som punkter i denna modell tar vi punkterna i Δ. En linje, eller Poincaré-linje när vi behöver vara tydliga, ska vara snittet med Δ och en cirkel eller rät linje i $\mathbb{C}$ som skär enhetscirkeln $\partial\Delta$ rätvinkligt, se figur,

och som kongruenstransformationer $\phi \in \mathcal{K}$ ska vi ta alla avbildningar som ges som

$$\phi(z) = \lambda \frac{z - \alpha}{1 - \bar{\alpha} z}$$

eller

$$\phi(z) = \bar{\lambda} \frac{\bar{z} - \bar{\alpha}}{1 - \alpha \bar{z}},$$

där $\alpha \in \Delta$ och $\lambda \in \partial\Delta$, dvs $|\alpha| < 1$ och $|\lambda| = 1$. Vi har att visa att dels dessa avbildningar ϕ verkligen är bijektioner av Δ och dels att A1 t.o.m. A5 är uppfyllda.

För $\alpha \in \Delta$ sätter vi

$$\phi_\alpha(z) = \frac{z - \alpha}{1 - \bar{\alpha} z}.$$

Denna avbildning är definierad för $z \neq 1/\bar{\alpha}$ (och $1/\bar{\alpha}$ ligger ju utanför Δ), speciellt i Δ, och vi har

Proposition 2.1. *Varje ϕ_α avbildar Δ bijektivt på sig själv och enhetscirkeln bijektivt på sig själv, och inversen är $\phi_{-\alpha}$.*

Bevis. En direkt kalkyl ger att (kom ihåg att $|z|^2 = z\bar{z}$)

$$|\phi_\alpha(z)|^2 = \frac{|z - \alpha|^2}{|1 - \bar{\alpha} z|^2} = \frac{|1 - \bar{\alpha} z|^2 - (1 - |z|^2)(1 - |\alpha|^2)}{|1 - \bar{\alpha} z|^2} = 1 - \frac{(1 - |z|^2)(1 - |\alpha|^2)}{|1 - \bar{\alpha} z|^2},$$

och högerledet är ≤ 1 och med likhet om och endast om $|z| = 1$, så ϕ_α tar Δ in i sig själv och enhetscirkeln in i sig själv. Nu är

$$\phi_\beta \circ \phi_\alpha(z) = \frac{\frac{z-\alpha}{1-\bar{\alpha}z} - \beta}{1 - \bar{\beta}\frac{z-\alpha}{1-\bar{\alpha}z}} = \frac{1+\bar{\alpha}\beta}{1+\alpha\bar{\beta}} \frac{z-\sigma}{1-\bar{\sigma}z},$$

där

$$\sigma = \frac{\alpha+\beta}{1+\bar{\alpha}\beta} = \phi_{-\alpha}(\beta).$$

Speciellt ser man att $\phi_\alpha \circ \phi_{-\alpha}(z) = z = \phi_{-\alpha} \circ \phi_\alpha(z)$ så ϕ_α är bijektiv med invers $\phi_{-\alpha}$. □

Vi kan nu visa

PROPOSITION 2.2. *$\mathcal{K}$ är en grupp, dvs en sammansättning av två avbildningar ur $\mathcal{K}$ är en ny avbildning i $\mathcal{K}$, och varje avbildning i $\mathcal{K}$ har en invers som också ligger i $\mathcal{K}$.*

BEVIS. Om $R_\lambda(z) = \lambda z$ och $S(z) = \bar{z}$ så kan varje element i $\mathcal{K}$ skrivas som $R_\lambda \circ \phi_\alpha$ eller som $S \circ R_\lambda \circ \phi_\alpha$. Vi har redan sett att ϕ_α är en bijektion av Δ, och R_λ och S är en rotation runt origo respektive spegling i x-axeln, och dessa är förstås bijektioner av Δ, och deras respektive inverser ligger i $\mathcal{K}$. Det räcker nu att visa att sammansättningar av element i $\mathcal{K}$ ligger i $\mathcal{K}$. Det följer från förra beviset att det, givet $\alpha, \beta \in \Delta$, finns $\lambda \in \partial\Delta$ och $\sigma \in \Delta$ sådana att $\phi_\beta \circ \phi_\alpha = R_\lambda \circ \phi_\sigma$. Man får nu det allmänna fallet genom att kombinera med relationerna

$$\phi_\alpha \circ R_\lambda = R_\lambda \circ \phi_{\bar{\lambda}\alpha}, \quad \phi_\alpha \circ S = S \circ \phi_{\bar{\alpha}}, \quad R_\lambda \circ S = S \circ R_{\bar{\lambda}},$$

vilkas verifikationer vi lämnar som övning. □

VERIFIKATION AV ATT AXIOMEN A1 TILL A5 ÄR UPPFYLLDA. Vi är nu rustade för att visa att axiomen A1 t.o.m. A5 är uppfyllda i vår modell. Notera först att varje $\phi \in \mathcal{K}$ är en Möbiusavbildning eller konjugatet av en sådan, och enligt proposition 1.1 alltså avbildar M-cirklar på M-cirklar och bevarar vinklar. Eftersom våra ϕ i $\mathcal{K}$ även tar Δ på Δ och $\partial\Delta$ på $\partial\Delta$, följer det att ϕ tar Poincaré-linjer på Poincaré-linjer i vår modell. Från och med nu låter vi linje betyda Poincaré-linje om inget annat sägs.

A1: Antag vi har två punkter i Δ varav en är origo. Då är det klart, se exempel 2.3 nedan, att det finns precis en linje genom dessa punkter, nämligen den euklidiska linjen genom dem snittad med Δ. Om nu a och b är godtyckliga i Δ så kan man använda avbildningen ϕ_a som tar a på origo. Inversa bilden, $\phi_a^{-1}(\ell)$, av linjen ℓ genom origo och $\phi_a(b)$ är nu en linje genom a och b, och dessutom är den entydig eftersom ℓ är det.

A2: Låt först ℓ vara x-axeln snittad med Δ. Då är ℓ det euklidiska intervallet $(-1,1)$, vilket är ordningsisomorft med $\mathbb{R}$ via en godtycklig kontinuerlig bijektion från ℓ till $\mathbb{R}$, t ex $x \mapsto \tan(\pi x/2)$. Om $\phi \in \mathcal{K}$ tar ℓ på sig själv är den en Möbiusavbildning (eller eventuellt konjugatet av en) som tar x-axeln på sig själv. Speciellt är ϕ en kontinuerlig bijektion av intervallet $(-1,1)$ på sig självt och bevarar därför ordningsrelationen. Mer explicit måste ϕ vara

$$\phi(z) = \pm\frac{z-\alpha}{1-\alpha z}$$

för något $\alpha \in (-1,1)$. En godtycklig linje ℓ' får sin ordningsrelation genom att man först avbildar den på linjen ℓ ovan. Den så erhållna ordningen på ℓ' beror inte på valet av avbildning eftersom ordningen på ℓ är invariant under avbildningar som tar ℓ på sig själv.

A3: Eftersom vi redan sett att elementen i $\mathcal{K}$ avbildar linjer på linjer och bevarar ordningsrelationerna, så räcker det att kontrollera axiomet A3 för linjen ℓ som är x-axeln snittad med Δ.

Vi definierar sidorna till ℓ som övre och undre halvcirkelskivan, och vi måste alltså kontrollera att två punkter ligger på samma sida om och endast om sträckan (dvs poincaré-sträckan) mellan dem inte skär ℓ (dvs det euklidiska intervallet $(-1,1)$).

Antag först att a och b båda ligger i, säg övre halvskivan, och låt ℓ' vara linjen genom dem. Om ℓ' inte skär ℓ är det klart. Antag därför att de skär varandra i punkten $c \in (-1,1)$. Avbildningen ϕ_c tar c på 0, ℓ på sig själv, övre halvskivan på sig själv och ℓ' på en linje genom 0, dvs en euklidisk linje genom 0. Bilderna a' och b' av a respektive b ligger nu på samma euklidiska stråle från origo och därför skär sträckan mellan a' och b' inte ℓ. Det följer att Poincaré-sträckan mellan a och b inte skär ℓ.

Omvänt, om a och b ligger på olika sidor om ℓ så är det klart att sträckan mellan dem skär ℓ.

Uppgift 2.1. Varför det?

A4 och A5: Vi har redan sett att $\mathcal{K}$ är en grupp och att elementen i $\mathcal{K}$ avbildar linjer på linjer och bevarar deras ordningsrelationer. Vi har nu att verifiera sista villkoret i A4. Som för motsvarande verifikation i avsnitt 4 i kapitel 4 kan vi reducera oss till fallet att a' är origo, strålen $a'b'$ är positiva x-axeln och att c' bestämmer, säg, övre halvskivan. Givet a, b, och c, tag nu först $\phi_a \in \mathcal{K}$. Den tar åtminstone a på origo. Om inte strålen ab hamnar på positiva x-axeln får man multiplicera ϕ med ett lämpligt $\lambda \in \partial\Delta$. Om nu inte c hamnar på övre halvskivan får man i så fall ordna detta genom att ta konjugatet av $\lambda\phi_a$. Detta visar existensen. Entydigheten följer av att det bara finns en enda $\phi \in \mathcal{K}$ (nämligen identiteten) som tar 0 på 0, positiva x-axeln på sig själv och övre halvskivan på sig själv. Om nämligen $\phi \in \mathcal{K}$ och $\phi(0) = 0$ så är $\phi(z) = \lambda z$ eller $\phi(z) = \lambda \bar{z}$. Om positiva x-axeln går på sig själv så måste λ vara 1. Om nu övre halvskivan går på sig själv så utesluter detta fallet $\phi(z) = \bar{z}$, och alltså är $\phi = I$. $\square$

Exempel 2.3. Att (den euklidiska) cirkeln $|z - m| = r$ skär cirkeln $\partial\Delta$ rätvinkligt vid p är detsamma som att de euklidiska radierna till de respektive cirklarna från p skär varandra rätvinkligt, och detta är, enligt Pythagoras sats och dess omvändning, ekvivalent med att $1+r^2 = |m|^2$, se figur. Poincaré-linjerna består alltså precis av (snittet med Δ av) sådana euklidiska cirklar, samt euklidiska linjer genom origo. $\square$

1

r

m

Snittet ℓ av enhetsskivan och cirkeln $|z - \sqrt{2}| = 1$ är en Poincaré-linje som inte går genom origo. Följaktligen finns gott om Poincaré-linjer genom origo som inte skär ℓ, så parallellaxiomet inte är uppfyllt i denna modell, se figur.

ℓ

$\sqrt{2}-1$ $\sqrt{2}$

3. Längd i Poincaré-modellen

Från den allmänna teorin för absolut geometri A vet vi att det finns ett väldefinierat längdbegrepp, givet att vi fixerar en enhetssträcka. Vi vill nu hitta en formel för längden $d(a,b)$ av sträckan ab mellan två punkter i Δ. Notera först att eftersom $R_\lambda \in \mathcal{K}$ så är $d(0,a) = d(0,a')$ om och endast om $|a| = |a'|$. Eftersom en godtycklig sträcka alltid kan förflyttas till en sträcka mellan 0 och a så börjar vi att titta på sådana. Om a är mittpunkten på sträckan $0, a'$ så ska längden av sträckan $0, a'$ vara dubbelt så stor som längden av sträckan $0, a$, dvs

$$d(0,a') = 2d(0,a). \tag{3.1}$$

Det är brukligt att normalisera längdmåttet så att

$$d(0,a)/|a| \to 2 \text{ då } |a| \to 0. \tag{3.2}$$

För att hitta a' kan vi till att börja med anta att $a = r$ ligger på positiva x-axeln. Avbildningen

$$\phi(z) = \frac{z+r}{1+rz} \in \mathcal{K}$$

tar positiva x-axeln, dvs strålen $0, r$, på sig själv, 0 på r, och följaktligen r på r'. Alltså är $r' = 2r/(1+r^2)$ så

$$|a'| = \frac{2|a|}{1+|a|^2}.$$

Om vi sätter

$$d(0,a)=\log\frac{1+|a|}{1-|a|} \tag{3.3}$$

är det nu lätt att verifiera att (3.1) och (3.2) är uppfyllda.

Uppgift 3.1. Visa att avståndet mellan punkterna a och b blir

$$d(a,b)=\log\frac{|1-\bar{a}b|+|b-a|}{|1-\bar{a}b|-|b-a|}.$$

Anmärkning 3.1. Skälet till att man har 2 istället för t ex 1 i (3.2) är bl a att många formler, som (3.3), (3.8) och (4.2), blir enklare. □

Om $A=d(0,a)$ så följer det från (3.3) att

$$|a|=\tanh(A/2). \tag{3.4}$$

Anmärkning 3.2. Vi påminner om att de hyperboliska trigonometriska funktionerna är definierade av

$$\sinh x=\frac{e^x-e^{-x}}{2},\quad \cosh x=\frac{e^x+e^{-x}}{2},$$
$$\tanh x=\frac{\sinh x}{\cosh x}=\frac{e^x-e^{-x}}{e^x+e^{-x}},\quad \coth x=\frac{\cosh x}{\sinh x}=\frac{e^x+e^{-x}}{e^x-e^{-x}}.$$

Vi lämnar som övning att visa att

$$\frac{1+\tanh^2(x/2)}{1-\tanh^2(x/2)}=\cosh x. \tag{3.5}$$

□

Från teorin för absolut geometri vet vi att egenskaperna (3.1) och (3.2) (och invariansen under kongruensavbildningar) entydigt bestämmer längden av sträckor. Det är rimligt att fråga sig hur man kan komma fram till (3.3) utgående från (3.1) och (3.2). Låt a vara fix och studera längden av sträckan från a till $a+h$ för små h. Genom att använda

$$\phi(z)=\frac{z-a}{1-\bar{a}z}$$

kan vi förflytta denna sträcka till sträckan från 0 till $h/(1-\bar{a}(a+h))$, och alltså är, jmf (3.2),

$$\lim_{h\to 0}\frac{d(a,a+h)}{|h|}=\lim_{|h|\to 0}\frac{d(0,h/(1-\bar{a}(a+h)))}{|h|}=$$
$$=\lim_{|h|\to 0}\frac{d(0,h/(1-\bar{a}(a+h)))}{|h/(1-\bar{a}(a+h))|}\frac{1}{|1-\bar{a}(a+h)|}=\frac{2}{1-|a|^2}.$$

Eftersom euklidiska avståndet av sträckan $a, (a+h)$ är ungefär $|h|$ ser vi att skalfaktorn för längder är $2/(1-|a|^2)$, dvs en liten sträcka vid a som har euklidisk längd ϵ, har ungefär Poincaré-längd $2\epsilon/(1-|a|^2)$. Om $[\alpha, \beta] \ni t \mapsto \gamma(t)$ är en någorlunda reguljär kurva i Δ, är det nu lätt att se att den har en väldefinierad Poincaré-längd som är

$$\int_\alpha^\beta \frac{2|\gamma'(t)|dt}{1-|\gamma(t)|^2}, \tag{3.6}$$

och speciellt om man tillämpar detta på sträckan mellan 0 och z så får man (3.3).

En cirkel med medelpunkt a och radie r i Poincaré-modellen är förstås mängden av alla punkter med Poincaré-avstånd precis r till punkten a. Det är klart att cirklar med medelpunkt i origo är euklidiska cirklar med medelpunkt i origo, och eftersom avbildningarna i $\mathcal{K}$ tar M-cirklar på M-cirklar så följer att en delmängd av Δ är en cirkel (dvs en Poincaré-cirkel) om och endast om den är en euklidisk cirkel.

Exempel 3.3. Betrakta tre punkter i Δ. Det följer att det finns en Poincaré-cirkel genom dessa tre punkter om och endast om de inte ligger på samma euklidiska linje och att den euklidiska cirkeln genom dem helt ligger i Δ. Låt C vara en euklidisk cirkel som inte helt ligger i Δ och som inte skär $\partial\Delta$ rätvinkligt. Om man tar tre punkter som ligger i snittet av Δ och C, så ligger de alltså inte på samma Poincaré-linje, men likväl finns det ingen Poincaré-cirkel som går genom dem. $\square$

Exempel 3.4. Vi vill beräkna omkretsen c av en cirkel γ med radie r. Efter eventuell förflyttning kan vi anta att γ har medelpunkt i origo. Dess euklidiska radie är då $\tanh(r/2)$, och enligt (3.6) är omkretsen

$$\frac{2 \cdot 2\pi \tanh(r/2)}{1-\tanh^2(r/2)},$$

som efter omskrivning blir $4\pi \cosh^2(r/2) \tanh(r/2) = 2\pi \sinh r$ och alltså

$$c = 2\pi \sinh r. \tag{3.7}$$

För stora radier r är omkretsen ungefär πe^r och växer därför mycket snabbt när $r \to \infty$; detta ska ses i kontrast till det euklidiska fallet där omkretsen är proportionell mot radien. $\square$

SATS 3.5. *Antag att A,B och C är längderna på kateterna respektive hypotenusan i en rätvinklig triangel. Då är*

$$\cosh A \cosh B = \cosh C. \tag{3.8}$$

Eftersom $\cosh x \approx 1 + x^2/2$ när x är litet, så säger (3.8) att för små längder $1 + C^2/2 \approx (1 + A^2/2)(1 + B^2/2) \approx 1 + A^2/2 + B^2/2$, vilket ju stämmer överens med Pythagoras sats.

BEVIS. Efter förflyttning kan vi anta att T har hörn i 0, a och ib, där $a = \tanh(A/2)$ och $b = \tanh(B/2)$. Avbildningen

$$\phi(z) = \frac{z - a}{1 - az}$$

tar a på 0 och ib på $\omega = (ib - a)/(1 - iab)$, se figuren i beviset av lemma 4.1. Om $c = |\omega|$ så är alltså $c = \tanh(C/2)$. Men vidare har vi att

$$c^2 = \frac{a^2 + b^2}{1 + a^2 b^2}.$$

Efter omskrivning är detta

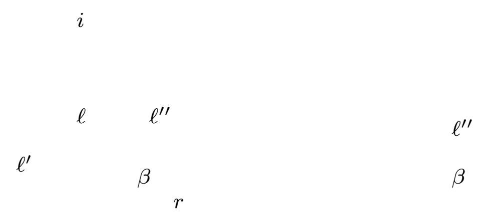
$$\frac{1 + a^2}{1 - a^2}\frac{1 + b^2}{1 - b^2} = \frac{1 + c^2}{1 - c^2},$$

och i ljuset av (3.5) följer nu (3.8). $\square$

EXEMPEL 3.6. Vi har redan sett att det genom en given punkt p utanför en linje ℓ finns många parallella linjer. Antag att ℓ' är normal till ℓ genom p. Från teorin för absolut geometri vet vi redan att linjen genom p som är normal till ℓ' är parallell med ℓ. Låt β vara infimum av alla vinklar β' sådana att linjen genom p med vinkel β' gentemot ℓ' är parallell med ℓ. Det visar sig att vinkeln β är kopplad till avståndet A mellan ℓ och p. Närmare bestämt har vi att

$$e^{-A} = \tan(\beta/2). \tag{3.9}$$

För att bevisa detta, kan vi anta att ℓ är y-axeln, ℓ' är x-axeln och att p är punkten $r > 0$. Gränsfallet måste vara, se figur,



linjen som går genom r och i. För att beräkna β flyttar vi denna till en linje genom origo genom transformationen $z \mapsto (z-r)/(1-rz)$, varvid r avbildas på 0 och i avbildas på punkten

$$\frac{i-r}{1-ri} = \frac{-2r+i(1-r^2)}{1+r^2},$$

så

$$\tan\beta = \frac{1-r^2}{2r}.$$

Men

$$\tan\beta = \frac{2\tan(\beta/2)}{1-\tan^2(\beta/2)}$$

och

$$e^{-A} = \frac{1-r}{1+r},$$

och sätter man samman dessa likheter får man (3.9). □

Exempel 3.7. Bestäm ekvationen för cirkeln med medelpunkt i $i/2$ och radie $\log 2$. Beräkna även dess omkrets.

Lösning: Imaginäraxeln är en diameter till cirkeln. Punkten $i/2$ har avstånd $\log 3$ till origo. Övre skärningspunkten med cirkeln och imaginäraxeln kommer att ha avstånd $\log 3 + \log 2 = \log 6$ till origo, så denna punkt har euklidiskt avstånd $\tanh(\log 6/2) = 5/7$ från origo och är alltså $5i/7$. Det är här praktiskt att notera att

$$\tanh(x/2) = \frac{e^x-1}{e^x+1}. \tag{3.10}$$

Den nedre skärningspunkten får avstånd $\log 3 - \log 2$ till origo och är därför $i/5$. Cirkelns ekvation är därför $|z - 16i/35| = 9/35$.

Eftersom radien är $\log 2$ ger formel (3.7) att omkretsen är $\pi \sinh \log 2 = 3\pi/2$. □

EXEMPEL 3.8. Bestäm radie och medelpunkt för cirkeln $|z-1/4| = 1/2$.

Lösning: Eftersom x-axeln skär cirkeln rätvinkligt så är sträckan $-1/4, 3/4$ en diameter till cirkeln. Längden av denna sträcka är

$$\log \frac{1+3/4}{1-3/4} + \log \frac{1+1/4}{1-1/4} = \log \frac{35}{3}.$$

Alltså är radien $r = \frac{1}{2} \log \frac{35}{3} = \log \sqrt{\frac{35}{3}}$. Medelpunkten kommer att vara den punkt x på positiva realaxeln sådan att $r = d(x, 3/4) = d(0, 3/4) - d(0, x)$. Alltså är

$$\log \frac{1+x}{1-x} = \log \frac{1+3/4}{1-3/4} - \log \sqrt{\frac{35}{3}} = \log \sqrt{\frac{21}{5}},$$

och löser man ut x ur denna likhet, t ex genom att använda (3.10), så får man att $x = \frac{13-\sqrt{105}}{8}$. □

4. Area i Poincaré-modellen

Vi övergår nu till att definiera areor. I euklidisk geometri har en rätvinklig triangel med kateterna A och B arean $AB/2$. Det visar sig att areabegreppet i det hyperboliska fallet är kopplat till defekten. Om man känner vinklarna i ett polygonområde, t ex en triangel, så känner man också dess area. Vi påminner om att defekten för en triangel T med innervinklar α, β och γ, definieras som

$$d(T) = \pi - (\alpha + \beta + \gamma).$$

Vi börjar med att undersöka defekten för små rätvinkliga trianglar.

LEMMA 4.1. *Om T är en rätvinklig triangel och kateternas längder är A och B så har vi att*

$$\frac{d(T)}{AB/2} \to 1 \text{ då } A, B \to 0.$$

BEVIS. Efter förflyttning kan vi anta att T har hörn i 0, a och ib, jmf figur. Låt vinklarna vid a och ib vara α respektive β. Avbildningen

$$\phi(z) = \frac{z-a}{1-az}.$$

tar a på 0, 0 på $-a$ och ib på

$$c = \frac{ib-a}{1-iab} = \frac{-a(1+b^2) + ib(1-a^2)}{1+a^2b^2}.$$

Nu är α vinkeln mellan denna punkt och negativa x-axeln så

$$\tan\alpha = \frac{b(1-a^2)}{a(1+b^2)},$$

och på samma sätt är

$$\tan\beta = \frac{a(1-b^2)}{b(1+a^2)}.$$

Notera att

$$d(T) = \pi - (\pi/2 + \alpha + \beta) = \pi/2 - (\alpha + \beta)$$

så $0 \le d(T) < \pi/2$. Vidare är

$$\tan d(T) = \tan(\pi/2 - (\alpha+\beta)) = \frac{1}{\tan(\alpha+\beta)} = \frac{1-\tan\alpha\tan\beta}{\tan\alpha + \tan\beta} =$$

$$= \frac{1 - \frac{1-a^2}{1+b^2}\frac{1-b^2}{1+a^2}}{\frac{b}{a}\frac{1-a^2}{1+b^2} + \frac{a}{b}\frac{1-b^2}{1+a^2}} = \frac{2ab}{1-a^2b^2}.$$

Alltså har vi att $\tan d(T)/(ab/2) \to 4$ när $a, b \to 0$. Nu är

$$\frac{d(T)}{AB/2} = \frac{d(T)}{\tan d(T)}\frac{\tan d(T)}{ab/2}\frac{ab/2}{AB/2},$$

och eftersom $A/a \to 2$ och $B/b \to 2$ då $a, b \to 0$ och $\tan x/x \to 1$ då $x \to 0$ så följer lemmat. $\square$

Speciellt ser vi att $d(T) > 0$ för små rätvinkliga T. Detta visste vi redan från teorin för absolut geometri eftersom parallellaxiomet inte är uppfyllt. Sedan tidigare vet vi också att defekten är summabel, dvs att om triangeln T är uppdelad i deltrianglar $T_1, ..., T_m$ så är

$$d(T) = \sum_{j=1}^{m} d(T_j). \tag{4.1}$$

Det är nu naturligt att definiera arean, $m(T)$, av T som

$$m(T) = d(T). \tag{4.2}$$

Då har vi:

- (i) $m(T) > 0$ för alla trianglar T.
- (ii) $m(T) = \sum_{j=1}^{m} m(T_j)$, om $T_1, ..., T_m$ är en uppdelning av triangeln T i deltrianglar.
- (iii) $m(T_1) = m(T_2)$ om T_1 och T_2 är kongruenta.
- (iv) $m(T) \approx AB/2$ om T är en liten rätvinklig triangel med katetlängderna A och B.

Man kan nu lätt utvidga areabegreppet till t ex polygonområden.

Anmärkning 4.2. För ett mer allmänt Riemann-mätbart område D kan vi sätta

$$m(D) = \int\int_D \frac{4dxdy}{(1-r^2)^2}, \quad r = \sqrt{x^2+y^2}. \tag{4.3}$$

Vi påstår att denna definition stämmer överens med vår tidigare definition (4.2) i fallet att D är en triangel. För att inse detta, notera att även vår nya definition uppfyller (i) och (ii). Eftersom varje triangel kan delas upp i godtyckligt små rätvinkliga deltrianglar, jmf proposition 3.8 i kapitel 4, räcker det därför att visa att även (iv) är uppfyllt. Men om T är en liten rätvinklig triangel vid z med katetlängder A och B, så är den approximativt en rätvinklig euklidisk triangel med sidlängder $A(1-|z|^2)/2$ och $B(1-|z|^2)/2$ och följaktligen har T euklidisk area ungefär $AB(1-|z|^2)^2/8$. Därför blir integralen i (4.3) approximativt $AB/2$, dvs $m(T)/(AB/2) \to 1$ då $A, B \to 0$. $\square$

Exempel 4.3. Beräkna arean av triangeln med hörn i 0, 1/2 och $i/2$.

Lösning: Låt α vara vinkeln vid 1/2. Då är vinkeln vid $i/2$ lika stor eftersom triangeln är likbent. Avbildningen

$$z \mapsto \frac{z-1/2}{1-z/2}$$

tar 1/2 på 0, 0 på något negativt reellt tal och $i/2$ på $(-10+6i)/17$. Om vinkeln mellan denna punkt och negativa x-axeln är α, så är $\tan\alpha = 3/5$. Alltså är

$$m(T) = d(T) = \pi - (\pi/2 + 2\alpha) = \pi/2 - 2\arctan(3/5).$$

Eftersom

$$\tan(\pi - (\pi/2 + \alpha + \alpha)) = 1/\tan 2\alpha = \frac{1-\tan^2\alpha}{2\tan\alpha} = 8/15,$$

kan arean också utryckas som $\arctan \frac{8}{15}$. $\square$

EXEMPEL 4.4. Antag att $0 < r < 1$. Finn z så att triangeln med hörn i 0, r och z är liksidig. Hur stor area kan en liksidig triangel ha?
Lösning: Eftersom avståndet mellan 0 och r är det samma som avståndet mellan 0 och z är det klart att $z = re^{i\theta}$, och vi vill bestämma θ så att avståndet mellan r och z blir lika stort som avståndet mellan 0 och r. Detta leder till att

$$\left| \frac{z - r}{1 - rz} \right| = r,$$

och löser man ut $z = re^{i\theta}$ så får man att $\cos\theta = \frac{1+r^2}{2}$. Arean av en liksidig triangel med sidan $\log \frac{1+r}{1-r}$ är alltså $\pi - 3\theta$; speciellt ser vi att arean kan fås godtyckligt nära π. $\square$

EXEMPEL 4.5. Man kan inse resultatet om arean i förra exemplet på ett enklare sätt. Efter förflyttning kan vi anta att hörnen i vår liksidiga triangel ligger på en cirkel med medelpunkt i origo. Eftersom den är liksidig kommer hörnen att ligga på tre radier från origo med vinkel $2\pi/3$ mellan varandra. Man inser nu, se figur, att vinklarna i triangeln kan fås godtyckligt små genom att välja en mycket stor cirkel. Följaktligen kan man få defekten att vara godtyckligt nära π.

α

På liknande sätt kan man inse att det finns trianglar med en föreskriven vinkel α som har area godtyckligt nära $\pi - \alpha$. $\square$

ANMÄRKNING 4.6. Det är naturligt att tala om punkterna på enhetscirkeln $\partial\Delta$ som mängden av punkter i oändligheten. Man ser lätt att det till varje par av punkter i oändligheten finns exakt en linje med dessa två punkter som ändpunkter, se övning 4.1. Man kan nu tala om trianglar med hörn i oändligheten. I själva verket kan man *definiera* area genom att föreskriva att en triangel med två hörn i oändligheten och en vinkel α ska ha arean $\pi - \alpha$. Att då arean för en godtycklig triangel blir lika med triangelns defekt följer sedan från följande figur:

γ

β

α

De tre hörnen i den stora triangeln ligger i oändligheten och alltså har stora triangeln arean π. De tre yttre deltrianglarna har areorna α, β respektive γ. Om T är den inre triangeln har vi alltså att att $\pi = m(T) + \alpha + \beta + \gamma$, dvs $m(T) = d(T)$. $\square$

UPPGIFT 4.1. Visa att det givet två punkter på $\partial\Delta$ finns exakt en euklidisk cirkel genom dessa som skär $\partial\Delta$ rätvinkligt.

5. Ytterligare övningar till kapitel 5

UPPGIFT 5.1. Verifiera (3.1) och (3.2) givet att man definierat $d(0, a)$ genom (3.3).

UPPGIFT 5.2. Visa (3.5).

UPPGIFT 5.3. Visa att $\cosh^2 x - \sinh^2 x = 1$ och att $1 - \tanh^2 x = 1/\cosh^2 x$. Vad blir $\cosh(x+y)$ och $\sinh(x+y)$ uttryckt i cosh och sinh av x och y?

UPPGIFT 5.4. Visa att om två trianglar båda har sidlängderna A, B och C så är de kongruenta.

UPPGIFT 5.5. Antag att två olika trianglar båda har vinklarna α, β och γ. Visa att de är kongruenta.

UPPGIFT 5.6. Bestäm medelpunkt och (längd av) radien i cirkeln $|z - 1/3| = 1/2$.

UPPGIFT 5.7. Bestäm medelpunkt och (längd av) radien i cirkeln $|z + i/4| = 1/2$.

UPPGIFT 5.8. Bestäm medelpunkt och (längd av) radien i cirkeln $|z - 16/35| = 9/35$.

UPPGIFT 5.9. Ange den cirkel (på formen $|z - a| = r$) som har radie $\log \sqrt{2}$ och medelpunkt $3 - 2\sqrt{2}$.

UPPGIFT 5.10. Antag $0 < a < 1$ och bestäm Poincaré-linjen (angiven på formen $|z - c| = r$) genom a som är symmetrisk m a p x-axeln.

UPPGIFT 5.11. Bestäm arean av triangeln med hörn i 0, 1/3 och $-i/3$.

UPPGIFT 5.12. Visa påståendet i exempel 4.5 om areor av trianglar med en vinkel α.

UPPGIFT 5.13. Låt P vara en polygon med m hörn och innervinklar $\alpha_1, \ldots, \alpha_m$. Sätt $\beta_j = \pi - \alpha_j$. Visa att arean av P är $\sum_{j=1}^{m} \beta_j - 2\pi$. Det underlättar om man antar att P är *konvex*, dvs att sträckan mellan två punkter i P helt ligger i P.

UPPGIFT 5.14. Beräkna arean av en cirkel med radie r.

UPPGIFT 5.15. Låt Q_t vara fyrhörningen med hörn i punkterna 0, t, it och $(1+i)t$, där $0 < t < 1/\sqrt{2}$. Beräkna arean $m(Q_t)$ då $t \to 1/\sqrt{2}$.

UPPGIFT 5.16. Bestäm arean av en rätvinklig likbent triangel vars ena katet bestäms av punkterna 1/2 och 5/7.

UPPGIFT 5.17. Hur långa kan kateterna vara i en rätvinklig triangel om en av de andra vinklarna är θ?

UPPGIFT 5.18. Visa att skärningen ℓ mellan Δ och $|z-(1+i)| = 1$ är en Poincaré-linje. Låt $\phi(z) = e^{i\theta}z$, och bestäm θ så att ℓ och $\phi(\ell)$ skär varandra rätvinkligt.

UPPGIFT 5.19. Visa att det finns en konstant C sådan att avståndet mellan medelpunkten i en liksidig fyrhörning och randen av fyrhörningen alltid är mindre än C. Bestäm den bästa konstanten.

UPPGIFT 5.20. Låt T vara en likbent triangel som är inskriven i en cirkel med radie $\log(1 + r)/(1 - r)$ sådan att ena sidan är en diameter till cirkeln. Bestäm triangelns area.

UPPGIFT 5.21. Beräkna arean av en liksidig triangel som är inskriven i cirkeln $|z| = 1/2$.

Några kommentarer

För en mer utförlig framställning av hyperbolisk geometri, se [**UP**]. Där finns även andra modeller beskrivna.

Lösningar och kommentarer till några av övningarna

Kapitel 1

Övning 2.2: Låt A vara "du kommer hit" och B vara "jag går på bio". Då säger påståendet formellt $\neg A \to B$. Negationen är då $\neg A \wedge \neg B$ d.v.s. "Du kommer inte hit men jag går inte på bio (ändå)". Många tycker nog att man egentligen menar $\neg A \leftrightarrow B$ och då blir negationen $(A \wedge B) \vee (\neg A \wedge \neg B)$ d.v.s. "Du kommer inte hit och jag går inte på bio eller du kommer hit och jag går på bio (ändå)". Man kan även tänka sig att tolka meningen som "Om du inte kommer hit så går jag på bio ändå" d.v.s. $(A \to B) \wedge (\neg A \to B)$ vilket är ekvivalent med B. Då blir förstås negationen $\neg B$ d.v.s. "jag går inte på bio (oavsett vad du gör)".

Övning 2.3: T ex "Om a och b är irrationella tal så är a^b irrationellt".

Övning 2.4: T ex "Det finns ett barn som inte är blåögt men som har blåögda föräldrar".

Övning 2.8: Antag att $(C \to A) \to B$ är falsk och $(\neg A \vee B) \wedge C$ är sann. Då är $C \to A$ och $\neg B$ sanna och C och $\neg A \vee B$ är sanna. Men ur detta följer att A är sann, och då får vi att B är sann. Detta är dock en motsägelse. Alltså är $(C \to A) \to B$ sann om $(\neg A \vee B) \wedge C$ är sann.

Övning 2.9: (i) och (iii) är tautologier. Obs att (i) är en implikation så om den inte är sann så är högerledet falskt. Men då måste C vara falsk och därför är $B \vee \neg C$ sann och därför är $A \to (B \vee \neg C$ sann och därför är vänsterledet falskt. Detta innebär att hela satsen är en tautologi.

Övning 3.1: Se övning 5.3

Övning 3.4: T ex $(\exists \epsilon)_{\epsilon > 0}(\delta)_{\delta > 0} \exists x \exists y \neg \phi$, där ϕ är som i exempel 3.7. Vi kan uttrycka detta som "Det finns ett $\epsilon > 0$ sådant att det till varje $\delta > 0$ går att hitta $x, y \in \mathbb{R}$ sådana att $|x - y| < \delta$ men $|f(x) - f(y)| \geq \epsilon$.

Övning 5.1: Det finns bara ett ändligt antal studenter registrerade på kursen. Någon av dem måste vara kortast; eventuellt finns flera som är lika korta. Alltså är formeln sann i denna tolkning. Observera att om man byter tolkningen av $A(x, y)$ mot "x är strikt längre än y" så är formeln falsk, eftersom ingen är strikt längre än sig själv.

Övning 5.3: Om M är en tolkning så svarar formeln $\phi = \phi(x, y)$ mot en tvåställig relation R på M. Vänsterledet uttrycker att det finns $m \in M$ sådant att $\langle m', m \rangle \in R$ för alla $m' \in M$. Högerledet uttrycker att det till varje $m' \in M$ finns något $m'' \in M$ sådant att $\langle m', m'' \rangle \in R$. Om vänsterledet är sant är också högerledet sant; man kan ju ta $m'' = m$. Beskriv detta med en figur i $M \times M$!

Övning 5.6: Se exempel 1.3 i kapitel 3.

Övning 5.7: Låt m vara en sluten formel som är en kontradiktion, t ex $\phi \wedge \neg \phi$ för någon sluten formel ϕ.

Antag först att vi kan fullständighetssatsen. Om Φ saknar modell så är är m en logisk konsekvens av Φ och enligt fullständighetssatsen går alltså m att härleda ur Φ, d.v.s. man kan härleda en motsägelse ur Φ.

Omvänt; antag att vi kan sats 5.22 och antag att ϕ är en logisk konsekvens av Φ. Om nu $\Phi \cup \{\neg \phi\} \vdash m$ så $\Phi \vdash \neg \phi \to m$ och eftersom satslogiska tautologier är härledningsregler så följer att $\Phi \vdash \phi$. Om vi inte kan härleda ϕ ur Φ så kan vi därför inte härleda en motsägelse är $\Phi \cup \{\neg \phi\}$

och alltså har denna satsmängd en modell. Detta strider dock mot att ϕ var en logisk konsekvens av Φ.

Övning 6.2: (ii), (iv), (viii), (xi), (xii), (xiii) är logiskt sanna.

Övning 6.3: Man kan definiera ϕ_m induktivt. Som ϕ_1 tar man förstås bara $\neg A_1$. Givet att man har hittat ϕ_{m-1} så kan man ta ϕ_m som

$$(\neg\phi_{m-1} \wedge A_m) \vee (\phi_{m-1} \wedge \neg A_m).$$

Man kan också utnyttja övning 6.5.

Övning 6.4: Varje sats är uppbyggd i steg enligt de syntaktiska reglerna. Om man vill uttrycka $A \wedge B$ kan man istället bilda den ekvivalenta satsen $\neg(\neg A \vee \neg B)$. Om man vill uttrycka $A \rightarrow B$ kan man istället uttrycka detta med den ekvivalenta satsen $B \vee \neg A$, och på samma sätt med $\leftrightarrow$.

Övning 6.5: För varje kombination där man har angett S bildar man $nA_1 \wedge nA_2 \wedge \ldots nA_m$ där nA_j betyder A_j om A_j är sann och $\neg A_j$ om A_j är falsk. Man tar sedan disjunktionen av alla dessa satser.

Övning 6.6: Alla tre är OK.

Övning 6.7: $\exists x \exists y (x \neg y)$.

Övning 6.8: $\forall x \forall y \forall z (z = x \vee z = y \vee y = x)$.

Övning 6.11: Ingen av dem!

Övning 6.12: (ii) och (iii). (i) uttrycker att f är likformigt kontinuerlig, vilket är ett starkare påstående, jmf övning 6.14 medan däremot (iv) (faktiskt) är sann för varje funktion f.

Övning 6.13: (iv).

Övning 6.14: Fixera x. Nu är

$$|x^2 - y^2| = |x-y||x+y| \leq |x-y|(|x| + |y-x| + |x|) \leq 2|x||x-y| + |x-y|^2 < \epsilon$$

om $2|x||x-y| < \epsilon/2$ och $|x-y|^2 < \epsilon/2$ d.v.s. om $|x-y| < \delta$ där

$$\delta = \min\left(\frac{\epsilon}{4|x|}, \sqrt{\frac{\epsilon}{2}}\right).$$

För att se att $f(x)$ inte är likformigt kontinuerlig, tag $\epsilon = 1$. Givet $\delta > 0$, tag $x = 1/\delta$ och $y = 1/\delta + \delta/2$. Då är $|x-y| + \delta/2 < \delta$ men

$$|f(x) - f(y)| = \left(\frac{1}{\delta} + \frac{\delta}{2}\right)^2 - \left(\frac{1}{\delta}\right)^2 = 1 + \frac{\delta^2}{4} > 1 = \epsilon.$$

Övning 6.15: (ii).

Övning 6.16: (iii) och (iv).

Övning 6.17: (iii).

Övning 6.18: (ii) och (iv).

Kapitel 2

Övning 1.3: $x \in b \setminus \bigcup_{y\in A} y \iff x \in b$ och $x \notin \bigcup_{y \in A} y \iff x \in b$ och $x \notin y$ för alla $y \in A$ $\iff x \in b \setminus y$ för alla $y \in A \iff x \in \bigcap_{y\in A} b \setminus y$.

Övning 1.4: Antag att $c \in a$. Enligt regularitetsaxiomet finns $b \in \{a, c\}$ sådant att $b \cap \{a, c\} = \emptyset$. Eftersom $c \in a$ så $b \neq a$. Alltså $b = c$ men då har vi att $a \notin c$.

Övning 2.1: Vi måste visa att $f(x) = f(x')$ medför att $x = x'$. Om $f(x) = f(x')$ så är $g(f(x)) = g(f(x'))$ dvs $g \circ f(x) = g \circ f(x')$. Eftersom $g \circ f$ är injektiv så är då $x = x'$.

Om f är injektiv så är $g \circ f$ injektiv omm g är injektiv på bilden av f. Om Y är strikt större än $f(X)$ kan alltså $g \circ f$ vara injektiv utan att g är det. Tag t ex $X = \{0\}$, $Y = \{0, 1\}$ och $Z = \{0\}$.

Om $g \circ f$ är surjektiv så är g surjektiv men inte nödvändigtvis f. Om $z \in Z$ så finns $x \in X$ sådant att $g \circ f(x) = z$, men då är ju $g(y) = z$ om $y = f(x)$.

Övning 2.2: Om det finns g sådan att $g \circ f = I_X$ och $f \circ g = I_Y$ så är f injektiv och surjektiv enligt övning 2.1.

Övning 2.3: $x \in f^{-1}(A \cap B) \iff f(x) \in A \cap B \iff f(x) \in A$ och $f(x) \in B \iff x \in f^{-1}(A)$ och $x \in f^{-1}(B) \iff x \in f^{-1}(A) \cap f^{-1}(B)$. På samma sät med $\cup$ istället för $\cap$.

Övning 2.4: $f(f^{-1}(E)) \subset E$ men likhet behöver inte gälla.

Övning 2.7: Varje ekvivalensrelation på X är ett element i $P(X \times X)$. Ekvivalensrelationerna är precis elementen i mängden $\{x \in P(X \times X);\ \forall y \forall z \forall w (y \in X \wedge z \in X \wedge w \in X \to \phi)\}$ där ϕ är formeln

$$\langle y, y\rangle \in x \wedge (\langle y, z\rangle \in x \to \langle z, y\rangle \in x) \wedge ((\langle y, z\rangle \in x \wedge \langle z, w\rangle \in x) \to \langle y, w\rangle \in x).$$

Övning 3.1: Vi definierar $\Phi\colon (A^B)^C \to A^{B\times C}$ enligt följande: Om $f\colon C \to A^B$ så är, för $c \in C$, $f(c)\colon B \to A$. Vi definierar $\Phi(f)\colon B \times C \to A$ genom att $\Phi(f)(b, c) = f(c)(b)$. Om $\Phi(f) = \Phi(f')$ betyder detta att $f(c)(b) = f'(c)(b)$ för alla $c \in C$ och $b \in B$, vilket innebär att $f(c) = f'(c)$ för alla $c \in C$ dvs $f = f'$. Alltså är Φ injektiv. Om $h\colon B \times C \to A$ är godtycklig, kan vi definiera en f som ovan genom att sätta $f(c)(b) = h(b, c)$. Då blir $h = \Phi(f)$ och alltså är Φ surjektiv.

Övning 3.3: Antag att $f\colon A \to B$ är en bijektion. Vi definierar en avbildning $\Phi\colon 2^B \to 2^A$ genom att om $\phi\colon B \to 2 = \{0, 1\}$ sätta $\Phi(\phi)(a) = \phi(f(a))$. Vi kollar först att Φ är injektiv. Antag att $\Phi(\phi) = \Phi(\phi')$. Då är $\phi(f(a)) = \phi'(f(a))$ för alla a. Eftersom f är surjektiv följer att $\phi = \phi'$. Om $\psi \in 2^A$ är given, definiera $\phi \in 2^B$ genom att $\phi(b) = \psi(f^{-1}(b))$. Då är $\Phi(\phi) = \psi$ och alltså är Φ surjektiv.

Övning 3.4: Påståendet är ekvivalent med: Om $\phi\colon n \to n$ är en injektion så är n en bijektion. Använd induktion.

Övning 3.5: Använd induktion över antalet element.

Övning 3.6: Antag $A = \{a_1, \ldots a_n\}$ och $B = \{b_1, \ldots, b_m\}$. Vill man definiera en $f\colon A \to B$ så kan till att börja med $f(a_1)$ väljas på m olika sätt. För varje sådant val har man m olika val av $f(a_2)$ etc. Man får totalt m^n olika möjligheter.

Övning 3.7: Antag att $f(A) \in A$. Då är $f(A) = f(B)$ för någon $B_0 \in P(M)$. Om f är injektiv så är $B_0 = A$. Nu får man en motsägelse för $f(A) \in A$ medför $f(A) \notin A$ och $f(A) \notin A$ medför $f(A) \in A$. Alltså kan inte f vara injektiv.

Övning 5.1: Om $A = \{a_0, a_1, \ldots\}$ och $B = \{b_0, b_1, \ldots\}$ så kan man t ex ta uppräkningen

$$\langle a_0, b_0\rangle, \langle a_1, b_0\rangle, \langle a_0, b_1\rangle, \langle a_2, b_0\rangle, \langle a_1, b_1\rangle, \langle a_0, b_2\rangle, \langle a_3, b_0\rangle, \ldots.$$

Övning 5.3: Kopiera argumentet i exempel 5.7!

Övning 5.4: Visa först att man har en bijektion $2^{\mathbb{N}} \times 2^{\mathbb{N}} \simeq 4^{\mathbb{N}}$, jmf exempel 3.3.

Övning 5.5: $\mathbb{R} \simeq 2^{\mathbb{N}} \simeq 2^{\mathbb{N}} \times 2^{\mathbb{N}} \simeq \mathbb{R} \times \mathbb{R}$.

Övning 6.1: Om $M = \{A_b;\ b \in B\}$ är en mängd av disjunkta icketomma mängder så kan vi definiera $f\colon \cup_{b\in B} A_b \to B$ genom att $f(x) = b$ om $x \in A_b$. Då blir f surjektiv, och om $g\colon B \to \cup_{b\in B} A_b$ sådan att $f \circ g = I_B$ så är bilden av g den sökta mängden.

Övning 6.2: Antag att M är oändlig. Enligt exempel 6.3 kan vi skriva M som en disjunkt union $M = A \cup M'$ där $A = \{a_0, a_1, \ldots\}$ är uppräknelig. Vi kan nu definiera en bijektion $f\colon M \to M \setminus \{a_0\}$ genom att $f(x) = x$ om $x \in M'$ och $f(a_k) = a_{k+1}$. Alltså är $M \simeq M \setminus \{a_0\}$. Att en ändlig mängd inte kan vara ekvipotent med en äkta delmängd till sig själv följer av övning 3.4.

Det kan noteras att man i första delen av denna lösning använde resultatet i exempel 6.3 som i sin tur bygger på det uppräkneliga urvalsaxiomet.

Övning 8.2: Jämför övning 2.7.

Övning 8.3: Vi kan anta att mängden är $\mathbb{N}$. Låt E_n vara mängden av alla delmängder till $\mathbb{N}$ som endast innehåller element ur $n = \{0, 1, \ldots, n-1\}$. Då består $M = \cup_{n\in\mathbb{N}} E_n$ av alla ändliga delmängder till $\mathbb{N}$. Eftersom varje E_n är ändlig följer det av proposition 5.6 att M är ändlig eller uppräknelig. Men M innehåller bl a alla $\{k\}$ som är uppräkneligt många så M är inte ändlig.

Övning 8.4: Antag att A har m element. Tag $a_0 \in A$ och sätt $a_1 = f(a_0)$, $a_2 = f(a_1)$ etc. Då måste två av elementen $a_0, \ldots, a_m$ vara lika; säg $a_\ell = a_{\ell+k}$. Då är $B = \{a_\ell, a_{\ell+1}, \ldots, a_{k-1}\}$ en icketom mängd med önskad egenskap.

Här är ett annat argument: Betrakta mängderna $A_0 = A$, $A_1 = f(A_0)$, $A_2 = f(A_1)$ etc. Vi har att $A_0 \supset A_1 \supset A_2 \supset \ldots$ och alla A_k är icketomma. Eftersom A_0 är ändlig finns det ett k sådant att $A_{k+1} = A_k$, jmf övning 8.18. Tag nu $B = A_k$.

Övning 8.5: Varje tal kan skrivas entydigt som $x = x'2^m3^n$ där x' inte är delbart med 2 eller 3. Då blir $A_0 = \{x;\ m \geq n\}$, $A_1 = \{x;\ m < n\}$, $B_0 = \{x;\ n \geq m\}$ och $B_1 = \{x;\ n < m\}$.

Övning 8.8: Sätt $A_0 = A$ och $A_k = B$ för $k \geq 1$. Då följer det av proposition 5.6 att $A \cup B$ är ändlig eller uppräknelig. Men eftersom $A \cup B \supset A$ och A är uppräknelig så kan inte $A \cup B$ vara ändlig. Man kan förstås också direkt hitta en uppräkning av elementen i $A \cup B$.

Övning 8.9: Allmänt har vi att $A^n = A \times A^{n-1}$. Med induktion följer att $\mathbb{R}^n \simeq \mathbb{R} \times \mathbb{R}^{n-1} \simeq \mathbb{R} \times \mathbb{R} \simeq \mathbb{R}$. Sista ekvivalensen följer av övning 5.5. Man kan även använda sats 7.10.

Se även lösningen av nästa övning!

Övning 8.10:

$\mathbb{R}^{\mathbb{N}} \simeq (2^{\mathbb{N}})^{\mathbb{N}} \simeq$ enligt övning 3.1 $\simeq 2^{\mathbb{N}\times\mathbb{N}} \simeq 2^{\mathbb{N}} \simeq \mathbb{R}$.

Denna metod fungerar även för övning 8.9.

Övning 8.11: $\mathbb{C} \simeq \mathbb{R}^2 \simeq \mathbb{R}$.

Övning 8.12: Obs att varje algebraiskt tal är rot till en polynomekvation med heltalskoefficienter. Låt M_n vara rötterna till alla polynom med heltalskoefficienter som ligger i $\{1, -1, 2, -2, \ldots n, -n\}$ och som har högst grad n. Det finns bara ändligt mänga sådana polynom och följaktligen är varje M_n ändlig. Unionen av dessa är uppräknelig och består av alla algebraiska tal.

Övning 8.13: Om $\mathbb{R} \setminus A$ vore uppräknelig så skulle $\mathbb{R} = A \cup (\mathbb{R} \setminus A)$ vara uppräknelig eftersom A är uppräknelig.

Övning 8.14: Kalla mängden M. Låt J beteckna mängden av jämna naturliga tal och U mängden av udda. Om $E \subset U$ och $X = J \cup E$ så innehåller $2X \cap X$ åtminstone alla tal som är delbara med 4 och är därför uppräknelig, dvs $X \in M$. Notera vidare att två olika $E \subset U$ ger upphov till två olika X. Alltså har vi en injektiv avbildning $P(U) \to M$, och eftersom $P(U)$ är överuppräknelig så är även M överuppräknelig.

Övning 8.15: Låt A vara mängden av bijektioner. Då är $A \subset \mathbb{N}^{\mathbb{N}} \simeq 2^{\mathbb{N}}$. Omvänt, kan vi definiera en avbildning $2^{\mathbb{N}} \to A$ på följande vis. Givet $f \in 2^{\mathbb{N}}$, dvs $f\colon \mathbb{N} \to \{0, 1\}$, definiera $g\colon \mathbb{N} \to \mathbb{N}$ genom att $g(2k) = 2k$ och $g(2k+1) = 2k+1$ om $f(k) = 0$ och $g(2k) = 2k+1$ och $g(2k+1) = 2k$ om $f(k) = 1$. Man ser lätt att detta blir en injektion $2^{\mathbb{N}} \to A$, och enligt Schröder–Bernsteins sats följer att $M \simeq 2^{\mathbb{N}}$.

Här är en alternativ konstruktion av an injektion $2^{\mathbb{N}} \to A$. Visa först att om $X \subset \mathbb{N}$ har minst två element så finns en bijektion $f\colon X \to X$ sådan att $f(x) \neq x$ för alla x. Om $X = \{a_0, \ldots, a_m\}$ är ändlig kan man ta $f(a_k) = a_{k+1}$ för $k < m$ och $f(a_m) = a_0$. Om $X = \{a_0, a_1, \ldots\}$ är uppräknelig kan man ta $f(a_0) = a_1$, $f(a_1) = a_0$, $f(a_2) = a_3$, $f(a_3) = a_2$ etc. Till varje $Y \subset \mathbb{N}$ sådan att $\mathbb{N} \setminus Y$ har minst två element kan vi alltså hitta en bijektion $f\colon \mathbb{N} \to \mathbb{N}$ sådan att $Y = \{x;\ f(x) = x\}$. Eftersom mängden av alla sådana delmängder Y är ekvipotent med $2^{\mathbb{N}}$, så får vi en injektion $2^{\mathbb{N}} \to A$.

Övning 8.16: Vi säger att $A \in P(P(M))$ är i H om elementen i A är disjunkta delmängder som alla består av två olika element ur M, t ex $A = \{\{x_0, x_1\}, \{x_2, x_3\}\}$ där $x_0, \ldots, x_3$ är fyra olika element ur M. Om A och A' är i H så säger vi att $A \leq A'$ helt enkelt om $A \subset A'$. På detta sätt blir H en partiellt ordnad mängd. Man ser lätt att varje totalt ordnad delmängd K till H har en majorant, genom att ta unionen av mängderna A i K. Enligt Zorns lemma finns ett maximalt element $A \in H$. Pga maximaliteten kan det vara högst ett element i M som inte finns med i något av paren i A. Vi definierar nu lätt en bijektion $M \to M$ utan fixpunkter, dvs utan att $f(x) = x$ för något $x \in M$. För varje par $\{x, x'\}$ sätter vi $f(x) = x'$ och $f(x') = x$. Om det fanns ett element x_0 som inte fanns med i något av paren så tar vi ett fixt par $\{x_1, x_2\}$ och sätter $f(x_k) = x_{k+1} <$ $k \leq 1$ och $f(x_2) = x_0$.

Övning 8.17: För varje (icketom) mängd A gäller att $A \simeq 2^A$ inte är sant. Om A är oändlig så är $A^2 \simeq A \times A \simeq A$. Om $2^A \simeq A^2$ måste alltså A vara ändlig. Säg att A har m element. Då ska $2^m = m^2$ och detta är bara sant om $m = 2$ eller $m = 4$.

Övning 8.18: Om $b_k = |B_k|$ betecknar antalet element i B_k så är b_k en avtagande följd av positiva naturliga tal. Därför finns ett N sådant att $b_k = b_N$ för alla $k \geq N$. Men då är $B_k = B_N$ för alla $k \geq N$ och följaktligen är $\cap_{k=0}^{\infty} B_k = B_N$.

Låt $B_k = \{k, k+1, k+2, \ldots\}$. Då är $B_0 \supset B_1 \supset \ldots$ men $\cap B_K = \emptyset$.

Övning 8.19: Tag fixt k och observera att

$$f_k(A_{k+1}) \supset f_k \circ f_{k+1}(A_{k+2}) \supset f_k \circ f_{k+1} \circ f_{k+2}(A_{k+3}) \supset \ldots$$

och att alla dessa mängder är icketomma. Enligt övning 8.18 följer att snittet B_k av dessa mängder är en icketom delmängd till A_k. Restriktionen av f_k till B_{k+1} en surjektion $B_{k+1} \to B_k$. Man kan nu ta ett godtyckligt element $a_0 \in B_0$. Till detta finns $a_1 \in B_1$ sådant att $f_0(a_1) = a_0$. Nu finns $a_2 \in B_2$ sådant att $f_1(a_2) = a_1$ etc.

Övning 8.20: Antalet rader i en sanningstabell för m satsvariabler är det samma som antalet funktioner från m till $\{S, F\} \simeq \{0, 1\}$ dvs 2^m. På varje rad kan vi välja S eller F; detta svarar mot alla olika funktioner från 2^m till $\{S, F\}$ vilket är $2^{(2^m)}$ stycken. Nu gäller bara att se att man till varje sådan funktion kan hitta en sats ϕ som är sann precis på de rader man önskar. Men detta är lätt. För varje rad för vilken vi vill att ϕ ska vara sann tar vi och bildar satsen

$$n_0A_0 \wedge n_1A_1 \wedge \ldots \wedge n_{m-1}A_{m-1}$$

där n_jA_j betyder A_j om A_j är sann på den raden och $\neg A_j$ annars. Man tar sedan disjunktionen av dessa satser.

Övning 8.21: Börja med att avbilda $[0, 1/2)$ på $[1/2, 1)$.

Ett annat sätt är att ta avbildningen sådan att $0 \mapsto 1/2$, $1/n \mapsto 1/(n+1)$, $n = 2, 3, 4, \ldots$ och $x \mapsto x$ för alla andra x.

Kapitel 3

Övning 1.2: Induktion över k. Om $n + 0 = m + 0$ så är $n = m$ enligt definition. Antag nu att $n + k = m + k$ medför att $n = k$ och antag att $n + k' = m + k'$. Enligt definition betyder detta att $(n + k)' = (m + k)'$ och enligt P4 alltså att $n + k = m + k$, och enligt induktionsantagandet alltså att $n = m$.

Övning 1.3: Vi använder induktion över n. Om $n = 0$ så har $n + x = m$ en lösning, nämligen $x = m$. Antag nu att vi vet att $n + x = m$ eller $m + x = n$ är lösbar. Om $m + k = n$ så är $m + k' = n'$. Antag att vi istället har att $n + k = m$. Om $k = 0$ så är $m = n$ och alltså är $m + 0' = n'$. Om $k \neq 0$ så är $k = \ell'$ och alltså är $n' + \ell = \ldots = m$.

Övning 1.5: Om alla modeller till T vore isomorfa med $\mathbb{N}$ så skulle Gödels formel ϕ vara sann i alla modeller till T, alltså en logisk följd av (axiomen för) T. Enligt fullständighetssatsen skulle därför finnas ett predikatlogiskt bevis för ϕ, vilket motsäger att ϕ inte är bevisbar.

Övning 4.1: Antag att m är en minsta rationell majorant till $M = \{x \in \mathbb{Q};\ x^2 < 2\}$. Vi vet att $(m + r)^2 - m^2 \to 0$ då $r \to 0$, där r löper över rationella tal. Om $m^2 < 2$ så finns alltså rationellt $r > 0$ sådant att även $(m + r)^2 < 2$, vilket strider mot att m var en majorant till M. Om $m^2 > 2$ så finns av samma skäl ett litet positivt rationellt r sådant att $(m - r)^2 > 2$. Men då är ju även $m - r$ en majorant till M, stridande mot att m var den minsta. Alltså är $m^2 = 2$.

Om man vill undvika att tala om gränsvärden kan man göra så här: Antag att $m^2 < 2$. Vi kan anta att $m < 2$. Om $r < 1$ har vi att $(m + r)^2 - m^2 = r(2m + r) < 5r$. Om vi tar $r = (2 - m^2)/5$ så är alltså $m + r > m$ men ändå $(m + r)^2 < 2$, vilket strider mot att m var en majorant till M. På liknande sätt inser man att vi inte kan ha att $m^2 > 2$.

Övning 4.3: Eftersom $(a_k) > 0$ så finns M och $q > 0$ sådana att $a_k \geq q$ on $k \geq M$. Eftersom $a_k - a'_k$ är en nollföljd så finns M' sådant att $|a_k - a'_k| < q/2$ om $k \geq M'$. Om $k \geq \max(M, M')$ har vi alltså att

$$a'_k > a_k - q/2 \geq q - q/2 = q/2.$$

Detta innebär att $(a'_k) > 0$.

Kapitel 4

Övning 1.1: Antag att $a \neq b$ och låt I_0 vara vår enhetssträcka. Som i det tidigare resonemanget följer det att ett ändligt antal, säg 2^m av sträckorna ab täcker I_0. Det följer att ab har längd åtminstone 2^{-m}.

Övning 1.2: Låt cirkeln C vara bestämd av medelpunkten m och radien r. Antag att m' är en annan punkt och låt ℓ vara linjen mm'. Eftersom m har lika avstånd (nämligen r) till de båda skärningspunkterna mellan ℓ och C så följer av (1.1) att m' har olika avstånd till skärningspunkterna. Alltså är inte m' en medelpunkt till C.

Övning 2.2: Man noterar först att mittpunkten på sträckan från origo till $(1,0)$ är $(1/2,0)$. Detta inses genom att sträckorna $(0,0),(1/2,0)$ och $(1/2,0),(1,0)$ är kongruenta via en translation i x_1-led. Man inser sedan att sträckan $(0,0),(2^{-k},0)$ har längd 2^{-k}. Det följer att sträckan från origo till $(t,0)$, för ett godtyckligt tal $t>0$, måste ha längden t. Genom att använda invariansen under rotationer kan man nu se att sträckan från origo till en punkt (c_1,c_2) har längd $\sqrt{c_1^2+c_2^2}$. Slutligen har vi att sträckan från a till b är kongruent med sträckan från origo till $b-a$.

Övning 3.2: Om en vinkel är rät eller trubbig så är dess yttervinkel rät eller spetsig, och därför är båda motstående vinklarna spetsiga, enligt sats 3.7.

Övning 3.1: Antag att $\wedge a = \wedge b$ i triangeln abc. Enligt det tredje kongruensfallet följer att abc är kongruent med bca och alltså är $|ac| = |bc|$.

Övning 3.3: Låt medelpunkten vara m och antag att p ligger på cirkeln. Om ℓ är linjen genom p som är vinkelrät mot radien pm så är ℓ en tangent (hypotenusan är längre än kateterna i en rätvinklig triangel). Antag att ℓ' är en annan linje genom p. Tag q på ℓ' sådan att mq blir vinkelrät mot ℓ'. Då är mpq en rätvinklig triangel och därför är är $|mq| < |mp|$, vilket medför att q ligger i cirkeln, och detta motsäger att ℓ' är en tangent.

Övning 3.4: Använd resultatet som föregår övningen!

Övning 4.1: Notera att $|bc| = |ac| - |ab|$ och använd först delen av satsen.

Övning 4.3: Kopiera beviset av det första och det andra likformighetsfallet och använd transversalsatsen samt det tredje kongruensfallet.

Övning 6.1: Bisektrisen till två strålar består av de punkter som har samma (minsta) avstånd till de båda strålarna. Givet en punkt på bisektrisen så finns det bara en punkt på varje stråle där detta minsta avstånd antas, nämligen i skärningen med normalen till strålen genom punkten.

Genom samma resonemang som för mittpunktsnormalerna finner man att det finns exakt en punkt p där bisektriserna skär varandra, nämligen den som har samma avstånd r till alla tre sidorna i triangeln. Cirkeln genom p med radie r kommer bara att skära triangeln i de tre punkter som har avstånd r till p.

Övning 6.4: Dra en höjd (en normal till en av sidorna som går genom det tredje hörnet) från en av de närliggande sidorna till vinkeln θ och tillämpa Pythagoras sats på de båda rätvinkliga trianglarna som uppstår.

Övning 6.5: Se avsnitt 1 i kapitel 5.

Övning 6.6: Eftersom en euklidisk transformation bevarar vinklar och bara gör en förändring av längdskalan med ett tal r är det klart att en given triangel avbildas på en likformig triangel. Två givna likformiga trianglar abc och $a'b'c'$ kan först flyttas med kongruensavbildningar så att a och a' hamnar i origo och strålarna ab och ac hamnar på strålarna $a'b'$ respektive $a'c'$. En lämplig dilatation avbildar nu den ena triangeln på den andra.

Övning 9.1: Fixera en punkt 0 och en stråle $0c$. Tag normalen i c till strålen. Från ordningen på normalen kan vi bestämma vilka vinklar med strålen $0c$ ovan som ena vinkelben, som är positiva respektive negativa. Vid definition av $\psi(a)$, avsätt sträckan så att strålen $0\psi(a)$ har större vinkel än strålen $0a$. Avbildningen ψ är inte en kongruensavbildning, för vore den, skulle sträckan $0a$ vara kongruent med sträckan $0\psi(a)$, men detta strider mot sats 3.9.

Kapitel 5

Övning 2.1: Om en cirkel C inte skär x-axeln, eller tangerar x-axeln, så undviker C helt övre eller undre halvplanet. Om C skär x-axeln i två punkter som båda ligger utanför enhetscirkeln så skär inte denna och C varandra alls.

Övning 5.4: Detta är precis det andra kongruensfallet i absolut geometri.

Övning 5.5: Efter förflyttning kan vi anta att vinkeln α ligger i origo och att ena sidan ligger på x-axeln och andra följaktligen ligger på strålen ℓ given av $t \mapsto te^{i\alpha}$. Enligt antagande finns en punkt r på positiva x-axeln sådan strålen ut från x-axeln från r med vinkel β mot x-axeln skär ℓ med vinkel γ. Om $r' > r$ och man tar strålen ut från r' med vinkel β så blir den uppkomna triangeln större än den förra och följaktligen defekten större, så vinkeln mot ℓ måste vara mindre.

Övning 5.7: Jämför exempel 3.8; $r = \log\sqrt{\frac{35}{3}}$, $m = \frac{13-\sqrt{105}}{8}$

Övning 5.8: Cirkeln går genom $1/5$ och $5/7$ och x-axeln är en diameter, eftersom den skär cirkeln vinkelrätt. Diameterns längd är alltså

$$\log\left(\frac{1+5/7}{1-5/7}\right) - \log\left(\frac{1+1/5}{1-1/5}\right) = \log 4,$$

så $r = \log 2$. Medelpunkten x ligger på positiva x-axeln och

$$\log\frac{1+x}{1-x} = \log\frac{1+1/5}{1-1/5} + \log 2,$$

vilket ger $x = 1/2$.

Övning 5.9: Eftersom $a = 3 - 2\sqrt{2}$ ligger på x-axeln är denna en diameter till cirkeln. Cirkeln måste gå genom de två punkter på x-axeln som har avstånd $r = \log\sqrt{2}$ till a. Punkterna är 0 och $1/3$. Cirkeln är alltså $|z - 1/6| = 1/6$.

Övning 5.11: Arean är $\pi/2 - 2\arctan(4/5) = \arctan(9/40)$.

Övning 5.13: Antag att $a_0, a_1, \ldots, a_n, a_{n+1}$ är en uppräkning av hörnen så att randen till P precis är unionen av alla sträckorna a_0a_1, a_1a_2, $\ldots a_na_{n+1}$,$a_{n+1}a_0$. Om P är konvex ligger sträckan a_na_0 helt i P så $m(P)$ är arean av den polygon P' vars hörn är $a_0, \ldots, a_n$ plus arean av triangeln $a_na_{n+1}a_0$. Även P' kommer att vara konvex. Använd nu induktion över antalet hörn.

Övning 5.16: Sträckan $1/2, 5/7$ är kongruent med sträckan $0, 1/3$. Eftersom triangeln är rätvinklig och likbent kan vi anta att den har hörn i $0, 1/3, i/3$. Arean är $\pi/2 - \arctan(4/5) = \arctan(9/40)$.

Övning 5.15: Triangeln T_t med hörn i 0, t och $(1+i)t$ har halva arean av Q_t. Vinkeln vid t går mot en rät vinkel när $t \to 1/\sqrt{2}$. Alltså $m(T_t) \to \pi - \pi/4 - \pi/2 = \pi/4$, så $m(Q_t) \to \pi/2$.

Övning 5.17: Gränsfallet är när tredje vinkeln är 0, t ex som i triangeln med hörn i 0, r samt $e^{i\theta}$ där $r > 0$ är valt så att vinkeln vid r är rät. Via avbildningen $\phi(z) = \frac{z-r}{1-rz}$ ser man att detta innebär att $\phi(e^{i\theta}) = i$. Löser man denna ekvation får man att $\cos\theta = 2r/(1+r^2)$. Sträckan mellan 0 och r, dvs katetens längd, är

$$\log\frac{1+r}{1-r} = \frac{1}{2}\log\frac{1+\frac{2r}{1+r^2}}{1-\frac{2r}{1+r^2}} = \frac{1}{2}\log\frac{1+\cos\theta}{1-\cos\theta}.$$

Övning 5.18: Två euklidiska cirklar med euklidiska radier 1 skär varandra rätvinkligt omm det euklidiska avståndet mellan deras medelpunkter är $\sqrt{2}$. Bilden av $|z-(1+i)| = 1$ under avbildningen $z \mapsto e^{i\theta}z$ är en cirkel med medelpunkt a på cirkeln $|z| = \sqrt{2}$, och θ ska väljas så att euklidiska avståndet mellan $1+i$ och $a = e^{i\theta}(1+i)$ är $\sqrt{2}$. Men då är 0, $1+i$ och a hörn i en euklidisk liksidig triangel, så $\theta = \pm\pi/3 + m2\pi$.

Övning 5.20: Man kan anta att triangeln har hörn i $\pm ir$ och r. Det räcker att beräkna arean av triangeln T med hörn i 0, r och ir. Den är likbent med två vinklar α där

$$\tan\alpha = \frac{1-r^2}{1+r^2}.$$

Alltså blir arean av hela triangeln

$$2d(T) = 2(\pi - \pi/2 - 2\alpha) = \pi - 4\alpha = \pi - \arctan\frac{1-r^2}{1+r^2}.$$

Litteraturförteckning

[DD] D. van Dalen: *Logic and Structure*, Springer Universitext (1997).

[OH] O. Hanner: *Logik, tal och geometri*, Kompendium Göteborg (1980).

[JK] J-K. Krivine: *Introduction to Set Theory*, D. Reidel Publ. Company.

[LÅL] L-Å. Lindahl: *En inledning till geometrin*, Kompendium Uppsala (1985).

[JL] J. Löfström: *Geometriska bilder av världen*, Kompendium Göteborg (1997).

[EM] E. Mendelson: *Introduction to Mathematical Logic*, D. Van Nostrand Company (1976).

[GM] G. Moore: *Zermelo's Axiom of Choice– its origins, development, and influence*, Springer Verlag 1982.

[UP] U. Persson: *Hyperbolic Geometry*, Göteborg (1998).

[IS] I. Stewart: *Galois Theory*, Chapman and Hall (1973).