Ge svar på tal: en inledande kurs i matematisk analys

David Rule

Höstterminen 2017

Kursens hemsida
(`www.davidrule.net/tata79/`)

Inlämningsuppgifter
(`www.davidrule.net/tata79/ex/`)

Innehåll

Senast ändrad: 10 november 2017.

1 Introduktion

I den här kursen ska vi helt enkelt börja om. Vi ska börja från enkla och förhoppningsvis rimliga idéer och bygga upp all trigonometri och analys därifrån. Ni har säkert läst mycket matematik i gymnasiet och är till och med ganska duktiga studenter, så varför skulle vi återvända till matematikens början? I gymnasiet har ni med all sannolikhet koncentrerat er på metoder, det vill säga på hur man kan komma fram till de rätta svaren till numeriska frågor. Men på universitetsnivå är vägen till svaret viktigare—vi lägger mer vikt på att argumentera och tänka logiskt än på att bara hålla med läraren eller facit i boken.

En skeptiker skulle kunna svara: ”Men i matematik är någonting antingen rätt eller fel och jag är duktig på att komma fram till det rätta svaret, så det är ingen idé att upprepa så mycket.” Till skeptikern säger jag att fastän du säkert förstår bra vad du har lärt dig är jag övertygad om att genom att lära dig logiken bakom får du en mycket djupare och intressantare kunskap än genom att bara lära dig vissa metoder.

Till det skulle kunna skeptikern svara: ”Jo, men jag är bara intresserad av tillämpningar av matte och är inte intresse av själva ämnet.” Den synpunkten förstår jag, men syftet med en

universitetsutbildning är att ge möjlighet att skaffa sig nya kunskaper och tillämpningar. Visst finns det vissa begrepp ni måste lära er för att klara särskilda kurser, men jag vill också lära er någonting allmänt tillämpbart, det vill säga, jag vill ge er ett sätt att tänka som kan tillämpas på vilket problem som helst.

Matematik är ett unikt ämne som befinner sig mellan konst och vetenskap. Det är mer kraftfullt än andra vetenskaper—des som är bevisat rätt i matematik är rätt för evigt. Matematiken är lika vacker som den vackraste konst, så jag tror ni kommer njuta av den här kunskapen.

2 Grundlägande koncept och verktyg

2.1 Logik och räkneoperationer

2.1.1 Logik

Framför allt är matematik ett sätt att tänka. Den är förstås ett ämne där man funderar på tal och mönster men en matematiker använder deduktiva slutledningar för att upptäcka sanningar om världen. För att förstå matematik är det viktigt att man förstår deduktion. I det här avsnittet tittar vi lite närmare på logiska argument, hur man bygger dem och hur de kan gå fel. Aven om det är svårare att komma i gång på detta sätt kommer man längre än man skulle kunna tro.

Vi kommer att betrakta många påståenden. Ett *påstående* är en mening som kan vara antingen sann eller falsk. Till exempel:

Jag har en dotter; (2.1)

Köpenhamn är Danmarks huvudstad; eller (2.2)

Köpenhamn ligger i Tanzania. (2.3)

Vi vet alla att (2.2) är sant och (2.3) är falskt. Sanningen av (2.1) beror på vem som säger det, men vid alla tillfällen är det antingen sant eller falskt.

Om man vet eller antar att några påståenden är sanna (så kallade *antaganden*) så kan man använda deduktiva slutledningar för att bygga nya påståenden som också är sanna när antagandena stämmer (och kallade *slutsatser*).

Sveriges riksdag ligger i Sveriges huvudstad.
Sveriges huvudstad kallas för Stockholm.
Därför ligger Sveriges riksdag i Stockholm.

Men vi måste vara försiktiga att undvika ogiltiga argument:

Alla fåglar lägger ägg.
Kackerlackor lägger ägg.
Därför är kackerlackor fåglar.

Att alla fåglar lägger ägg medför inte att alla djur som lägger ägg är fåglar. Ibland kan det vara svårt att se problemet i ett argument:

Alla människor som föddes i Paris föddes i Frankrike.
Jean Paige föddes i Paris.
Därför föddes Jean Paige i Frankrike.

Det ser rimligt ut, men Jean Paige föddes faktiskt i USA. Problemet ligger i hur man tolkar ordet ”Paris”. I första påstående menar vi Frankrikes huvudstad, men i det andra påståendet pratar vi om Paris, Illinois. Båda antagandena är sanna men enkla att misstolka.

Från enkla påståenden kan man bygga mer komplicerade påståenden: Om två påstående är sanna kan vi lägga ihop dem med ordet ”och”:

> Jag bor på landet.
>
> Jag jobbar i stan.
>
> Därför bor jag på landet och jobbar i stan.

Om det finns två eller flera alternativa kan man använder ”eller”:

> Jag tar antingen busen eller tunnelbanan till jobbet. (2.4)

Negation är ett förnekande av ett påstående. Till exempel, negationen av (2.3) är:

> Köpenhamn ligger utanför Tanzania.

Ibland när vi försöker förneka ett påstående ser vi att det inte var så tydligt som vi trodde. Till exempel är negationen av (2.1) kunde vara:

> Jag har ingen dotter; (2.5)

eller,

> Jag har antingen ingen dotter eller mer än en dotter.

Rent logiskt är (2.5) negationen av (2.1). Men förvirring kan uppstå på grund hur vi brukar tolka saker som (2.1). I vardagspråk skulle det vara konstigt att gömma man har flera barn än man säger, men rent logiskt är det helt möjligt. Inom logik (och matematik) måste vi hålla oss till det som verkligen sägas och slipper de underförstådda delar! Jämför (2.1) med det följande påståendet:

> Martin har en penna.

Här skulle ingen tänker det var konstigt om Martin hade flera pennor även om han bara sa han har en. Det samma gäller när man använder ”eller”: (2.4) utesluter inte att man tar både busen och tunnelbana till jobbet. Om du skriver någonting du tror kan misstolkas så är det bäst att förtydliga.

För två påståenden A och B skriver vi $A \implies B$ för att säga A *medför* B. Det är ekvivalent med att säga B *medförs av* A, eller $B \impliedby A$. Men observera att det är inte det samma som $B \implies A$: Att jag parkerade fel medför att jag får en bot, men att jag fick en bot medför inte att jag hade parkerat fel. Om $A \implies B$ och $B \implies A$ skriver vi $A \iff B$.

För ett påstående A skriver vi $\neg A$ för negationen av A. Till exempel:

$A =$ ”Varje land har en flagga.”

$\neg A =$ ”Det finns ett land som inte har någon flagga.”

Påståendet $\neg B \implies \neg A$ är *kontrapositionen* av $A \implies B$. Både påståendena är logiska ekvivalent, det vill säga $A \implies B$ om och endast om $\neg B \implies \neg A$. Till exempel:

> Om tåget är försenat ringer jag till frugan;

är ekvivalent med

> Om jag har inte ringt till frugan så är tåget inte försenat.

2.1.2 Hur man skriver argument inom matematik

Man kan säga att matematik är en tillämpning av logik. Varje påstående i matematik följs av ett logiskt argument från grundläggande begrepp, så kallade *axiom* som man antar utan frågor. I nästa avsnitt kommer vi titta lite närmare på axiomen som räcker för att räkna med tal, men först tänker vi på hur man beskriver våra matematiska tänkar för andra.

Det skulle vara oerhört tråkigt och långsamt att bevisa alla nya påstående från axiomen, så istället bygger vi upp olika användbara påståenden som vi noggrant bevisar med hjälp av axiom tillsammans med påståendena vi har bevisat tidigare. I princip kan alla påståenden bevisas från axiomen fast vi behöver inte upprepa ett så långt argument för att visa det. På så sätt sparar man tid, men ännu viktigare är att man skapar någon slags intuition över ämnet. Man kan tänka att verkligheten är ett mörkt rum och matematisk logik är en ficklampa som visar oss vad som finns i rummet. Genom att skriva ner påståendena ritar vi en karta över rummet. Ficklampan är jättebra att avslöja enskilda föremål men kartan ger oss en bättre bild över hela rummet.

Till skillnad från en ficklampa kan det vara svårt att avslöja vad som finns i rummet. Till exempel, några geometriska problem som betraktades i antiken löstes först under 1800-talet.[1] Viktiga påståenden kallas för *satser*. Det kan vara både smidigt och nödvändigt att införa *definitioner* som precisera vad ett tekniskt ord eller begrepp egentligen betyder. Det gör det lättare att prata om idéer precis som ett bra ordförråd gör det enklare att uttrycka sig.

Nu går vi genom ett exempel, som förhoppningsvis gör idéer tydligare. Vi antar att man redan är van vid enkla räkneoperationer med tal och förstår likhet och olikhet mellan tal (men återkommer till dem i avsnitt 2.2 nedan). I figur 1 finns det en påminnelsetabell om grundläggande notation vi använder oss av genom texten. Vi börjar med en definition.

	Begrepp	Notation	Betydelse
Relationer	Likhet	$a = b$	'a är lika med b'
	Olikhet	$a \neq b$	'a är inte lika med b'
		$a < b$	'a är strängt mindre än b'
		$a > b$	'a är strängt större än b'
		$a \leq b$	'a är mindre än eller lika med b'
		$a \geq b$	'a är större än eller lika med b'
Operationer	Addition	$a + b$	'a plus b'
	Subtraktion	$a - b$	'a minus b'
	Multiplikation	ab, $a \times b$ eller $a \cdot b$	'a gånger b'
	Division	a/b, $\frac{a}{b}$ eller $a \div b$	'a delat med b'
	Kvadrering	a^2	'a gånger a'

Figur 1: Notation för att anteckna relationer mellan och operationer på två tal a och b. Man använder också parentes '(' och ')' för att markera en operatorprioritet som skiljer sig från den vanlig: multiplikation och division först, sen addition och subtraktion.

Definition 2.1. *Givet två positiva heltal n och m kallar vi r resten av n delat med m om*

$$n = qm + r \quad \textit{för något positiva heltal } q \tag{2.6}$$

där r är ett heltal sådant att $0 \leq r < m$.[2]

[1] Till exempel, cirkelns kvadratur (`https://sv.wikipedia.org/wiki/Cirkelns_kvadratur`) och vinkelns tredelning (`https://sv.wikipedia.org/wiki/Vinkelns_tredelning`).

[2] När man skriva två olikhet eller likhet tillsammans det betyder att båda gäller samtidigt: så här till exempel $0 \leq r$ och $r < m$.

Terminologien i definition 2.1 är lite tydligare om vi skriver (2.6) i den ekvivalenta formen $n/m = q + r/m$.

Nu vet vi precis betydelsen av ordet *rest* och kan gå vidare och undersöka egenskapen hos tal. Näst upptäcker vi en sats som är långt ifrån uppenbar utan ett bevis. Det handlar om begreppet rest vi har definierat i definition 2.1 samt de grundläggande begrepp positiva tal, olikhet och division.

Sats 2.2. *Betrakta tre positiva heltal n_1, n_2 och m sådana att:*

1. *n_1 delat med m har rest 1; och*

2. *n_2 delat med m har rest 1.*

Då är resten av produkten $n_1 n_2$ delat med m också 1.

För att dela upp texten lite tydligare använder vi oss av en rubrik i kursivstil "*Bevis*" som säger beviset börjar. Och slutar med en lite fyrkant "□" som säger beviset är klart.

Bevis. Eftersom n_1 delat med m har rest 1 vet vi från (2.6) att det finns ett positivt heltal q_1 så att

$$n_1 = q_1 m + 1.$$

Eftersom n_2 delat med m har rest 1 vet vi också från (2.6) att det finns ett positivt heltal q_2 så att

$$n_2 = q_2 m + 1.$$

Därför kan vi dra slutsatsen att

$$n_1 n_2 = (q_1 m + 1)(q_2 m + 1) = (q_1 q_2 m + q_1 + q_2)m + 1,$$

så (2.6) i Definition 2.1 uppfylles med $n = n_1 n_2$, $q = (q_1 q_2 m + q_1 + q_2)$ och $r = 1$. Dessutom eftersom 1 är, till exempel, resten av n_1 delat med m så är $0 \leq 1 < m$. Därför uppfyller 1 alla krav från definition 2.1 att vara resten av $n_1 n_2$ delat med m. □

Observera att rent logiskt är det okej att använda antagandena från satsen i sitt bevis, men inte slutsatsen! Den måste bevisas utan att använda någonting vi har inte bevisat tidigare. I beviset har vi använt oss av bara egenskapen (2.6) från definition 2.1 som gäller enligt satsens antaganden 1 och 2.

Till sist observera att vi skriver matematik i hela meningar. Visst använder vi också mycket matematisk notation för att skriva snabbt och tydligt, men det kan inte ersätta vanliga ord helt och hållet. Ord hjälper oss att förstå vad det är man antar, hur man räknar fram det och vad man kan dra för slutsatser. Utan den förklaringen skulle argumentet vara otydligt. Det samma gäller när man löser uppgifter eller problem i arbetslivet. Det är helt okej att skriva korta anteckningar för att själv förstå problem eller förklara för en vän medan ni sitter och pratar, men det räcker inte som en fullständigt lösning om man inte kan läsa den och förstå utan föregående kunskap om problemet. För att veta om det du har skrivit räcker som en fullständigt lösning, tänk om du skulle kunna vara nöjd om du läste den som en lösning i en lärobok. Där har man ingen möjlighet att utvidga lösningen med ytterligare förklaringar, så allt måste finnas redan i texten. Färdigheten att förklara sina tankar skriftligt är en viktigt del av en universitetsutbildning.

2.2 Längdmått och de reella talen

Vi har redan pratat om tal och använt oss av räkneoperationer men nu vänder vi tillbaka och funderar över vad det är vi antar om de begrepp samt talen vi räknar med. Vi ställer frågan: Vad är tal?

Det är ett svårt begrepp att definiera som det ofta är när ett begrepp är så fundamentalt. Precis som en ordbok aldrig lyckas att definiera ett ord utan att lita till andra ord, det verkar som vi inte kan klara det med tal heller. Trots att vi inte precis kommer definiera tal kan vi få bra intuition på följande sätt.

För att komma igång snabbt men fortfarande med tillräcklig noggrannhet ska vi använda oss av geometriska koncept. Vi föreställer oss en rät linje och en punkt på linjen som vi kallar för noll (tecknas 0). Linjen fortsätter för evigt till både vänster och höger om punkten 0. Nu tar vi en annan punkt på linjen, till höger om 0 och kallar den för ett (tecknas 1). Punkten som har samma avstånd till höger om 1 som 1 ligger till höger om 0 kallas två (2). Och på så sätt kan vi fortsätta och hitta alla punkter som representerar *icke negativa heltal* $0, 1, 2, 3, \ldots$. Icke negativa heltal $0, 1, 2, 3, \ldots$ kallas också för *naturliga tal*.[3] På liknande sätt representeras vänster om 0 alla *negativa heltal* $-1, -2, -3, \ldots$.

De *reella talen* ska definieras som alla punkter som ligger på linjen, som vi kallar *reella linjen*. Det vill säga inte bara de hela talen $\cdots -2, -1, 0, 1, 2, 3, \ldots$ utan också alla punkter som ligger emellan. Vår definition är på något sätt bristfällig. Till exempel antar vi redan nu att man vet vad punkt, linje, avstånd, och så vidare, betyder. Men förhoppningsvis har ni en bra känsla för de geometriska begreppen och vi kommer att utveckla idéerna mer under kursens gång.

Till reella tal läggs två kompositionsregler, addition (som tecknas '+') och multiplikation (som tecknas antingen '$\times$', '$\cdot$' eller bara genom att skriva två tal bredvid varandra). Till skillnad från de reella talen är det lättare att lägga en stabil grund för de reglerna. Vi undersöker de tillsammans med deras motsatser, subtraktion och division ($-$, och $\div$ eller /) i avsnitt 2.2.1.

2.2.1 Ett axiomsystem för de reella talen

I det här avsnitt kommer vi fundera på de grundläggande axiom vi antar om aritmetik med reella tal samt olikhet och likhet. Med de där axiomen kan man vid behöv kontrollera att de vanliga räkneroperationerna vi använder oss av stämmer. Det kan vara lätt att glömma grundläggande räknerreglar så de funkar också som en nyttig påminnelse. Fast de ser så uppenbara ut vill vi anteckna det här: Senare i kursen upptäcker vi så kallade *komplexa tal* och då ser vi att några axiom för reella tal inte stämmer för komplexa tal och därför påverkar det hur man får räkna med dem.

I reella tal finns det två överallt definierade kompositionsregler, addition (tecknas $+$) och multiplikation (tecknas $\times$, $\cdot$ eller bara genom att skriva två tal bredvid varandra). En ordningsrelation $<$ definieras också. Reella tal tillsammans med kompositionsreglerna och ordningsrelationen uppfyller de följande axiomen.

I **De algebraiska axiomen:**

(a) $a + b = b + a$ och $ab = ba$ för alla reella tal a och b (de kommutativa lagarna);

(b) $(a + b) + c = a + (b + c)$ och $(ab)c = a(bc)$ (de associativa lagarna);

(c) $a(b + c) = ab + ac$ för alla reella tal a, b och c (den distributiva lagen);

(d) Det finns två olika tal 0 och 1 så att $a + 0 = a$ och $a \times 1 = a$ för alla reella tal a (existens av neutrala element);

[3] Man måste vara lite försiktig eftersom definitionen av naturliga tal varierar beroende på författaren. Ibland räknas inte talet 0 som ett naturligt tal.

(e) Till varje $a \neq 0$ finns det inversa element $-a$ och a^{-1} så att $a + (-a) = 0$ och $a \times a^{-1} = 1$ (existens av invers).[4]

II **Ordningsaxiom:**

(a) För godtyckliga a och b gäller en och endast en av möjligheterna $a < b$, $a = b$ och $a > b$ (trikotomi);

(b) $a < b$ och $b < c$ medför att $a < c$ (transitiva lagen);

(c) $a < b$ medför att $a + c < b + c$ för alla reella tal c;

(d) $a < b$ och $0 < c$ medför att $ac < bc$.

Det finns ytterligare ett par axiom men det ska vi ta upp i avsnitt 3.2.3. Med axiomen ovan tillsammans med dem i III i avsnitt 3.2.3 kan vi komma fram till alla de reella talens egenskaper. Att skriva $a > b$ har samma betydelse som $b < a$. Vi använder också notationen $a \leq b$ för att säga antingen $a < b$ eller $a = b$ samt $a \geq b$ för att säga antingen $a > b$ eller $a = b$.

Observera i I(e) skriver vi om existens av inversa element istället för subtraktion och division. Antagandet att addition och multiplikation funkar som i I och II tillsammans med existensen av inversa element gör att vi kan definiera subtraktion och division.

Definition 2.3. *Givet två reella tal a och b definieras operationerna substraktion och division enligt:*

1. $a - b := a + (-b)$;[5] *respektive*

2. $a/b := a(b^{-1})$ *om* $b \neq 0$.

Det vill säga subtraktion av b definieras som addition av inversa elementet $-b$ och division med b definieras som multiplikation med inversa elementet b^{-1}.

Axiomen I och II räcker för att komma fram till alla räknerreglarna vi är vana vid från skolan. Speciellt får man göra lika räkneoperationer på både leden av en likhet och behålla likheten.

Sats 2.4. *För reella tal a, b och c medför påståendet $a = b$ att*

1. $a + c = b + c$;

2. $a - c = b - c$;

3. $ac = bc$; *och*

4. $a/c = b/c$ *om* $c \neq 0$.

Innan vi bevisar satsen är det bra att tala om hur bevisningen går till. Det är ett motsägelsebevis. Ett *motsägelsebevis* är ett bevis där vi antar att satsen vi vill bevisa är felaktig och sen kommer fram till en motsägelse. Därifrån drar vi slutsatsen att antagandet är fel och alltså är satsen rätt![6]

[4] Som skrivet ovan använder vi också notationen $1/a = a^{-1}$ samt $a/b = a \times b^{-1}$.

[5] Tecknet ':=' betyder att vänsterledet definieras som högerledet.

[6] Det finns en stor filosofisk debatt kring metoden, men det är någonting för en annan kurs.

Bevis. Vi bevisar bara 1 och lämnar de andra som övning åt läsaren.

Enligt II(a) har vi tre möjligheter och precis en av de måste stämma: Antingen $a+c < b+c$, $a+c > b+c$ eller $a+c = b+c$. Om vi lyckas utesluta de två olikheterna så är likhet den enda möjligheten.

Först antar vi att $a+c < b+c$. Enligt I(e) får vi använda II(c) med talet $-c$ och får

$$(a+c)+(-c) < (b+c)+(-c) \tag{2.7}$$

Då är vänsterledet

$$(a+c)+(-c) \underset{\text{I(b)}}{\uparrow}= a+(c+(-c)) \underset{\text{I(e)}}{\uparrow}= a+0 \underset{\text{I(d)}}{\uparrow}= a \tag{2.8}$$

och högerledet är

$$(b+c)+(-c) \underset{\text{I(b)}}{\uparrow}= b+(c+(-c)) \underset{\text{I(e)}}{\uparrow}= b+0 \underset{\text{I(d)}}{\uparrow}= b. \tag{2.9}$$

Då får vi skriva om (2.7) som $a < b$ vilket är en motsägelse till $a = b$ enligt II(a).

Sen antar vi att $a+c > b+c$. Enligt I(e) får vi använda II(c) med talet $-c$ och får

$$(a+c)+(-c) > (b+c)+(-c) \tag{2.10}$$

Enligt (2.8) och (2.9) får vi skriva om (2.10) som $a > b$ vilket är också en motsägelse till $a = b$ enligt II(a).

Därför är den enda möjligheten kvar från II(a) att $a+c = b+c$ och 1 är bevisat. □

Nedan har vi flera exempel på enkla räkneregler som följer av axiomen.

Exempel 2.5. *Bevisa att $a0 = 0$ för alla reella tal a.*

Enligt I(d) (med $a = 0$) är $0 = 0+0$ och så $a0 = a(0+0) = a0+a0$ enligt I(c). Sedan medför I(b), I(d) och I(e) att $0 = a0+(-(a0)) = (a0+a0)+(-(a0)) = a0+(a0+(-(a0))) = a0$.

Exempel 2.6. *Bevisa att $(b+c)a = ba+ca$ för alla reella tal a, b och c.*

Enligt I(a) är $(b+c)a = a(b+c)$ och $a(b+c) = ab+ac$ enligt I(c). Sedan medför I(a) att $ab+ac = ba+ca$. Därför kan vi dra slutsatsen att $(b+c)a = ba+ca$.

Exempel 2.7. *I(e) säger att för ett givet reella tal a finns det minst ett tal $-a$ så att $a+(-a) = 0$. Bevisa att $-a$ är det enda tal som uppfyller det kriteriet. Det vill säga, bevisa att $b = -a$ är det enda talet så att*

$$a+b = 0. \tag{2.11}$$

Anta att $b = b_1$ och $b = b_2$ är två tal som uppfyller (2.11). Vi vill bevisa att $b_1 = b_2$. Enligt I(a) vet vi att det medför att $b_1+a = b_2+a$ och sen medför I(d) och I(b) att $b_1 = b_1+(a+(-a)) = (b_1+a)+(-a) = (b_2+a)+(-a) = b_2+(a+(-a)) = b_2$. Därför är $b_1 = b_2$.

Exempel 2.8. *Bevisa att $a(-b) = -(ab)$ för alla reella tal a och b.*

Eftersom vi vet från exempel 2.7 att inversen $-(ab)$ är unik, räcker det att visa $ab+a(-b) = 0$. Enligt I(c), I(d) och exempel 2.5 vet vi att $ab+a(-b) = a(b+(-b)) = a0 = 0$

Exempel 2.9. *Bevisa att $a < b$ och $c < 0$ medför att $bc < ac$.*

Observera först att enligt I(e) finns det ett $-c$ så att $c+(-c) = 0$. Sedan om $c < 0$ är $0 < -c$ enligt II(c). Därför får vi använda II(d) och säga att $a < b$ medför $a(-c) < b(-c)$. Enligt exempel 2.8 är $-(ac) < -(bc)$ och sen $bc = bc+ac-ac < bc+ac-bc = ac$.

2.2.2 Heltalspotenser

Precis som multiplikation är upprepad addition kan vi också upprepa multiplikation. Vi definierar

$$a^n := \underbrace{aa\ldots a}_{n \text{ gånger}} \tag{2.12}$$

för reella tal a och positiva heltal n. I synnerhet är $a^1 = a$, $a^2 = aa$ (*a i kvadrat*) och $a^3 = aaa$ (*a i kubik*).

Vi definierar också

$$a^{-n} := (a^{-1})^n \quad \text{och} \quad a^0 = 1 \tag{2.13}$$

för reella tal $a \neq 0$ och positiva heltal n. (Observera att vi har definierat a^{-1} på två sätt: först i I(e) från avsnitt 2.2.1 som inversa element till a och enligt (2.13) som $(a^{-1})^1$. Men de två uttrycken är lika med varandra.)

I början känns det lite konstigt att ha tre olika definitioner för a^n beroende på vilket fall från II(a) n hamnar i. Följande sats gör definitionerna lite mer rimliga. Tyvärr kräver satsens bevis många fall att gå genom på grund av definitionernas olika former.

Sats 2.10. *Givna godtyckliga nollskilda reella tal a och b, och hela tal n och m de följande regler gäller.*

1. $a^n a^m = a^{n+m}$,
2. $a^n a^{-m} = a^{n-m}$,
3. $a^n b^n = (ab)^n$,
4. $(a^n)^{-1} = a^{-n}$, *och*
5. $(a^n)^m = a^{nm}$.

Bevis. Vi bevisar bara 3 och 4 och lämnar de andra som en övning för läsaren.

För att bevisa 3 har vi tre fall: $n = 0$, $n > 0$ och $n < 0$.

Fall 1: $n = 0$. När $n = 0$ är $a^n = a^0 = 1$, $b^n = b^0 = 1$ och $(ab)^n = (ab)^0 = 1$ enligt definitionen (2.13). Så vi får att $a^n b^n = 1 \times 1 = 1 = (ab)^n$.

Fall 2: $n > 0$. Då får vi att

$$a^n b^n = \underbrace{aa\ldots a}_{n \text{ gånger}} \underbrace{bb\ldots b}_{n \text{ gånger}}$$

enligt (2.12). Genom att använda I(a) i avsnitt 2.2.1 $(n-1)$ gånger får vi att

$$\underbrace{aa\ldots a}_{n \text{ gånger}} \underbrace{bb\ldots b}_{n \text{ gånger}} = (ab) \underbrace{aa\ldots a}_{(n-1) \text{ gånger}} \underbrace{bb\ldots b}_{(n-1) \text{ gånger}}.$$

Genom att använda I(a) i avsnitt 2.2.1 $n-2$ gånger får vi att

$$(ab) \underbrace{aa\ldots a}_{(n-1) \text{ gånger}} \underbrace{bb\ldots b}_{(n-1) \text{ gånger}} = (ab)(ab) \underbrace{aa\ldots a}_{(n-2) \text{ gånger}} \underbrace{bb\ldots b}_{(n-2) \text{ gånger}}.$$

På samma sätt kan vi fortsätta och kommer fram till att

$$a^n b^n = \underbrace{(ab)(ab)\ldots(ab)}_{n \text{ gånger}} = (ab)^n.$$

Fall 3: $n < 0$. Vi bevisar först 3 om $n = -1$. Då får vi att $a^n = a^{-1}$, $b^n = b^{-1}$ och $(ab)^n = (ab)^{-1}$, så vi måste bevisa att inversen till ab är $a^{-1}b^{-1}$. Men

$$(ab)(a^{-1}b^{-1}) = a(ba^{-1})b^{-1} = a(a^{-1}b)b^{-1} = (aa^{-1})(bb^{-1}) = 1 \times 1 = 1$$

enligt I(a), I(b) och I(e) i avsnitt 2.2.1. Det säger att $(ab)^{-1} = a^{-1}b^{-1}$.

Sätt $p = -n$ för allmängiltiga $n < 0$ så $p > 0$. Då har vi att $a^n = a^{-p} = (a^{-1})^p$, $b^n = b^{-p} = (b^{-1})^p$ och $(ab)^n = (ab)^{-p} = ((ab)^{-1})^p$. Men vi har precis bevisat att $(ab)^{-1} = a^{-1}b^{-1}$ så $(ab)^n = (a^{-1}b^{-1})^p$. Så vill vi bevisa att $(a^{-1})^p(b^{-1})^p = (a^{-1}b^{-1})^p$ men det är precis som i det första fallet fast med a^{-1} och b^{-1} istället för a och b, som redan är bevisat.

Nu bevisar vi 4. Om $n = 0$ är likheten klar, så vi antar först att $n > 0$. Definitionen (2.13) säger att $a^{-n} = (a^{-1})^n$ så vi vill bevisa att $(a^n)^{-1} = (a^{-1})^n$, det vill säga att inversen till a^n är $(a^{-1})^n$. Men enligt I(a) i avsnitt 2.2.1 har vi att

$$a^n(a^{-1})^n = a^{n-1}(aa^{-1})(a^{-1})^{n-1} = a^{n-1} \times 1 \times (a^{-1})^{n-1} = a^{n-1}(a^{-1})^{n-1}.$$

Genom att tillämpa samma argument $(n-1)$ gånger får vi att $a^n(a^{-1})^n = 1$, som säger att $a^{-n} = (a^{-1})^n$ vilket var vad som krävdes. □

2.2.3 Talbeteckningssystem för heltal

Det finns spår av att människor har använt sig av aritmetik och utvecklat talbeteckningssystem sedan forntiden och den äldste verkar vara cirka 35 000 år gammal [2, s. 3]. Men den äldsta civilisationen som vi har bra historiska bevis angående deras användning av matematik är Mesopotamien. Vi brukar kalla det för Babylonisk matematik trots att Babylonien bara var en del av Mesopotamien. Mesopotamien betyder *Tvåflodslandet* på grekiska och var området runt floderna Eufrat och Tigris i dagens Irak som hade sitt ursprung kring 3 500 f.Kr. [8, p. 2].

Men vi ska återkomma till det babyloniska talsystemet för att det på något sätt är ganska avancerat och det är mer naturligt att först tänka på det egyptisk talsystem som utvecklades fem hundra år senare cirka 3 000 f.Kr. Det börjar ganska naturligt: Talet ett skrivas som ett lodrätt streck 𓏺, och för två till tio upprepar man strecket samma antal gånger som talet, så fyra är 𓏺𓏺𓏺𓏺 till exempel. I princip kunde man förtsättar i evighet och på så sätt representera vilket tal som helst. Men det blir snabbt klart att det är opraktiskt. Därför skapade de andra tecken som representerade högre tal. Till exempel 𓎆 stod för tio och 𓍢 för hundra och man skulle skriva

𓏺𓏺𓏺𓏺𓏺𓏺𓏺𓎆𓎆𓎆𓍢𓍢𓍢𓍢𓍢𓍢𓍢𓍢𓍢

för 947. Observera att vi skev talet baklänges jämfört med det system vi använder idag, men det gjorde ingenting for fornegyptierna, man kunde skriva åt både håll eller till och med i kolonner. Spegelbilden av en tecken kunde även skrivas med samma betydelse [2, s. 10]. Varje potens av tio hade sitt eget tecken upp till en million (se figur 2).

10^0	10^1	10^2	10^3	10^4	10^5	10^6
𓏺	𓎆	𓍢	𓆼	𓂭	𓆐	𓁨

Figur 2: De fornegyptitiska tecknen för potens av tio från ett till en miljon.

En nackdel med systemet var att det inte fanns något tecken för 10^7. Det skulle då vara jobbigt att skriva tal större än tio millioner. Medan man kunde hitta på sådana tecken vid behov måste man sluta någonstans, så ett sådant system kan aldrig ha ett tecken för alla heltal. Det leder oss till det babyloniska talsystemet. De hade till och med 59 teckna för tal från 1 till 59 (se figur 3).

	1	2	3	4	5	6	7	8	9
10	11	12	13	14	15	16	17	18	19
20	21	22	23	24	25	26	27	28	29
30	31	32	33	34	35	36	37	38	39
40	41	42	43	44	45	46	47	48	49
50	51	52	53	54	55	56	57	58	59

Figur 3: De babyloniska tecknen från och med ett till och med femtionio.

Logiken bakom tecknen är ganska lik den bakom de egyptiska, med ett streck 𒁹 för ett och en kil 𒌋 för tio. Men vad som är intressant är hur de skrev tal större än 59. Till exempel skrev de 64 så här:

𒁹 𒐜.

De skrev tecken för ett bredvid en för fyra. Men att ettan står till vänster om fyran betyder att det representerar 1×60 istället för ett och tecknen tillsammans representerade

$$1 \times 60^1 + 4 \times 60^0 = 64.$$

På så sätt i princip kunde man skriva vilket tal som helst och alla skulle förstå. Till exempel,

$$18184 = 5 \times 60^2 + 3 \times 60^1 + 4 \times 60^0$$

så man skriva 𒐝 𒐙 𒐜 . Ett problem med systemet däremot var att skilja mellan exempelvis

$$1 \times 60^2 + 0 \times 60^1 + 8 \times 60^0 \quad \text{och} \quad 1 \times 60^1 + 8 \times 60^0,$$

eftersom de i början inte hade något sätt att markera att talform 60^1 saknade. För att lösa det lämnade de ett mellanslag mitt i mellan så de skrev

𒁹 𒐕 , respektive 𒁹 𒐕 .

Senare införde de tecknet 𒑱 som hade också markerat slutet på en mening istället för ett mellanslag fast de aldrig använde den i slutet av ett tal! [7, p. 12]

Talbeteckningssystemet vi använder oss av idag kallas for det hindu-arabiska talsystemet. Det har sitt ursprung i Indien (medan de äldsta inskrifter med alla systemets drag fanns i Kambodja, år 683 e. Kr. [8, p.217]) och sen infördes först av persiska och arabiska matematiker och till slut i Europa under 800-talet. Precis som i det babyloniska talsystemet är placeringen av tecknen viktigt men en troligare förklaring är att man inspirerats av ett kinesiskt stav-talsystem [8, p. 217]. Vi har bara tio tecken (1, 2, 3, 4, 5, 6, 7, 8, 9 och 0) — att införa tecknet 0 var en innovation eftersom för första gången tolkades det som ett eget tal. Annars funkar det på samma sätt som det babyloniska systemet, till exempel

$$4308 = 4 \times 10^3 + 3 \times 10^2 + 0 \times 10^1 + 8 \times 10^0.$$

Det spelar ingen stor roll vilket tal b vi tar potenser av i ett talbeteckningssystem men vi behöver alltid b tecken (det hindu-arabiska talsystemet har tio, och det babyloniska talsystemet hade 59 plus ett mellanslag). Talet b kallas för *basen* av talsystemet. Till exempel, i det ternära talsystemet är $b = 3$. Man har tecknen 0, 1 och 2 och skriver

$$2 \times 3^4 + 1 \times 3^3 + 1 \times 3^2 + 0 \times 3^1 + 2 \times 3^0.$$

som 21102 fast det skrivs 200 i det (decimala) hindu-arabiska talsystemet.

2.2.4 Decimalutveckling

Det nästa problemet vi ska behandla är hur vi kan identifiera eller markera varje reellt tal. Ett sätt är med decimalutveckling som vi nu beskriver. Ett givet tal ligger någonstans på linjen som beskrevs här ovanför. I synnerhet ligger det mellan minst två angränsande heltal n och $n+1$. Vi preciserar att ett heltal m ligger mellan $m-1$ och m samt m och $m+1$. En decimalutveckling för talet börjar med n sen skriver vi en punkt[7] (.) sen fortsätter vi så här.

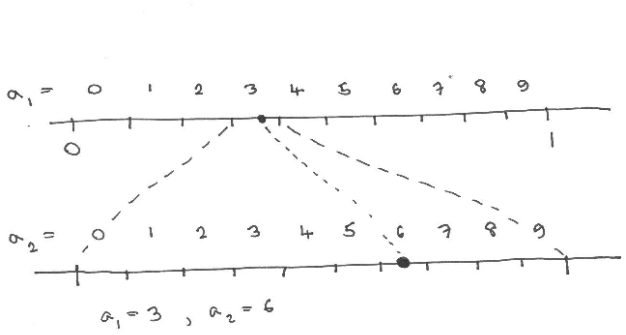
Först delar vi längden mellan n och $n+1$ i tio delar och märker delarna från 0 till 9. Vi preciserar att en randpunkt tillhör båda delarna som möter randpunkten. Det reella talet måste ligga i minst en del så vi väljer en av dem, säg del a_1. Så är a_1 den första siffran i decimalutvecklingen. Sen tar vi del a_1 och delar den på samma sätt i tio delar. Om igen måste det reella talet ligga i minst en del så vi väljer en, säg del a_2. Så är a_2 den andra siffran i decimalutvecklingen. Vi kan upprepa proceduren hur många gånger som helst och sen skriver vi decimalutvecklingen som

$$n.a_1a_2a_3a_4\ldots$$

Exempel 2.11. *I figur 4 vill vi räkna ut de första två siffrorna i decimalutvecklingen av talet som markeras med* •. *Självklart ligger det mellan* 0 *och* 1 *så* $n = 0$ *och sen delar vi avståndet mellan* 0 *och* 1 *i tio delar. Vi ser att talet ligger i del* 3. *Det betyder att* $a_1 = 3$. *Sen delar vi del* 3 *i tio delar och ser att talet ligger i del* 6 *som betyder att* $a_2 = 6$. *På så sätt ser vi att utvecklingen börjar som*

$$0.36\ldots.$$

Vi kan så klart fortsätta så länge vi vill, antagande att man kan zooma så mycket man vill.



Figur 4: Först ser vi att talet markerat med • ligger i del 3 av de tio delarna mellan 0 och 1, och sedan, när vi zoomar in, ligger det i del 6 av del 3.

Vi kan tolka en decimalutveckling som en upprepad mätning av avståndet från 0. Vid första mätningen är vi nöjda med en tiondels exakthet, nästa steg vill vi göra tio gånger bättre, och

[7] I några länder inklusive Sverige kan man använda ett komma (,) istället för punkt, men för att undvika förväxling när vi senare skriver koordinater använder vi en punkt.

så vidare. På detta sätt följer det samma mönster som heltal, fast nu betraktar vi negativa potenser. Ett tal har en decimalutveckling som börjar $3.26\ldots$ säger att det kan skrivs som

$$3 \times 10^0 + 2 \times 10^{-1} + 6 \times 10^{-2} + \ldots$$

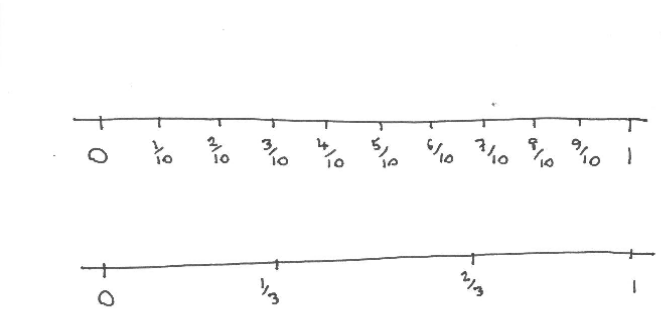
där $\ldots$ är en summa av fler negativa potenser av 10. Mer precist, det säger att talet ligger någonstans mellan

$$3 \times 10^0 + 2 \times 10^{-1} + 6 \times 10^{-2} \quad \text{och} \quad (3 \times 10^0 + 2 \times 10^{-1} + 6 \times 10^{-2}) + 10^{-2}.$$

2.2.5 Rationella tal

Ett annat sätt att markera vissa tal är att skriva ett bråk. Till exempel, när vi delar strecket mellan 0 och 1 i tio delar, finns det ett tal som ligger på högerkanten av del 0 och på vänsterkanten av del 1. Talet kallas för 1/10 eftersom det ligger en tiondel från 0. Om vi istället delar strecket i tre, kallas talet som ligger på hörgerkanten av del 1 och vänsterkanten av del 2 för 2/3. Se figur 5. Observera att det finns flera sätt att markera samma tal. Till exempel $1 = 10/10 = 3/3$ och $1/3 = 2/6$, och så vidare.

Sådana tal kallas för *rationella tal* och definieras som alla tal som kan skrivas som n/m där n och m är heltal.



Figur 5: Talen 0, 1/10, 2/10, 3/10, 4/10, 5/10, 6/10, 7/10, 8/10, 9/10 och 10/10 = 1, och 0 = 0/3, 1/3, 2/3 och 3/3 = 1 tecknat som bråk.

Exempel 2.12. *Skriv decimalutvecklingen $x = 0.6666\ldots$ (det vill säga $a_k = 6$ för alla positiva heltal k) som ett bråk.*

Om vi multiplicerar båda sidor av likheten $x = 0.6666\ldots$ med 10 får vi $10x = 6.6666\ldots$. Men

$$6.6666\ldots = 6 + 0.6666\ldots = 6 + x.$$

Alltså är $10x = 6 + x$ som i sin tur säger att $9x = 6$. Om vi delar båda sidor med 6 kommer vi fram till att

$$x = \frac{6}{9} = \frac{2}{3}.$$

Exempel 2.13. *Skriv decimalutvecklingen $x = 0.125125125\ldots$ (det vill säga $a_{1+3k} = 1$, $a_{2+3k} = 2$, $a_{3+3k} = 5$ för alla positiva heltal k) som ett bråk.*

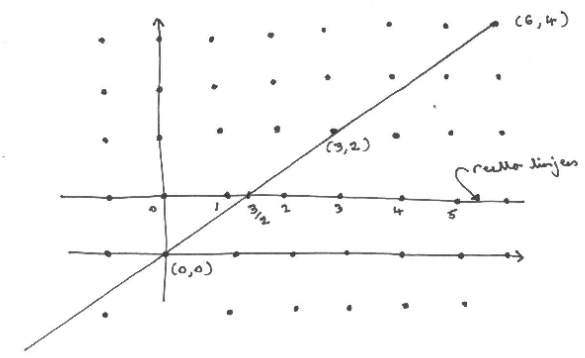
Om vi multiplicerar båda sidor av likheten $x = 0.125125125\ldots$ med 1000 får vi $1000x = 125.125125125\ldots$. Men

$$125.125125125\ldots = 125 + 0.125125125\ldots = 125 + x.$$

Alltså är $1000x = 125 + x$ *som i sin tur säger att* $999x = 125$*. Om vi delar båda sidor med* 999 *kommer vi fram till att*

$$x = \frac{125}{999}.$$

Det finns ett bra sätt att visualisera rationella tal — se figur 6. Vi ritar två axlar, en parallellt med den horisontella reella linjen men en enhet lägre, och den andra lodrät så att den skär den reella linjen i punkten 0. Punkten där de två axlarna skär varandra kallas för origo. En godtycklig punkt i planet kan identifieras genom två tal: Det första är avståndet med tecken a från punkten till den lodräta axeln, med plustecken om punkten ligger till höger om axeln och minustecken om den ligger till vänster; Det andra är avståndet med tecken b från punkten till den horisontella axeln med plustecken om punkten ligger ovanför axeln och minustecken om den ligger under. Talen kallas för *(kartesiska) koordinater* och skrivs (a, b). Nu markerar vi alla punkter som har koordinaterna (n, m) där både n och m är hela tal.



Figur 6: Rationella tal n/m kan tolkas som punkter med koordinater (n, m). Alla koordinater som representerar samma rationella tal ligger på linjen som skär r och origo. Till exempel, $3/2 = 6/4$.

Givet ett rationellt tal n/m rita en rät linje genom punkten med koordinaterna (n, m) och origo (som har koordinaterna $(0, 0)$). Vi ser att linjen mellan origo och (n, m) skär den reella linjen i punkten n/m. Omvänt ser vi att linjen som skär origo och ett reellt tal r, skär en punkt (n, m) där n och m båda är hela tal endast om $r = n/m$.

Tror du att alla reella tal är rationella? Det vill säga, kan alla reella tal r skrivas som $r = n/m$ för hela tal n och m? Det funderar vi på lite till senare.

2.3 Mängder och följder

2.3.1 Mängder

I avsnitt 2.2 har vi diskuterat olika samlingar av tal. I matematik kallas en samling av någonting för en *mängd*. I princip kan en mängd vara en samling av vad som helst, till exempel alla äpplen på ett träd, alla röda bilar i Sverige eller alla människor som bor i Helsingfors. Men för det mesta är vi intresserade av talmängder. Någonting som tillhör en mängd kallas för ett *element* i mängden och parenteserna '{' och '}' används för att lista alla element som tillhör en mängd: Till exempel, om mängden M innehåller talen 1, 2, 3, 5, och 7, kan man skriva

$$M = \{1, 2, 3, 5, 7\}.$$

Vi kan teckna påståendet att 5 är ett element av M genom att skriva $5 \in M$. För att beteckna att någonting inte tillhör till en mängd, skriver man till exempel $4 \notin M$. För de flesta mängder

som vi redan har nämnt finns det en välkänd notation. Naturliga tal tecknas som $\mathbf{N}$.[8] Det vill säga

$$\mathbf{N} = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, \dots\}.$$

Alla heltal tecknas som

$$\mathbf{Z} = \{\dots, -10, -9, -8, -7, -6, -5, -4, -3, -2, -1, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, \dots\},$$

Bokstaven $\mathbf{Z}$ kommer här från det tyska ordet *Zahlen*. Positiva heltal tecknas som

$$\mathbf{Z}_+ = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, \dots\}.$$

Det är användbart med lite utökad notation för att precisera en mängd och säga någonting som 'alla tal som uppfyller de här villkoren'. Till exempel, *jämna tal* är heltal n så att $n = 2k$ för ett heltal k. Det kan man skriva som

$$\{\dots, -10, -8, -6, -4, -2, 0, 2, 4, 6, 8, 10, \dots\} = \{n \in \mathbf{Z} \mid n = 2k \text{ och } k \in \mathbf{Z}\}.$$

Alltså betyder '|' 'så att'. Ibland skriver man ':' istället för '|'. *Udda tal* är då

$$\{n \in \mathbf{Z} \mid n = 2k + 1 \text{ och } k \in \mathbf{Z}\}.$$

Med den här notationen kan vi definiera mängden av alla *primtal* som

$$\{n \in \mathbf{Z}_+ \mid \text{om } n = pq \text{ med } p, q \in \mathbf{Z}_+ \text{ då } p, q \in \{1, n\}\},$$

det vill säga, primtal är alla tal $n \in \mathbf{Z}_+$ som inte har några faktorer utom 1 och n själv. Rationella tal är

$$\mathbf{Q} = \{a/b \mid a, b \in \mathbf{Z}\}.$$

Kom ihåg att för ett givet $r \in \mathbf{Q}$ finns det många möjliga a och b i $\mathbf{Z}$ så att $r = a/b$, men vi räknar dem alla som samma element i $\mathbf{Q}$.

Reella tal som vi försökte definiera i avsnitt 2.2 tecknas $\mathbf{R}$. En viktig typ av mängd är ett *intervall*. Ett intervall är alla reella tal mellan två tal $a, b \in \mathbf{R}$, $a < b$. Vi har olika notation beroende på om a och/eller b tillhör intervallet:

$$\begin{aligned}
[a, b] &:= \{x \in \mathbf{R} \mid a \leq x \leq b\};\\
[a, b) &:= \{x \in \mathbf{R} \mid a \leq x < b\};\\
(a, b] &:= \{x \in \mathbf{R} \mid a < x \leq b\};\\
(a, b) &:= \{x \in \mathbf{R} \mid a < x < b\}.
\end{aligned}$$

Längden av intervallet ovan definieras lika med $b - a$. Att skriva till exempel $a \leq x \leq b$ betyder att x uppfyller både $a \leq x$ och $x \leq b$ och liknande för de andra olikheterna. Vi betecknar också

$$\begin{aligned}
[a, \infty) &:= \{x \in \mathbf{R} \mid a \leq x\};\\
(a, \infty) &:= \{x \in \mathbf{R} \mid a < x\};\\
(\infty, b] &:= \{x \in \mathbf{R} \mid x \leq b\};\\
(\infty, b) &:= \{x \in \mathbf{R} \mid x < b\}.
\end{aligned}$$

För två mängder E och M säger vi att E är en *delmängd* till M och skriver $E \subseteq M$ om $x \in E$ medför att $x \in M$ för alla $x \in M$. Uppenbarligen är alla intervall delmängder av $\mathbf{R}$, så

[8] Kom ihåg fotnot 3.

vi kan till exempel skriva $(-1, 2) \subseteq \mathbf{R}$. Vi säger att två mängder A och B är lika och skriver $A = B$ om $A \subseteq B$ och $B \subseteq A$; det är det samma som att säga $x \in A$ om och endast om $x \in B$.

Vi säger att E är en *äkta delmängd* till M och skriver $E \subset M$ om $E \subseteq M$ men E inte lika med M.

Det finns också operationer på mängder som kallas *snitt* och *union*. För två mängder A och B definieras *snittet* av A och B

$$A \cap B := \{x \,|\, x \in A \text{ och } x \in B\}.$$

Unionen av A och B definieras

$$A \cup B := \{x \,|\, x \in A \text{ eller } x \in B\}.$$

Differensen mellan två mängder A och B skrivs $A \setminus B$ och definieras som

$$A \setminus B := \{x \,|\, x \in A \text{ och } x \notin B\}.$$

Om vi bara är intresserade av mängder som är delmängder av $\mathbf{R}$ (det vill säga att vårt universum är $\mathbf{R}$) definierar vi komplementet till en mängd A som

$$A^c := \mathbf{R} \setminus A. \tag{2.14}$$

Vad 'universum' är kan bero på kontexten och $\mathbf{R}$ i definitionen (2.14) ersätts ibland med ett annat 'universum', till exempel med $\mathbf{R}^2$ som definieras nedan.

Sats 2.14. *Givna godtyckliga mängder A, B och C är*

1. $(A \cap B) \cap C = A \cap (B \cap C)$,
2. $(A \cup B) \cup C = A \cup (B \cup C)$,
3. $(A \cup B)^c = A^c \cap B^c$ *och*
4. $(A \cap B)^c = A^c \cup B^c$.

Bevis. Beviset lämnas som en övning för läsaren. □

Del 1 och 2 i sats 2.14 säger att det inte är tvetydigt att skriva $A \cap B \cap C$ eller $A \cup B \cup C$ (utan parenteser). För en samling av mängder A_j indexerade med $j = 1, 2, \ldots, N$ skriver vi

$$\bigcup_{j=1}^{N} A_j := \{x \,|\, x \in A_j \text{ för minst ett } j = 1, 2, \ldots, N\}$$

och

$$\bigcap_{j=1}^{N} A_j := \{x \,|\, x \in A_j \text{ för alla } j = 1, 2, \ldots, N\}.$$

I avsnitt 2.2.5 ritade vi ordnade par av heltal i planet. Vi kan förstås också betrakta ordnade par av reella tal. Vi skriver

$$\mathbf{R}^2 := \{(x, y) \,|\, x, y \in \mathbf{R}\}$$

mängden av alla ordnade par av reella tal.

Mängden som inte inhåller något element alls kallas för den *tomma mängden* och tecknas $\emptyset$.

Definition 2.15. *Låt $A \subseteq \mathbf{R}$ vara en icketom mängd. Man säger att mängden A är uppåt begränsad om det finns $C \in \mathbf{R}$ så att*

$$a \leq C \quad \text{för alla } a \in A.$$

Talet C kallas för en övre begränsning av A. Man säger att mängden A är nedåt begränsad om det finns $c \in \mathbf{R}$ så att

$$a \geq c \quad \text{för alla } a \in A.$$

Talet c kallas för en undre begränsning till A.

Exempel 2.16. *Visa att 4 är en undre begränsning till $B = \{y \in \mathbf{R} \,|\, y = x^2 - 4x + 9, x \in \mathbf{R}\}$.*

Ett godtyckligt $y \in B$ har formen $y = x^2 - 4x + 9$ för något $x \in \mathbf{R}$ men $x^2 - 4x + 9 = (x-2)^2 + 5 \geq 5 \geq 4$ så $y \geq 4$ för alla $y \in B$.

Definition 2.17. *Låt $A \subseteq \mathbf{R}$ vara en icketom mängd. Om det finns ett tal $u \in \mathbf{R}$ så att*

$$a \leq u \quad \text{för alla } a \in A. \tag{2.15}$$

och

$$\text{till varje } \varepsilon > 0 \text{ finns det ett } a \in A \text{ så att } u - \varepsilon < a \tag{2.16}$$

kallas u för en minsta övre begränsning (eller supremum) till A. Om u finns skriver man $u = \sup A$. Om det finns ett tal $\ell \in \mathbf{R}$ så att

$$\ell \leq a \quad \text{för alla } a \in A.$$

och

$$\text{till varje } \varepsilon > 0 \text{ finns det ett } a \in A \text{ så att } a < \ell + \varepsilon$$

kallas ℓ för en största undre begränsning (eller infimum) till A. Om ℓ finns skriver man $\ell = \inf A$.

Om vi jämför (2.15) och (2.16) med definitionen 2.15 ser vi att (2.15) är ekvivalent med påståendet

$$u \text{ är en övre begränsning av } A \tag{2.17}$$

och (2.16) är ekvivalent med påståendet

$$\text{varje tal mindre än } u \text{ är inte en övre begränsning av } A. \tag{2.18}$$

Därför är talet u en minsta övre begränsning till A om och endast om både (2.17) och (2.18) är uppfyllde. På samma sätt är ℓ en största undre begränsning till A om och endast om

$$\ell \text{ är en undre begränsning av } A$$

och

$$\text{varje tal större än } \ell \text{ är inte en undre begränsning av } A.$$

Om en mängd A definieras med hjälp av ett index eller en parameter t, till exempel $A = \{y \,|\, y = f(t), t \in I\}$ för en given funktion $f\colon I \to \mathbf{R}$ och en given mängd $I \subseteq \mathbf{R}$, då skriver man

$$\sup A = \sup_{t \in I} f(t) \quad \text{och} \quad \inf A = \inf_{t \in I} f(t).$$

Exempel 2.18. *Visa att $\sup A = 0$ för $A = \{x \in \mathbf{R} \,|\, x \leq 0\}$.*

Det är självklart att $x \leq 0$ enligt A:s definition så (2.15) är uppfylld. Betrakta $\varepsilon > 0$. Då är $x = -\varepsilon/2 \leq 0$ så $x \in A$ och $0 - \varepsilon < -\varepsilon/2 = x$ så (2.16) är uppfylld.

Sats 2.19. *En mängd har högst en supremum. En mängd har högst en infimum.*

Bevis. Vi bevisar bara det första påståendet och lämnar det andra som en övning för läsaren.

Anta att det finns två olika supremum $u_1 < u_2$ till en mängd A. Vi får tillämpa (2.16) med $\varepsilon = (u_2 - u_1)/2$ för att få reda på att det finns $a \in A$ så att $u_1 < (u_2+u_1)/2 = u_2-(u_2-u_1)/2 < a$. Men (2.15) medför att $u_1 \geq a$ så vi har hittat en motsägelse till hypotesen att det finns två supremum till A. □

Sats 2.19 ½. *Låt I vara ett intervall och $a < b$ vara reella tal. Om*

$$I = (a,b), [a,b], [a,b), (a,b], (-\infty, b) \text{ eller } (-\infty, b]$$

är $\sup I = b$. *Om*

$$I = (a,b), [a,b], [a,b), (a,b], (a, \infty) \text{ eller } [a, \infty)$$

är $\inf I = a$.

Bevis. Vi bevisar satsen i fallet $I = (a,b)$ och lämnar det andra som en övning för läsaren.

Först visa vi att $\sup I = b$. Eftersom intervallet I är lika med mängden av alla x sådan att

$$a < x < b$$

följer det direkt att b är en övre begränsning av I, det vill säga $x \leq b$ för alla $x \in I$. Vi måste också visa att ett godtyckligt tal b_- strängt mindre än b inte är en övre begränsning av I. Vi kan skriva $b_- = b - \varepsilon$ för positivt $\varepsilon > 0$ och betraktar två fall: $0 < \varepsilon < 2(b-a)$ och $\varepsilon \geq 2(b-a)$.

Om $0 < \varepsilon < 2(b-a)$ får vi att

$$a = b - (b-a) < b - \frac{\varepsilon}{2} < b - 0 = b.$$

Så talet $x := b - \frac{\varepsilon}{2}$ är ett element i I. Dessutom är

$$b_0 = b - \varepsilon < b - \frac{\varepsilon}{2} = x$$

så b_- är inte en övre begränsning av I.

Om $\varepsilon \geq 2(b-a)$ är

$$b_- = b - \varepsilon \leq b - 2(b-a) = a + (a-b) < a + 0 = a$$

och därför ligger b_- till vänster av intervallet. Därför är det lätt att visa b_- inte är en övre begränsning av I: Betrakta $x := (a+b)/2$ till exempel. Vi ser att $x \in I$ och $b_- < x$.

Sammanfattningsvis har vi visat att b är en övre begränsning och varje tal mindre än b inte är en övre begränsning. Därför är b den minsta övre begränsningen av I.

Nu visa vi att $\inf I = a$. Eftersom intervallet I är lika med mängden av alla x sådan att

$$a < x < b$$

följer det direkt att a är en undre begränsning av I, det vill säga $a \leq x$ för alla $x \in I$. Vi måste också visa att ett godtyckligt tal a_+ strängt större än a inte är en undre begränsning av I. Vi kan skriva $a_+ = a + \varepsilon$ för positivt $\varepsilon > 0$ och betraktar två fall: $0 < \varepsilon < 2(b-a)$ och $\varepsilon \geq 2(b-a)$.

Om $0 < \varepsilon < 2(b-a)$ får vi att

$$a = a + 0 < a + \frac{\varepsilon}{2} < a + (b-a) = b.$$

Så talet $x := a + \frac{\varepsilon}{2}$ är ett element i I. Dessutom är

$$x = a + \frac{\varepsilon}{2} < a + \varepsilon = a_+$$

så a_+ är inte en undre begränsning av I.

Om $\varepsilon \geq 2(b-a)$ är

$$a_+ = a + \varepsilon \geq a + 2(b-a) = b + (b-a) > b + 0 = b$$

och därför ligger a_+ till höger av intervallet. Därför är det lätt att visa a_+ inte är en undre begränsning av I: Till exempel ser vi att $x := (a+b)/2$ är så att $x \in I$ och $b_- < x$.

Sammanfattningsvis har vi visat att a är en undre begrånsning och varje tal större än a inte är en undre begränsning. Därför är a den största undre begränsningen av I. □

2.3.2 Ytterligare egenskaper hos reella tal

Enligt sats 2.19 vet vi att en mängd inte kan ha flera än ett supremum, men vilka mängder har ett supremum? För att hjälpa oss svara på den frågan ska vi lägga till ett par axiom.

III **Ytterligare axiomen för reella tal:**

(a) Om en icketom mängd $A \subseteq \mathbf{R}$ är begränsad uppåt så finns det en minsta övre begränsning u för A. Man skriver

$$u = \sup_{a \in A} a \quad \text{eller} \quad u = \sup A.$$

(fullständighetsaxiomet)

(b) $0 \leq a < 1/n$ för alla $n \in \mathbf{Z}_+$ medför det att $a = 0$ (den arkimediska egenskapen).

Det är klart att det första nya axiomet har någonting att göra med frågan. Det ger ett tillräckligt villkor för att veta när en mängd har ett supremum, men det kan också tolkas som påståendet att reella tal inte har något slags 'hål' — det undersöker vi vidare i avsnitt 3.2.3. Det andra axiomet säger att det inte finns något 'oändligt litet tal', men vi kan också använda det för att visa att vissa mängder saknar supremum.

Exempel 2.20. *Mängden $\mathbf{Z}_+$ av positiva heltal har inget supremum.*

Anta att det finns ett supremum u för $\mathbf{Z}_+$. Vi vet att $n \leq u$ för alla $n \in \mathbf{Z}_+$ enligt (2.15)*. I synnerhet $0 < 1 \leq u$ så $0 < 1/u$. Nu räcker det att skriva kontrapositionen av III(b): $a \neq 0$ medför att*

$$\textit{antingen } a < 0 \textit{ eller det finns ett } n \in \mathbf{Z}_+ \textit{ så att } 1/n < a.$$

Om vi betraktar kontrapositionen av III(b) med $a = 1/u$. Eftersom $1/u \neq 0$ och $1/u < 0$ inte stämmer vet vi att det finns ett $n \in \mathbf{Z}_+$ så att $1/n < 1/u$. Därför är $u < n$ för samma n som är en motsägelse till (2.15)*.*

Vid första ögonkastet ser den arkimediska egenskapen lite konstig ut. Men man förstår lite bättre varför den behövs genom att tänka på ett exempel när det inte gäller.

Exempel 2.21. *Betrakta mängden*

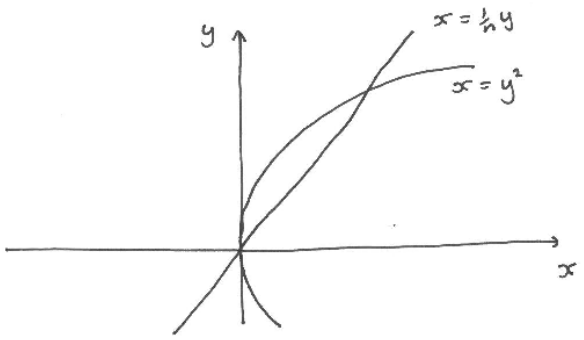
$$P = \{p(y) \,|\, p(y) = ay + by^2 \textit{ för } a, b \in \mathbf{R}\}$$

av alla linjära och andragradspolynom (se avsnitt 2.5.2) i variabeln y som går genom origo tillsammans med addition och subtraktion definierad som i (2.29)*. (Vi bryr oss inte om multiplikation och division här.) Vi definierar en ordning $\lesssim$ genom följande regel. För $p_1(y) = a_1y+b_1y^2$ och $p_2(y) = a_2y + b_2y^2$ säger vi att*

$$p_1(y) \lesssim p_2(y) \text{ om } \begin{cases} \text{antingen } a_1 < a_2, \\ \text{eller } a_1 = a_2 \text{ och } b_1 < b_2. \end{cases}$$

Observera att enligt avsnitt 2.2.5 kan vi representera ett reellt tal α som förstagradspolynomet (det vill säga linjen) $q_\alpha(y) = \alpha y$ och $\alpha_1 < \alpha_2$ om och endast om $q_{\alpha_1}(y) \lesssim q_{\alpha_2}(y)$. Dessutom uppfyller mängden P axiomen I och II i avsnitt 2.2.1 som gäller addition och subtraktion med $\lesssim$ istället för $<$.

Däremot uppfyller P inte III(b) eftersom till exempel polynomet $p(y) = y^2$ är så att $0 \lesssim p(y) \lesssim q_{1/n}(y)$ för alla $n \in \mathbf{Z}_+$ fast $p(y)$ inte är nollfunktionen. Se figur 7.



Figur 7: Reella tal tolkas här som lutningar till de räta linjer som går genom origo. Om man lägger till grafer av kvadratiska polynom där somliga har en lutning i origo som är mindre (sedd från sidan) än alla räta linjer med positiva lutningar.

2.3.3 Följder och induktion

En *följd* är en mängd av tal indexerad med positiva heltal. Det vill säga att det är en vanlig mängd som har också en ordning på elementen i mängden precis som positiva heltal har en ordning. En följd tecknas med parenteserna '(' och ')' istället för '{' och '}' för att skilja den från en vanlig mängd som inte har någon ordning.

Exempel 2.22. *Några exempel:*

1. *Positiva heltal* $(1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, \dots)$ *är sig i själv en följd med vanlig numerisk ordning;*

2. $(2, 4, 6, 8, 10, 12, 14, 16, \dots)$ *är följden av alla positiva jämna tal i numerisk ordning;*

3. $(1, 4, 9, 25, 36, 49, 64, \dots)$ *är följden av alla positiva heltal i kvadrat.*

Det kan vara smidigt att skriva en formel a_n för den n:e termen i en följd. Man skriver följden $(a_n)_{n\in\mathbf{Z}_+}$ eller $(a_n)_n$. I exempel 2.22 har vi $a_n = n$, $a_n = 2n+2$ respektive $a_n = n^2$ (som också kan skrivas direkt som $(n)_n$, $(n + 2)_n$ respektive $(n^2)_n$). Kom ihåg att i avsnitt 2.2.4 pratade vi om decimalutvecklingar $0.a_1a_2a_3\dots$ av tal mellan 0 och 1. Vi kan tolka sådana decimalutvecklingar $0.a_1a_2a_3\dots$ som följden $(a_n)_n$ — en följd där varje term $a_n \in \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$.

Man kan också definiera en följd rekursivt. Det vill säga att man definierar a_n med hjälp av föregående $a_0, a_1, \ldots, a_{n-1}$. Följande exempel är en följd som definieras rekursivt.

Exempel 2.23. *Definiera en följd vars n:e term tecknas* $n!$ *och uttalas 'n-fakultet' genom att sätta* $0! = 1$ *och sen*

$$n! = n(n-1)!$$

för varje $n \in \mathbf{Z}_+$. *De första sex termerna är*

$$0! = 1, \quad 1! = 1, \quad 2! = 2, \quad 3! = 6, \quad 4! = 24, \quad 5! = 120, \quad \textit{och} \quad 6! = 720.$$

Nu inför vi lite notation för en produkt av flera tal. Produkten av tal $a_n, a_{n+1}, \ldots, a_m$ tecknas

$$\prod_{j=n}^{m} a_j := a_n a_{n+1} \ldots a_m.$$

Med denna notation kan vi bevisa att följden från exempel 2.23 har formeln

$$n! = \begin{cases} 1 & \text{om } n = 0, \\ \prod_{j=1}^{n} j & \text{om } n \in \mathbf{Z}_+. \end{cases} \tag{2.19}$$

För att bevisa (2.19) ska vi använda den metod som kallas *induktion*. Det är en metod man kan använda sig av när man har ett påstående $P(n)$ som beror på positiva heltal n. I vårt fall är påståendet att (2.19) stämmer för det positiva heltalet n. Metoden har två steg:

1. Först kontrollerar man att påståendet är sant för $n = 0$;
2. Sen bevisar man för varje $m \in \mathbf{Z}_+$ att om påståendet är sant för $n = m$ då är det också sant när $n = m + 1$.

Då kan man dra slutsatsen att påståendet $P(n)$ är sant för alla $n \in \mathbf{Z}_+$.

För att bevisa (2.19) kollar vi först att (2.19) är sant om $n = 0$. Formeln säger att $0! = 1$ och det stämmer med definitionen i exempel 2.23, så första steget är bevisat. Sen betraktar vi $m \in \mathbf{Z}_+$ och antar att

$$m! = \begin{cases} 1 & \text{om } m = 0, \\ \prod_{j=1}^{m} j & \text{om } m \in \mathbf{Z}_+. \end{cases} \tag{2.20}$$

Vi vill bevisa att

$$(m+1)! = \prod_{j=1}^{m+1} j. \tag{2.21}$$

(Observera att $m + 1 \in \mathbf{Z}_+$ om $m \in \mathbf{N}$.) Vi vet från definitionen i exempel 2.23 att

$$(m+1)! = (m+1)m!$$

och från (2.20) får vi att

$$(m+1)m! = (0+1)1 = 1 = \prod_{j=1}^{1} j$$

om $m = 0$ och

$$(m+1)m! = (m+1)\prod_{j=1}^{m} j = 1 = \prod_{j=1}^{m+1} j$$

om $m \in \mathbf{Z}_+$. Tillsammans bevisar dessa likheter (2.21), och därmed är, enligt induktion, (2.19) sant för alla $n \in \mathbf{Z}_+$.

Det finns ett nyttigt kombinotoriskt sätt att tolka talet $n!$. Om en mängd innehåller n element, på hur många sätt kan man då ordna elementen i mängden? Man kan ordna elementen genom att lägga ut dem i en linje på ett bord. Om man gör det så väljer man först ett av de n elementen, så finns det n möjliga val för det första. Sen väljer man ett av de $(n-1)$ elementen som är kvar, så finns det $(n-1)$ möjliga val för det andra element, och så vidare. Så kan man ordna mängden på

$$n(n-1)(n-2)\dots 1 = n!$$

sätt.

2.3.4 Aritmetiska och geometriska följder

Två typer av följder är viktiga i den här kursen och de är även viktiga i matematik i allmänhet.

Den första är en *aritmetisk följd*. Det är en följd $(a_n)_n$ så att

$$a_{n+1} - a_n =: d \tag{2.22}$$

är oberoende av n. Det betyder att differensen mellan två angränsande element alltid är densamma.

Exempel 2.24. *Följande följder är aritmetiska:*

1. $(1, 2, 3, 4, 5, 6, 7, \dots)$
2. $(2, 4, 6, 8, 10, 12, 14, \dots)$
3. $(3, 10, 17, 24, 31, 38, 45, \dots)$
4. $(60, 66, 72, 78, 84, 90, 96, \dots)$

Vad är d i de ovanstående exemplen?

Vi kan bevisa att en aritmetisk följd $(a_n)_n$ uppfyller formeln

$$a_n = a_1 + d(n-1) \tag{2.23}$$

för alla $n \in \mathbf{Z}_+$. För att bevisa (2.23) kan man använda induktion. Vi kan snabbt kontrollera att

$$a_1 = a_1 + d(1-1)$$

så (2.23) stämmer om $n = 1$. Sen antar vi att (2.23) gäller för $n = 1, 2, \dots, k$ och betraktar a_{k+1}. Från (2.22) med $n = k$ vet vi att $a_{k+1} = a_k + d$ och sen kan vi använda oss av (2.23) med $n = k$ och säga

$$a_{k+1} = a_k + d = a_1 + d(k-1) + d = a_1 + d((k+1)-1).$$

Nu har vi bevisat att (2.23) stämmer för $n = k+1$. Enligt induktion drar vi slutsatsen att (2.23) stämmer för alla $n \in \mathbf{Z}_+$.

Den andra följdtypen vi vill betrakta är en *geometrisk följd*. Den är en följd som uppfyller

$$a_{n+1} = a_n r$$

för alla $n \in \mathbf{Z}_+$ och något $r \in \mathbf{R}$.

Exempel 2.25. *Följande följder är geometriska följder:*

1. $(1, 1/2, 1/4, 1/8, 1/16, 1/32, 1/64, \dots)$

2. $(1, 2, 4, 8, 16, 32, 64, \dots)$

3. $(5, 25, 125, 625, 3125, 15625, 78125, \dots)$

4. $(6, 2, 2/3, 2/9, 2/18, 2/54, 2/162, \dots)$

Vad är r i de ovanstående exemplen?

På liknande sätt som beviset av (2.23) kan man bevisa att

$$a_n = a_1 r^{n-1}$$

för alla $n \in \mathbf{Z}_+$.

2.4 Summor och kombinatorik

2.4.1 Summor

Axiomen I(a) och I(b) i avsnitt 2.2.1 säger att det inte spelar någon roll i vilken ordning man summerar tre eller flera tal. Det är användbart att ha en beteckning för summan av många termer:

$$\sum_{i=n}^{m} a_i := a_n + a_{n+1} + \cdots + a_m$$

för givna $n, m \in N$ och $a_i \in \mathbf{R}$ för $i = n, n+1, \dots, m$. Notationen i vänsterledet betyder att alla a_i summeras med $i \in \mathbf{Z}_+$ från och med n till och med m.

Sats 2.26. *För $n \leq m$, reella tal $a_i, b_i \in \mathbf{R}$ för $i = n, n+1, \dots, m$, och $c \in \mathbf{R}$ har vi att*

$$\sum_{i=n}^{m}(a_i + b_i) = \sum_{i=n}^{m} a_i + \sum_{i=n}^{m} b_i$$

och

$$\sum_{i=n}^{m}(ca_i) = c\sum_{i=n}^{m} a_i.$$

Exempel 2.27. *Carl Friedrich Gauss är en berömd matematiker som bidragit till flera ämnen inom matematik och fysik, till exempel talteori, analys och geometri. Enligt traditionen fick den unge Gauss följande problem som straff en gång när han varit stygg i skolan: Addera alla heltal mellan ett och hundra. Det vill säga, räkna*

$$1 + 2 + \cdots + 100 = \sum_{i=1}^{100} i.$$

Läraren blev förvånad när Gauss löste problemet på några minuter. Gauss löste problemet på följande sätt: Han skrev summan två gånger, första gången som vanligt och sen andra gången nedanför den första fast baklänges

$$\begin{array}{rcccccccccccc} 1 & + & 2 & + & 3 & + & \dots & + & 98 & + & 99 & + & 100 \\ + \ 100 & + & 99 & + & 98 & + & \dots & + & 3 & + & 2 & + & 1. \end{array}$$

Han observerade att om man adderar ett par tal där ett tal står direkt ovanför det andra får man alltid svaret 101*. Eftersom det finns* 100 *sådana par, är hela summan* 101 *gånger* 100*, och lika med* $2\sum_{i=1}^{100} i$*. Alltså*

$$\sum_{i=1}^{100} i = \frac{101 \times 100}{2} = 5050.$$

Vi kan upprepa metoden för summan $\sum_{i=1}^{n} i$ *och får*

$$\sum_{i=1}^{n} i = \frac{n(n+1)}{2}. \tag{2.24}$$

Vi kan använda oss av (2.24) för att räkna ut summan av några termer i en aritmetisk följd. Observera först att

$$\sum_{i=1}^{n}(i-1) = \sum_{j=0}^{n-1} j = \sum_{j=1}^{n-1} j = \frac{n(n-1)}{2} \tag{2.25}$$

enligt (2.24). Betrakta sedan en godtycklig aritmetisk följd $(a_i)_i$. Den uppfyller (2.23) så

$$\sum_{i=1}^{n} a_i = \sum_{i=1}^{n}(a_1 + d(i-1)) = \sum_{i=1}^{n} a_1 + d\sum_{i=1}^{n}(i-1) = na_1 + d\frac{n(n-1)}{2}$$

när vi använder oss av (2.25) i sista steget.

Exempel 2.28. *Nu bevisar vi att*

$$\sum_{i=1}^{n} i^2 = \frac{n(n+1)(2n+1)}{6}. \tag{2.26}$$

Vi använder oss av induktion. Först kan vi kontrollera att (2.26) *stämmer när* $n = 1$*. Då har vi att*

$$\sum_{i=1}^{1} i^2 = 1 \quad \textit{och} \quad \frac{n(n+1)(2n+1)}{6} = \frac{1(1+1)(2+1)}{6} = 1,$$

så (2.26) *stämmer för* $n = 1$*. Nu antar vi att det finns ett tal* $m \in \mathbf{Z}_+$ *så att* (2.26) *gäller för* $n = m$ *(vi vet att det finns minst ett sådant* m *eftersom vi precis har bevisat likheten* (2.26) *då* $n = 1$*) och försöker bevisa* (2.26) *i fallet* $n = m + 1$*: Vi räknar*

$$\sum_{i=1}^{m+1} i^2 = \sum_{i=1}^{m} i^2 + (m+1)^2 \underset{\substack{\uparrow \\ \textit{enligt antagandet ovan}}}{=} \frac{m(m+1)(2m+1)}{6} + (m+1)^2.$$

Men

$$\begin{aligned}\frac{m(m+1)(2m+1)}{6} + (m+1)^2 &= \frac{m(m+1)(2m+1) + 6(m+1)^2}{6} = \frac{(m+1)\left(m(2m+1) + 6(m+1)\right)}{6} \\ &= \frac{(m+1)\left(2m^2 + 7m + 6\right)}{6} = \frac{(m+1)(m+2)(2m+3)}{6} \\ &= \frac{(m+1)((m+1)+1)(2(m+1)+1)}{6}.\end{aligned}$$

Därför

$$\sum_{i=1}^{m+1} i^2 = \frac{(m+1)((m+1)+1)(2(m+1)+1)}{6},$$

som säger (2.26) *gäller med* $n = m + 1$.

Så vi har bevisat att

$$\textit{om } (2.26) \textit{ gäller med } n = m \textit{ för något } m \in \mathbf{Z}^+ \textit{ så gäller det med } n = m + 1. \tag{2.27}$$

Vi har också kontrollerat att (2.26) *gäller för* $n = 1$. *Så vi drar slutsatsen att* (2.26) *gäller för alla* $n \in \mathbf{Z}_+$.

Betrakta en geometrisk följd $(ar^{i-1})_i$ med kvoten r. Vi kan använda induktion för att bevisa

$$\sum_{i=1}^{n} ar^{i-1} = a\frac{1-r^n}{1-r}$$

2.4.2 Binomialsatsen

I exempel 2.23 definierade vi fakultet. Nu definierar vi *binomialkoefficienten* som

$$\binom{n}{k} := \frac{n!}{k!(n-k)!}$$

för $n, k \in \mathbf{N}$ men $k \leq n$. Den kanske verkar som en konstig definition men precis som för fakultet finns det ett kombinatoriskt sätt att tolka den. Binomialkoefficienten kan tolkas som antalet sätt att välja ut k element från en mängd av n element om man inte tar hänsyn till i vilken ordning man väljer dem. Varför det?

Först frågar vi: hur många sätt finns det att välja k element från en mängd av n element om man tar hänsyn till i vilken ordning man väljer dem? Precis som vi argumenterade för fakultet är det $n(n-1)(n+2)\dots(n-(k-1))$. Varje val är en ordnad mängd av k element. Men om vi inte vill ta hänsyn till ordningen, har vi räknat varje möjlighet $k!$ gånger. Därför är antalet sätt att välja ut k element från en mängd av n element

$$\frac{n(n-1)\dots(n-(k-1))}{k!} = \frac{n!}{k!(n-k)!} = \binom{n}{k}$$

när man inte tar hänsyn till i vilken ordning man väljer dem.

Sats 2.29 (Pascals identitet)**.** *För positiva hela tal* n *och* k *så att* $k \leq n$ *har man att*

$$\binom{n}{k} + \binom{n}{k-1} = \binom{n+1}{k}$$

Sats 2.30 (Binomialsatsen)**.** *För* $x, y \in \mathbf{R}$ *och* $n \in \mathbf{Z}_+$ *har man att*

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^{n-k}y^k. \tag{2.28}$$

Bevis. Vi använder induktion. Om $n = 1$ är högerledet av (2.28)

$$\sum_{k=0}^{1} \binom{1}{k} x^{1-k}y^k = x + y$$

som stämmer med vänsterledet. Då antar vi att (2.28) stämmer för ett givet $n \in \mathbf{Z}_+$ och sen räknar vi ut att

$$\begin{aligned}
(x+y)^{n+1} &= (x+y)(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^{n+1-k}y^k + \sum_{k=0}^{n} \binom{n}{k} x^{n-k}y^{k+1} \\
&= x^{n+1} + \sum_{k=1}^{n} \binom{n}{k} x^{n+1-k}y^k + y^{n+1} + \sum_{k=0}^{n-1} \binom{n}{k} x^{n-k}y^{k+1} \\
&= x^{n+1} + y^{n+1} + \sum_{k=1}^{n} \binom{n}{k} x^{n+1-k}y^k + \sum_{k=1}^{n} \binom{n}{k-1} x^{n+1-k}y^k \\
&= x^{n+1} + y^{n+1} + \sum_{k=1}^{n} \left(\binom{n}{k} + \binom{n}{k-1} \right) x^{n+1-k}y^k \\
&\underset{\text{enligt lemma 2.29}}{\overset{\uparrow}{=}} x^{n+1} + y^{n+1} + \sum_{k=1}^{n} \binom{n+1}{k} x^{n+1-k}y^k \\
&= \sum_{k=0}^{n+1} \binom{n+1}{k} x^{n+1-k}y^k.
\end{aligned}$$

Enligt induktion är satsen bevisad. □

2.5 Funktioner, koordinatsystem och grafer

2.5.1 Funktioner, defintionsmängder och värdemängder

En *funktion* f är en regel som till varje element x i en mängd A ger ett element $f(x)$ i en annan mängd B. Till exempel, ta A lika med mängden av alla bilar i Sverige och B likamed alfabetet, sen kan vi definiera en funktion som för en bil anger första siffran i bilens registreringsskylt. Vi skriver $f\colon A \to B$ för att notera att funktionen f skickar element i A till element i B, och kallar A funktionens *definitionsmängd* och B funktionens *målmängd*. Observera att funktionen kallas för f medan elementen i B som f skickar $x \in A$ till, noteras $f(x)$.

Exempel 2.31. *1. Ta $A = B = \mathbf{R}$ sen kan man definiera en funktion $f\colon \mathbf{R} \to \mathbf{R}$ enligt formeln $f(x) = x^2 + 2x + 1$ för alla $x \in \mathbf{R}$.*

2. Ta $A = [0,1]$ och $B = \mathbf{R}$. Definiera sen en funktion $g\colon [0,1] \to \mathbf{R}$ enligt formeln $g(x) = 6x + 9$ för alla $x \in [0,1]$.

3. Ta $A = \mathbf{R}$ och $B = \mathbf{Z}$. Sen kan man definiera en funktion $t\colon \mathbf{R} \to \mathbf{Z}_+$ enligt regeln

$$t(x) \text{ är det minsta heltal } n \text{ så att } x \leq n.$$

Ibland skriver man $x \mapsto f(x)$ för att slippa använda ett namn för funktionen. Så de första två exemplen ovan kunde skrivas $x \mapsto x^2 + 2x + 1$ för $x \in \mathbf{R}$ och $x \mapsto 6x + 9$ för $x \in [0,1]$.

Två funktioner $f\colon A \to B$ och $g\colon C \to D$ sägs vara *lika* om $A = C$ och $f(x) = g(x)$ för alla $x \in A$.

En funktion som är ganska användbar kallas för *absolutbeloppet* funktionen $x \mapsto |x|$. Den har definitionsmängden $\mathbf{R}$ och definieras enligt formeln

$$|x| = \begin{cases} x & \text{om } x \geq 0; \\ -x & \text{om } x < 0. \end{cases}$$

(Observera att $-x$ är positiv om x är negativ!)

Sats 2.32 (Triangelolikheten)**.** *Olikheten*

$$|x+y| \leq |x| + |y|$$

gäller för alla $x, y \in \mathbf{R}$.

Bevis. Om både x och y är ickenegativa har vi att

$$|x+y| = x + y = |x| + |y|.$$

Om både x och y är ickepositiva har vi att

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

Om x och y har olika tecken får vi anta att $|x| > |y|$ eftersom om det inte stämmer får vi byta roller på x och y. Observera att $x \leq |x|$ för vilket x som helst. Om x är positivt och y är negativt medför det att

$$|x+y| = x + y \leq |x| + |y|.$$

Om y är positivt och x är negativt har vi att

$$|x+y| = -x - y \leq |x| + |y|.$$

□

För en funktion $f\colon A \to B$ kallas mängden

$$\{f(x) \mid x \in A\}$$

för funktionens *värdemängd*. Observera att värdemängden är en delmängd av målmängden B.

Ibland är det användbart att summera två funktioner, samt att multiplicera dem, ta differensen eller dela den ena med den andra. För två funktioner $f\colon A \to \mathbf{R}$ och $g\colon A \to \mathbf{R}$ som har samma definitionsmängder skriver man $f+g$ för summan, $f-g$ för differensen, fg för produkten och f/g för kvoten. Uttrycken definieras punktvis som

$$\begin{aligned} (f+g)(x) &= f(x) + g(x), \\ (f-g)(x) &= f(x) - g(x) \quad \text{och} \\ (fg)(x) &= f(x)g(x) \end{aligned} \tag{2.29}$$

för $x \in A$. Endast om $g(x) \neq 0$ för alla $x \in A$ definieras f/g som

$$(f/g)(x) = f(x)/g(x)$$

för $x \in A$.

Låt nu $f\colon A \to B$ och $g\colon B \to C$ vara två funktioner. Observera att definitionsmängden av g är lika med värdemängden av f så man kan bilda en funktion $g \circ f\colon A \to C$ som kallas *sammansättningen* av f och g genom

$$(g \circ f)(x) = g(f(x))$$

för alla $x \in A$.

Sats 2.33. *Låt* $f\colon A \to B$, $g\colon B \to C$ *och* $h\colon C \to D$. *Då är* $h \circ (g \circ f) = (h \circ g) \circ f$.

Bevis. Båda funktionerna har definitionsmängden A. Vi har också att

$$h \circ (g \circ f) = h((g \circ f)(x)) = h(g(f(x)))$$

och

$$(h \circ g) \circ f = (h \circ g)(f(x)) = h(g(f(x)))$$

för alla $x \in A$. □

2.5.2 Polynom

Funktioner $p\colon \mathbf{R} \to \mathbf{R}$ som har formen

$$p(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n = \sum_{k=0}^{n} a_k x^k$$

för något $n \in \mathbf{N}$ och $a_k \in \mathbf{R}$ $(k = 0, 1, \ldots, n)$ kallas för *polynom*.[9] Om $a_n \neq 0$ kallas n polynomets *grad* vilket skrivs som $n = \deg(p)$. Speciellt om $n = 2$ kallas polynomet ett andragradspolynom. Graden av det *triviala polynomet*, det vill säga polynomet som är identiskt noll ($p(x) = 0$ för alla $x \in \mathbf{R}$), är odefinierad. Ibland preciserar man att de är polynom med reella koefficienter för att alla $a_k \in \mathbf{R}$.[10]

Sats 2.34. *Om p och q är polynom så är $p + q$, pq och $p \circ q$ polynom. Dessutom*

1. $\deg(p + q) \leq \max\{\deg(p), \deg(q)\}$,[11]
2. $\deg(pq) = \deg(p) + \deg(q)$ *och*
3. $\deg(p \circ q) = \deg(p)\deg(q)$

för alla icke-triviala p och q.

Sats 2.35. *Till varje par av polynom p och q med $q(x) \neq 0$ för något $x \in \mathbf{R}$ finns det polynom k och r så att*

$$\begin{aligned} &p(x) = q(x)k(x) + r(x) \text{ för alla } x \in \mathbf{R}, \text{ och} \\ &r \text{ är antingen trivialt eller uppfyller } \deg(r) < \deg(q). \end{aligned} \tag{2.30}$$

Dessutom bestäms k och r entydigt av p och q. Ibland kallas polynomet k för kvoten och polynomet r för resten.

Observera att (2.30) har samma form som ekvation (2.6) för heltal. Sats 2.35 säger att samma division är möjlig för polynom på samma sätt som den vi är vana vid för heltal.

Innan vi bevisar satsen är det nyttigt att gå genom ett exempel. Jonas Månsson vid Lunds universitet har spelat in en bra film om polynomdivision. För att titta på den scanna QR-koden i figur 8 eller skriv adressen direkt i en webbläsare.

Figur 8: Polynomdivision (`https://youtu.be/DnyiLJkQq8o`)

Bevis av sats 2.35. Om $\deg(p) < \deg(q)$ kan vi ta $k(x) = 0$ och $r(x) = p(x)$ for alla $x \in \mathbf{R}$.

Om $\deg(p) \geq \deg(q)$ skriv

$$p(x) = \sum_{k=0}^{\deg(p)} a_k x^k \quad \text{and} \quad q(x) = \sum_{k=0}^{\deg(q)} b_k x^k$$

där $a_{\deg(p)} \neq 0$ och $b_{\deg(q)} \neq 0$. Därför om $k_1(x) := (a_{\deg(p)}/b_{\deg(p)})x^{\deg(p)-\deg(q)}$ är

$$p_1(x) := p(x) - q(x)k_1(x)$$

[9] Vi är lite slarviga när vi skriver x^0 i första termen i summan. Egentligen är x^0 odefinerad om $x = 0$, men här (bara) tolkar vi det lika med 1 för alla x.

[10] I avsnitt 4.3.3 betraktar vi polynom med koefficienter som tillhör till en annan mängd än $\mathbf{R}$.

[11] Notationen $\max\{a, b\}$ betyder den största talet av a och b.

ett polynom av grad strängt mindre än $\deg(p)$. Om $\deg(p_1) < \deg(q)$ har vi bevisat satsen med $k = k_1$ och $r = p_1$.

Om $\deg(p_1) \geq \deg(q)$ kan upprepa argumentet ovan med p_1 istället för p och får ett polynom k_2 så att

$$p_2(x) := p_1(x) - q(x)k_2(x)$$

ett polynom av grad strängt mindre än $\deg(p_1)$. Om $\deg(p_2) < \deg(q)$ har vi att

$$p_2(x) = p(x) - q(x)k_1(x) - q(x)k_2(x) = p(x) - q(x)(k_1(x) + k_2(x))$$

och därför bevisat satsen med $k = k_1 + k_2$ och $r = p_2$.

Om $\deg(p_2) \geq \deg(q)$ kan vi upprepa argumentet igen och får k_3 och p_3. Vi fortsätter processen tills vi hittar något m så att antingen $\deg(p_m) < \deg(q)$ eller $p_m = 0$ och då har vi bevisat satsen med $k = k_1 + k_2 + \cdots + k_m$ och $r = p_m$. □

Vi säger att ett polynom p är *delbart* med ett annat polynom q om (2.30) gäller med $r(x) = 0$ för alla $x \in \mathbf{R}$. Ett tal $c \in \mathbf{R}$ kallas för en *rot* till polynomet p om det löser ekvationen $p(c) = 0$.

Sats 2.36. *Talet $c \in \mathbf{R}$ är en rot till polynomet p om och endast om p är delbart med $x \mapsto x - c$.*

Bevis. Om c är en rot till p så är $p(c) = 0$. Om vi tillämpar sats 2.35 med $q(x) = x - c$ får vi att

$$p(x) = k(x)(x - c) + r(x)$$

för några polynom k och r och alla $x \in \mathbf{R}$. Dessutom är r antingen trivialt eller uppfyller $\deg(r) < 1$, så $r(x) = a$ för någon konstant $a \in \mathbf{R}$. Om vi sätter $x = c$ får vi

$$0 = p(c) = k(c)(c - c) + a = a$$

så $r(x) = a = 0$ och p är därför delbart med $x \mapsto x - c$.

Om p är delbart med $x \mapsto x - c$ kan vi skriva

$$p(x) = k(x)(x - c)$$

för alla $x \in \mathbf{R}$ och därför är $p(c) = k(c)(c - c) = 0$ □

Sats 2.37. *Om $p(x) = \sum_{k=0}^{n} a_k x^k$ är ett polynom av grad högst n och det finns tal $x_0 < x_1 < \cdots < x_n$ så att $p(x_j) = 0$ för alla $j = 0, 1, \ldots, n$ så är $p(x) = 0$ för alla $x \in \mathbf{R}$.*

Bevis. Vi bevisar satsen med hjälp av induktion över polynomets grad.

Först betraktar vi $n = 0$. Då är

$$p(x) = a_0 \quad \text{för alla } x \in \mathbf{R} \tag{2.31}$$

och

$$p(x_0) = 0 \quad \text{för ett givet } x_0 \in \mathbf{R}. \tag{2.32}$$

Om vi tar $x = x_0$ i (2.31) medför det tillsammans med (2.32) att $a_0 = p(x_0) = 0$. Genom att sätta $a_0 = 0$ in (2.31) får vi att $p(x) = a_0 = 0$ för alla $x \in \mathbf{R}$ och satsen är bevisad i fallet $n = 0$.

Nu antar vi att satsen gäller om $n = m$ för ett visst $m \in \mathbf{N}$ och betraktar ett polynom p av grad $m + 1$ tillsammans med tal $x_0 < x_1 < \cdots < x_m < x_{m+1}$ så att $p(x_j) = 0$ för alla $j = 0, 1, \ldots, m, m + 1$. Eftersom $p(x_{m+1}) = 0$ medför sats 2.36 att

$$p(x) = k(x)(x - x_{m+1}) \tag{2.33}$$

där k är ett polynom av grad högst m. Men eftersom $x_j \neq x_{m+1}$ för alla $j = 0, 1, \ldots, m$ vet vi att $k(x_j) = 0$ för alla $j = 0, 1, \ldots, m$. Eftersom vi antog att satsen gäller om $n = m$ kan vi tillämpa den till k och dra slutsatsen att $k(x) = 0$ för alla $x \in \mathbf{R}$. Enligt (2.33) är då $p(x) = 0$ för alla $x \in \mathbf{R}$ som är satsens slutsats där $n = m + 1$.

Enligt induktion har vi bevisat satsen för alla $n \in \mathbf{N}$. □

Sats 2.38. *Om* $p(x) = \sum_{k=0}^{n} a_k x^k$ *är ett polynom och* $p(x) = 0$ *för alla* $x \in \mathbf{R}$ *så är* $a_k = 0$ *för alla* $k = 0, 1, \ldots, n$.

Bevis. En gång till bevisar vi satsen med hjälp av induktion över n.

Först betraktar vi $n = 0$. Då är

$$p(x) = a_0 \quad \text{för alla } x \in \mathbf{R}.$$

Eftersom vi också vet att $p(x) = 0$ för alla $x \in \mathbf{R}$ så är $a_0 = 0$ satsen är bevisat i fallet $n = 0$.

Nu antar vi att satsen gäller om $n = m$ för ett visst $m \in \mathbf{N}$ och betraktar ett polynom $p(x) = \sum_{k=0}^{m+1} a_k x^k$ av grad $m + 1$ så att $p(x) = 0$ för alla $x \in \mathbf{R}$. I synnerhet är $p(0) = 0$ så $0 = p(0) = a_0 + \sum_{k=1}^{m+1} a_k 0^k = a_0$. Därför är

$$p(x) = \sum_{k=1}^{m+1} a_k x^k = \left(\sum_{k=1}^{m+1} a_k x^{k-1} \right) x. \tag{2.34}$$

Polynomet $q(x) := \sum_{k=1}^{m+1} a_k x^{k-1}$ har grad m och eftersom $p(x) = 0$ är $q(x) = 0$ för alla $x \in \mathbf{R} \setminus \{0\}$ enligt (2.34). Men då medför sats 2.37 att $q(x) = 0$ för alla $x \in \mathbf{R}$. Enligt induktionsantagande är alla q:s koefficienter lika med noll, det vill säga $a_k = 0$ för alla $k = 1, 2, \ldots, m + 1$.

Enligt induktion har vi bevisat satsen för alla $n \in N_0$. □

2.5.3 Koordinatsystem och grafer

I avsnitt 2.2.5 byggde vi en bild av alla rationella tal. Vi ritade punkter i ett plan med hjälp av axlar och koordinater. Planet kallas för *koordinatplanet* och axlarna *koordinataxlarna*. Alla punkter (x, y) i $\mathbf{R}^2$ representeras av en punkt i koordinatplanet precis som $x \in \mathbf{R}$ representeras av en punkt på en linje. Koordinatplanet är lika användbart för att visualisera funktioner. Vi gör det genom att rita en funktions graf. Om $D \subseteq \mathbf{R}$ definierar vi *grafen* av en funktion $f\colon D \to \mathbf{R}$ att vara mängden

$$\operatorname{graf}(f) := \{(x, y) \mid x \in D \text{ och } y = f(x)\}.$$

Sen kan vi rita mängden $\operatorname{graf}(f)$ genom att representera ett element $(x, y) \in \operatorname{graf}(f)$ som punkten i planet som ligger på ett avstånd (med tecken) x till vänster om den lodräta axeln och ett avstånd (med tecken) y ovanför den vågräta axeln.[12] Se figur 9.

Grafer är förstås inte de enda mängder vi kan representera på koordinatplanet. I princip kan alla delmängder av $\mathbf{R}^2$ representeras som en bild i koordinatplanet.

2.5.4 Monotonicitet

För att förstå hur en funktion beter sig är det användbart att definiera monotonicitet. En funktion $f\colon D \to \mathbf{R}$ som har definitionsmängden $D \subseteq \mathbf{R}$ kallas för *växande* på en mängd $E \subseteq D$ om

$$x, y \in E \text{ och } x < y \text{ medför det att } f(x) \leq f(y)$$

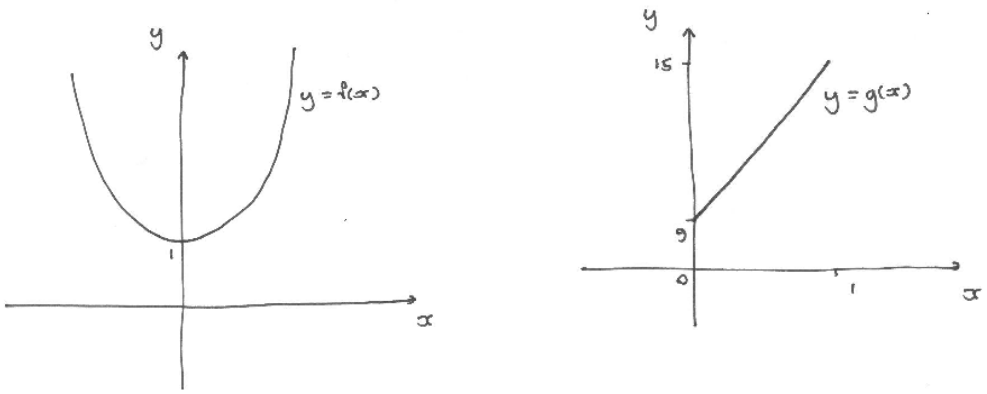
och *avtagande* på E om

$$x, y \in E \text{ och } x < y \text{ medför det att } f(x) \geq f(y).$$

En funktion $f\colon D \to \mathbf{R}$ som har definitionsmängden $D \subseteq \mathbf{R}$ kallas för *strängt växande* på E om

$$x, y \in E \text{ och } x < y \text{ medför det att } f(x) < f(y)$$

[12] Kom ihåg att ett negativt avstånd till vänster (respektive ovanför) betyder till höger (under).



(a) Grafen av $f\colon \mathbf{R} \to \mathbf{R}$ definierad enligt formeln $f(x) = x^2+2x+1$ för alla $x \in \mathbf{R}$.

(b) Grafen av $g\colon [0,1] \to \mathbf{R}$ definierad enligt formeln $g(x) = 6x + 9$ för alla $x \in [0,1]$.

Figur 9: Här har vi ritat graferna av funktionerna från exempel 2.31.

och *strängt avtagande* på E om

$$x, y \in E \text{ och } x < y \text{ medför det att } f(x) > f(y).$$

Om en funktion är (strängt) växande, respektive (strängt) avtagande på hela definitionsmängden D säger man bara att funktionen är (strängt) växande, respektive (strängt) avtagande.

Exempel 2.39. *1. Funktionen $f\colon \mathbf{R} \to \mathbf{R}$ definierad som $f(x) = x^2$ för alla $x \in \mathbf{R}$ är strängt växande på $[0, \infty)$ eftersom $0 \leq x < y$ medför att $f(x) = x^2 < y^2 = f(y)$.*

2. På $(-\infty, 0]$ är funktionen f strängt avtagande eftersom $x < y \leq 0$ medför att $f(x) = x^2 > y^2 = f(y)$.

3. Exempel 1 och 2 medför att f varken är växande eller avtagande (på $\mathbf{R}$).

Sats 2.40. *Låt $A, B, C \subseteq \mathbf{R}$, $f\colon A \to B$ och $g\colon B \to C$.*

1. Om både f och g är strängt växande då är $(g \circ f)$ strängt växande.

2. Om både f och g är strängt avtagande då är $(g \circ f)$ strängt växande.

3. Om f är strängt växande och g är strängt avtagande då är $(g \circ f)$ strängt avtagande.

4. Om f är strängt avtagande och g är strängt växande då är $(g \circ f)$ strängt avtagande.

Påståendena gäller fortfarande om man stryker 'strängt' från både hypoteserna och slutsatserna.

Bevis. Här visar vi bara 2 och lämnar de andra som en övning för läsaren.

Då antar vi att både f och g är strängt avtagande. Det innebär att $x_1 < x_2$ medför $f(x_2) < f(x_1)$ och det i sin tur medför att $(g \circ f)(x_2) = g(f(x_2)) > g(f(x_1)) = (g \circ f)(x_1)$. □

3 Geometri och reella tal

Nu ska vi ta begreppet reella tal vidare och fundera på hur geometri kan hjälpa oss att förstå det bättre. Geometri kan beskrivas axiommässigt precis som vi började att göra för de reella talen i I och II i avsnitt 2.2.1. Det är en spännande historia men det är inte riktigt syftet med kursen. Hellre vill vi lita till vår geometriska intuition för att motivera och bygga vidare på axiomsystemet för de reella talen. Det möjliggör att vi kan definiera kvadratrötter, exponential- och logaritmfunktioner, samt utveckla trigonometri.

3.1 Former, längd och area

3.1.1 Avstånd, längd och vinkel

I avsnitt 2.2 ritade vi en bild av de reella talen som en oänligt lång linjal och i avsnitt 2.2.5 och 2.5.3 utvecklade vi bilden till koordinateplanet. Där har vi redan pratat om de lodräta och horisontella axlarna, men här vill vi precisera och generalisera begreppen av längd och vinkel.

Vi föreställer oss ett plan till sammans med en linje. På linjen markera man reella talen precis som vi beskrev i avsnitt 2.2. Om vi tillåter linjen att flytta sig runt i planet utan att böja sig eller deformeras så får vi använda den för att mäta avståndet mellan punkter i planet. Vi betecknar[13] avståndet mella två punkter A och B

$$d(A, B).$$

Med hjälp av det här avståndet kan vi definiera vår första form: Mängden $\{P \in \mathbf{R}^2 : \ell(M, P) = r\}$ av alla punkter som ligger ett avstånd r från en given punkt M kallas för en *cirkel* med *medelpunkten* M och *radien* r. Se figur 10a.

En cirkel är ett exempel på en *kurva*, trots att det är inte helt självklart från definitionen ovan. Här ge vi ingen precis definition av en kurva, men begreppet vi vill fånga är det matematiskt analog av att rita med en penna på papper. Som en första försök kan vi säga en *kurva* beskrivs av en funktion $\gamma\colon [0, 1] \to \mathbf{R}^2$. Se figur 10b. Man kan tänka att funktionen γ beskriver inte bara kurvan på pappret (som är funktionens värdemängd) utan också hur man rita den. I synnerhet ärver kurvan en ordning på sina element från ordningen som redan finns på intervallet $[0, 1]$. Självklart finns det flera funktioner som beskriva det samma kurvan.

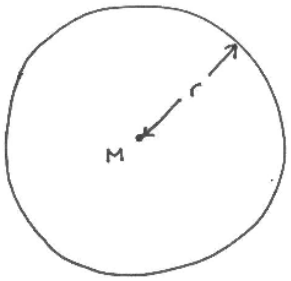
Trots att det är lätt att föreställa sig hur man rita en cirkel är det inte lätt att hitta ett uttryck för en funktion γ som beskriver en cirkel. Men i avsnitt 3.2 kommer vi göra förarbetet för begreppet som behovs — kvadratrötter — som vi slutlingen definierar i avsnitt 3.3. Vi introducera även andra notation att beskriva en cirkel i avsnitt 3.4.1.

När vi ritar en cikel slutar vi på samma punkt som vi började i — det vill säga, $\gamma(0) = \gamma(1)$. En sådan kurva kallas för en *sluten kurva*.

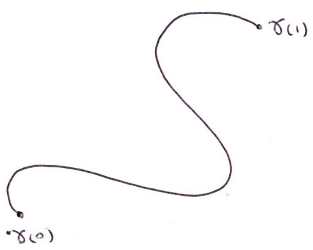
Det är inte alldeles uppenbart hur vi kan mäta längden av en kurva: Vi vet bara hur det skulle funkar för en rät linje där vi tar längden av vara avståndet mellan de två ändpunkterna. Andra kurvor är mer komplicerade. Men vi kan uppskata längden genom att bryta upp kurvan i små streck och uppskatta längden av varje streck men en rät linje. Först välj ut tal $a_0, a_1, a_2, \ldots, a_{N-1}, a_N$ från $[0, 1]$ så att $a_0 = 0$, $a_N = 1$ och $a_{i-1} < a_i$ för $i = 1, 2, \ldots, N$ — det kallas för en *partition* eller *indelning* av intervallet $[0, 1]$. Då ger uppskatningen av kurvans längd av

$$\sum_{i=1}^{N} d(\gamma(a_{i-1}), \gamma(a_i)). \tag{3.1}$$

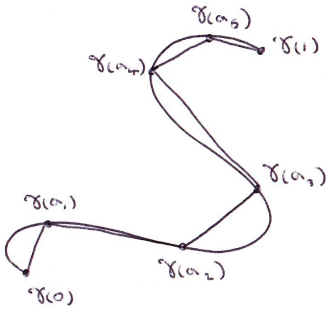
[13]Det finns en formel för avståndet mellan två punkter som är en funktion av punkternas koordinator, men det kan vi bara göra med hjälp av Pythagoras sats (sats 3.2) och kvadratrötter från avsnitt 3.3.1.



(a) En cirkel med medelpunkten M och radien r.



(b) En kurva som beskrivs av en funktion γ.



(c) En uppskattning av en kurvas längd med hjälp av en partition.

Figur 10: Cirklar och kurvor.

Eftersom vi underskattar längden av varje streck är (3.1) alltid en underskattning av kurvans längd. Men genom att dela upp kurvan i tillräkligt sma streck känns det som vi kan uppskatta längden precis hur exakt vi vill. Därför definierar vi kurvans längd att vara

$$\sup_{\{a_0,a_1,a_2,\ldots,a_N\}} \sum_{i=1}^{N} \ell(\gamma(a_{i-1}), \gamma(a_i))$$

där supremum tas över alla partitioner $\{a_0, a_1, a_2, \ldots, a_N\}$ av intervallet $[0, 1]$.

Defintionen är ganska svårhanterat men i de flesta fall kan vi undvika att använda den direkt och lita istället på symmetri, våra geometriska intuition, och så vidare.

När två räta linjer möter varandra i en punkt kan vi tala om *vinkeln* mellan de två linjerna med hjälp av en cirkel. Betrakta två linjer, den ena mellan A och B som kallas för AB, och den andra mellan B och C, som kallas för BC. Rita en cirkel med radien 1 och medelpunkten B. Vinkeln definieras som längden θ av den delen av cirkeln som ligger mellan AB och BC. Se figur 11a.

I synnerhet om A, B och C ligger på en gemensam rät linje är vinkeln hälften av cirkelns omkrets (figur 11b). Vi betecknar längden π.[14] Med den här definitionen kallas mätenheten av vinklar *radianer*. Tidigare har ni mött mätenheten *grader*, då vinkeln ovan skulle har varit 180° (180 grader), men i matematik är radianer mer praktiskt. En vinkel som är hälften så stor är då $\pi/2$ radianer. Vinkeln $\pi/2$ dyker upp ganska ofta så vi döper den till *rät vinkel* (figur 11c).

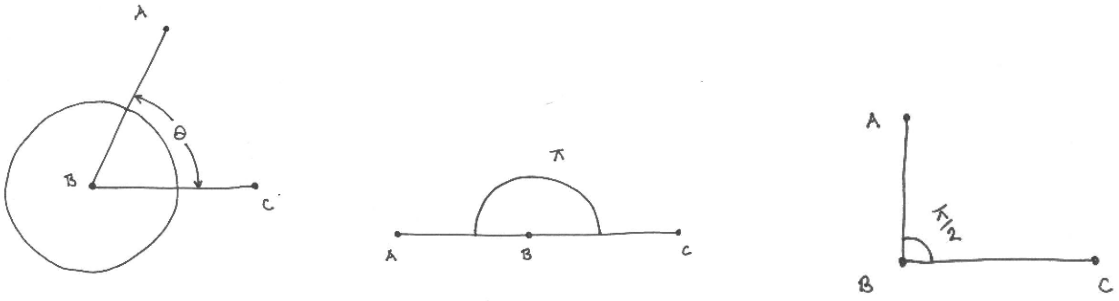
3.1.2 Former och area

Här betraktar vi ett par viktiga former i geometrin. En *triangel* skapas av tre punkter A, B och C i planet tillsammans med tre olika räta linjer som börjar och slutar i punkterna A, B och C. Se figur 12a. Observera att det bara finns ett sätt att dra tre olika linjer mellan tre punkter. En triangel som har en rät vinkel mellan två sidor kallas för en *rätvinklig triangel*.

En *fyrhörning* skapas av fyra punkter A, B, C och D med fyra linjer som börjar och slutar i punkterna A, B, C och D på så sätt att linjerna inte skär varandra och varje punkt är början eller slutet på två linjer. Se figur 12b. Punkterna A, B, C och D kallas för *hörnpunkter* och linjerna kallas för *sidor*.

En *rektangel* är en fyrhörning där vinklarna mellan linjerna i varje hörnpunkt är lika med $\pi/2$ radianer. En *kvadrat* är en rektangel där alla sidor har samma längd.

[14] Det är ett tal mellan 333/106 och 355/113 fast vi inte bevisar det här.

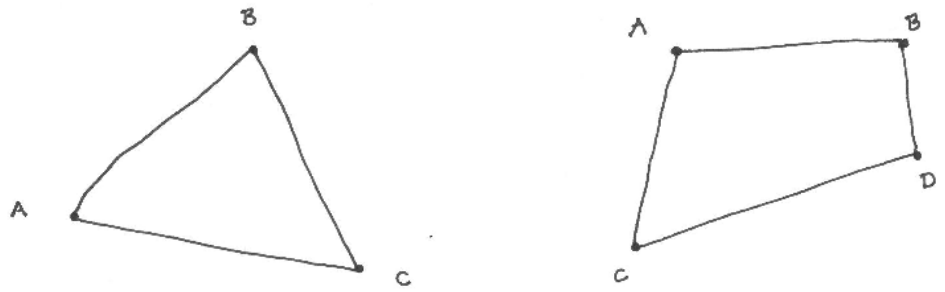


(a) En vinkel θ mellan linjerna AB och BC.

(b) Vinklen π radianer mellan AB och BC.

(c) En rät vinkel mellan AB och BC är like med $\pi/2$.

Figur 11: Vinklar och definitionen av π.



(a) En triangel med hörnpunkterna A, B och C. Kan du bevisa att det bara finns ett sätt att dra tre olika linjer mellan tre punkter? Tips: Tänk på binomialkoefficienter.

(b) En fyrhörning med hörnpunkterna A, B, C och D. På vilka andra sätt kan man rita frya räta linjer mellan punkterna om man glömmer bort villkoren att linjerna inte skär varandra och varje punkt är början eller slutet på två linjer?

Figur 12: Två former.

Varje form delar planet i två disjunkta mängder: *interiören* och *exteriören* av formen. Ibland smälter vi samman en form med sin interiör. Vi kan mäta storleken av en form genom att mäta *arean* av sin interiör. Arean av en form är analog till längden av en kurva och egentligen mätar storleken av delmängder av $\mathbf{R}^2$.[15]

Definition 3.1. *Arean $|F|$ av en mängd $F \subseteq \mathbf{R}^2$ är ett ickenegativt tal som uppfyller följande regler:*

1. *Arean av en rektangel R med sidlängderna a och b är $|R| = ab$ och den tomma mängden $\emptyset$ har arean $|\emptyset| = 0$;*

2. *För två mängder F_1 och F_2, om $F_1 \subseteq F_2$ så är $|F_1| \leq |F_2|$;*

[15]Det finns mängder vilkas area inte kan definieras men den där diskussionen är utöver kursen.

3. Om $F_1, F_2, F_3, \ldots$ är en följd av parvis icke överlappande mängder så gäller

$$\left|\bigcup_j F_j\right| = \sum_j |F_j|.$$

Utifrån denna definition av area kan man bevisa att arean av en rätvinklig triangel med kortsidorna av längderna a och b är $ab/2$. Med lite mer arbete (och trigonometri från avsnitt 3.4 gör det betydligt enklare) kan man visa att arean av en cirkel med radien r är πr^2.

3.2 Pythagoras sats och irrationella tal

3.2.1 Pythagoras sats och en praktisk fråga

En av de äldsta satserna i matematiken är Pythagoras sats. Den handlar om rätvinkliga trianglar. Trots att den döptes efter Pythagoras är det oklart om det var Pythagoras som först upptäckte den. Förmodligen har satsen upptäckts av olika matematiker flera gånger under historiens gång. Vi vet att babyloniska matematiker kände till algebraiska samband mellan tripplar av heltal redan tusen år innan Pythagoras föddes men vi inte vet om de kände till sambandet med trianglar. Det finns även en kinesisk avhandling samtida med Pythagoras som innehåller ett bevis av satsen.

Sats 3.2 (Pythagoras sats)**.** *Betrakta en rätvinklig triangel som har sidor med längden a, b och c. Anta att sidan mittemot den räta vinkeln har längden c.*[16] *Då är*

$$a^2 + b^2 = c^2. \tag{3.2}$$

Bevis. Betrakta en kvadrat med sidorna av längden $a + b$. Markera en punkt på toppsidan som har längden a från vänsterkanten. Gör det samma på de tre andra sidorna och rita en kvadrat som har de fyra punkterna som hörnpunkter. Se figur 13.

Mittemellan de två kvadraterna finns fyra rätvinkliga trianglar som alla har sidor med längderna a, b och c.

Vi kan räkna ut arean av den störst kvadraten på två olika sätt. Det första är med den vanliga formeln för arean av en kvadrat med sidorna med längden $a+b$. Då är arean $(a+b)^2$. Det andra sättet är att addera arean av den mindre kvadraten c^2 och arean av de fyra trianglarna $ab/2$. Det vill säga arean är $c^2 + 4(ab/2)$. Eftersom båda uttrycken för arean måste vara lika får vi att

$$a^2 + 2ab + b^2 = (a + b)^2 = c^2 + 4(ab/2) = c^2 + 2ab.$$

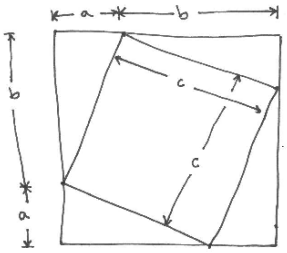
Det medför att $a^2 + b^2 = c^2$. □

Pythagoras sats leder till en naturlig fråga. Om vi vet längden på två sidor av en rätvinklig triangel, till exempel a och b i satsen, kan vi då skriva om likheten (3.2) för c? Låt oss tänka lite mer konkret genom ett exempel: om vi hade en rätvinklig triangel med två korta sidor av längden 1, kan vi då räkna ut längden av hypotenusan? Vi kan ställa frågan med andra ord: kan vi lösa ekvationen $1 + 1 = c^2$, eller

$$2 = c^2 \tag{3.3}$$

för c? Det verkar inte så svårt, men i nästa avsnitt börjar vi med att visa att det inte är helt enkelt.

[16] Sidan mittemot den räta vinkeln kallas för *hypotenusan* och de två andra sidorna kallas för *kateterna*.



Figur 13: En kvadrat med sidor av längden c inom en kvadrat med sidor av längden $a + b$.

3.2.2 Lösningar c till (3.3) är irrationella

Nu kommer vi fram till nästa viktiga sats. Den är ungefär lika gammal som Pythagoras sats och är lika vacker idag som den dagen då den upptäcktes.[17]

Observera att om man adderar två rationella tal får man ett rationellt tal. Man kan dra samma slutsats om man subtraherar, multiplicerar eller dividerar de två talen (antagande förstås vid division att det andra är skilt från 0). Det finns alltså inget sätt att använda aritmetik för att börja med ett rationellt tal och räkna fram till ett tal som inte är rationellt. Ett tal som inte är rationellt kallas för *irrationellt*. Den följande satsen säger då att vi inte kan räkna fram en lösning c till (3.3) genom att bara använda aritmetiska operationer (det vill säga genom addition, subtraktion, multiplikation och division).

Sats 3.3. *Om $c^2 = 2$ så är c irrationellt.*[18]

Innan vi bevisar satsen är det bäst att bevisa en hjälpsats:

Om man vet att ett heltal i kvadrat c^2 är jämnt så måste c också vara jämnt. (3.4)

Hur bevisar man (3.4)? Det räcker att visa sin kontraposition: Om c inte är jämt så kan c^2 inte vara jämnt.

Om c är inte jämnt så är det udda och kan skrivas som $c = 2n + 1$ för något heltal n. Det medför att $c^2 = (2n - 1)^2 = 2(2n^2 - n) + 1$ så c^2 är också udda, det vill säga c^2 är inte jämnt!

Bevis av sats 3.3. Vi antar att c är rationellt, det vill säga att $c = n/m$ för $n, m \in \mathbf{Z}$. Vi kan också anta att n och m inte har någon gemensam delare eftersom om vi kan skriva $c = n/m$ och sedan stryker vi alla gemensamma delare mellan n och m.

Därav får vi att $2 = c^2 = n^2/m^2$ som medför att $2m^2 = n^2$. Så n^2 är jämnt och enligt (3.4) medför det att n är jämnt och kan då skrivas som $n = 2\ell$ för något $\ell \in \mathbf{Z}$. Nu kan vi sätta $n = 2\ell$ i $2m^2 = n^2$ och få $2m^2 = 4\ell^2$ som medför att $m^2 = 2\ell^2$. Det innebär att m^2 är jämnt och därför enligt (3.4) är m också jämt.

Vi har bevisat att både n och m är jämna och det innebär att 2 delar både n och m. Det är en motsägelse till att n och m inte har någon gemensam delare, därför är c inte rationellt. □

Om det inte finns något enkelt sätt att hitta lösningar till (3.3) hur kan vi då lösa den?

[17] G.H. Hardy [5, s. 92] sa om satsen samt om Euklides bevis att det finns oändligt många primtal: They are 'simple' theorems, simple both in idea and in execution, but there is no doubt at all about their being theorems of the highest class. Each is as fresh and significant as when it was discovered—two thousand years have not written a wrinkle on either of them.

[18] Observera att c är en lösning till (3.3).

3.2.3 Det finns en unik positiv lösning till (3.3)

Nu försöker vi att hitta en lösning till (3.3). Eftersom c representerar en längd, vill vi hitta en positiv lösning c till (3.3). Dessutom kommer vi till och med att visa denna lösning är unik bland positiva tal.

Metoden vi kommer använder oss av att hitta en lösning är att dela upp positiva tal $\mathbf{R}^+ = \{x \,|\, x > 0\}$ i två delar. Vi definierar

$$M_- = \{x \in \mathbf{R}^+ \,|\, x^2 < 2\} \quad \text{and} \quad M_+ = \mathbf{R} \setminus M_- = \{x \in \mathbf{R}^+ \,|\, x^2 \geq 2\}. \tag{3.5}$$

Alla tal $x \in \mathbf{R}^+$ tillhör antingen till M_- eller M_+, men inte både M_- och M_+. På så sätt har vi delat upp $\mathbf{R}^+$ i två disjunkta delar. Vi skulle gissa att en lösning c till (3.3) finns där M_- och M_+ möter varandra och det är nu vad vi försöker visa.

Sats 3.4. *Låta $c = \sup M_-$. Då är c en lösning till* (3.3)*, det vill säga $c^2 = 2$.*

Bevis. Först noterar vi att M_- inte är tomt: till exempel $1 \in M_-$ ty $1^2 < 2$. Dessutom är M_- begränsad uppåt: Om $x \geq 2$ så är $x^2 \geq 2^2 \geq 2$ och därför måste $x < 2$ för alla $x \in M_-$. Enligt III(a) finns det en minsta övre begränsning till M_-, så c är väldefinierat och är så att $1 \leq c \leq 2$.

Enligt (2.16) för varje $0 < \varepsilon \leq 1$ finns det $x \in M_-$ så att $c - \varepsilon < x$. Eftersom $c \geq 1$ och $\varepsilon \leq 1$ är $c - \varepsilon \geq 0$ så $(c-\varepsilon)^2 < x^2$, och eftersom $x \in M_-$ är $x^2 < 2$.

$$\begin{aligned} (c-\varepsilon)^2 < x^2 < 2 &\implies (c-\varepsilon)^2 < 2 \\ &\iff c^2 - 2c\varepsilon + \varepsilon^2 < 2 \\ &\iff c^2 - 2 < 2c\varepsilon - \varepsilon^2 = (2c-\varepsilon)\varepsilon. \end{aligned}$$

Eftersom $c \leq 2$ och $\varepsilon > 0$ så gäller

$$c^2 - 2 < (2c-\varepsilon)\varepsilon < (2 \times 2 - 0)\varepsilon < 5\varepsilon \tag{3.6}$$

för alla $\varepsilon > 0$.

Enligt (2.15) är talet c också en övre begränsning till M_- och därför om $\varepsilon > 0$ är $x \leq c < c+\varepsilon$ för alla $x \in M_-$, så

$$\begin{aligned} c + \varepsilon \in M_+ &\implies (c+\varepsilon)^2 \geq 2 \\ &\iff c^2 + 2c\varepsilon + \varepsilon^2 \geq 2 \\ &\iff c^2 - 2 \geq -2c\varepsilon - \varepsilon^2 = -(2c+\varepsilon)\varepsilon \end{aligned}$$

Vi har valt $\varepsilon \leq 1$ och vi vet att $c \leq 2$ så

$$c^2 - 2 \geq -(2c+\varepsilon)\varepsilon > -(2 \times 2 + 1)\varepsilon = -5\varepsilon \tag{3.7}$$

för alla $0 < \varepsilon \leq 1$. Uppskattningar (3.6) och (3.7) tillsammans medför att

$$0 \leq |c^2 - 2| < 5\varepsilon$$

för alla $0 < \varepsilon \leq 1$. Nu får vi välja $\varepsilon = 1/(5n)$ för godtyckliga $n \in \mathbf{Z}_+$ så

$$0 \leq |c^2 - 2| < 1/n$$

för godtyckliga $n \in \mathbf{Z}_+$. Därför enligt III(b) i avsnitt 2.3.2 är $|c^2 - 2| = 0$. Därifrån ser vi att $c^2 = 2$. □

Sats 3.5. *Talet c definierat i sats 3.4 är det enda positivt tal som löser* (3.3).

Bevis. Anta att det finns två olika positiva lösningar c_1 och c_2 till (3.3), så $c_1^2 = c_2^2 = 2$. Vi får också anta att

$$c_1 < c_2. \tag{3.8}$$

(Annars byter vi bara plats av c_1 och c_2.) Vi kan räkna ut att

$$0 = 2 - 2 = c_2^2 - c_1^2 = (c_2 - c_1)(c_2 + c_1) \tag{3.9}$$

Men eftersom både c_1 och c_2 är positiva är $(c_2 + c_1) > 0$ och från (3.8) är $(c_2 - c_1) > 0$, så

$$(c_1 - c_2)(c_1 + c_2) > 0$$

som är en motsägelse till (3.9) och därför kan det inte finnas två lösningar till (3.3). □

3.3 Inversa funktioner

Nu får vi backa ett steg och tänka på vad vi har gjort i avsnitt 3.2.3. Vi tog en funktion f och ett tal d och visade att det finns precis en positiv lösning c till $f(c) = d$ — i avsnitt 3.2.3 var $f(x) = x^2$ och $d = 2$. I det här avsnittet vill vi fundera över några andra funktioner $f\colon A \to B$ där A och B är delmängder av $\mathbf{R}$. Vad behövs för att få en unik lösning $c \in A$ till ekvationen

$$f(c) = d \tag{3.10}$$

för varje $d \in B$?

Vi kan dela upp vad vi frågar efter i två delar. Först vill vi för varje $d \in B$ finna en lösning c till (3.10). Om det är möjligt kallas f för en *surjektiv* funktion. Sist vill vi till varje $d \in B$ finna högst ett $c \in A$ så att (3.10) gäller. Om det stämmer kallas f för en *injektiv* funktion. En surjektiv och injektiv funktion kallas för *bijektiv* (eller *inverterbar*).

Om $f\colon A \to B$ är bijektiv, definierar man en *invers* funktion $f^{-1}\colon B \to A$ enligt

$$f^{-1}(d) = c$$

för varje $d \in B$, där c är den unika lösningen till (3.10).

3.3.1 Kvadratrötter

Om vi går tillbaka till avsnitt 3.2.3 vi kan se att det var ingenting speciellt med $d = 2$ och vi kan upprepa argumentet för vilket positivt tal d som helst. Det vill säga, vi får ersätta mängderna M_- och M_+ i (3.5) med

$$M_-(d) = \{x \in \mathbf{R}^+ \,|\, x^2 < d\} \quad \text{and} \quad M_+(d) = \mathbf{R} \setminus M_-(d) = \{x \in \mathbf{R}^+ \,|\, x^2 \geq d\}.$$

Vi kan upprepa argumentet i avsnitt 3.2.3 för vilket positivt tal d som helst och bevisa följande satsen.

Sats 3.6. *Låt $c = \sup M_-(d)$. Då är c den unika lösning till $c^2 = d$ bland positiva tal.*

Dessutom om $d = 0$ är $c = 0$ den enda lösningen till $c^2 = 0$. (Hur bevisar man det?) Låt mängden av ickenegativa tal skrivas som $\mathbf{R}_0^+ := \{x \in \mathbf{R} \,|\, x \geq 0\}$. Därför har vi visat att $f\colon \mathbf{R}_0^+ \to \mathbf{R}_0^+$ definierad enligt

$$f(x) = x^2 \tag{3.11}$$

för alla $x \geq 0$ är bijektiv och har då en invers $f^{-1}\colon \mathbf{R}_0^+ \to \mathbf{R}_0^+$. Inversen f^{-1} är ganska praktisk så vi använder en särskild notation som kallas *kvadratrot*: Vi skriver

$$\sqrt{y} := f^{-1}(y)$$

för alla $y \geq 0$ där f definieras enligt (3.11).

De följande exemplen visar att när man försöker invertera en funktion är det viktigt att tänka noga på både definitionsmängden och målmängden.

Exempel 3.7. *Funktionen* $f\colon \mathbf{R} \to \mathbf{R}$ *definierad enligt* (3.11) *för alla* $x \in \mathbf{R}$ *är varken surjektiv (det finns inget* x *så att* $x^2 = -4$*) eller injektiv (*$(-2)^2 = 2^2$*). Om* $f\colon \mathbf{R} \to \mathbf{R}_0^+$ *definieras enligt* (3.11) *för alla* $x \in \mathbf{R}$ *så är den surjektiv men inte injektiv. Som sagt ovan definieras* $f\colon \mathbf{R}_0^+ \to \mathbf{R}_0^+$ *enligt* (3.11) *för alla* $x \in \mathbf{R}_0^+$ *så den är både surjektiv och injektiv, därför bijektiv.*

3.3.2 Rationella potenser

Med bara ett fåtal ändringar får vi generalisera argumentet i avsnitt 3.2.3 ett steg till. Låt $n \in \mathbf{Z}_+$ och d vara ett positivt tal. Vi ersätter mängderna M_- och M_+ i (3.5) med

$$M_-^n(d) = \{x \in \mathbf{R}^+ \mid x^n < d\} \quad \text{and} \quad M_+^n(d) = \mathbf{R} \setminus M_-^n(d) = \{x \in \mathbf{R}^+ \mid x^n \geq d\}$$

och sen arbetar man ungefär som i avsnitt 3.2.3 för att visa den följande generaliseringen av sats 3.6.

Sats 3.8. *Låta* $c = \sup M_-^n(d)$*. Då är* c *den unika lösningen till* $c^n = d$ *bland positiva tal.*

Återigen om $d = 0$ är $c = 0$ den enda lösningen till $c^n = 0$ så det visas att funktionen $f_n\colon \mathbf{R}_0^+ \to \mathbf{R}_0^+$ definierad enligt $f_n(x) = x^n$ för alla $x \geq 0$ är bijektiv. Vi definierar då

$$\sqrt[n]{y} := f_n^{-1}(y)$$

för alla $y \geq 0$. Speciellt kallas $\sqrt[3]{y}$ för *kubikroten* ur y.

Nu får vi även utöka definitionen av potens från heltal till rationella tal genom definitionen

$$a^{m/n} := \sqrt[n]{a^m} \tag{3.12}$$

för alla $a > 0$, $m \in \mathbf{Z}$ och $n \in \mathbf{Z}_+$.[19] I synnerhet får vi skriva $\sqrt[n]{a}$ som $a^{1/n}$. Man får kontrollera att om $a, b > 0$ så gäller sats 2.10 för rationella n och m. Vi skriver om satsen för rationella tal här nedan.

Sats 3.9. *Givna godtyckliga* $a, b > 0$ *och* $r, s \in \mathbf{Q}$ *har vi att*

1. $a^r a^s = a^{r+s}$,
2. $a^r a^{-s} = a^{r-s}$,
3. $a^r b^r = (ab)^r$,
4. $(a^r)^{-1} = a^{-r}$, *och*
5. $(a^r)^s = a^{rs}$.

Bevis. Här bevisar vi 3 och lämnar de andra som en övning för läsaren.

Först ska vi bevisa att

$$\sqrt[n]{ab} = \sqrt[n]{a}\sqrt[n]{b} \tag{3.13}$$

för alla positiva rationella a och b och $n \in \mathbf{Z}_+$. Vänsterledet är lika med den unika positiva lösningen x till ekvationen

$$x^n = ab. \tag{3.14}$$

[19] Observera att här har vi inte utökat definitionen av rationella potenser för negativa a fast definitionen för heltalspotenser (2.12) även gäller för $a < 0$. Definitionen kräver att rationella tal först skrivs med en positiv nämnare.

Vi vet också att $\sqrt[n]{a} = y$ och $\sqrt[n]{a} = z$ där y respektive z är den unika positiva lösningen till

$$y^n = a \quad \text{respektive} \quad z^n = b.$$

Vi vill bevisa att $x = yz$. Men enligt 3 från sats 2.10 är

$$(yz)^n = y^n x^n = ab$$

så yz är också en lösning till (3.14). Som en produkt av två positiva tal är yz också positiva. Eftersom det finns högst en positiv lösning till (3.14) så måste $x = yz$, alltså är (3.13) bevisad.

Vi får skriva $r \in \mathbf{Q}$ som $r = m/n$ för $n \in \mathbf{Z}_+$ och $m \in \mathbf{Z}$. Enligt definitionen (3.12) och (3.13) är

$$a^r b^r = a^{m/n} b^{m/n} = \sqrt[n]{a^m}\sqrt[n]{b^m} = \sqrt[n]{a^m b^m}.$$

I sin tur är

$$\sqrt[n]{a^m b^m} = \sqrt[n]{(ab)^m} = (ab)^{m/n} = (ab)^r$$

enligt 3 från sats 2.10 och återigen definitionen (3.12). Så är 3 bevisad. □

3.3.3 Andragradspolynom

Med verktyget kvadratrot kan vi räkna ut en formel för lösningar (eller *rötter*) till ekvationen

$$ax^2 + bx + c = 0 \tag{3.15}$$

för $a \neq 0$ och $b, c \in \mathbf{R}$. Det vill säga en formel för *nollställen* $x \in \mathbf{R}$ till andragradspolynomet $ax^2 + bx + c$. Först måste vi tänka på vad som händer när definitionsmängden av f i (3.11) är $\mathbf{R}$ istället för $\mathbf{R}_0^+$.

Vi vet redan att f är växande på $\{x \,|\, x \geq 0\}$. På liknande sätt visar man att f är avtagande på $\{x \,|\, x \leq 0\}$. Precis som i avsnitt 3.2.3 kan man visa att det inte kan finnas två eller flera ickepositiva lösningar x till $x^2 = d$ och vi räknar ut direkt att $(-\sqrt{d})^2 = d$ så $-\sqrt{d}$ är den ickepostiva lösningen vi letar efter. Alltså vet vi att det finns exakt två lösningar x till $x^2 = d$ om $d > 0$, ingen lösning om $d < 0$ och exakt en lösning om $d = 0$ eftersom då är $\sqrt{d} = -\sqrt{d} = 0$.

För att nu lösa (3.15) observera att

$$ax^2 + bx + c = a\left(x + \frac{b}{2a}\right)^2 - \frac{b^2}{4a} + c$$

så (3.15) är ekvivalent med

$$\left(x + \frac{b}{2a}\right)^2 = \frac{b^2 - 4ac}{4a^2}. \tag{3.16}$$

Det finns lösningar till (3.16) om och endast om $b^2 - 4ac \geq 0$. I fallet $b^2 - 4ac \geq 0$ är (3.16) ekvivalent med

$$\left(x + \frac{b}{2a}\right) = \pm\sqrt{\frac{b^2 - 4ac}{4a^2}}$$

enligt argumentet ovan. Här betyder $\pm$ att vi egentligen har två möjliga ekvationer för x, antingen den med plustecken eller den med minustecken (fast båda ekvationerna är de samma om $b^2 - 4ac = 0$). Alltså får vi att (3.15) har just två lösningar

$$x = -\frac{b}{2a} \pm \sqrt{\frac{b^2 - 4ac}{4a^2}} \quad \text{om } b^2 - 4ac > 0,$$

precis en lösning

$$x = -\frac{b}{2a} \quad \text{om } b^2 - 4ac = 0,$$

och ingen lösning alls om $b^2 - 4ac < 0$.

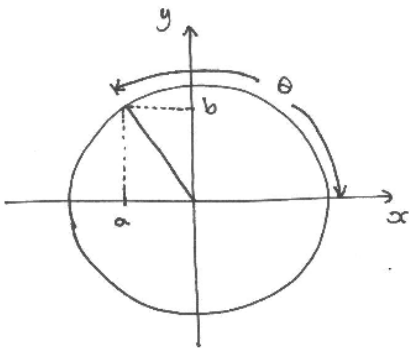
3.4 Trigonometri

3.4.1 Definitioner

Ordet 'trigonometri' kommer ifrån det grekiska ordet för 'triangel' (trígōnos) och 'linjal' (métron) och handlar om relationer mellan längder och vinklar i trianglar. Trots det är det smidigt att definiera trigonometriska funktioner med hjälp av en enhetscirkel, det vill säga en cirkel med radien 1, med medelpunkten i origo. Anledningen till det är att vi då får ta $\mathbf{R}$ som definitionsmängden för de två första funktionerna vi definierar: *cosinus* och *sinus*. Funktionerna $\cos\colon \mathbf{R} \to \mathbf{R}$ och $\sin\colon \mathbf{R} \to \mathbf{R}$ definieras som reella tal

$$\cos\theta = a \quad \text{och} \quad \sin\theta = b$$

från figur 14 för varje $\theta \in \mathbf{R}$.



Figur 14: $\cos\theta$ definieras som den längd (med tecken) längs x-axeln man får när man går moturs runt enhetscirkeln på ett avstånd θ. $\sin\theta$ definieras som den längd (med tecken) längs y-axeln man får när man går moturs runt enhetscirkeln på ett avstånd θ.

Observera att $\cos\theta = 0$ om och endast om $\theta = \pi/2 + n\pi$ för $n \in \mathbf{Z}$. Med

$$A = \{\theta \in \mathbf{R} \,|\, \theta \neq \pi/2 + n\pi \text{ för något } n \in \mathbf{Z}\}$$

kan man definiera funktionen *tangens*; $\tan\colon A \to \mathbf{R}$ enligt

$$\tan\theta = \frac{\sin\theta}{\cos\theta}.$$

3.4.2 Trigonometriska formler

Kom ihåg från avsnitt 3.1.1 att vi betecknade längden runt hälften av enhetscirkelns omkrets π, därför har hela omkretsen längden 2π. Figur 14 övertygar oss om riktigheten hos följande formler:

$$\cos(-\theta) = \cos\theta; \tag{3.17}$$
$$\sin(-\theta) = -\sin\theta; \tag{3.18}$$
$$\cos(\theta + \pi/2) = -\sin\theta; \quad \text{och} \tag{3.19}$$
$$\sin(\theta + \pi/2) = \cos\theta. \tag{3.20}$$

Från Pythagoras sats (sats 3.2) får vi formeln

$$\cos^2\theta + \sin^2\theta = 1 \tag{3.21}$$

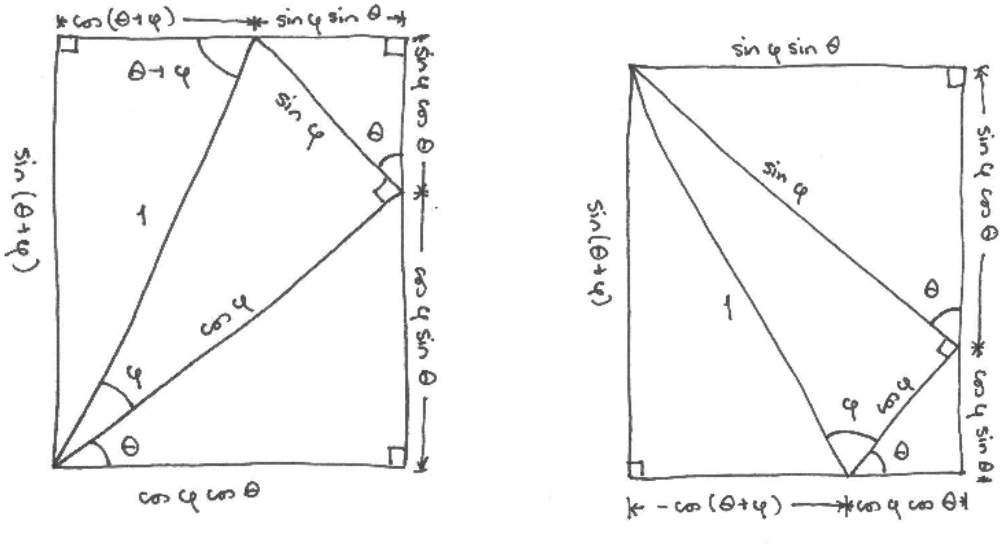
för alla $\theta \in \mathbf{R}$.[20] Formlerna

$$\cos(\theta+\varphi) = \cos\theta\cos\varphi - \sin\theta\sin\varphi \quad \text{och} \tag{3.22}$$
$$\sin(\theta+\varphi) = \sin\theta\cos\varphi + \cos\theta\sin\varphi, \tag{3.23}$$

som gäller för alla $\theta, \varphi \in \mathbf{R}$, bevisas med hjälp av figur 15. Först bevisar vi (3.22) och (3.23) för $\theta \geq 0$, $\phi \geq 0$ och $\theta + \varphi \leq \pi$. Om $\theta + \varphi \leq \pi/2$, se figur 15a och om $\pi/2 < \theta + \varphi \leq \pi$, se figur 15b. Formeln (3.22) bevisas genom att observera att längden av rektangelns ovansida i figuren är lika med nedansidans längd. Och (3.23) bevisas genom att observera att längden av vänster sida av rektangeln är lika med den högra sidans längd. För godtyckliga θ och φ använder man (3.19) och (3.20) tillsammans med det särskilda fallet av (3.22) och (3.23) som vi bevisat. Till exempel, om $\pi/2 < \theta < \pi$ och $\pi/2 < \varphi < \pi$ då är $0 < \theta - \pi/2 < \pi/2$ och $0 < \varphi - \pi/2 < \pi/2$ så

$$\begin{aligned}
\sin(\theta+\phi) &= \sin((\theta-\pi/2)+(\varphi-\pi/2)+\pi) \\
&\underset{(3.20)}{\uparrow} \\
&= \cos((\theta-\pi/2)+(\varphi-\pi/2)+\pi/2) \underset{(3.19)}{\uparrow} = -\sin((\theta-\pi/2)+(\varphi-\pi/2)) \\
&\underset{(3.23)}{\uparrow} \\
&= -\sin(\theta-\pi/2)\cos(\varphi-\pi/2) - \cos(\theta-\pi/2)\sin(\varphi-\pi/2) \\
&\underset{(3.19) \text{ och } (3.20)}{\uparrow} \\
&= \cos\theta\sin\varphi + \sin\theta\cos\varphi
\end{aligned}$$

när (3.23) kan användas här eftersom vi redan har bevisat att det gäller då $0 < \theta - \pi/2 < \pi/2$ och $0 < \varphi - \pi/2 < \pi/2$.



(a) Fallet $\theta + \varphi \leq \pi/2$. (b) Fallet $\pi/2 < \theta + \varphi \leq \pi$.

Figur 15: Med de här bilderna bevisas (3.22) och (3.23) i fallet $\theta \geq 0$, $\phi \geq 0$ och $\theta + \varphi \leq \pi$. För andra fall kan man använda det här fallet tillsammans med (3.19) och (3.20).

Utifrån (3.17)–(3.23) kan man bevisa alla andra användbara trigonometriska formler. Till exempel

$$\cos(\theta-\varphi) = \cos\theta\cos(-\varphi) - \sin\theta\sin(-\varphi) = \cos\theta\cos\varphi + \sin\theta\sin\varphi$$

[20]Det är brukligt att skriva $(\cos\theta)^n$ som $\cos^n\theta$, samt $\sin^n\theta = (\sin\theta)^n$ för positiva heltal n.

enligt (3.22), (3.17) och (3.18). På liknande sätt får man

$$\sin(\theta - \varphi) = \sin\theta\cos(-\varphi) + \cos\theta\sin(-\varphi) = \sin\theta\cos\varphi - \cos\theta\sin\varphi$$

utifrån (3.23), (3.17) och (3.18). Samt

$$\sin(2\theta) = \sin(\theta + \theta) = \sin\theta\cos\theta + \cos\theta\sin\theta = 2\sin\theta\cos\theta$$

enligt (3.23) och

$$\cos(2\theta) = \cos(\theta + \theta) = \cos\theta\cos\theta - \sin\theta\sin\theta = \cos^2\theta - \sin^2\theta \tag{3.24}$$

enligt (3.22). Från (3.24) och (3.21) får man

$$\cos^2\theta = \frac{1 + \cos(2\theta)}{2}$$
$$\sin^2\theta = \frac{1 - \cos(2\theta)}{2}.$$

Man får också

$$\tan(\theta + \varphi) = \frac{\sin(\theta + \varphi)}{\cos(\theta + \varphi)} = \frac{\sin\theta\cos\varphi + \cos\theta\sin\varphi}{\cos\theta\cos\varphi - \sin\theta\sin\varphi} = \frac{\tan\theta + \tan\varphi}{1 - \tan\theta\tan\varphi}$$

enligt definitionen av tangens, (3.22) och (3.23) om vi antar att varken θ, φ eller $\theta + \varphi$ är lika med $\pi/2 + n\pi$ för något $n \in \mathbf{Z}$.

Sats 3.10 (Trigonometriska olikheter)**.** *För $\theta \in (0, \pi/2)$ är*

$$\sin\theta < \theta < \tan\theta.$$

Bevis. Betrakta den fjärdedel av en enhetscirkel som ritas i figur 16. Låt T_1 vara den triangel med hörnpunkterna A, B och E som ritas i figur 16a, T_2 vara den triangel med hörnpunkterna B, C och E och T_3 vara den triangel med hörnpunkterna A, C och D. Dessutom låt W vara den delen av enhetscirkeln mellan linjen från A till C och linjen från A till E. Enligt figur 16a ser vi att

$$T_1 \cup T_2 \subset W \subset T_3$$

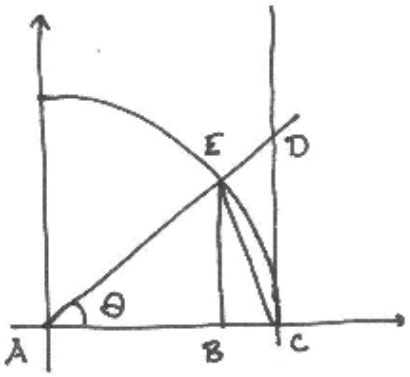
så genom att ta arean av mängderna och använda definition 3.1 får vi att

$$|T_1 \cup T_2| < |W| < |T_3|.$$

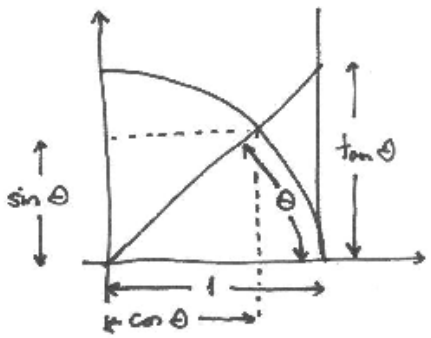
Med hjälp av figur 16b räknar vi ut areorna ovan och får att

$$\frac{\cos\theta\sin\theta}{2} + \frac{(1 - \cos\theta)\sin\theta}{2} < \frac{\theta}{2} < \frac{\tan\theta}{2}$$

som kan föreklas till $\sin\theta < \theta < \tan\theta$. □



(a) En fjärdedel av en enhetscirkel, en vinkel $\theta \in (0, \pi/2)$, några räta linjer och punkterna A, B, C, D och E.



(b) Samma fjärdedelen av en enhetscirkel med längderna $\sin\theta$, $\cos\theta$ och $\tan\theta$ markerade.

Figur 16: En fjärdedel av en enhetscirkel och en vinkel $\theta \in (0, \pi/2)$.

3.4.3 Arcusfunktioner

De trigonometriska funktionerna sinus, cosinus och tangens är inte bijektiva så det är inte möjligt att definiera inversa funktioner för dem. De är i synnerhet inte injektiva; till exempel

$$\sin\theta = \sin(\theta + 2k\pi), \quad \cos\theta = \cos(\theta + 2k\pi) \quad \text{och} \quad \tan\theta = \tan(\theta + \pi)$$

för alla $\theta \in \mathbf{R}$ och $k \in \mathbf{Z}_+$. Om vi emellertid begränsar definitionsmängden på ett lämpligt sätt är funktionerna bijektiva och, åtminstone till de begränsade funktionerna finns inversa funktioner.

I avsnitt 3.3 innebar det mycket arbete att bevisa $f\colon \mathbf{R}_0^+ \to \mathbf{R}_0^+$ definierad enligt formeln $f(x) = x^2$ för alla $x \in \mathbf{R}_0^+$ var bijektiv (i synnerhet injektiv) eftersom vi inte hade någon geometrisk bild att lita till. Här kan vi vara lite slarviga eftersom vi har bilden i figur 14 och trigonometriska funktioner är definierade som längder i bilden. Egentligen är det samma procedur att mäta en längd som vi gick igenom i avsnitt 3.2.3.

Det är klart från geometrin (till exempel, figur 14) att om vi begränsar definitionsmängden av sinus till $[-\pi/2, \pi/2]$ är den injektiv. Då har sinus värdemängden $[-1, 1]$. Samtidigt är det klart att cosinus är injektiv om vi begränsar definitionsmängden till $[0, \pi]$ och har värdemängden $[-1, 1]$, och att tangens (se figur 16) är injektiv om vi begränsar definitionsmängden till $(-\pi/2, \pi/2)$ och har värdemängden $\mathbf{R}$. Alltså definierar vi

$$\begin{aligned} &\arcsin\colon [-1,1] \to [-\pi/2, \pi/2], \\ &\arccos\colon [-1,1] \to [0,\pi], \quad \text{respektive} \\ &\arctan\colon \mathbf{R} \to (-\pi/2, \pi/2) \end{aligned}$$

som inversa funktioner till

$$\begin{aligned} &\sin\colon [-\pi/2, \pi/2] \to [-1,1], \\ &\cos\colon [0,\pi] \to [-1,1], \quad \text{respektive} \\ &\tan\colon (-\pi/2, \pi/2) \to \mathbf{R}. \end{aligned}$$

4 Exponentialfunktion, logaritm och komplexa tal

4.1 Exponentialfunktion

4.1.1 Ränta på ränta

Om man är tillräckligt rik för att spara pengar finns det två bra anledningar att sätta pengarna på ett bankkonto. Den första är att banken har stöldförsäkring och jättebra kassaskåp. Den andra är att banken ger ränta så man kan bli ännu rikare! Banken har råd att betala ränta eftersom den investerar de pengarna den samlar in från alla sina kunder. Banken erbjuder sig att dela behållningen med kunderna. Här ska vi fundera på vad som händer när banken betalar ränta och vad som händer när man låter räntan stå kvar på bankkontot, så att vi får ränta på ränta.

Betrakta ett bankkonto som ger $100x\%$ ränta per år och banken gör n delbetalningar jämnt fördelade under året. Det betyder att varje n:e del av året multipliceras beloppet på kontot med $(1 + x/n)$ och, eftersom det händer n gånger under året, har man efter ett år $(1 + x/n)^n$ gånger så mycket mer pengar på kontot än man ursprungligen satte in.

Nu ställer vi ett par frågor. Frågorna handlar om balansen mellan ökningsfaktorn $(1 + x/n)$ som sjunker med n och antalet delbetalningar n som förstås ökar med n.

1. Är det bättre för kunden att banken gör flera delbetalningar?

2. Finns det någon övre begräns beroende bara på x och inte på n för hur mycket banken ska betala i ränta?

För att svara på frågorna behöver vi en hjälpsats.

Sats 4.1 (Bernoullis olikhet)**.** *För alla reella tal $x \geq -1$ och alla $n \in \mathbf{Z}_+$ får man att*

$$(1 + x)^n \geq 1 + nx \tag{4.1}$$

Dessutom är olikheten sträng (det vill säga är '$>$') om $x \neq 0$ och $n \geq 2$.

Bevis. Vi ger ett induktionsbevis. Först kollar vi vad som händer då $n = 1$: $(1+x)^n = (1+x) \geq (1 + x) = 1 + nx$ så (4.1) stämmer om $n = 1$. Sen antar vi att (4.1) stämmer för $n = m$ där $m \in \mathbf{Z}_+$ och betraktar fallet $n = m + 1$:

$$(1 + x)^{m+1} = (1 + x)^m(1 + x) \geq (1 + mx)(1 + x) = 1 + mx + x + mx^2 \geq 1 + (m + 1)x$$

så (4.1) stämmer för $n = m+1$ och (4.1) är bevisad. I vilket steg använde vi hypotesen $x \geq -1$?

Sträng olikhet följer också av ett induktionsbevis men lämnas som en övning för läsaren. □

Med hjälp av sats 4.1 kan vi svara jakande på första frågan. Ju fler delbetalningar banken gör desto mer ränta får kunden.

Sats 4.2. *För alla $x \in \mathbf{R}$ och $n \in \mathbf{Z}_+$ så att $n > |x|$ är*

$$\left(1 + \frac{x}{n}\right)^n \leq \left(1 + \frac{x}{n+1}\right)^{n+1}.$$

Bevis. Det räcker att bevisa

$$\frac{\left(1 + \frac{x}{n+1}\right)^{n+1}}{\left(1 + \frac{x}{n}\right)^n} \geq 1.$$

Enligt sats 4.1 har vi att

$$\begin{aligned}
\frac{\left(1+\frac{x}{n+1}\right)^{n+1}}{\left(1+\frac{x}{n}\right)^{n}} &= \frac{\left(1+\frac{x}{n+1}\right)^{n}}{\left(1+\frac{x}{n}\right)^{n}}\left(1+\frac{x}{n+1}\right)\\
&= \frac{\left(1+\frac{x}{n}+\frac{x}{n+1}-\frac{x}{n}\right)^{n}}{\left(1+\frac{x}{n}\right)^{n}}\left(1+\frac{x}{n+1}\right)\\
&= \left(\frac{\left(1+\frac{x}{n}\right)+\left(\frac{x}{n+1}-\frac{x}{n}\right)}{1+\frac{x}{n}}\right)^{n}\left(1+\frac{x}{n+1}\right)\\
&= \left(1+\left(\frac{\frac{x}{n+1}-\frac{x}{n}}{1+\frac{x}{n}}\right)\right)^{n}\left(1+\frac{x}{n+1}\right)\\
&= \left(1-\left(\frac{x}{(n+x)(n+1)}\right)\right)^{n}\left(1+\frac{x}{n+1}\right)\\
&\geq \left(1-\left(\frac{nx}{(n+x)(n+1)}\right)\right)\left(1+\frac{x}{n+1}\right)\\
&= 1+\frac{x}{n+1}-\frac{nx}{(n+x)(n+1)}-\frac{nx^2}{(n+x)(n+1)^2}\\
&= 1+\frac{(n+x)(n+1)x-(n+1)nx-nx^2}{(n+x)(n+1)^2}\\
&= 1+\frac{x^2}{(n+x)(n+1)^2} \geq 1
\end{aligned}$$

eftersom $\frac{x}{(n+x)(n+1)} \leq 1$ och $1+\frac{x}{n+1} \geq 0$ om $n > |x|$. □

Vi kan också svara ja på den andra frågan.

Sats 4.3. *Det finns en funktion* $B\colon \mathbf{R} \to \mathbf{R}$ *så att*

$$\left(1+\frac{x}{n}\right)^{n} \leq B(x).$$

för alla $x \in \mathbf{R}$ *och* $n \in \mathbf{Z}_+$ *som uppfyller* $n \geq |x|$.

Bevis. Vi räknar ut att

$$\left(1+\frac{x}{n}\right)\left(1-\frac{x}{n}\right) = \left(1-\left(\frac{x}{n}\right)^{2}\right) \leq 1$$

så

$$\left(1+\frac{x}{n}\right)^{n} \leq \left(1-\frac{x}{n}\right)^{-n}.$$

Enligt sats 4.2 vet vi att

$$\left(1-\frac{x}{n}\right)^{n} \geq \left(1-\frac{x}{n_0}\right)^{n_0}$$

för $n \geq n_0 > |x|$. Därför får vi sätta $n_0 = \lceil |x| \rceil + 1$ där $\lceil |x| \rceil$ är det minsta heltalet större än eller lika med $|x|$ och får att

$$\left(1-\frac{x}{n}\right)^{-n} \leq \left(1-\frac{x}{\lceil |x| \rceil + 1}\right)^{-(\lceil |x| \rceil + 1)}$$

och då

$$\left(1+\frac{x}{n}\right)^{n} \leq \left(1-\frac{x}{\lceil |x| \rceil + 1}\right)^{-(\lceil |x| \rceil + 1)} := B(x).$$

□

4.1.2 Exponentialfunktion

Nu är vi bereda att definiera en funktion $\exp\colon \mathbf{R} \to \mathbf{R}$ som kallas för *exponentialfunktionen*. Först definierar vi till varje $x \in \mathbf{R}$ en följd $(\exp_n(x))_n$ enligt

$$\exp_n(x) = \begin{cases} 0 & \text{om } n \leq |x|, \\ \left(1 + \frac{x}{n}\right)^n & \text{om } n > |x|. \end{cases} \tag{4.2}$$

För varje givet $x \in \mathbf{R}$ är $(\exp_n(x))_n$ så klart en icketom mängd och uppåt begränsad (enligt sats 4.3) så enligt supremumaxiomet (III(a) i avsnitt 2.3.2) finns det en minsta övre begränsning, som betecknas

$$\exp(x) = \sup_{n \in \mathbf{Z}_+} \exp_n(x). \tag{4.3}$$

Speciellt benämnar vi

$$e := \exp(1).$$

Sats 4.4. *Exponentialfunktionen* $\exp\colon \mathbf{R} \to \mathbf{R}$ *har följande egenskaper:*

1. $\exp(x) > 0$ *för alla* $x \in \mathbf{R}$*;*
2. $\exp(x) \geq 1 + x$ *för alla* $x \in \mathbf{R}$*;*
3. $\exp(x)\exp(-x) = 1$ *för alla* $x \in \mathbf{R}$*;*
4. $\exp(x) \leq 1/(1-x)$ *för alla* $x < 1$*; och*
5. $x \mapsto \exp(x)$ *är en växande funktion.*

Bevis. Observera först att

$$\exp_n(x) > 0$$

om $n > |x|$ och eftersom $\exp(x)$ är en övre begränsning av $(\exp_n(x))_n$ är $\exp(x) > 0$ och 1 bevisad.

Enligt sats 4.1 är

$$\exp_n(x) = \left(1 + \frac{x}{n}\right)^n \geq 1 + x$$

för $n > |x|$, så eftersom $\exp(x)$ är en övre begränsning av $(\exp_n(x))_n$ är $\exp(x) \geq 1 + x$ och 2 bevisad.

För att bevisa 3 betrakta godtyckliga $n, m \in \mathbf{Z}_+$, och skriv $N = \max\{n, m\}$ och $M = \min\{n, m\}$. Eftersom $\exp_n(x)$ är växande i variabeln n är

$$\exp_n(x)\exp_m(-x) = \exp_N(x)\exp_N(-x) = \left(1 + \frac{x}{N}\right)^N \left(1 - \frac{x}{N}\right)^N = \left(1 - \frac{x^2}{N^2}\right)^N \leq 1$$

för $N > |x|$, så $\exp(x)\exp(-x) \leq 1$. Sen enligt sats 4.1 har vi att

$$\exp_n(x)\exp_m(-x) \geq \exp_M(x)\exp_M(-x) = \left(1 - \frac{x^2}{M^2}\right)^M \geq 1 - \frac{x^2}{M}$$

för $M > |x|$, så $\exp(x)\exp(-x) \geq 1 - \frac{x^2}{M}$ för alla $M > |x|$. Vi kan sammanställa dessa slutsatser och säger

$$0 \leq 1 - \exp(x)\exp(-x) \leq \frac{x^2}{M}$$

för alla $M > |x|$. Enligt den arkimediska egenskapen III(b) i avsnitt 2.3.2 är då $1 - \exp(x)\exp(-x) = 0$ och 3 är bevisad.

Man kan bevisa 4 med hjälp av 2 och 3:

$$\exp(x) = \frac{1}{\exp(-x)} \leq \frac{1}{1-x}$$

om $x < 1$.

Om $x > y$ och $n > |x| + |y|$ så är

$$\exp_n(x) = \left(1 + \frac{x}{n}\right)^n \geq \left(1 + \frac{y}{n}\right)^n = \exp_n(y)$$

så $\exp(x) \geq \exp(y)$ och 5 är bevisad. □

Sats 4.5. *För alla $x \in \mathbf{R}$ och $r \in \mathbf{Q}$ är*

$$\exp(rx) = (\exp(x))^r \tag{4.4}$$

och speciellt är

$$\exp(r) = e^r. \tag{4.5}$$

Bevis. Vi bevisar bara (4.5) och lämnar (4.4) som en övning för läsaren.

Eftersom $\exp_n(0) = 1$ för alla $n \in \mathbf{Z}_+$ är $\exp(0) = 1$ och $(\exp(1))^0 = 1$ oavsett värdet $\exp(1) > 0$, så vi har bevisat (4.5) om $r = 0$.

Nu betraktar vi $r > 0$. Då får talet $r \in \mathbf{Q}$ skrivas som $r = \ell/m$ för $\ell, m \in \mathbf{Z}_+$ och man får att

$$\exp_p(\ell/m)^m = \left(1 + \frac{(\ell/m)}{p}\right)^{pm} = \left(1 + \frac{1}{mk}\right)^{mk\ell} = \left(1 + \frac{1}{q}\right)^{q\ell} = \exp_q(1)^\ell \tag{4.6}$$

för $p = \ell k$ och $q = mk$ med $k \in \mathbf{Z}_+$.

Betrakta $A_\ell := \{p \,|\, p = \ell k \text{ för } k \in \mathbf{Z}_+\}$. Mängden A_ℓ är en delmängd av $\mathbf{Z}_+$ så att till varje $n \in N$ finns det en $p \in A_\ell$ så att $p \geq n$. Enligt sats 4.2 vet vi också att

$$\exp_p(\ell/m)^m$$

är växande i p för $p \in A_\ell$. Dessa två fakta medför att

$$\sup_{p \in A_\ell} \exp_p(\ell/m)^m = \sup_{n \in \mathbf{Z}_+} \exp_n(\ell/m)^m = \exp(\ell/m)^m. \tag{4.7}$$

Samma argument tillämpat på $A_m := \{q \,|\, q = mk \text{ för } k \in \mathbf{Z}_+\}$ ger oss att

$$\sup_{q \in A_m} \exp_q(1)^\ell = \sup_{n \in \mathbf{Z}_+} \exp_n(1)^\ell = \exp(1)^\ell. \tag{4.8}$$

Enligt (4.6) är mängderna

$$\{\exp_p(\ell/m)^m \,|\, p \in A_\ell\} \quad \text{och} \quad \{\exp_q(1)^\ell \,|\, q \in A_m\}$$

lika med varande, så (4.7) och (4.8) medför att

$$\exp(\ell/m)^m = \exp(1)^\ell. \tag{4.9}$$

Ekvation (4.9) säger att $f_m(\exp(\ell/m)) = \exp(1)^\ell$ när f_m definieras i avsnitt 3.3.2. Vi vet från 1 i sats 4.4 att $\exp(\ell/m) > 0$, så definitionen (3.12) medför att $\exp(\ell/m) = \exp(1)^{\ell/m}$ och (4.5) är bevisad för $r > 0$.

Nu betraktar vi $r < 0$. I så fall får talet $r \in \mathbf{Q}$ skrivas som $r = -\ell/m$ för $\ell, m \in \mathbf{Z}_+$ och istället för (4.6) får man att

$$\exp_p(-\ell/m)^m = \left(1 - \frac{(\ell/m)}{p}\right)^{pm} = \left(1 - \frac{1}{mk}\right)^{mk\ell} = \left(1 - \frac{1}{q}\right)^{q\ell} = \exp_q(-1)^\ell$$

för $p = \ell k$ och $q = mk$ med $k \in \mathbf{Z}_+$. Genom samma argument som för $r > 0$ bevisar man att

$$\exp(-\ell/m)^m = \exp(-1)^\ell = \exp(1)^{-\ell}$$

där sista likheten följer från 3 i sats 4.4 med $x = 1$ och beviset slutar på samma sätt som för $r > 0$. □

När man ser sats 4.7 förstår man direkt varför vi döpte exponentialfunktion så. Men vad som är ännu mer intressant är att exponentialfunktionen definieras inte bara för rationella tal utan för alla reella tal! Denna punkt återkommer vi till i avsnitt 4.2.

Sats 4.6. *För alla $x, y \in \mathbf{R}$ är*

$$\exp(x + y) = \exp(x)\exp(y). \tag{4.10}$$

Bevis. Med hjälp av sats 3.9, del 1 och sats 4.7 är det enkelt att bevisa (4.10) då $x, y \in \mathbf{Q}$: Då har vi att

$$\exp(x + y) = e^{x+y} = e^x e^y = \exp(x)\exp(y).$$

Men tyvärr är e^{x+y}, e^x och e^y inte definierade om $x, y \in \mathbf{R} \setminus \mathbf{Q}$ så då måste vi jobba mer.

För godtyckliga x och y kan vi uppskatta talen både uppifrån och nedifrån med följder av rationella tal, till exempel med avrundade decimalutvecklingar. Vi väljer två följder av rationella tal, den ena $(A_n)_n \subseteq \mathbf{Q}$ som är en övre begränsning av x och den andra $(a_n)_n \subseteq \mathbf{Q}$ som är en undre begränsning av x så att uppskattningen blir bättre och bättre:

$$a_n \le x \le A_n \quad \text{och} \quad 0 \le A_n - a_n \le \frac{1}{n}. \tag{4.11}$$

för alla $n \in \mathbf{Z}_+$. På samma sätt väljer vi två följder $(B_n)_n \subseteq \mathbf{Q}$ och $(b_n)_n \subseteq \mathbf{Q}$ som uppskattar y uppifrån och nedifrån:

$$b_n \le x \le B_n \quad \text{och} \quad 0 \le B_n - b_n \le \frac{1}{n}. \tag{4.12}$$

för alla $n \in \mathbf{Z}_+$.

Sen betraktar vi $\exp(x + y)/(\exp(x)\exp(y))$. Eftersom exp är en växande funktion (enligt sats 4.4, del 5) medför (4.11) och (4.12) att

$$\frac{\exp(x+y)}{\exp(x)\exp(y)} \le \frac{\exp(A_n + B_n)}{\exp(a_n)\exp(b_n)} \underset{\uparrow\,(4.5)}{=} \frac{e^{A_n+B_n}}{e^{a_n}e^{b_n}} \underset{\uparrow\,\text{sats 3.9}}{=} e^{A_n - a_n + B_n - b_n} \underset{\uparrow\,\text{sats 4.4, del 4}}{\le} \frac{1}{1 - (A_n - a_n + B_n - b_n)} \underset{\uparrow\,(4.11)}{\le} \frac{1}{1 - 2/n}$$

för alla $n \ge 3$. På samma sätt medför (4.11) och (4.12) att

$$\frac{\exp(x+y)}{\exp(x)\exp(y)} \ge \frac{\exp(a_n + b_n)}{\exp(A_n)\exp(B_n)} \underset{\uparrow\,(4.5)}{=} \frac{e^{a_n+b_n}}{e^{A_n}e^{B_n}} \underset{\uparrow\,\text{sats 3.9}}{=} e^{a_n - A_n + b_n - B_n} \underset{\uparrow\,\text{sats 4.4, del 2}}{\ge} 1 + (a_n - A_n + b_n - B_n) \underset{\uparrow\,(4.12)}{\ge} 1 - 2/n$$

för alla $n \in \mathbf{Z}_+$. Alltså drar vi slutsatsen att

$$-\frac{2}{n-2} \le -\frac{2}{n} = (1 - 2/n) - 1 \le \frac{\exp(x+y)}{\exp(x)\exp(y)} - 1 \le \frac{1}{1-2/n} - 1 = \frac{2/n}{1-2/n} = \frac{2}{n-2}$$

för alla $n \geq 3$, så

$$\left| \frac{\exp(x+y)}{\exp(x)\exp(y)} - 1 \right| \leq \frac{2}{n-2}$$

för alla $n \geq 3$ och enligt den arkimediska egenskapen (III(b) i avsnitt 2.3.2) är

$$\left| \frac{\exp(x+y)}{\exp(x)\exp(y)} - 1 \right| = 0$$

så (4.10) är bevisat för godtyckliga $x, y \in \mathbf{R}$. □

4.2 Den naturliga logaritmfunktionen och irrationella potenser

4.2.1 Inversen till exponentialfunktionen

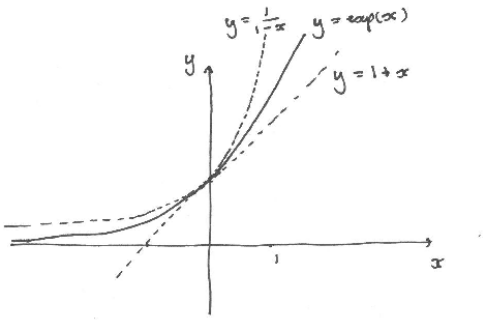
Nu vet vi ganska mycket om exponentialfunktionen. Tack vare sats 4.4 vet vi att den är växande och att grafen sitter ovanför grafen för $x \mapsto 1 + x$ och nedanför grafen för $x \mapsto 1/(1-x)$ om $x < 1$. Det innebär att vi kan skissa en graf för hur funktionen ser ut. Se figur 17. I synnerhet ser vi att exponentialfunktionens värdemängd är intervallet $(0, \infty)$. Utöver del 5 av sats 4.4 är exponentialfunktionen faktiskt strängt växande. (Varför? Tänk på del 2 av sats 4.4 och sats 4.6.) Det medför att exponentialfunktionen är injektiv, så vi kan dra slutsatsen att funktionen är bijektiv. Det vill säga att exponentialfunktionen är inverterbar och vi kallar inversfunktionen för den *naturliga logaritmen*:

$$\ln\colon (0, \infty) \to \mathbf{R}$$

definieras enligt

$$\ln(x) = y \text{ då } y \text{ är den unika lösningen till } \exp(y) = x$$

för varje givet $x \in (0, \infty)$. Varje egenskap hos exponentialfunktionen kan skrivas om som en egenskap hos den naturliga logaritmen.



Figur 17: Exponentialfunktionens graf ligger under grafen av $x \mapsto 1/(1-x)$ för $x < 1$ och över grafen av funktionen $x \mapsto x + 1$ samt x-axeln. Den är också växande.

Sats 4.7. *Den naturliga logaritmen* $\ln\colon (0, \infty) \to \mathbf{R}$ *har följande egenskaper:*

1. $x \mapsto \ln(x)$ *är en växande funktion;*
2. $\ln(e) = 1$*;*
3. $\ln(a) \leq a - 1$ *för alla* $a \in (0, \infty)$*;*
4. $\ln(a) \geq (a-1)/a$ *för alla* $a \in (0, \infty)$*;*

5. $\ln(ab) = \ln(a) + \ln(b)$ *för alla* $a, b \in (0, \infty)$*; och*

6. $\ln(a^r) = r \ln(a)$ *för alla* $a \in (0, \infty)$ *och* $r \in \mathbf{Q}$.

Bevis. 1. Vi vet att exp är växande enligt 5 av sats 4.4. Definitionen av växande säger att $x < y$ medför att $\exp(x) < \exp(y)$. Sätt $\exp(x) = a$ och $\exp(y) = b$. Vi vill bevisa att $a < b$ medför att $x < y$. Anta att $x \geq y$ då antingen är $x = y$ eller $x > y$. Om $x = y$ är $a = b$ eftersom exp är injektiv. Om $x > y$ så är $a > b$ eftersom exp är växande. Därför kan vi dra slutsatsen att $x \geq y$ medför att $a \geq b$ som säger att $a < b$ medför att $x < y$.

2. Eftersom $\exp(1) = e$ så är $\ln(e) = 1$.

3. Betrakta olikheten i del 2 av sats 4.4 och sätt $\exp(x) = a$ (så $x = \ln a$). Då säger den att

$$a \geq 1 + \ln(a)$$

som medför att $\ln(a) \leq a - 1$.

4. Betrakta olikheten i del 4 av sats 4.4 och sätt $\exp(x) = a$ för $x < 1$, så $a < e$. Den säger att

$$a \leq 1/(1 - \ln(a))$$

vilket medför att $(a-1)/a \leq \ln(a)$. Om $a \geq e$ kan man kontrollera direkt att olikheten stämmer: Vi vet att $\ln(a) \geq \ln(e) = 1$ enligt del 1 och $(a-1)/a \leq 1$, så $\ln(a) \geq 1 \geq (a-1)/a$.

5. Betrakta likheten (4.10) hos sats 4.6 och sätt $\exp(x) = a$ och $\exp(y) = b$. Satsen säger att

$$\exp(\ln(a) + \ln(b)) = ab$$

och det medför att $\ln(ab) = \ln(a) + \ln(b)$.

6. Betrakta (4.4) hos sats 4.5 och sätt $\exp(x) = a$. Satsen säger då att

$$\exp(r \ln(a)) = a^r$$

och det medför att $\ln(a^r) = r \ln(a)$. □

4.2.2 Irrationella potenser

Observera att vi kan använda exponentialfunktionen och den naturliga logaritmfunktionen för att skriva om en rationell potens av ett reellt tal: Om vi tillämpar exponentialfunktionen på likheten i sats 4.7, del 6 får vi att

$$a^r = \exp(r \ln(a)) \tag{4.13}$$

för alla $a \geq 0$ och $r \in \mathbf{Q}$. Medan vänsterledet av (4.13) är odefinierat om $r \in \mathbf{R} \setminus \mathbf{Q}$, är högerledet definierat för alla $r \in \mathbf{R}$! Det vill säga att (4.13) ger oss ett smidigt sätt att utöka begreppet a^r till irrationella potenser r. Alltså definierar vi (om)

$$a^x := \exp(x \ln(a)) \tag{4.14}$$

för $a > 0$ och $x \in \mathbf{R}$. Ekvationen (4.13) säkerställer att de två definitionerna (3.12) och (4.14) stämmer överens om $x \in \mathbf{Q}$.

Ännu en gång får vi upprepa sats 2.10, eller närmare bestämt utöka sats 3.9 till irrationella potenser. Men med definitionen (4.14) är beviset mycket enklare än beviset av sats 3.9.

Sats 4.8. *Med givna godtyckliga* $a, b > 0$ *och* $x, y \in \mathbf{R}$ *har vi att*

1. $a^x a^y = a^{x+y}$,

2. $a^x a^{-y} = a^{x-y}$,

3. $a^x b^x = (ab)^x$,

4. $(a^x)^{-1} = a^{-x}$, *och*

5. $(a^x)^y = a^{xy}$.

Bevis. 1. Enligt (4.14) är

$$a^x a^y = \exp(x\ln(a))\exp(y\ln(a)) \underset{\substack{\uparrow \\ \text{sats 4.6}}}{=} \exp(x\ln(a) + y\ln(a)) = \exp((x+y)\ln(a)) = a^{x+y}.$$

2. Enligt (4.14) är

$$a^x a^{-y} = \exp(x\ln(a))\exp(-y\ln(a)) \underset{\substack{\uparrow \\ \text{sats 4.6}}}{=} \exp(x\ln(a) - y\ln(a)) = \exp((x-y)\ln(a)) = a^{x-y}.$$

3. Enligt (4.14) är

$$\begin{aligned} a^x b^x &= \exp(x\ln(a))\exp(x\ln(b)) \underset{\substack{\uparrow \\ \text{sats 4.6}}}{=} \exp(x\ln(a) + x\ln(b)) \\ &= \exp(x(\ln(a) + \ln(b))) \underset{\substack{\uparrow \\ \text{sats 4.7, del 5}}}{=} \exp(x\ln(ab)) = (ab)^x. \end{aligned}$$

4. Enligt (4.14) är

$$(a^x)^{-1} = \exp(-\ln(a^x)) = \exp(-\ln(\exp(x\ln(a)))) = \exp(-(x\ln(a))) = \exp(-x\ln(a)) = a^{-x}.$$

5. Enligt (4.14) är

$$(a^x)^y = \exp(y\ln(a^x)) = \exp(y\ln(\exp(x\ln(a)))) = \exp(y(x\ln(a))) = \exp(xy\ln(a)) = a^{xy}.$$

□

4.3 Komplexa tal

Vi har använt en linje som en modell för de reella talen. Till den har vi lagt axiom som i synnerhet beskriver de aritmetiska operationer man kan utföra med dem. Vi har också utnyttjat planet för att representera par av reella tal, det vill säga element i $\mathbf{R}^2$. Vi har till exempel ritat grafer och andra delmängder av $\mathbf{R}^2$ med hjälp av koordinataxlarna (se avsnitt 2.5.3). Men vi har ännu inte tänkt på hur de aritmetiska operationerna kunde se ut när de utförs med punkter i planet istället för punkter på en linje.

Därför ställer vi en fråga:

Finns det ett sätt att utöka aritmetiska operationer från den reella linjen till planet? (4.15)

För att vara mer exakt, betraktar vi den reella linjen som delmängden

$$\{(x, y) \mid x \in \mathbf{R} \text{ och } y = 0\} \tag{4.16}$$

i planet $\mathbf{R}^2$ (se figur 18), så punkten $(x, 0) \in \mathbf{R}^2$ representerar de reella talen $x \in \mathbf{R}$. I (4.15) frågar vi om det är möjligt att definiera aritmetiska operationer (addition och multiplikation) för element i $\mathbf{R}^2$ så att:

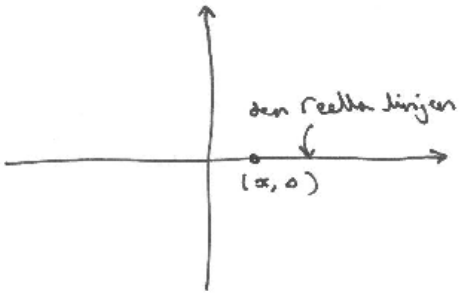
1. Definitionerna stämmer överens med de vanliga aritmetiska operationerna med reella tal för delmängden (4.16);

2. Definitionerna medför att axiomen I(a)–I(e) i avsnitt 2.2.1 gäller för godtyckliga $a, b, c \in \mathbf{R}^2$.[21]

I synnerhet I(e) säger att de neutrala elementen $0, 1 \in \mathbf{R}$, nu skrivna som $(0,0), (1,0) \in \mathbf{R}^2$, måste uppfylla

$$\begin{aligned}(x,y) + (0,0) &= (x,y) \quad \text{och} \\ (x,y)(1,0) &= (x,y)\end{aligned} \tag{4.17}$$

för alla $(x, y) \in \mathbf{R}^2$.



Figur 18: Vi identifierar x-axeln med den reella linjen och varje reellt tal x med punkten $(x, 0)$.

4.3.1 Kartesiska koordinater och addition i planet

Observera att när vi skriver om hur addition och multiplikation av reella tal sedd som mängden (4.16) får vi

$$\begin{aligned}(x,0) + (a,0) &:= (x+a,0) \quad \text{och} \\ (x,0)(a,0) &:= (xa,0)\end{aligned}$$

för alla reella $x, a \in \mathbf{R}$. Därför är det naturligt att gissa definitionen

$$(x,y) + (u,v) := (x+u, y+v) \tag{4.18}$$

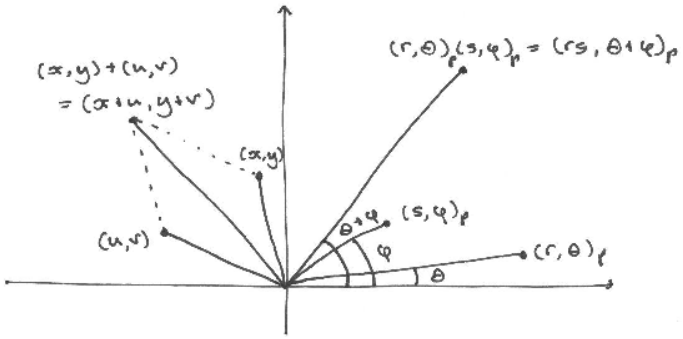
för addition i $\mathbf{R}^2$. Se figur 19. I (4.18) är additionen i $\mathbf{R}^2$ definierad med hänsyn till de aritmetiska operationerna i $\mathbf{R}$. Därför är det möjligt att verifiera axiom i $\mathbf{R}^2$ under hypotesen att de stämmer i $\mathbf{R}$. Med valet (4.18) kan vi snabbt kontrollera att alla axiom som bara gäller addition stämmer:

$$(x,y) + (u,v) = (x+u, y+v) = (u+x, v+y) = (u,v) + (x,y)$$

och

$$\begin{aligned}((x,y) + (u,v)) + (r,s) &= (x+u, y+v) + (r,s) \\ &= (u+x+r, v+y+s) \\ &= (x,y) + (u+r, v+s) = (x,y) + ((u,v) + (r,s)),\end{aligned}$$

[21] Vi insisterar inte på att man kan jämföra två element i $\mathbf{R}^2$. Det vill säga att vi inte letar efter någon utökning av begreppet olikhet som redan gäller för reella tal. Man kan faktiskt bevisa att det är omöjligt att definiera olikhet i det här sammanhanget, men det ligger utanför kursen. I synnerhet axiomen II(a)–II(d) i avsnitt 2.2.1 och axiom III(a)–III(b) i avsnitt 2.3.2 gäller inte här.



Figur 19: På vänstersidan finns en illustration av addition i planet. På högersidan finns en illustration av multiplikation i planet.

samt

$$(x, y) + (0, 0) = (x + 0, y + 0) = (x, y)$$

för alla $(x, y), (u, v) \in \mathbf{R}^2$, så att det som gäller addition i I(a), I(b) och I(d) stämmer. Det finns också ett additivt inverst element: Eftersom

$$(x, y) + (-x, -y) = (x - x, y - y) = (0, 0)$$

kallar vi $-(x, y) := (-x, -y)$ och därmed stämmer det som gäller addition i I(e).

4.3.2 Polära koordinater och multiplikation i planet

Man skulle kunna tänka på liknande sätt när det gäller multiplikation: Vi skulle gissa att

$$(x, y)(u, v) = (xu, yv)$$

är en definition för multiplikation. Men tyvärr går det inte eftersom då är $(x, y)(1, 0) = (x, 0)$ som betyder att (4.17) inte stämmer för $y \neq 0$. Så vi måste fundera lite noggrannare på det.

Inledningvis betraktar vi ett nytt koordinatsystem, som kallas för det *polära koordinatsystemet*. Varje punkt identifieras med hjälp av två tal, precis som i det kartesiska koordinatsystemet, men här är första talet avståndet r från punkten till origo och det andra talet θ är vinkeln mellan x-axeln och strecken mellan punkten och origo. Se figur 20. Vi skriver

$$(r, \theta)_p$$

för att markera att koordinaterna är polära och inte kartesiska. Likheterna

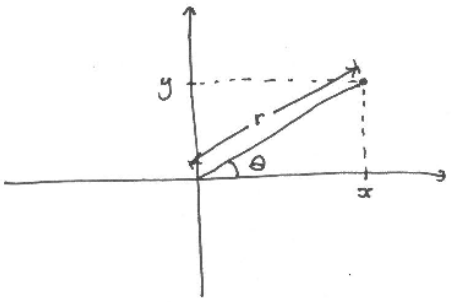
$$x = r \cos\theta \quad \text{och} \quad y = r \sin\theta$$

ger sambandet mellan en punkts kartesiska koordinater (x, y) och polära koordinater $(r, \theta)_p$.

Till skillnad från det kartesiska koordinatsystemet finns det flera sätt att anteckna samma punkt, till exempel

$$(0, 0)_p = (0, \theta)_p \quad \text{och} \quad (r, \theta)_p = (r, \theta + 2n\pi)_p$$

för alla $\theta \in \mathbf{R}$ och $n \in \mathbf{Z}$.



Figur 20: En punkt i planet har de kartesiska koordinaterna (x, y) och de polära koordinaterna $(r, \theta)_p$.

Nu går vi tillbaka till multiplikation i planet. Observera att multiplikation för den reella linjen i polära koordinater ser ut som

$$\begin{aligned}(r,0)_p(s,0)_p &= (rs,0)_p,\\ (r,0)_p(s,\pi)_p &= (rs,\pi)_p,\\ (r,\pi)_p(s,0)_p &= (rs,\pi)_p, \quad \text{respektive}\\ (r,\pi)_p(s,\pi)_p &= (rs,2\pi)_p\end{aligned}$$

för två positiva reella tal $(r,0)_p$ och $(s,0)_p$, ett positivt tal $(r,0)_p$ och ett negativt tal $(s,\pi)_p$, ett negativt tal $(r,0)_p$ och ett positivt tal $(s,\pi)_p$, respektive två negativa tal $(r,\pi)_p$ och $(s,\pi)_p$. Det leder till att vi kan gissa definitionen

$$(r,\theta)_p(s,\varphi)_p := (rs,\theta+\varphi)_p$$

för multiplikation i planet. Se figur 19. För att kontrollera att axiomen stämmer för den här multiplikationen är det praktiskt att först räkna ut hur den ser ut i kartesiska koordinater. Betrakta godtyckliga punkter $(x,y) = (r,\theta)_p$ och $(u,v) = (s,\varphi)_p$. Då är

$$\begin{aligned}(x,y)(u,v) &= (r\cos\theta, r\sin\theta)(s\cos\varphi, s\sin\varphi)\\ &= (rs\cos(\theta+\varphi), rs\sin(\theta+\varphi))\\ &= ((r\cos\theta)(s\cos\varphi) - (r\sin\theta)(s\sin\varphi), (r\cos\theta)(s\sin\varphi) + (r\sin\theta)(s\cos\varphi))\\ &\quad \uparrow \text{ (3.22) och (3.23)}\\ &= (xu - yv, xv + yu).\end{aligned} \tag{4.19}$$

Utifrån (4.19) kan vi kontrollera att alla återstående axiom stämmer:

$$(x,y)(u,v) = (xu - yv, xv + yu) = (ux - vy, vx + uy) = (u,v)(x,y)$$

och

$$\begin{aligned}((x,y)(u,v))(r,s) &= (xu - yv, xv + yu)(r,s)\\ &= ((xu - yv)r - (xv + yu)s, (xv + yu)r + (xu - yv)s)\\ &= (x(ur - vs) - y(us + vr), y(ur - vs) + x(us + vr))\\ &= (x,y)(ur - vs, us + vr) = (x,y)((u,v)(r,s)),\end{aligned}$$

samt

$$(x,y)(1,0) = (1x - 0y, 0x + 1y) = (x,y)$$

för alla $(x,y),(u,v) \in \mathbf{R}^2$, så vad gäller multiplikation i I(a), I(b) och I(d) stämmer det. Dessutom kan vi också kontrollera att

$$\begin{aligned}(x,y)((u,v)+(r,s)) &= (x,y)(u+r,v+s)\\ &= (x(u+r)-y(v+s),x(v+s)+y(u+r))\\ &= ((xu-yv)+(xr-ys),(xv+yu)+(xs+yr))\\ &= (xu-yv,xv+yu)+(xr-ys,xs+yr)\\ &= (x,y)(u,v)+(x,y)(r,s),\end{aligned}$$

så I(c) stämmer. Det enda axiom som står kvar är existensen av en multiplikativ invers i I(e). Vi senarelägger till (4.21) kontroll av den så att vi kan införa lite notation. Kom ihåg från (4.19) att vi identifierade punkten $(1,0)$ i planet med det reella talet 1. Nu identifierar vi också $(0,1)$ med bokstaven i. Eftersom

$$(x,y) = (x,0)(1,0)+(y,0)(0,1)$$

för alla $x,y \in \mathbf{R}$ kan vi identifiera

$$\text{en punkt i planet } (x,y) \text{ med } x+iy.$$

Med denna notation är all aritmetisk räkning enklare att komma ihåg. Nu är aritmetisk räkning precis det samma som räkning med reella tal eftersom talsystemet uppfyller samma axiom. Om vi bara kommer ihåg att

$$i^2 = -1 \quad \text{eftersom enligt (4.19) är} \quad (0,1)(0,1) = (-1,0)$$

är det lätt att se

$$\begin{aligned}(x+iy)+(u+iv) &= (u+v)+i(y+v) \quad \text{och}\\ (x+iy)(u+iv) &= ux+xvi+yui+yvi^2\\ &= ux+xvi+yui+yv(-1) = (ux-yv)+(xv+yu)i\end{aligned} \tag{4.20}$$

precis som sagt i (4.18) och (4.19) med tyngre notation. Vi kallar $\mathbf{R}^2$ rustad med notationen $(1,0)=1$ och $(0,1)=i$, och additions- och multiplikationsreglar (4.20) för det *komplexa planet*. Det betecknas

$$\mathbf{C} := \{x+iy \,|\, x,y \in \mathbf{R}\}.$$

4.3.3 Aritmetik med komplexa tal

Betrakta figur 21. För $z = x+iy \in \mathbf{C}$ definierar vi *konjugatet* till z att vara

$$\overline{z} = x-iy$$

och *absolutbeloppet*

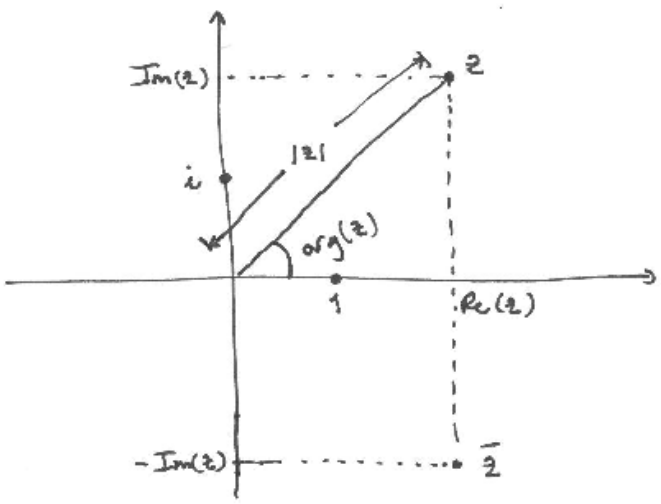
$$|z| = \sqrt{x^2+y^2}.$$

Realdelen av z är $\Re(z) := x$ och $\Im(z) := y$ kallas för *imaginärdelen* av z. Om $\Re(z) := 0$ kallas z för *imaginärt*. *Argumentet* av z definieras som en vinkel $\arg(z) := \theta$ som uppfyller

$$\tan\theta = \frac{b}{a}.$$

Observera att

$$z\overline{z} = (x+iy)(x-iy) = x^2+y^2 = |z|^2.$$



Figur 21: Ett par begrepp för komplexa tal: konjugatet, absolutbeloppet, realdelen, imaginärdelen och argumentet, samt talen 1 och i i det komplexa planet.

Om $z \neq 0$ medför det att

$$z\frac{\overline{z}}{|z|^2} = 1 \tag{4.21}$$

så vi har hittat en multiplikativ invers

$$z^{-1} = \frac{\overline{z}}{|z|^2} = \frac{x - iy}{x^2 + y^2}$$

till z för alla $z \neq 0$ och I(e) är äntligen bekräftat.

Sats 4.9. *För alla komplexa tal z och w har vi att*

1. $\overline{zw} = \overline{z}\,\overline{w}$,
2. $\overline{z + w} = \overline{z} + \overline{w}$,
3. $|zw| = |z||w|$,
4. $|z + w| \leq |z| + |w|$ *och*
5. $\arg(zw) = \arg(z) + \arg(w)$.

Som sagt i avsnitt 4.3.2 följer alla beräkningar som gäller addition, multiplikation, subtraktion och division samma reglar som gäller för reella tal eftersom komplexa tal uppfyller samma axiom I(a)–I(e). Dessutom definierar vi heltalspotsener för komplexa tal på samma sätt. Det vill säga att (2.12) och (2.13) gäller för $a \in \mathbf{C}$, samt sats 2.10 för $a, b \in \mathbf{C}$.

Men kom ihåg att kvadratrötter är någonting utöver aritmetik så vi måste återvända till frågan i samband med komplexa tal. Det går dock lite fortare den här gången eftersom vi kan dra nytta av allt arbete vi har gjort med reella tal. Betrakta likheten

$$z^2 = w \tag{4.22}$$

för $z, w \in \mathbf{C}$. För ett givet $w \in \mathbf{C}$ vill vi hitta alla lösningar $z \in \mathbf{C}$ till (4.22). Man kan skriva $z = x + iy$ och $w = u + iv$ för $x, y, u, v \in \mathbf{R}$ och då kommer (4.22) att vara

$$x^2 - y^2 + 2xyi = (x + iy)^2 = u + iv.$$

Därför är

$$x^2 - y^2 = u \quad \text{och} \tag{4.23}$$
$$2xy = v \tag{4.24}$$

Man kan förstås lösa de två ekvationerna för x och y men det är lite smidigare att också ta absolutbeloppet av (4.22). Därav får man likheten

$$x^2 + y^2 = \sqrt{u^2 + v^2} \tag{4.25}$$

så från (4.23) och (4.25) får man att

$$x^2 = \frac{u + \sqrt{u^2 + v^2}}{2} \quad \text{och}$$
$$y^2 = \frac{\sqrt{u^2 + v^2} - u}{2}.$$

Observera att högerleden är alltid definierade och aldrig negativa. Det medför att vi har (högst) två möjliga x och också (högst) två möjliga y:

$$x = \pm\sqrt{\frac{u + \sqrt{u^2 + v^2}}{2}} \quad \text{and} \quad y = \pm\sqrt{\frac{\sqrt{u^2 + v^2} - u}{2}}.$$

Vid första anblicken ser det ut som det finns totalt fyra möjliga lösningar $z = x + iy$ till (4.22), två för x och två för y. Men tänk på vad (4.24) säger om tecknen av x och y: Om $v > 0$ så är både x och y positiva eller negativa; Om $v < 0$ så är ett av x och y positivt och det andra negativt. Det medför att vi har (högst) två lösningar till (4.22). Formeln för lösningarna är

$$z = \begin{cases} \pm\left(\sqrt{\frac{u+\sqrt{u^2+v^2}}{2}} - i\sqrt{\frac{\sqrt{u^2+v^2}-u}{2}}\right) & \text{om } v < 0; \\ \pm\left(\sqrt{\frac{u+\sqrt{u^2+v^2}}{2}} + i\sqrt{\frac{\sqrt{u^2+v^2}-u}{2}}\right) & \text{om } v \geq 0. \end{cases}$$

4.3.4 Komplexa exponentialfunktionen

Här vill vi motivera en definition för den komplexa exponentialfunktionen. Kom ihåg från avsnitt 4.1.2 att definitionen av $\exp(x)$ för reella tal x var supremum av uttrycket

$$\left(1 + \frac{x}{n}\right)^n$$

för $n > |x|$. Eftersom vi nu också kan betrakta komplexa tal verkar det rimligt att bara sätta in ett komplext tal istället för x och plötsligt har vi utökat exponentialfunktionen till komplexa tal. Problemet med den där idén är att vi inte har något begrepp om supremum över komplexa tal. Begreppet som vi behöver är istället *gränsvärde*. Här kommer vi inte att ge en precis definition men ett *gränsvärde* är det tal som elementen i en följd närmar sig längs följden. I avsnitt 4.1.2 kunde vi ha jobbat med gränsvärden istället för suprema och den skulle fungera för komplexa

tal också. Men eftersom gränsvärden hör till nästa kurs gör vi bara en liten handviftning här för att motivera en trigonometrisk definition.

Betrakta ett imaginärt tal $i\theta$ ($\theta \in \mathbf{R}$). Vi vill räkna ut hur

$$z_n := \left(1 + \frac{i\theta}{n}\right)^n \tag{4.26}$$

ser ut för jättestora $n \in \mathbf{Z}_+$.

Först uppskattar vi absolutbellopet $|z_n|$ underifrån:

$$|z_n| = \left|\left(1 + \frac{i\theta}{n}\right)^n\right| \underset{\text{sats 4.9, del 3}}{\uparrow}{=} \left|1 + \frac{i\theta}{n}\right|^n = \left(1 + \frac{\theta^2}{n^2}\right)^{n/2} \underset{\text{sats 4.1}}{\uparrow}{\geq} 1 + \frac{\theta^2}{2n} \geq 1$$

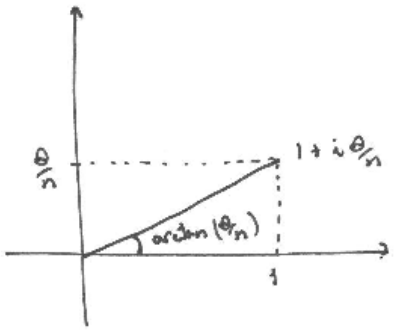
för alla jämna $n > |\theta|$. Sen uppskattar vi det uppifrån:

$$|z_n| = \left(1 + \frac{\theta^2}{n^2}\right)^{n/2} = \left(1 + \frac{\theta^2}{n^2}\right)^{n^2/(2n)} \underset{(4.2)}{\uparrow}{=} \exp_n(\theta^2)^{1/(2n)} \underset{(4.3)}{\uparrow}{\leq} \exp(\theta^2)^{1/(2n)} = \sqrt[2n]{\exp(\theta^2)}$$

för alla $n > |\theta|$. Man kan tänka sig att ju större n är, desto närmare den $(2n)$:e roten av $\exp(\theta^2)$ blir till 1. Därför är det rimligt att behöva

$$|e^{i\theta}| = 1. \tag{4.27}$$

Nu uppskattar vi argumentet i (4.26). Se figur 22. Tack vare sats 3.10 är $\arctan\theta \leq \theta$ för



Figur 22: Det komplexa talet $1 + i\theta/n$ har absolutbeloppet $\sqrt{1 + \theta^2/n^2}$ och argumentet $\arctan(\theta/n)$.

$\theta \in [0, \pi/2)$. Därför kan vi uppskatta $\arg(z_n)$ uppifrån för tillräckliga stora n:

$$\arg(z_n) \underset{\text{sats 4.9, del 5}}{\uparrow}{=} n \arg(1 + i\theta/n) = n \arctan(\theta/n) \leq \theta.$$

Man kan också tänka sig att $\arctan(\theta/n)$ är bara så lite mindre än θ/n för stort n, så att det är rimligt att behöva

$$\arg(e^{i\theta}) = \theta. \tag{4.28}$$

Med motivationen för (4.27) och (4.28) i åtanke definierar vi

$$e^{i\theta} := \cos(\theta) + i\sin(\theta) \tag{4.29}$$

för $\theta \in \mathbf{R}$. Formel (4.29) kallas för *Eulers formel* och leder till den följande representationen för komplexa tal i polär form:

$$z = r\cos\theta + ir\sin\theta = re^{i\theta}$$

där $r = |z|$ och $\theta = \arg(\theta)$.

Man kan sätta ett komplext tal $z = x + iy$ i (4.26) istället för $i\theta$ för att motivera definitionen

$$e^z := e^x e^{iy} = e^x(\cos(y) + i\sin(y))$$

för alla $z = x + iy \in \mathbf{C}$ $(x, y \in \mathbf{R})$. Återigen, eftersom komplexa tal följer samma aritmetiska axiom som reella tal, gäller satserna 4.5 och 4.6 för $x, y \in \mathbf{C}$ (och $\exp(z) = e^z$ för $z \in \mathbf{C}$). Med de här definitionerna är följande sats enkel att bevisa.

Sats 4.10. *För alla $\theta \in \mathbf{R}$ är*

1. $e^{i\pi} + 1 = 0$ *(Eulers identitet),*
2. $(\cos\theta + i\sin\theta)^n = \cos(n\theta) + i\sin(n\theta)$ *(de Moivres formel),*
3. $\cos\theta = (e^{i\theta} + e^{-i\theta})/2$*, och*
4. $\sin\theta = (e^{i\theta} - e^{-i\theta})/2i$*.*

Eulers identitet (del 1) är särdeles vackert eftersom den innehåller tre operationer som är väsentliga för vår resa—addition, multiplikation och exponentiering—samt fem tal med mycket olika ursprung—det additionsneutrala elementet 0, det multiplikationsneutrala elementet 1, den halva enhetscirklens längd π, den naturliga basen e och den komplexa enheten i.

Referenser

[1] A. Asratian, B.O. Turesson, A. Björn, *Diskret matematik*, Kompendium, Linköpings universitet, 2013.

[2] Carl B. Boyer & Uta C. Merzbach, *A History of Mathematics*, third edition, John Wiley & Sons, New Jersey 2011.

[3] I.M. Copi and C. Cohen, *Introduction to Logic*, trettonde upplagan, Pearson Education, New Jersey 2009.

[4] G. Forsling and M. Neymark, *Matematisk analys: En variabel*, andra upplagan, Liber AB, Stockholm 2011.

[5] G.H. Hardy, *A Mathematician's Apology*, Cambridge University Press, Canto Ed. 1992.

[6] C. Hyltén-Cavallius and L. Sandgren, *Matematisk analys II*, Håken Ohlssons Boktryckeri, Lund 1964.

[7] Robert Kaplan, *The Nothing That Is: A Natural History of Zero*, Oxford University Press, Oxford 1999.

[8] Victor J. Katz, *A History of Mathematics: An introduction*, HarperCollins College Publishers, New York 1993.

[9] Henrik Petersson, *Undersökande matematik: Differentierade problem*, Studentliteratur AB, Lund 2017.

[10] S. Stolt, *Geometri: euklidisk och icke-euklidisk*, Berlingska Boktryckeri, Lund 1968.

[11] M. Löwing, *Grundläggande geometri*, Studentlitteratur AB, Lund, 1964.