Ge svar på tal

En introduktion till matematisk analys

DAVID RULE

Studentlitteratur

Art.nr 43524-SB
ISBN 978-91-44-17914-8
Upplaga 1:1

studentlitteratur.se
Studentlitteratur AB, Lund

Omslagslayout: Jens Martin
Omslagsbild: Trustees of the British Museum/David Einar

INNEHÅLL

Förord 7

Till studenten 7
Till läraren 8
Omslaget 10
Tack 10

DEL I **Verktyg, notation och begrepp**

KAPITEL 1 **Logik och aritmetik 13**

1.1 Logik och räkneoperationer 13
Logik 13
Hur man skriver argument inom matematik 18
1.2 Upprepade aritmetiska operationer 26
Heltalspotenser 26
Summor 28
Talbeteckningssystem för heltal 35
Talbeteckningssystem för icke-heltal 37
1.3 Sammanfattning och uppgifter 41

KAPITEL 2 **Verktyg för bevisföring 43**

2.1 Mängder, följder och induktion 43
Mängder 43
Följder och induktion 47
Kombinatorik, Binomialsatsen och andra summor 53
2.2 Reella tal 56
Egenskaper hos delmängder av reella tal 57
Reella tal som oändliga decimalutvecklingar 66
2.3 Sammanfattning och uppgifter 81

KAPITEL 3 **Funktioner och former 87**

3.1 Funktioner: begrepp och egenskaper 87
Funktioner, definitionsmängder och värdemängder 87
Koordinatsystem och grafer 90
Monotonicitet 91
Polynom 94
3.2 Former, längd och area 100
Linjer, bågar och vinklar 101
Trianglar, fyrhörningar och area 104
Pythagoras sats 110
3.3 Sammanfattning och uppgifter 112

DEL II **Elementära funktioner, trigonometri och komplexa tal**

KAPITEL 4 **Kvadratrötter och andra inversa funktioner 117**

4.1 Inversa funktioner och irrationella tal 117
Inversa funktioner 117
Irrationella tal och en följd för potensfunktioner 121
4.2 En ny algebraisk operation: rot av tal 124
Kvadratrötter 124
Andragradspolynom 128
Rationella potenser 130
4.3 Sammanfattning och uppgifter 134

KAPITEL 5 **Trigonometri 137**

5.1 Definitioner och formler 137
Definitioner 137
Trigonometriska formler 138
5.2 Olikheter och arcusfunktioner 143
Trigonometriska olikheter 143
Arcusfunktioner 144
5.3 Sammanfattning och uppgifter 148

KAPITEL 6 **Den naturliga exponentialfunktionen och dess invers 151**

6.1 Exponentialfunktion 151
Ränta på ränta 151
Exponentialfunktionen 155
6.2 Den naturliga logaritmfunktionen och irrationella potenser 171
Inversen till exponentialfunktionen 171
Irrationella potenser 175
6.3 Sammanfattning och uppgifter 177

KAPITEL 7 **Komplexa tal 179**

7.1 Algebraiska operationer 179
Aritmetik med komplexa tal 180
Andragradspolynom och komplexa tal 184
7.2 Funktioner och geometri 187
Komplexa exponentialfunktionen 187
Komplex geometri 193
Avbildningar av komplexa funktioner 197
7.3 Sammanfattning och uppgifter 200

BILAGA A **Lösningsförslag 203**

BILAGA B **Det grekiska alfabetet 235**

BILAGA C **Ytterligare material 237**

C.1 Ett axiomsystem för de reella talen 237
C.2 En mängd utan den arkimediska egenskapen 240
C.3 Kurvor och längd 241
C.4 Motivation bakom komplexa tal 243

Litteraturförteckning 249

Sakregister 251

FÖRORD

Go, little book, go, little tragedy!…
So pray I God, that none miswrite thee,
Nor thee mismetre for default of tongue!
And read whereso thou be, or elles sung,
That thou be understanden, God I 'seech!
But yet to purpose of my rather speech.

Geoffrey Chaucer [4, s. 722]

Till studenten

Syftet med den här boken är att ge en grundläggande introduktion till matematisk analys. Du har kanske redan hunnit läsa innehållsförteckningen och med all sannolikhet märkt att du känner till det mesta av vad som finns där. Det kan vara vilseledande: boken är inte en repetition av det du har läst tidigare. Däremot kan den erfarenhet och intuition du har byggt upp vara till stor nytta. Så vad är då skillnaden?

I grundskolan och gymnasiet har du koncentrerat dig på metoder, det vill säga, på hur man kan komma fram till de rätta svaren till numeriska frågor. Mindre betoning låg på varför en viss metod fungerar eller hur man kommer fram till en viss metod från början. En sådan infallsvinkel är nödvändig när man börjar läsa matematik. Barn lär sig genom exempel som har sin grund i deras närmaste omgivning. Deras logiska tänkande motsvarar hur de tänker när de lär sig ett språk: de lyssnar, repeterar och bygger upp en intuition som fungerar bra i vardagslivet. Det kallas för *induktiv slutledning*.

Men i grunden är matematik inget språk. Matematik är en övning i *deduktiv slutledning*. Det krävs att vi preciserar begreppen vi pratar om så gott vi kan, att vi är tydliga med vilka regler som gäller och i vilka sammanhang, och sedan bevisar slutsatserna utifrån dessa begrepp och regler. Det ska vi försöka göra i den här boken.

Det är ingen enkel uppgift, men det är absolut värt mödan. Du vet säkert redan att matematik används i en rad olika ämnen för att både formulera teorier och förstå följderna av teorierna. Men att koncentrera sig på deduktiv slutledning är att sätta tankeprocessen i första rummet. På 1600-talet utvecklades matematisk analys för att lösa problem som uppstod under den naturvetenskapliga revolutionen. Den matematiska analysen är fortfarande mycket användbar, och att förstå den på ett grundläggande sätt innebär att man även kan bygga matematiska verktyg som kan lösa de problem som uppstår idag – problem som ingen har löst tidigare.

Som en bonus kommer vi att se att matematik är ett både vackert och kreativt ämne. Det är ett under att så mycket struktur finns i den matematiska världen och att vi kan härleda den med enbart våra tankar. Att kunna sluta sig till vissa matematiska fakta är inte alltid enkelt och långt ifrån naturligt för oss människor som i grunden är intuitiva och inte logiska varelser. För detta krävs både disciplin och kreativitet i samma utsträckning.

Till sist vill jag säga ett par allmänna ord om hur man läser matematik. Ingen kan läsa och förstå matematisk text med samma hastighet som man läser en roman. Informationen i texten är vanligtvis mycket mer kompakt än vanlig löpande text, och det är inte ovanligt att man behöver kontrollera vad som sägs i texten med egna beräkningar och anteckningar, så var medveten om det, när du planerar dina studier. Man behöver inte heller bli frustrerad när man kör fast. Visst vill man förstå det man inte förstår, men den tiden man är fast är inget slöseri. På lång sikt drar man nytta av att kämpa med idéer och koncept, och det är i den kampen som man lär sig på riktigt. Prioritera förståelse före mängdträning. Mängdträning är dåligt anpassad för eftertanke och reflektion och det är genom reflektion som man blir en bättre problemlösare. Det är även en vanlig missuppfattning att man antingen har en begåvning för matematik eller inte. Det kan jag direkt säga är falskt: alla – inklusive du – kan förbättra sig genom att spendera tid med ämnet.

Till läraren

Under några decennier har lärare kämpat med demografiska förändringar i studentkohorten. En oundviklig följd av att flera läser kurser på universitetsnivå är att vi ser en större variation i studenternas förkunskaper. Alla lärare måste förstås hitta sin eget sätt att bemöta den utmaningen.

En vanlig reaktion från lärare är att ta bort innehåll såsom bevis och mer abstrakta moment från sina kurser eftersom det kan upplevas som onödigt för att studenterna ska kunna svara på samma uppgifter och skriva samma tentor. Många lärare delar upp sina kurser i teori och tillämpningar och ser en minskning av den teoretiska delen som en nödvändig kompromiss för den nya situationen. Men jag är inte bekväm med den idén. Jag har svårt att se en uppdelning där en sats är teori och en följdsats är tillämpning som hjälpsamt. Är en historia mer gripande eller lärorik om man bara berättar slutet? Den här boken är mitt försök att hitta ett alternativ.

Jag minns själv från när jag var student att jag märkte en stor skillnad mellan matematik i skolan och matematik på universitetsnivå. Resonemanget var plötsligt centralt. Begrepp blev mer än bara ord och hade ibland konstiga tekniska definitioner. Jag såg för första gången logik som var klockren att följa men ändå kreativ och magisk. Nu förstår jag att det jag egentligen upplevde var paradigmskiftet från induktiv till deduktiv slutledning. Så småningom lärde jag mig att det var deduktiv slutledning som möjliggjorde att jag kunde lösa problem, även många problem som inte hade någonting direkt med matematik att göra. Så det är framför allt det som jag vill betona i boken. Jag är förstås inte den första att inse detta. E. Nardi [12] beskriver det så här:

> This transition signifies the quintessential cognitive shift that learners are expected to undertake in the first few months of university studies. Therefore in a sense, and given their secondary school mathematical background, these students' transition from school to university mathematics stands as a metaphor for the transition from concreteness to abstraction and from empirical/inductive explanation to deductive proof and axiomatics.

Den här boken är inte bara en introduktion till matematisk analys, utan den är även en introduktion till deduktiv slutledning. Vad är samspelet mellan intuition och den formella behandlingen? Skiljer sig orden inom matematik från vardagligt språk? Hur kommunicerar vi matematiska argument? Alla behöver arbeta med sådana frågor och framför allt i början av texten har jag diskuterat dem explicit. Jag tror att studenter på lång sikt gynnas av dessa diskussioner och blir mer självständiga problemlösare. För att kunna göra det har jag inte kunnat behandla lika mycket material som jag annars skulle ha gjort, men det vi diskuterar är djupare diskuterat.

Många studenter förstår att $x > 5 \implies x^2 > 25$ fast inte att $x > 5 \implies x^2 \geq 25$. Eller skriver de 4^{3^2} när de menar $(4^3)^2 = 4^6$ och inte $4^{(3^2)} = 4^9$. Men deras misstag brukar inte sakna motivering.

– Men jag vet att x^2 *inte* kan vara tjugofem!

– Man skriver $2 + 3 + 4$ och $2 \times 3 \times 4$, eller? Varför inte 4^{3^2}?

Det viktiga är att vi diskuterar det som vi gamlingar fick lista ut själva. Vi måste göra det som har varit implicit i undervisningen explicit.

Jag undviker mallar för till exempel hur man redovisar induktionsbevis. Det är bättre att studenterna ser många olika beskrivningar av samma logik och sedan skriver induktionsbevis med egna ord. På så sätt har studenterna en större chans att förstår budskapet och läraren har en bättre chans att tidigt fånga upp missförstånd. Jag hoppas att du som lärare lyckas övertyga dina studenter att reflektera över sådana frågor och att boken blir ett bra stöd i det arbetet.

Omslaget

På omslaget visas en del av Rhindpapyrusen [15] som är en fornegyptisk text från 1500-talet f.Kr. skriven i hieratisk skrift. Den är en kopia av en förlorad äldre text. Hieratisk skrift härrör från samma hieroglyfskrift som vi tittar närmare på på sidan 35 men gick fortare att skriva. Det som syns på omslaget är de så kallade uppgifterna 50–52. De förklarar hur man kan beräkna arean av olika slags fält: en cirkel, en triangel och en trunkerad triangel (jämför med sats 3.15 och 3.16). [14]

Tack

Till sist vill jag tacka alla studenter och kollegor som har hjälpt till att förbättra texten. Speciellt vill jag tacka Ingela Dellby, Sofie Abrahamson, Ulf Janfalk och Jesper Thorén för deras hjälp med texten, Åke Engsheden, Nils Billing och Wolfgang Staubach för deras stöd med omslaget och Arpan Ghosh för många intressanta diskussioner om matematik. Jag står förstås också i skuld till min fru, Elizabeth, som väljer att dela sitt liv med mig och pojkarna och som öppnat mina ögon för omvärlden.

Linköping, juni 2021

David Rule

DEL I

Verktyg, notation och begrepp

Logik och aritmetik

1.1 Logik och räkneoperationer

Framför allt är matematik ett sätt att tänka. Det 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 bli fel. Även om det är svårare att komma i gång på detta sätt kommer man längre än man skulle kunna tro.

LOGIK

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

Jag har en dotter; (1.1)

Köpenhamn är Danmarks huvudstad; eller (1.2)

Köpenhamn ligger i Tanzania. (1.3)

Vi vet alla att (1.2) är sant och (1.3) är falskt. Sanningen av (1.1) 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* eller *axiom*) så kan man använda deduktiva slutledningar för att bygga nya påståenden (så kallade *slutsatser*) som också är sanna när antagandena är sanna:

Sveriges riksdag ligger i Sveriges huvudstad.

Sveriges huvudstad kallas för Stockholm.

Därför ligger Sveriges riksdag i Stockholm.

Men vi måste vara försiktiga så att vi kan 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 vad vi tror ordet ”Paris” syftar på. I första påståendet menar vi Frankrikes huvudstad, men i det andra påståendet pratar vi om Paris, Illinois. Båda antagandena är sanna men lätta att misstolka.

Från enkla påståenden kan man bygga mer komplicerade påståenden. Man kan lägga ihop två påståenden 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 alternativ kan man använda ”eller”:

Jag äter potatis eller ris till middag. (1.4)

Om det finns en härledning av en slutsats B från ett antagande A, kan vi säger att A medför B. Påståendet ”A medför B” kallas för en *implikation* och kan skrivas ”$A \implies B$”. Observera att det är inte det samma som ”B medför A”: Om jag har en syster har jag ett syskon, men om man endast vet att jag har ett syskon får man inte dra slutsatsen att jag har en syster – mitt syskon kan vara en bror. Om $A \implies B$ och $B \implies A$ säger vi att A och B är *ekvivalenta*: Man kan säga ”A om och endast om B” eller ”A är ekvivalent med B” och skriver ”$A \iff B$”. Till exempel är påståendet ”B medförs av A” ekvivalent med ”A medför B” (och är därför också en implikation). Vi kan även säga ”B om A” för att säga ”B medförs av A” och skriver ”$B \impliedby A$”.

Negation är ett förnekande av ett påstående. Till exempel, negationen av (1.3) är:

Köpenhamn ligger inte i Tanzania.

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

$$A = \text{"Varje land har en flagga."}$$
$$\neg A = \text{"Det finns ett land som inte har någon flagga."}$$

Ibland när vi försöker förneka ett påstående ser vi att det inte var så tydligt som vi trodde. Till exempel negationen av (1.1) kunde vara:

Jag har ingen dotter; (1.5)

eller,

Jag har antingen ingen dotter eller mer än en dotter.

Rent logiskt är (1.5) negationen av (1.1). Men förvirring kan uppstå på grund av hur vi brukar tolka saker som (1.1). I vardagsspråk skulle det vara konstigt att glömma att man har flera barn än man säger, men rent logiskt säger (1.1) ”Jag har minst en dotter” så det är helt möjligt att ha fler. Det finns massor av exempel från vardagslivet på när vi använder logiskt språk på ett icke-logiskt sätt: För många finns det en viss tröst i det faktum att

”Om du älskade mig skulle du aldrig glömma min födelsedag”

inte är någon logisk implikation. Inom logik (och matematik) måste vi hålla oss till det som verkligen sägs och slippa det underförstådda så mycket vi kan! Jämför (1.1) med följande påstående:

Martin har en penna.

Här skulle ingen tänka att det var konstigt om Martin hade flera pennor även om han sa han bara har en. I det här fallet stämmer vardagstolkningen överens med den logistiska. Skillnader kan också uppstå i användning av ordet ”eller” jämfört med talspråket: Rent logiskt utesluter (1.4) inte att man äter både potatis och ris, fast det inte är så ofta man äter båda i samma rätt. Inom logiken uttrycks den *svaga disjunktionen* – det vill säga, minst ett av alternativen – med ”eller” medan den *starka disjunktionen* – precis ett av alternativen – uttrycks med ”antingen…eller”. Om du skriver någonting du tror kan misstolkas så är det bäst att förtydliga.

Påståendet ”$\neg B \implies \neg A$” är *kontrapositionen* av ”$A \implies B$”. Båda påståendena är logiskt ekvivalenta, det vill säga ”$A \implies B$” om och endast om ”$\neg B \implies \neg A$”. Till exempel:

Om batteriet tar slut funkar inte min mobiltelefon;

är ekvivalent med

Om min mobiltelefon funkar så har batteriet inte tagit slut.

Nu ska vi gå igenom ett par exempel. För att dela upp texten lite tydligare när vi skriver svaret använder vi oss av en rubrik i kursivstil ”*Lösning*” som säger att svaret börjar. Och slutar med en liten fyrkant ”□” som säger att svaret är klart.

Exempel 1.1 (Negation) Vilket av följande påståenden är negationen av

”Alla fåglar äter frön”? (1.6)

(a) ”Ingen fågel äter frön.”
(b) ”Det finns minst en fågel som äter frön.”
(c) ”Alla frön äter fåglar.”
(d) ”Det finns minst en fågel som inte äter frön.”

Lösning. Vi vill lista ut vilket påstående är ekvivalent med ”Inte alla fåglar äter frön”, det vill säga negationen av (1.6).

Om ingen fågel äter frön så är (1.6) falskt, men det är inte det enda sättet som (1.6) kan vara falskt på. Till exempel om det finns både fåglar som äter frön och fåglar som inte äter frön så är (1.6) falskt. Det vill säga (a) $\implies \neg$(1.6) men den motsatta implikationen är falsk. Därför är (a) inte negationen av (1.6).

Om alla fåglar äter frön så finns det minst en fågel som äter frön och i så fall är både (1.6) och (b) sanna. Om det både finns fåglar som äter frön och fåglar som inte äter frön, då är (1.6) falskt och (b) sant. Därför har sanningen av (b) inget att göra med sanningen av (1.6) och därför kan (b) inte vara negationen av (1.6).

Påståendet (c) låter lite konstigt. Vi har bara bytt plats på subjektet och objektet i satsen. Vad frön gör med fåglar har inget direkt logiskt att göra med vad fåglar gör med frön och därför är (c) inte negationen av (1.6).

Det finns bara en möjlighet kvar. Om det finns minst en fågel som inte äter frön så är (1.6) falskt. Om (1.6) är falskt så finns det minst en fågel som inte äter frön. Därför är påståendet (d) negationen av (1.6). □

Observera att det kan finnas olika omskrivningar av samma påstående. Påståendet ”Inte alla fåglar äter frön” är negationen av (1.6) precis som (d). Alla påståenden som är ekvivalenta med ¬(1.6) kallas för en negation av (1.6).

Exempel 1.2 (Kontraposition) Ge kontrapositionen av påståendet

> ”Om man körde från Linköping till Stockholm på mindre än 1 timme och 40 minuter har man brutit mot hastighetsbegränsningen.” (1.7)

Lösning. Vi kan skriva (1.7) i formen $A \implies B$ där

$$A = \begin{cases} \text{"Man körde från Linköping till Stockholm på mindre} \\ \text{än 1 timme och 40 minuter."} \end{cases}$$

och

$$B = \text{"Man har brutit mot hastighetsbegränsningen."}$$

Vi behöver skriva negationen av både A och B:

$$\neg A = \begin{cases} \text{"Man körde från Linköping till Stockholm på 1 timme} \\ \text{och 40 minuter eller längre."} \end{cases}$$

$$\neg B = \text{"Man har inte brutit mot hastighetsbegränsningen."}$$

Därför är kontrapositionen av (1.7)

> ”Om man inte har brutit mot hastighetsbegränsningen körde man från Linköping till Stockholm på 1 timme och 40 minuter eller längre.” □

Här får man även fundera över vad som är implicit i frågeställningen. Ett sätt att inte uppfylla A är att inte köra från Linköping till Stockholm alls. Ingår det i hur vi formulerade $\neg A$?

TABELL 1.1 Notation för att anteckna relationer mellan tal (a och b) samt operationer på tal. Man använder också parentes $(\dots)$ (och ibland även $[\dots]$ och $\{\dots\}$) för att markera en operatorprioritet som skiljer sig från vanlig: multiplikation och division först, sedan addition och subtraktion.

	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	$a^2 = a \times a$	"a i kvadrat"

HUR MAN SKRIVER ARGUMENT INOM MATEMATIK

Man kan tänka sig att den matematiska verkligheten är ett mörkt rum, och logik och matematiskt resonemang 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 på att avslöja enskilda föremål, det vill säga de matematiska detaljerna, men kartan ger oss en bättre bild över hela rummet och därmed en bättre känsla för ämnet. När man skriver lösningar till matematiska problem eller på andra sätt vill förklara matematiska idéer för andra behöver man kommunicera både detaljerna och intuitionen på ett förståeligt sätt. Det är ingen enkel uppgift och kan vara en helt separat färdighet från att förstå själva matematiken.

Till skillnad från en vanlig ficklampa kan det vara svårt inom matematiken att avslöja vissa föremål i rummet. Till exempel lyckades ingen lösa vissa geometriska problem – såsom cirkelns kvadratur och vinkelns tredelning – som först betraktades under antiken, förrän under 1800-talet. Det är en tecken på att matematiken också är ett kreativt ämne. Att komma på hur man argumenterar för att ett visst faktum gäller, kan vara helt annorlunda än att följa någon annans argument. Men även om många känner att det kan vara svårt, kan alla bli bättre genom att öva och spendera tid med ämnet.

Varje påstående i matematik härleds av grundläggande begrepp, så kallade *axiom* som man antar utan frågor. I bilaga C.1 kan man titta på de axiom som räcker för att beskriva alla egenskaper hos tal, men fullständig härledning av materialet från axiomen är inte huvudsyftet med boken. Däremot vill vi koncentrera oss på både hur man härleder påståenden på ett hyfsat stabilt logiskt sätt och hur man redovisar matematiska argument.

Vår utgångspunkt är att man redan är van vid de enkla *aritmetiska operationerna* – addition, subtraktion, multiplikation och division – och förstår likheter och olikheter mellan tal. I tabell 1.1 finns det en påminnelsetabell om den grundläggande notation vi använder oss av i texten. Under vår resa kommer vi att upptäcka ytterligare operationer som kommer att komplettera det vi här kallar för *aritmetik*.

Enkla regler

Betrakta dessa följande enkla påståenden som ni kan ha använt er av någon gång i livet:

$$(a+b)c = ac + bc, \tag{1.8}$$

$$(a+b)^2 = a^2 + b^2 \quad \text{och} \tag{1.9}$$

$$a < b \implies ac < bc \tag{1.10}$$

där a, b och c är godtyckliga tal.

Likheten (1.8) är en enkel aritmetisk regel och vi tänkte inte förklaras den ytterligare.[1] Vi passar bara på att betona lite terminologi för summor och produkter: Ett element i en summa kallas för en *term* medan ett element i en produkt kallas för en *faktor*.

Likheten (1.9) är enkel – det finns inget enklare sätt att skriva om en operation på en summa än att sätta den på varsin term, eller hur? Men tyvärr är det inte sant. Om varken a eller b är noll får vi inte likhet i (1.9). Det kan vi bevisa med hjälp av riktiga aritmetiska regler: Vi vet att

$$(a+b)^2 = (a+b)(a+b) = (a+b)a + (a+b)b = a^2 + b^2 + 2ab,$$

så (1.9) är sant om och endast om $a^2 + b^2 + 2ab = a^2 + b^2$. Sista likheten är ekvivalent med $2ab = 0$ och den gäller om och endast om minst en av a och b är noll. Eftersom (1.9) bara gäller vid exceptionella omständigheter är det ingen idé att betrakta det som en giltig regel.

1 Det är inte precis ett axiom enligt bilaga C.1, men det följer av dem: se exempel C.6.

Ibland är (1.10) sant, men inte alltid. Till exempel $5 < 9$ men

$$5(-4) = -20 \quad \text{och} \quad 9(-4) = -36$$

så $5(-4) \nless 9(-4)$. Implikationen (1.10) är sann så länge $c > 0$. Eftersom (1.10) gäller i en ganska bred belägenhet är det bra att komma ihåg det. Men varje gång vi använder oss av (1.10) är det oerhört viktigt att vi både kontrollerar att hypotesen $c > 0$ gäller och påpekar det i vår redovisning av argumentet.

Tecknet $\nless$ betyder "ej mindre än" och i allmänhet betyder ett streck genom ett tecken negationen av tecknets ursprungliga betydelse.

Lösningar till ekvationer

Ibland använder vi giltiga argument och regler men går ändå vilse. Det är inte alltid enkelt att förstå vad det är vi har bevisat. Ett vanligt exempel är när vi löser ekvationer. Betrakta följande system av två ekvationer:

$$\begin{aligned} &x + 4 = 3 \quad \text{och} \\ &x - 4 = 5. \end{aligned} \tag{1.11}$$

Ett tal x som uppfyller båda ekvationerna kallas för en *lösning* av systemet. Vi kan addera båda vänsterleden och resultatet måste vara lika med summan av högerleden: det ger oss

$$\begin{aligned} (x + 4) + (x - 4) &= 3 + 5 \\ \Longrightarrow \quad 2x &= 8 \end{aligned}$$

och därifrån drar vi slutsatsen att $x = 4$. Kort sagt har vi bevisat att

$$(1.11) \Longrightarrow x = 4.$$

Villkoret $x = 4$ är ett *nödvändigt villkor* för (1.11). Kan vi dra slutsatsen att $x = 4$ är en lösning till (1.11)? Med andra ord: Är $x = 4$ ett *tillräckligt villkor* för (1.11)?

Om $A \Longrightarrow B$ är B ett nödvändigt villkor för A och A är ett tillräckligt villkor för B.

Nej, det är det inte: Vi kan kolla direkt om $x = 4$ är en lösning genom att sätta in det i (1.11). Vi kollar att $x = 4$ medför att $x + 4 = 8 \neq 3$ och även $x - 4 = 0 \neq 5$ så (1.11) stämmer inte alls. Vi har bevisat att

$$x = 4 \not\Longrightarrow (1.11).$$

För att dra slutsatsen att $x = 4$ inte är en lösning till (1.11) räcker det förstås att x inte uppfyller en av de två ekvationerna i (1.11).

Har någonting gått fel då? Man kan kolla alla aritmetiska operationer vi har använt oss av och vi har inte gjort något fel. Problemet ligger i hur vi har tolkat våra slutsatser. Att skriva ”(1.11) $\implies$ $x = 4$” säger egentligen att

> ”Om det finns en lösning x till (1.11) så måste den vara 4.”

Genom att beräkna med (1.11) antar vi implicit att det finns minst ett x som uppfyller (1.11). Så 4 är bara en kandidat till en lösning x: inget annat tal har en chans att vara en lösning. Men i det här fallet finns det ingen lösning alls till (1.11) och det har vi bekräftat genom att sätta in $x = 4$ (som vi nu vet är den enda möjligheten) i (1.11).

Om vi löser varje ekvation separat är det kanske enklare att se: $x + 4 = 3 \implies x = -1$ och $x - 4 = 5 \implies x = 9$. Eftersom vi inte kan ha $x = -1$ och $x = 9$ samtidigt finns det ingen lösning som uppfyller båda ekvationerna.

Vad händer om vi byter ut ordet ”och” i (1.11) mot ”eller”? Är $x = 4$ en lösning då?

Exempel 1.3 Hitta alla lösningar till följande system av ekvationer:

$$\begin{aligned} x(x+4) &= -4 \quad \text{och} \\ x(x-4) &= 12. \end{aligned} \tag{1.12}$$

Lösning. Vi beräknar precis som ovan och adderar båda ekvationer:

$$\begin{aligned} x(x+4) + x(x-4) &= -4 + 12 \\ \implies 2x^2 &= 8 \\ \implies x^2 &= 4. \end{aligned}$$

Vi ska prata mer om lösningarna x till ekvationer av formen $x^2 = c$ i kapitel 4, men just nu kan vi ta för givet att alla lösningar till $x^2 = 4$ är $x = 2$ och $x = -2$: Det har ni säkert sett i skolan. Därför har vi bevisat

$$(1.12) \implies x = 2 \text{ eller } x = -2.$$

Vi testar om de är lösningar genom att sätta in dem i (1.12): Om $x = -2$ är $x(x + 4) = (-2)(-2 + 4) = -4$ och $x(x - 4) = (-2)(-2 - 4) = 12$ så löser $x = -2$ (1.12). Om $x = 2$ är $x(x + 4) = 2(2 + 4) = 12$ och $x(x - 4) = 2(2 - 4) = -4$ så gäller inte (1.12).

Slutsatsen är då att $x = -2$ är den enda lösningen till (1.12). □

Definitioner och satser

Det skulle vara oerhört tråkigt och långsamt att bevisa alla nya påståenden från enkel aritmetik, så i stället bygger vi upp olika användbara påståenden som vi noggrant bevisar med hjälp av grundläggande aritmetik tillsammans med de påståenden vi har bevisat tidigare. Ett viktigt påstående i matematik kallas för en *sats*. Genom att skriva upp och bevisa satser sparar man tid, men minst lika viktigt är att man skapar något slags intuition för ämnet. Många slutsatser är enklare att se som följder av allmänna regler och principer än att bevisa direkt från axiomen i enskilda fall. Vi formulerar också *definitioner* för att införa nya begrepp eller ord som är nyttiga i ett visst sammanhang. På så sätt är det tydligare vad vi menar när vi använder ämnesspecifika ord. Det är enklare att kontrollera vad sägs och det minskar förvirringen och tvetydigheten.

Nu går vi igenom ett exempel, som förhoppningsvis gör idéerna tydligare. Vi börjar med en definition.

Definition 1.4 Givet två heltal $m \neq 0$ och n kallar vi r *resten* av n delat med m om

$$n = qm + r \quad \text{för något positivt heltal } q \tag{1.13}$$

där r är ett heltal sådant att $0 \leq r < m$. I fallet $r = 0$ säger vi att n är *(jämnt) delbart* med m.

När man skriver två olikheter eller likheter tillsammans betyder det att båda gäller samtidigt: så här till exempel är $0 \leq r$ och $r < m$.

Definition 1.4 ger en precis mening till ordet *rest*. Terminologin är lite tydligare om vi skriver (1.13) i den ekvivalenta formen $n/m = q + r/m$. Då ser vi att n/m är hela talet q plus bråket r/m. Vi vet även att $0 \leq r/m < 1$ från villkoret $0 \leq r < m$, så resten bidrar till den mindre delen av summan, åtminstone när $q \neq 0$.

Nu vet vi exakt betydelsen av ordet *rest* och kan gå vidare och undersöka begreppet hos tal. Härnäst utreder vi en sats som är långt ifrån uppenbar utan ett bevis. Det handlar om begreppet rest, som vi har definierat i definition 1.4 samt de grundläggande begreppen positiva tal, olikhet och division.

Sats 1.5 *Betrakta tre positiva heltal n_1, n_2 och m sådana att:*

(a) n_1 delat med m har rest 1*; och*

(b) n_2 delat med m har rest 1.

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

Precis som vi gjorde i exemplen ovan, börjar vi beviset med en rubrik i kursivstil ”*Bevis*” och i slutet kommer en liten fyrkant ”□” som säger att beviset är klart.

Bevis. Eftersom n_1 delat med m har rest 1 vet vi från (1.13) att det finns ett heltal q_1 så att

$$n_1 = q_1 m + 1.$$

Eftersom n_2 delat med m har rest 1 vet vi också från (1.13) att det finns ett 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 (1.13) i Definition 1.4 uppfylls med $n = n_1 n_2$, $q = (q_1 q_2 m + q_1 + q_2)$ och $r = 1$ och alla är förstås heltal. 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 1.4 på 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 slutsatsen kan inte användas! Den måste bevisas utan att använda någonting vi inte har bevisat tidigare. Hos beviset har vi bara använt oss av antagandena (a) och (b). Egenskaper av produkten $n_1 n_2$ är det som måste visas utifrån antagandena.

Kvantifikatorer

Ofta dyker uttryck som ”för alla”, ”för något” och ”för inget” upp i vårt arbete. Sådana uttryck specificerar antalet objekt en viss slutsats gäller. Till exempel drog vi slutsatsen att *inga* lösningar finns till (1.11), och i (1.6) betraktade vi ett påstående som gällde *alla* fåglar. Sådana uttryck kallas för *kvantifikatorer*.

I exempel 1.1 såg vi ett exempel på relationerna mellan kvantifikatorer när vi tar negationen av ett påstående. Om $P(x)$ är påståendet ”Fågel x äter frön” är påståendet (1.6) ”För alla x är $P(x)$ sant” och negationen är ”Det finns minst ett x sådant att $P(x)$ är falskt”. Tabell 1.2 ger exempel

TABELL 1.2 Exempel på relationer mellan kvantifikatorer när man tar negationen av ett påstående $P(x)$ som beror på x.

Påstående	Negation
För alla x är $P(x)$ sant	Det finns minst ett x sådant att $P(x)$ är falskt
För minst ett x är $P(x)$ sant	Det finns inget x sådant att $P(x)$ är sant
För inget x är $P(x)$ sant	Det finns minst ett x sådant att $P(x)$ är sant

på påståenden och dess negationer som innehåller kvantifikatorer. I allmänhet är påståendet ”För inget x är $P(x)$ sant” förstås ekvivalent med ”För alla x är $P(x)$ falskt”.

Vilken kvantifikator som finns i ett påstående inverkar på hur vi rimligen kan gå till väga för att bevisa det. Allmänt sett måste ett bevis för ett påstående som innehåller ”för alla” genomföras med ett allmänt argument, men om påståendet innehåller ”för något” räcker det att vi hittar ett särskilt exempel där påståendet gäller.

Exempel 1.6 Visa att

$$(x+5)^2 - (x+3)^2 - 16 = 4x$$

gäller för *alla* tal x.

Lösning. Eftersom vi vill visa att likheten gäller oavsett vilket tal x är, måste vi bevisa det med ett allmänt argument som inte är beroende av värdet av x. Vi kan förstås räkna ut algebraiskt:

$$\begin{aligned}(x+5)^2 - (x+3)^2 - 16 &= (x^2+10x+25) - (x^2+6x+9) - 16\\ &= x^2 + 10x + 25 - x^2 - 6x - 9 - 16\\ &= 4x.\end{aligned}$$

Argumentet gäller för vilket x som helst och därmed har vi bevisat likheten för alla möjliga värden på x. □

Det skulle inte räcka att sätta in ett värde på x och sedan beräkna – då skulle vi bara ha visat den för de specifika värdena av x. Det är en sak att jobba med ett särskilt värde av x för att förstå problemet, men när man kommer till att skriva beviset för exempel 1.6 måste man säkerställa att alla beräkningar fungerar oavsett värdet av x.

Exempel 1.7 Betrakta likheten

$$(x-1)^2-(x-3)^2-16=0. \tag{1.14}$$

(a) Visa att *det finns minst ett* tal x som löser (1.14).
(b) Visa att (1.14) *inte* kan gälla för *alla* tal x.

Lösning. (a) På ett sätt kan man säga att den här uppgiften är enklare än exempel 1.6. I stället för att behöva visa någonting för alla möjliga tal x behöver vi bara hitta ett x sådant att likheten gäller. Men i praktiken kan det vara svårare att visa att någonting inte nödvändigtvis gäller för alla möjliga val av x, utan bara i vissa fall. För att bevisa att det finns minst ett x, räcker det att hitta ett exempel där (1.14) gäller.

Vi provar att sätta in $x=6$ i vänsterledet:

$$(x-1)^2-(x-3)^2-16=(6-1)^2-(6-3)^2-16=25-9-16=0.$$

Så vi har visat direkt att $x=6$ är en lösning till (1.14). Om det gäller för andra x eller inte spelar ingen roll – vi har redan löst uppgiften.

(b) Uppgiften påstår att likheten inte kan gälla för alla x. Enligt tabell 1.2 måste vi visa att det finns minst ett x sådant att likheten är falsk. Precis som i del (a) räcker det att vi hittar ett exempel på ett x som inte ger likhet i (1.14).

Vi provar att sätta in $x=2$ i vänsterledet:

$$\begin{aligned}(x-1)^2-(x-3)^2-16&=(2-1)^2-(2-3)^2-16\\&=1-1-16=-16\neq 0.\end{aligned}$$

Så vi har visat direkt att $x=2$ inte är en lösning till (1.14).

Eftersom vi inte har förklarat hur man skulle kunna gissa $x=6$ eller $x=2$ i del (a) respektive (b) kan det kännas som att redovisningen av lösningen är bristande, även om den rent logiskt är fullständig.[2] Ett mer intuitivt sätt att lösa uppgiften är att först skriva om (1.14) till en ekvivalent form: Vi kan räkna ut

$$(x-1)^2-(x-3)^2-16=(x^2-2x+1)-(x^2-6x+9)-16=4x-24$$

så (1.14) är ekvivalent med

$$4x-24=0$$

2 Jämför med analysen av (1.11).

och därmed är det tydligt att x är en lösning till (1.14) om och endast om $x = 6$. Därför finns det en lösning x till (1.14) men långt ifrån alla tal x uppfyller (1.14). □

Hela meningar

Observera till sist att vi skriver matematik i hela meningar. Visst använder vi också mycket matematisk notation för att skriva snabbt och tydligt, men det kan inte helt och hållet ersätta vanliga ord – såsom kvantifikatorer. Ord hjälper oss att förstå vad det är man antar, hur man går från ett steg till ett annat och vilka slutsatser man drar. Utan den förklaringen skulle argumentet vara otydligt. Detsamma gäller när man redovisar uppgifter och problem i arbetslivet. Det är helt okej att skriva korta anteckningar för att själv förstå ett problem eller förklara för en vän medan man sitter och pratar, men det räcker inte som en fullständig lösning om man inte kan läsa och förstå utan föregående kunskap om problemet. För att veta om det du har skrivit räcker som en fullständig lösning, kan du fundera på om du skulle 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 mycket viktigt.

1.2 Upprepade aritmetiska operationer

Utifrån de enkla aritmetiska operationerna byggs en hel värld av matematisk struktur. Genom att upprepa de grundläggande operationerna avslöjar vi andra begrepp och strukturer. Här tar vi en första titt på *potenser* – sett som upprepad addition – men begreppet återkommer flera gånger genom boken.

HELTALSPOTENSER

När man multiplicerar ett tal a med ett heltal n kan man tolka produkten an som en upprepad addition av a med sig själv:

$$an = \underbrace{a + a + \cdots + a}_{n \text{ gånger}}.$$

Nu vänder vi oss till upprepad multiplikation av ett tal med sig själv. Vi har redan upphöjt tal i kvadrat, det vill säga multiplicerat tal med sig själva, men nu vill vi utvidga begreppet. Vi definierar

$$a^n := \underbrace{aa\dots a}_{n \text{ gånger}} \tag{1.15}$$

för 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*).

Tecknet ":=" betyder att vänsterledet definieras som högerledet.

Vi definierar också

$$a^{-n} := \frac{1}{a^n} \quad \text{och} \quad a^0 = 1 \tag{1.16}$$

för tal $a \neq 0$ och positiva heltal n.

I början känns det lite konstigt att ha tre olika definitioner för a^n beroende på värdet av n. Följande sats visar att de olika definitionerna hänger ihop och därför blir definitionerna lite mer rimliga. Tyvärr är även satsens bevis uppdelade i många fall på grund av de olika fallen som uppstår i definitionen av potens.

Observera att vi inte har definierat noll upphöjt till noll. Anledningen att inte göra det motsvarar anledningen att inte definiera noll delat med noll.

Sats 1.8 *För givna godtyckliga nollskilda tal a och b, och hela tal n och m gäller följande regler.*

(a) $a^n a^m = a^{n+m}$,
(b) $a^n a^{-m} = a^{n-m}$,
(c) $a^n b^n = (ab)^n$, *och*
(d) $(a^n)^m = a^{nm}$.

Bevis. Vi bevisar bara (c) och lämnar de andra som en övning för läsaren. För att bevisa (c) 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 (1.16). Så vi får att $a^n b^n = 1 \times 1 = 1 = (ab)^n$.

Fall 2: $n > 0$.

Då får vi att

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

enligt (1.15). Det spelar ju inte någon roll i vilken ordning man multiplicerar talen[3] så genom att flytta b:t längst till vänster $(n-1)$ platser åt vänster får vi att

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

Genom att flytta nästa b:n $(n-2)$ platser till vänster får vi att

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

Och om vi fortsätter på samma sätt $(n-3)$ gånger till kommer vi fram till att

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

Fall 3: $n < 0$.

Sätt $p = -n$, så $p > 0$. Då kan vi säga

$$a^n b^n = a^{-p}b^{-p} \underset{\substack{\uparrow \\ (1.16)}}{=} \left(\frac{1}{a^p}\right)\left(\frac{1}{b^p}\right) = \frac{1}{a^p b^p}$$

$$\underset{\substack{\uparrow \\ \text{fall 2}}}{=} \frac{1}{(ab)^p} \underset{\substack{\uparrow \\ (1.16)}}{=} (ab)^{-p} = (ab)^n,$$

som önskat. □

SUMMOR

Likaväl som för multiplikation spelar det inte någon roll i vilken ordning man summerar tre eller flera tal.[4] Det är praktiskt att ha en beteckning för summan av många termer: För ett heltal n kan vi betrakta summan av n tal $a_1, a_2, \ldots, a_n$. Det skriver vi som

Som ovan använder vi här en ellipsis (…) för att indikera att summan innehåller fler termer än vad vi explicit har skrivit.

$$a_1 + a_2 + \cdots + a_n =: \sum_{j=1}^{n} a_j.$$

3 Jämför med axiom C.1(a) och C.1(b) i bilaga C.1

4 Jämför med axiom C.1(a) och C.1(b) i bilaga C.1

Bokstaven j kallas för *summationsindex* och får i princip vara vilken bokstav som helst, så länge vi skriver den bokstav vi vill summera under summatecknet $\sum$. Samma notationen får användas i fallet vi vill summera från indexet m till n ($m \leq n$):

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

Exempel 1.9 Här ger vi några exempel för att göra notationen lite tydligare.

(a) Summan $1 + 2 + 3 + 4 + 5 + 6$ kan skrivas som

$$\sum_{j=1}^{6} j \quad \text{eller baklänges} \quad \sum_{k=1}^{6}(7-k)$$
$$\text{eller med andra index} \quad \sum_{m=5}^{10}(m-4).$$

(b) Summan av jämna tal mellan 29 och 39, det vill säga $30 + 32 + 34 + 36 + 38$, kan skrivas som

$$\sum_{j=15}^{19} 2j \quad \text{eller} \quad \sum_{j=0}^{4}(30+2j).$$

(c) Vi kan skriva

$$\begin{aligned}\sum_{j=4}^{9} 2^j &= 2^4 + 2^5 + 2^6 + 2^7 + 2^8 + 2^9\\ &= 16 + 32 + 64 + 128 + 258 + 516 = 1014.\end{aligned}$$

Följande sats är en följd av vanliga aritmetiska regler.

Sats 1.10 *För $n \leq m$, tal a_j och b_j ($j = n, n+1, \ldots, m$), och ett tal c har vi att*

$$\sum_{j=n}^{m}(a_j + b_j) = \sum_{j=n}^{m} a_i + \sum_{j=n}^{m} b_j \tag{1.17}$$

och

$$\sum_{j=n}^{m}(ca_j) = c\sum_{j=n}^{m} a_j. \tag{1.18}$$

Vi har även

$$\sum_{j=n}^{m} \sum_{i=k}^{\ell} b_{i,j} = \sum_{i=k}^{\ell} \sum_{j=n}^{m} b_{i,j} \tag{1.19}$$

där $k \leq \ell$ och $b_{i,j}$ är tal ($j = n, \ldots, m$ och $i = k, \ldots, \ell$).

Likheten (1.18) säger att man kan faktorisera en gemensam faktor c från en summa och (1.17) och (1.19) är två olika sätt att ändra ordningen på termerna i en summa.

Exempel 1.11 Vi ska med hjälp av sats 1.10 visa att

$$\sum_{k=5}^{30} \left(4 + \frac{2k}{3}\right) = \frac{520}{3} + \frac{2}{3} \sum_{j=1}^{26} j.$$

Lösning. Vi kan till exempel börja med att använda (1.17) och (1.18), och räkna ut

$$\sum_{k=5}^{30} \left(4 + \frac{2k}{3}\right) = \sum_{k=5}^{30} 4 + \sum_{k=5}^{30} \frac{2}{3} k = \sum_{k=5}^{30} 4 + \frac{2}{3} \sum_{k=5}^{30} k.$$

Vi observerar att alla termerna i summan $\sum_{k=5}^{30} 4$ är desamma, det vill säga lika med 4. Därför vet vi att summan är en produkt av 4 med antalet termer. Det finns 26 termer och därför är $\sum_{k=5}^{30} 4 = 4 \times 26 = 104$ och

$$\sum_{k=5}^{30} \left(4 + \frac{2k}{3}\right) = 104 + \frac{2}{3} \sum_{k=5}^{30} k.$$

Det stämmer inte med det uttryck vi vill ha av två anledningar. För det första är de första termerna olika – 104 är inte lika med $520/3$ – och sedan är gränserna för summationsindexet olika – $5 \neq 1$ och $30 \neq 26$. Att vi summerar i k i stället för j är inget bekymmer eftersom vi är fria att döpa indexet om vi så skulle önska.

För att beräkna vidare använder vi oss av ett *variabelbyte* i summationsindexet. Vi vill att indexet börjar med 1 i stället för 5 så vi testar att sätta $j = k - 4$. När $k = 5$ så är $j = 5 - 4 = 1$ och när $k = 30$ är $j = 30 - 4 = 26$, så variabelbytet ser lovande ut. För att genomföra variabelbytet behöver vi bara ersätta varje instans av k med sitt uttryck i j, som förstås är $k = j + 4$. Därför är

$$\sum_{k=5}^{30} k = \sum_{j=1}^{26} (j + 4).$$

Nu kan vi fortsätta beräkna återigen med hjälp av (1.17) och (1.18):

$$\begin{aligned}104 + \frac{2}{3}\sum_{k=5}^{30} k &= 104 + \frac{2}{3}\sum_{j=1}^{26}(j+4) = 104 + \frac{2}{3}\left(\sum_{j=1}^{26} j + \sum_{j=1}^{26} 4\right) \\ &= 104 + \frac{2}{3}\left(\sum_{j=1}^{26} j + 104\right) = 104 + \frac{208}{3} + \frac{2}{3}\sum_{j=1}^{26} j \\ &= \frac{520}{3} + \frac{2}{3}\sum_{j=1}^{26} j \end{aligned}$$

□

Carl Friedrich Gauss är en berömd matematiker som bidragit till flera grenar inom matematik och fysik, till exempel talteori, analys och geometri. Enligt en gammal sägen fick den unge Gauss följande problem som straff en gång när han varit stygg i skolan. Läraren blev förvånad när Gauss löste problemet på några minuter!

Exempel 1.12 Addera alla heltal mellan ett och hundra. Det vill säga, räkna ut

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

Lösning. Gauss löste problemet på följande sätt: Han skrev summan två gånger, första gången som vanligt och sedan andra gången nedanför den första fast baklänges

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

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 samtidigt lika med $2\sum_{k=1}^{100} k$. Alltså

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

□

Man kan även tillämpa metoden från exempel 1.12 på summan $\sum_{k=1}^{n} k$ – där man summerar de första n positiva heltalen. Vi formulerar resultatet som en sats men lämnar beviset som en övning för läsaren. En alternativ metod att bevisa satsen presenteras i exempel 2.7.

Sats 1.13 *För varje positivt heltal n är*

$$\sum_{k=1}^{n} k = \frac{n(n+1)}{2}.$$

Men hjälp av sats 1.13 kan vi lätt beräkna utifrån exempel 1.11 att

$$\begin{aligned}\sum_{k=5}^{30}\left(4+\frac{2k}{3}\right) &= \frac{520}{3}+\frac{2}{3}\sum_{j=1}^{26} j = \frac{520}{3}+\frac{2}{3}\left(\frac{26(26+1)}{2}\right) \\ &= \frac{520}{3}+\frac{702}{3}=\frac{1222}{3}\end{aligned}$$

Här använder vi förstås sats 1.13 med $n = 26$.

I nästa sats betraktar vi en annan typ av summa.

Sats 1.14 *För givna tal a och r är*

$$\sum_{j=1}^{n} ar^{j-1} = a\frac{1-r^n}{1-r} \tag{1.20}$$

under antagandet att r är varken 0 *eller* 1.

Anmärkning 1.15 Innan vi bevisar satsen funderar vi på vad vi kan säga om summan när antagandet r är varken 0 eller 1 *inte* är uppfyllt.

Notera att högerledet av (1.20) har ingen chans att vara lika med summan i fallet $r = 1$ eftersom då är nämnaren lika med noll och operationen att dela med noll är odefinierad.

(a) Observera att vi redan har betraktat fallet $r = 1$ i exempel 1.11. Enligt (1.15) och (1.16) vet vi att $(1)^{j-1} = 1$ för alla positiva heltal j och därför är

$$\sum_{j=1}^{n} a(1)^{j-1} = \sum_{j=1}^{n} a = an.$$

Vi har helt enkelt en annan formel i fallet $r = 1$.

(b) Fallet $r = 0$ är lite mer subtilt. Nu är själva summan odefinierad eftersom första termen innehåller faktorn 0^0. Även om första termen var definierad skulle alla följande termer vara noll och en formel skulle inte vara så nyttig i alla fall.

Bevis (sats 1.14). Vi döper

$$S_n = \sum_{j=1}^{n} ar^{j-1}$$

för varje positivt heltal n och observerar att

$$\begin{aligned} S_{n+1} = \sum_{j=1}^{n+1} ar^{j-1} &= a + ar + ar^2 + \cdots + ar^{n-1} + ar^n. \\ &= S_n + ar^n \end{aligned}$$

Dessutom är

$$\begin{aligned} S_{n+1} = \sum_{j=1}^{n+1} ar^{j-1} &= a + ar + ar^2 + \cdots + ar^n \\ &= a + r(a + ar + ar^2 + \cdots + ar^{n-1}) \\ &= a + rS_n. \end{aligned}$$

Därför är $S_n + ar^n = S_{n+1} = a + rS_n$ som medför att

$$S_n = a\frac{1-r^n}{1-r}. \qquad \square$$

Vi ska nu titta på ett exempel som använder sats 1.14.

Exempel 1.16 Skriv om summan

$$\sum_{j=1}^{34} 36\left(\frac{4}{6}\right)^{j+3} 2^j$$

till en skillnad mellan två tal.

Lösning. Först förenklar vi uttrycket för termerna med hjälp av räkne-reglerna i sats 1.8:

$$\begin{aligned} 36\left(\frac{4}{6}\right)^{j+3} 2^j = 36\left(\frac{4}{6}\right)^{j-1}\left(\frac{4}{6}\right)^4 2^{j-1}2 &= \frac{36 \times 4^4 \times 2}{6^4}\left(\frac{4 \times 2}{6}\right)^{j-1} \\ &= \frac{128}{9}\left(\frac{4}{3}\right)^{j-1}. \end{aligned}$$

Nu vet vi att

$$\sum_{j=1}^{34} 36 \left(\frac{4}{6}\right)^{j+3} 2^j = \sum_{j=1}^{34} \frac{128}{9} \left(\frac{4}{3}\right)^{j-1}$$

och därför vill vi tillämpa sats 1.14 med $a = 128/9$, $r = 4/3$ och $n = 34$. Vi får att

$$\sum_{j=1}^{34} \frac{128}{9} \left(\frac{4}{3}\right)^{j-1} = \frac{128}{9} \left(\frac{1 - \left(\frac{4}{3}\right)^{34}}{1 - \frac{4}{3}}\right) = -\frac{128}{3} + \frac{128}{3} \left(\frac{4}{3}\right)^{34}.$$

Därför är

$$\sum_{j=1}^{34} 36 \left(\frac{4}{6}\right)^{j+3} 2^j = \frac{128}{3} \left(\frac{4}{3}\right)^{34} - \frac{128}{3} = \frac{2^{75}}{3^{35}} - \frac{128}{3} \qquad \square$$

och summan är skillnaden mellan $2^{75}/3^{35}$ och $128/3$.

Anmärkning 1.17 När man börjar räkna ut summor som i exempel 1.16 är det ett vanligt misstag att till exempel skriva

$$\sum_{j=1}^{34} 36 \left(\frac{4}{6}\right)^{j+3} 2^j = 36 \left(\sum_{j=1}^{34} \left(\frac{4}{6}\right)^{j+3}\right) \left(\sum_{j=1}^{34} 2^j\right).$$

Att bryta ut gemensamma faktorn 36 är helt rätt (se (1.18)), men summan av en produkt är inte produkten av summorna! Ett enkelt exempel visar att det inte är en giltig räkneregel:

$$\sum_{j=1}^{2} 2^j 3^j = 2 \times 3 + 2^2 3^2 = 42 \quad \text{men}$$

$$\left(\sum_{j=1}^{2} 2^j\right)\left(\sum_{j=1}^{2} 3^j\right) = (2 + 2^2)(3 + 3^2) = 72,$$

så vänstra leden är inte lika.

TALBETECKNINGSSYSTEM FÖR HELTAL

Centralt för hur man kan både kommunicera och förstå matematik är det beteckningssystem man använder sig av för att skriva tal. Ett talbeteckningssystem är ett utslag av abstraktionen av matematik men också avgörande för hur enkelt eller svårt det kan vara att utreda mönster och egenskaper hos tal.

Det finns spår av att människor har använt sig av aritmetik och utvecklat talbeteckningssystem sedan forntiden och de äldsta verkar vara cirka 35 000 år gamla [3, s. 3]. Här tittar vi lite närmare på det *egyptiska talsystemet* som utvecklades cirka 3 000 f.Kr. Det börjar ganska naturligt: Talet *ett* skrivs som ett lodrätt streck 𓏺, och för två till tio upprepar man strecket samma antal gånger som talet, så till exempel skrivs fyra

𓏺𓏺𓏺𓏺.

I princip kunde man fortsätta i evighet och på så sätt representera vilket tal som helst. Men det blir snabbt klart att det är opraktiskt. Därför skapade man andra tecken som representerade högre tal. Till exempel stod 𓎆 för tio och 𓍢 för hundra och man skrev

𓏺𓏺𓏺𓎆𓍢𓍢𓍢𓍢

för 413. Observera att de skrev talet baklänges jämfört med det system vi använder idag, men det gjorde ingenting for fornegyptierna, då de kunde skriva åt båda hållen eller till och med i kolonner. Spegelbilden av ett tecken kunde även skrivas med samma betydelse [3, s. 10]. Varje potens av tio hade sitt eget tecken upp till en miljon (se tabell 1.3).

TABELL 1.3 De fornegyptiska tecknen för potens av tio från ett till en miljon.

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

En nackdel med systemet var att det inte fanns något tecken för 10^7. Då var det besvärligt att skriva tal större än tio miljoner. Man kunde hitta på sådana tecken vid behov men man måste sluta någonstans. Däremot var addition ganska enkelt att genomföra: Man kunde helt enkelt samla ihop alla tecken och sedan byta ut eventuella grupper av tio 𓏺 mot 𓎆, grupper av tio 𓎆 mot 𓍢, och så vidare.

Det talbeteckningssystem vi använder oss av idag kallas för det *hindu-arabiska talsystemet*. Det har sitt ursprung i Indien (medan de äldsta inskrifterna med alla systemets drag fanns i Kambodja, år 683 e.Kr. [11, p.217]) och sedan infördes det av persiska och arabiska matematiker och därefter i Europa under 800-talet. Till skillnad från det egyptiska talsystemet är placeringen av tecknen viktigt.[5] Vi har bara tio tecken: 1,

5 Det stämde även för det tidigare babyloniska systemet, men det är mer sannolikt att

2, 3, 4, 5, 6, 7, 8 och 9, och sedan även 0 – att införa tecknet 0 var en innovation eftersom det för första gången tolkades som ett eget tal. I stället för att upprepa tecknen som man gjorde i det egyptiska talsystemet, använder man sig av tecknen 0–9 för att representera hur många ental, tiotal, hundratal, och så vidare, vi har. Placeringen av tecknet visar vad det representerar – ental, tiotal och så vidare – och ental placeras på höger sida, sedan tiotal till vänster om ental, hundratal till vänster om tiotal, och så vidare. Till exempel, femton är fem ental plus tio, det vill säga $\mathbf{5} \times 1 + \mathbf{1} \times 10$ och skrivs därför som

$$15.$$

Om vi skrev 51 i stället skulle det representera fem tiotal och ett ental, det vill säga $\mathbf{1} \times 1 + \mathbf{5} \times 10$ eller femtioett. Trettio är $\mathbf{0} \times 1 + \mathbf{3} \times 10$ och skrivs därför som 30. Observera: vi skrev 0 till höger om 3 för att indikera att 3 representerar tiotal och inte ental, även om själva talet kan skrivas som en summa av bara tiotal. I allmänhet skriver man

$$d_n d_{n-1} \ldots d_1 d_0,$$

där varje d_j är ett tecken 1, 2, 3, 4, 5, 6, 7, 8, 9 eller 0 ($j = 0, \ldots, n$), för att representera summan

$$\sum_{j=0}^{n} d_j \times 10^j. \tag{1.21}$$

Till exempel, i talet 53072 är $n = 4$, och $d_0 = 2$, $d_1 = 7$, $d_2 = 0$, $d_3 = 3$ och $d_4 = 5$ ty

$$53072 = \mathbf{2} \times 10^0 + \mathbf{7} \times 10^1 + \mathbf{0} \times 10^2 + \mathbf{3} \times 10^3 + \mathbf{5} \times 10^4.$$

Det hindu-arabiska talsystemet kallas också för *det decimala talsystemet* eftersom talen skrivs som summor av tiopotenser. Metoden fungerar lika bra om man byter ut 10 i (1.21) mot ett annat tal b: Det är möjligt att bevisa att varje heltal får skrivas på ett unikt sätt som

Observera att b^j är en potens av b men j:an i d_j är ett index.

$$\sum_{j=0}^{n} d_j \times b^j$$

det hindu-arabiska talsystemet inspirerats av ett kinesiskt stav-talsystem än det babyloniska [11, s. 217].

där varje d_j är ett av talen $0, 1, 2, \ldots, b-1$ ($j = 0, \ldots, n$). Vi ger inte beviset här men faktumet innebär att man kan införa ett sådant *positionssystem* för vilket tal b som helst, så länge man har b tecken för talen $0, 1, 2, \ldots, b-1$.

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 till exempel

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

som 21102 fast det skrivs 200 i det (decimala) hindu-arabiska talsystemet. Man kan även markera basen av ett skrivet tal med en nedsänkt b: Till exempel $21102_3 = 200_{10}$. Andra vanliga baser innefattar 5, 8, 12 och 16. Det babyloniska talsystemet hade $b = 60$ och lever kvar i hur vi mäter tid idag – därför går det 60 sekunder på en minut och 60 minuter på en timme.[6] Vilken bas man väljer delvis beror på vad det är man vill räkna ut med talsystemet. Delbarhet kan vara en viktig aspekt. Vid första anblicken kan basen 60 kännas som ett konstigt val, men det är praktiskt att vi lätt kan dela en timme med bland annat 2, 3 och 4!

TALBETECKNINGSSYSTEM FÖR ICKE-HELTAL

När man börjar lära sig aritmetik brukar man sysslar med heltal, och framför allt positiva heltal. Man ser inget problem när man adderar eller multiplicerar heltal – svaret man får är alltid ett annat (positivt) heltal. Man märker snart att vissa subtraktioner inte går att räkna ut med bara positiva heltal: Begreppet negativa tal krävs för att subtrahera ett större tal från ett mindre och motsvarar skuld i ekonomi eller att åka baklänges i geometri. Notationsmässigt är negativa tal inget större bekymmer och markeras med ett minustecken ($-$) till vänster om motsvarande positivt tal. Till exempel, om vi har en skuld på 5 kronor på ett konto kan man skriva saldot som -5.

Rationella tal

På samma sätt som med subtraktion går vissa divisioner inte att utföra med bara heltal. Om man delar två tårtor mellan sex personer så får varje person två sjättedelar. Man kan inte dela dem jämnt mellan sex personer så att alla får hela tårtor – det krävs bråk! För två sjättedelar skriver man

6 Närmare sagt hade det 59 tecken plus ett mellanslag.

2/6. Här kallas 6 för bråkets *nämnare* och för 2 bråkets *täljare*. Egentligen har vi bara skrivit den aritmetiska operationen vi vill utföra – vi skrev ”två delat med sex” – så då undrar man: På vilket sätt har vi räknat ut divisionen här? Några tycker att man alltid ska ”förenkla” ett bråk, det vill säga stryka alla gemensamma faktorer i nämnaren och täljaren:

$$\frac{2}{6} = \frac{1 \times 2}{3 \times 2} = \frac{1}{3}.$$

Men det viktigaste här är att man ser att olika divisioner kan leda till samma svar, och därför finns det flera sätt att skriva ett och samma tal som ett bråk.

Kom ihåg att operationen att dela med noll är odefinierad.

Alla tal som kan skrivas som ett bråk av två heltal kallas för *rationella tal*. Även heltal är rationella tal, ty varje heltal kan skrivas som heltalet delat med 1. Ett tal som inte är rationellt kallas för *irrationellt*. Det är inte orimligt att undra om det över huvud taget finns irrationella tal. Det kommer vi att titta närmare på i kapitel 4.

Exempel 1.18 Vilket rationellt tal i följande par är större?

(a) 112/96 eller 6/5?

Lösning. Vi inför tecknet $\vdash$ för att ersätta <, > eller = och beräknar

$$\frac{112}{96} \vdash \frac{6}{5} \iff \frac{112 \times 5}{96} \vdash 6 \iff 112 \times 5 \vdash 6 \times 96$$
$$\iff 560 \vdash 576.$$

Eftersom 560 < 576 måste $\vdash$ vara < och därför är 112/96 < 6/5. □

(b) −182/130 eller −6/5?

Lösning. Vi använder tecknet $\vdash$ igen för att ersätta <, > eller = och beräknar

$$-\frac{182}{130} \vdash -\frac{6}{5} \iff -\frac{182 \times (-5)}{130} \dashv 6$$
$$\iff -182 \times (-5) \dashv 6 \times 130 \iff 910 \dashv 780.$$

Eftersom 910 > 780 måste $\dashv$ vara > och därför är −182/130 < −6/5. Observera att när man multiplicerar båda sidor av en olikhet med ett negativt tal, har olikheten man får motsatt riktning till den ursprungliga olikheten och därför har vi bytt $\vdash$ till $\dashv$ för att representera att den eventuella olikheten har vänt sig efter att vi multiplicerade med −5. □

Ändliga decimalutvecklingar

Vissa rationella tal går att representera som så kallade *ändliga decimalutvecklingar*. Kom ihåg att vi skrev alla heltal som summan (1.21). Vi kan utöka notationen så att den omfattar alla tal av formen

$$\sum_{j=-m}^{n} d_j \times 10^j, \tag{1.22}$$

där n och m är icke-negativa heltal och varje d_j är en *siffra*, det vill säga ett tecken 1, 2, 3, 4, 5, 6, 7, 8, 9 eller 0 ($j = -m, \ldots, n$). Det vill säga, vi tillåter även negativa tiopotenser i summan. Talet skrivs då som

$$d_n d_{n-1} \ldots d_0{,}d_{-1} \ldots d_{-m+1} d_{-m},$$

Det enda talet som varken är positivt eller negativt är noll. Därför är icke-negativa heltal lika alla positiva heltal tillsammans med noll.

där kommatecken alltid skrivs mellan d_0 och d_{-1} och kallas för ett *decimalkomma*.[7]

Negativa tal skrivs på samma sätt förutom att man lägger till ett minustecken framför talet. Det representerar en summa precis som (1.22) men man kan tänka att d_j väljs från −1, −2, −3, −4, −5, −6, −7, −8, −9 och 0 i stället för positiva tal, och sedan har det gemensamma minustecknet faktoriserats.

Exempel 1.19 Skriv ⅛ som en ändlig decimalutveckling.

Lösning. Vi får använda oss av lång division som man lär sig i skolan. Uppställningen kallas ofta för *liggande stolen*.

$$\begin{array}{r|l}
\multicolumn{1}{r}{0{,}125} & \\
\cline{1-1}
1{,}000 & 8 \\
\cline{2-2}
0{,}800 & \\
\cline{1-1}
0{,}200 & \\
0{,}160 & \\
\cline{1-1}
0{,}040 & \\
0{,}040 & \\
\cline{1-1}
0{,}000 &
\end{array}$$

Därför är ⅛ = 0,125. □

7 I vissa språk skriver man en punkt (det vill säga $d_n d_{n-1} \ldots d_0 . d_{-1} \ldots d_{-m+1} d_{-m}$) i stället för ett komma.

Exempel 1.20 Skriv 0,225 som ett bråk.

Lösning. Enligt (1.22) är

$$0{,}225 = 2 \times 10^{-1} + 2 \times 10^{-2} + 5 \times 10^{-3}$$

och vi kan beräkna att

$$\begin{aligned} 2 \times 10^{-1} + 2 \times 10^{-2} + 5 \times 10^{-3} &= \frac{2}{10} + \frac{2}{100} + \frac{5}{1000} \\ &= \frac{200 + 20 + 5}{1000} = \frac{225}{1000} = \frac{9}{40}. \end{aligned}$$

□

Exempel 1.21 Skriv ⅓ som en ändlig decimalutveckling.

Lösning. Vi räknar ut med hjälp av lång division:

$$\begin{array}{l|l}
\multicolumn{1}{l}{0{,}333\ldots} & \\
\hline
1{,}000 & 3 \\
\cline{2-2}
0{,}900 & \\
\cline{1-1}
0{,}100 & \\
0{,}090 & \\
\cline{1-1}
0{,}010 & \\
0{,}009 & \\
\cline{1-1}
0{,}001 & \\
\vdots &
\end{array}$$

Vi ser att första icke-noll siffra i resten är alltid ett. Den enda skillnaden från ett steg till ett annat är att ettan förflyttas ett steg till höger. Algoritmen kommer aldrig att ta slut och därför har vi inte lyckats skriva ⅓ som en ändlig decimalutveckling! □

Att det inte ens är möjligt att skriva ⅓ som en ändlig decimalutveckling bevisar vi inte här, men vi kommer att prata mer om decimalutvecklingar som aldrig tar slut i avsnitt 2.2.

1.3 Sammanfattning och uppgifter

I kapitel 1 har vi:

- diskuterat grundläggande begrepp inom logik såsom påståenden, antaganden, slutsatser, implikation, ekvivalens, negation och kontraposition;
- diskuterat grundläggande matematiska begrepp såsom aritmetiska operationer och deras räkneregler, vad som menas med att lösa en ekvation, och hur vi delar upp matematisk text med hjälp av definitioner, satser, bevis, med mera;
- definierat heltalspotenser och visat utifrån definitionen vissa räkneregler, samt räknat ut formler för vissa summor;
- sett hur vi kan på olika sätt representera heltal samt beräknat med rationella tal och ändliga decimalutvecklingar.

UPPGIFTER

1.1 Betrakta följande två påståenden.

(a) ”Om $x \geq 5$ så är $x^2 \geq 25$.”

(b) ”Alla $x \geq -5$ uppfyller olikheten $x^2 \geq 25$.”

Det ena är sant och det andra är falskt. Vilket är falskt? Bevisa det. Motivera utifrån grundläggande egenskaper om tal att det andra är sant. Varför funkar inte samma argument för att bevisa att det falska påståendet är sant?

1.2 Förklara, utifrån grundläggande egenskaper om tal, följande implikationer.

(a) $x \geq 3 \implies x^2 \geq 9$

(b) $x^2 \geq 20 \impliedby x \geq 5$

(c) $x > 5 \implies x(x-2) > 15$

1.3 Skriv kontrapositionen till varje påstående i uppgift 1.2.

1.4 Visa att följande implikationer är felaktiga.

(a) $x \geq 5 \impliedby x^2 \geq 25$

(b) $x > 5 \impliedby x(x-2) > 15$

1.5 Skriv negationen av följande påståenden.

(a) Det finns (minst) ett heltal n sådant att $n^2 - 3n + 1 < 0$.

(b) För alla tal x och y är $x + y = y + x$.

(c) $x > 8 \implies x^2 - 14x + 48 > 0$.

1.6 Ett heltal n kallas för *jämnt* om $n = 2k$ för något annat heltal k. Om n inte är jämnt så kallas det *udda*.

(a) Vilka av 1^2, 2^2, 3^2, 4^2 och 5^2 är jämna?

(b) Undersök ett tillräckligt villkor för heltalet n för att n^2 ska vara jämnt.

(c) Undersök ett nödvändigt och tillräckligt villkor för ett heltal n för att n^2 ska vara jämnt.

1.7 Visa att om

(a) n_1 delat med 82 har rest 49, och

(b) n_1 delat med 82 har rest 9

så har $n_1 n_2$ delat med 82 rest 31.

1.8 Ta fram, med bevis, villkor för tal a och b, och heltal n så att $a < b$ medför $a^n < b^n$.

1.9 Låt r och s vara rationella tal.

(a) Undersök om $r + s$ är rationellt eller irrationellt. Motivera ditt svar.

(b) Undersök om rs är rationellt eller irrationellt. Motivera ditt svar.

1.10 (a) Skriv de decimala heltalen 7, 17, 12 och 32 i det binära talsystemet (det vill säga i bas 2).

(b) Skriv de decimala heltalen 7, 17, 12 och 32 i det ternära talsystemet (det vill säga i bas 3).

1.11 (a) Bevisa med hjälp av sats 1.14 att

$$\sum_{j=-m}^{n} d_j 10^j < 10^{n+1} \tag{1.23}$$

där varje d_j kan vara en av siffrorna 1, 2, 3, 4, 5, 6, 7, 8, 9 och 0.

(b) Vad säger (1.23) om vi tolkar summan i vänsterledet som ett decimaltal? Använd din tolkning för att skriva ett alternativt bevis av (1.23). Förklara varför det är rimligt att högerledet i (1.23) inte beror på m utifrån din tolkning.

Verktyg för bevisföring

2.1 Mängder, följder och induktion

Med den stabila grunden av logik och aritmetik i fickan kan vi gå vidare och titta på begrepp och verktyg som är matematikens arbetshästar. Som mål för avsnittet bevisar vi binomialsatsen (sats 2.13). Satsen samlar avsnittets idéer: Den är ett uttryck för kombinatoriska idéer och dess bevis är ett bra exempel på den kraftfulla bevisföringstekniken – matematisk induktion.

MÄNGDER

Vi har redan diskuterat vissa samlingar av tal, till exempel heltal och rationella 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, skriver man

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

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 $\mathbf{N}$.[1] Det vill säga

$$\mathbf{N} = \{0, 1, 2, 3, 4, 5, \ldots\}.$$

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

Alla heltal tecknas som

$$\mathbf{Z} = \{\dots, -5, -4, -3, -2, -1, 0, 1, 2, 3, 4, 5, \dots\}.$$

Bokstaven **Z** kommer här från det tyska ordet *Zahlen*. Positiva heltal tecknas som

$$\mathbf{Z}_+ = \{1, 2, 3, 4, 5, \dots\}.$$

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 sådana att $n = 2k$ för ett heltal k. Det kan man skriva som

$$\{n \in \mathbf{Z} \,|\, n = 2k \text{ för något } k \in \mathbf{Z}\}.$$

Alltså betyder ”|” ”sådan att”. Ibland skriver man ”:” i stället för ”|”. *Udda tal* är då

$$\{n \in \mathbf{Z} \,|\, n = 2k + 1 \text{ för något } k \in \mathbf{Z}\}.$$

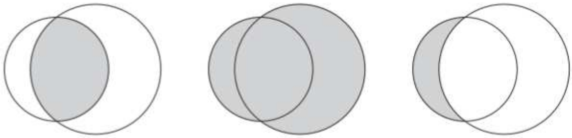
Med den här notationen kan vi definiera mängden av alla *primtal* som

$$\left\{ n \in \mathbf{Z}_+ \;\middle|\; \begin{array}{l} n > 1 \text{ och om } n = pq \text{ med } p, q \in \mathbf{Z}_+ \\ \text{då är } p \text{ eller } q \text{ lika med } n \end{array} \right\},$$

det vill säga, primtal är alla positiva heltal $n > 1$ som inte har några faktorer utom 1 och n själv. Rationella tal är

$$\mathbf{Q} = \{n/m \,|\, n, m \in \mathbf{Z} \text{ och } m \neq 0\}.$$

Här kan vi förtydliga vad vi menar med mängden **Q**. Den är mängden av alla tal som kan skrivas som n/m där n och m är heltal (och m är förstås inte lika med noll). Det är inte samma sak som mängden av alla möjliga kvoter av heltal eftersom många olika kvoter är enligt enkel aritmetik lika. Till exempel är $\frac{1}{4}$, $\frac{3}{12}$ och $\frac{7}{28}$ olika kvoter men vi kan beräkna att de är lika med varandra, och därför betraktas som samma element i **Q**. Även när vi skriver talet som decimalutvecklingen 0,25 är det fortfarande ett rationellt tal eftersom det *kan* med hjälp av aritmetik skrivas om till en kvot av två heltal.



(a) Snittet $A \cap B$. (b) Unionen $A \cup B$. (c) Differensen $A \setminus B$.

FIGUR 2.1 Resultatet (i grått) av operationer på mängderna A (den mindre skivan) och B (den större).

En viktig typ av mängd är ett *intervall*. För två tal a och b med $a < b$ är ett intervall alla tal mellan a och b. Vi har olika notation beroende på om a och/eller b tillhör intervallet:

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

Längden av intervallen ovan definieras lika med $b - a$. Vi betecknar också

Kom ihåg att $a \le x \le b$ betyder att x uppfyller både $a \le x$ och $x \le b$, och liknande för de andra olikheterna.

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

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 $x \in M$. Uppenbarligen är intervallet $[1,3]$ en delmängd av $[0,\infty)$, så vi kan till exempel skriva $[1,3] \subseteq [0,\infty)$. Vi säger att två mängder A och B är lika och skriver $A = B$ om $A \subseteq B$ och $B \subseteq A$; det är detsamma som att säga ”$x \in A$ om och endast om $x \in B$”.

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

Det finns också operationer på mängder som kallas *snitt* och *union*. Vi illustrerar operationerna i figur 2.1 där vi representerar mängder som skivor i planet. För två mängder A och B definieras *snittet* av A och B som

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

Unionen av A och B definieras som

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

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

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

Exempel 2.1 Låt $A = \{3, 2, 14\}$, $B = \{3, 2, 7, 9\}$ och $C = \{3, 7, 9, 14, 17\}$. Då är:

(a) $A \setminus B = \{14\}$;
(b) $B \setminus A = \{7, 9\}$;
(c) $A \cup B = \{2, 3, 7, 9, 14\}$;
(d) $(A \cup B) \cup C = \{2, 3, 7, 9, 14, 17\}$;
(e) $A \cup (B \cup C) = \{2, 3, 7, 9, 14, 17\}$;
(f) $(A \cup B) \cap C = \{3, 7, 9, 14\}$; och
(g) $A \cup (B \cap C) = \{2, 3, 7, 9, 14\}$.

Följande sats beskriver några enklare egenskaper av snitt, union och differens.

Sats 2.2 *Givna godtyckliga mängder A, B och C är*

(a) $(A \cap B) \cap C = A \cap (B \cap C)$,
(b) $(A \cup B) \cup C = A \cup (B \cup C)$,
(c) $C \setminus (A \cup B) = (C \setminus A) \cap (C \setminus B)$ *och*
(d) $C \setminus (A \cap B) = (C \setminus A) \cup (C \setminus B)$.

Bevis. Vi bevisar (d) och lämnar beviset av de andra delarna som en övning för läsaren.

Att x tillhör $C \setminus (A \cap B)$ betyder att x tillhör C men varken A eller B. Det är detsamma som att säga x tillhör C men inte A, och C men inte B. I sin tur är det samma sak som att säga x tillhör $(C \setminus A) \cup (C \setminus B)$. Därför innehåller $C \setminus (A \cap B)$ och $(C \setminus A) \cup (C \setminus B)$ samma element och därmed är (d) bevisad. □

Del (a) och (b) i sats 2.2 säger att det inte spelar någon roll var man skriver parenteser när man tar unionen eller när man tar snittet av tre (eller fler) mängder. Därför är det inte tvetydigt att skriva $A \cap B \cap C$ eller $A \cup B \cup C$ (utan parenteser) eftersom det skulle ha samma betydelse oavsett var man väljer att skriva parenteserna. Vi såg ett exempel på detta i del (d) och (e) av exempel 2.1.

Följande exempel visar att detsamma inte kan sägas om differensen.

Exempel 2.3 Återigen tar vi $A = \{3, 2, 14\}$, $B = \{3, 2, 7, 9\}$ och $C = \{3, 7, 9, 14, 17\}$.

(a) Med hjälp av exempel 2.1(a) kan vi beräkna att $(A \setminus B) \setminus C$ inte innehåller något element. Mängden som inte innehåller något element alls kallas för den *tomma mängden* och tecknas $\varnothing$. Därför är

$$(A \setminus B) \setminus C = \varnothing.$$

(b) Vi kan först beräkna att $B \setminus C = \{2\}$ och därför är

$$A \setminus (B \setminus C) = \{3, 14\}.$$

Att vi inte behöver skriva parenteser när vi tar antingen snittet eller unionen av tre eller fler mängder betyder att vi kan introducera motsvarade notation till $\sum$ för summor. För en samling av mängder A_j indexerade med $j = 1, 2, \ldots, N$ skriver vi

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

och

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

Vi kommer att använda oss av notationen i kapitel 3 när vi betraktar arean av en mängd.

FÖLJDER OCH INDUKTION

En *följd* är en mängd av tal där elementen har en ordning. Vanligtvis är elementen indexerade med antingen positiva heltal eller naturliga tal men andra indexeringar kan förekomma.[2] En följd tecknas med parenteserna ”(” och ”)” i stället för ”{” och ”}” för att skilja den från en vanlig mängd som inte har någon ordning.

2 Jämför till exempel fakultet på sidorna 51–52 med trunkerade decimalutvecklingar på sidan 67.

Exempel 2.4 Några exempel på följder:

(a) Positiva heltal $(1, 2, 3, 4, 5, \dots)$ är i sig en följd med vanlig numerisk ordning;

(b) $(2, 4, 6, 8, 10, \dots)$ är följden av alla positiva jämna tal i numerisk ordning;

(c) $(1, 4, 9, 16, 25, \dots)$ 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 hela följden som $(a_n)_{n\in\mathbf{Z}_+}$ eller $(a_n)_n$ där ett nedsänkt n utanför parenteserna indikerar att n är följdens index. I exempel 2.4 är $a_n = n$, $a_n = 2n$ respektive $a_n = n^2$ formeln för den n:e termen. Och hela följden skrivs som

$$(n)_n, \quad (2n)_n \quad \text{respektive} \quad (n^2)_n.$$

En följd $(a_n)_n$ som uppfyller

$$a_m \le a_{m+1} \tag{2.1}$$

för alla $m \in \mathbf{Z}_+$ kallas för *växande* och en som uppfyller

$$a_m \ge a_{m+1} \tag{2.2}$$

för alla $m \in \mathbf{Z}_+$ kallas för *avtagande*. Alla följder i exempel 2.4 är då växande. Följden

$$(1, -1, 1, -1, 1, -1, 1, -1, \dots) = ((-1)^n)_n$$

är varken växande eller avtagande eftersom den varken uppfyller (2.1) eller (2.2) för alla m. Om en följd är växande eller avtagande kallas den för en *monoton* följd.

Aritmetiska följder

En *aritmetisk följd* är en följd $(a_n)_n$ som har egenskapen att

$$a_{n+1} - a_n =: d \tag{2.3}$$

är oberoende av n. Det vill säga, differensen mellan två angränsande element är en konstant d.

Exempel 2.5 Följande följder är aritmetiska:
(a) $(1, 2, 3, 4, 5, 6, 7, \dots)$
(b) $(2, 4, 6, 8, 10, 12, 14, \dots)$
(c) $(3, 10, 17, 24, 31, 38, 45, \dots)$
(d) $(60, 66, 72, 78, 84, 90, 96, \dots)$
Vad är d i de ovanstående exemplen?

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

$$a_n = a_1 + d(n-1) \tag{2.4}$$

för alla $n \in \mathbf{Z}_+$. Se exempel 2.8 nedan.

Metoden från exempel 1.12 gav oss en formel – i sats 1.13 – för summan av de första n elementen i den aritmetiska följden $(1, 2, 3, 4, 5, \dots)$. I exempel 2.7 nedan presenterar vi en annan metod för att bevisa den.

Matematisk induktion

Matematisk induktion är en metod man kan använda sig av när man vill bevisa sanningen av ett påstående $P(n)$ som beror på ett positivt heltal n.[3] Det förankras i ett axiom som kallas för *induktionsaxiomet* eller *induktionsprincipen*:

Axiom 2.6 (Induktionsaxiomet) *Låt $P(n)$ vara ett påstående som beror på ett positivt heltal n. Om*

(a) påståendet $P(1)$ är sant och
(b) för varje $m \in \mathbf{Z}_+$ har man implikationen $P(m) \implies P(m+1)$,

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

Induktionsaxiomet är framför allt ett antagande om strukturen av positiva heltal. Det säger att vi från talet 1 kan nå alla andra positiva heltal genom att upprepade gånger hoppa från det ena till det nästföljande heltalet. Det kan jämföras med fallande dominobrickor: (a) säger att vi kan knuffa omkull första dominot; och (b) säger att dominobrickor står tillräckligt nära varandra för att få en kedjereaktion och alla faller. Ibland kallas del (a) för *basfallet* och del (b) för *induktionssteget*.

Nu tillämpar vi axiomet i ett par exempel som illustrerar metoden.

3 Det är inte samma sak som induktiv slutledning och matematiskt induktion är faktiskt ett exempel på deduktiv slutledning som vi pratade om i förordet.

Exempel 2.7 Använd induktion för att ge ett annat bevis av sats 1.13: Det vill säga bevisa genom matematisk induktion att

$$\sum_{k=1}^{n} k = \frac{n(n+1)}{2} \tag*{$(2.5)_n$}$$

för varje $n \in \mathbf{Z}_+$.

Lösning. Först kontrollerar vi att $(2.5)_1$ är sant: Om $n = 1$ är

$$\sum_{k=1}^{n} k = \sum_{k=1}^{1} k = 1 = 1(1+1)/2 = n(n+1)/2$$

så är $(2.5)_1$ sant.

Nu visar vi implikationen $(2.5)_m \implies (2.5)_{m+1}$: Vi beräknar

$$\begin{aligned}\sum_{k=1}^{m+1} k &= \sum_{k=1}^{m} k + (m+1) \underset{\substack{\uparrow \\ (2.5)_m}}{=} \frac{m(m+1)}{2} + (m+1) \\ &= \frac{m(m+1) + 2(m+1)}{2} = \frac{(m+1)(m+2)}{2},\end{aligned}$$

som är $(2.5)_{m+1}$. Därför, enligt induktionsaxiomet, är $(2.5)_n$ bevisat för varje $n \in \mathbf{Z}_+$. □

Exempel 2.8 Använd matematisk induktion för att bevisa (2.4) för en aritmetisk följd $(a_n)_n$.

Lösning. Uppgiften här är att härleda (2.4) från (2.3) eftersom (2.3) är definitionen att $(a_n)_n$ är en aritmetisk följd.

Vi kan snabbt kontrollera att

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

så (2.4) är sant om $n = 1$. Sedan antar vi att (2.4) gäller för $n = k$ och betraktar a_{k+1}. Från (2.3) med $n = k$ vet vi att $a_{k+1} = a_k + d$ och sedan kan vi använda oss av (2.4) med $n = k$ och säga

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

som är (2.4) med $n = k+1$. Vi har därför visat att (2.4) med $n = k$ medför (2.4) med $n = k+1$. Enligt induktionsaxiomet drar vi slutsatsen att (2.4) är sant för alla $n \in \mathbf{Z}_+$. □

Det är värt att jämföra exempel 2.7 och 2.8 med vårt tidigare bevis av sats 1.8(c). Det kan kännas som att ett induktionsbevis av (2.4) är överflödigt och ett bevis som det av sats 1.8(c) skulle vara lämpligare i exempel 2.8. Men fördelen med ett induktionsbevis är att metoden är systematisk och lätt kan tillämpas på mindre självklara saker såsom $(2.5)_n$ som vi såg i exempel 2.7. Lärdomen är att en abstraktion kan vara kraftfull. Metoden befäster våra logiska tankar och vi kommer längre än med mer naiva intuitiva argument. Kan du ge ett induktionsbevis av sats 1.8(c)? Jämför det med beviset vi gav direkt efter satsen.

Geometriska följder

Nästa följdtyp vi vill betrakta är en *geometrisk följd*. Den är en följd $(a_n)_n$ som uppfyller

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

för alla n och något $r \in \mathbf{R}$. I fallet då inga a_n är noll, är kvoten mellan två angränsande element lika med r och därför oberoende av n.

Exempel 2.9 Följande följder är geometriska följder:

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

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

(c) $(5, 25, 125, 625, 3\,125, 15\,625, 78\,125, \dots)$

(d) $(6, 2, 2/3, 2/9, 2/27, 2/81, 2/243, \dots)$

Vad är r i de ovanstående exemplen?

På liknande sätt som beviset av (2.4) kan man bevisa att

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

för alla $n \in \mathbf{Z}_+$. I sats 1.14 och anmärkning 1.15 har vi även formler för summan av de första n elementen i en geometrisk följd.

Fakultet

Vi definierar en följd vars n:te term tecknas $n!$ och uttalas ”n-fakultet” genom formeln

$$n! := 1 \times 2 \times \cdots \times n$$

Inom matematik är ordet *nollte* ordningstalet till talet noll. Det används också i fysik i till exempel *Termodynamikens nollte huvudsats*.

för varje $n \in \mathbf{Z}_+$. För den här följden är det bra att även definiera en *nollte* term. Vi sätter $0! = 1$. Det innebär att följden egentligen är indexerad med naturliga tal i stället för positiva heltal, men det gör ingen stor skillnad. Vi kunde lika väl definiera $0!$ som den första, $1!$ som den andra, och så vidare, men det ger inte så mycket mer än en lättglömd indexering. Vi kan till exempel räkna ut att

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

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 har man formeln

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

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

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

sätt.

Exempel 2.10 Man kan lätt se att följden $(n!)_{n\in\mathbf{N}}$ är växande. Men vi noterar även att den växer väldigt fort i n. Till exempel är $10!$ lika stort som antalet sekunder på sex veckor! Vi vet att det går 60 sekunder på en minut, 60 minuter på en timme, 24 timmar på ett dygn och sju dagar på en vecka. Därför finns det

$$\begin{aligned} &60 \times 60 \times 24 \times 7 \times 6 \\ &= (2 \times 2 \times 3 \times 5) \times (2 \times 2 \times 3 \times 5) \times (2 \times 2 \times 2 \times 3) \times 7 \times (2 \times 3) \\ &= 1 \times 2 \times 3 \times 4 \times 5 \times 6 \times 7 \times 8 \times 9 \times 10 = 10! \end{aligned}$$

sekunder på sex veckor.

KOMBINATORIK, BINOMIALSATSEN OCH ANDRA SUMMOR

Än så länge har vi använt matematisk induktion för att bevisa formler som antingen var relativt enkla eller i fallet när vi hade ett alternativt bevis. Nu använder vi induktion för att bevisa en formel som vi inte har träffat på tidigare.

Exempel 2.11 Bevisa att

$$\sum_{i=1}^{n} i^2 = \frac{n(n+1)(2n+1)}{6}. \tag{2.6}$$

Lösning. Vi använder oss av induktion. Först kan vi verifiera (2.6) när $n = 1$. Då har vi att

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

så (2.6) är sant för $n = 1$.

Nu antar vi att $m \in \mathbf{Z}_+$ är ett tal sådant att (2.6) gäller för $n = m$. Vi vet att det finns minst ett sådant m eftersom vi precis har bevisat likheten (2.6) då $n = 1$. Men hjälp av antagandet försöker vi bevisa (2.6) i fallet $n = m + 1$: Vi beräknar

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

Men

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

Därför är

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

som säger att (2.6) gäller för $n = m + 1$.

Så vi har bevisat att

> om (2.6) gäller för $n = m$ för något $m \in \mathbf{Z}^+$ så gäller (2.6) för $n = m + 1$.

Vi har också kontrollerat att (2.6) gäller för $n = 1$. Så med hjälp av axiom 2.6 drar vi slutsatsen att (2.6) gäller för alla $n \in \mathbf{Z}_+$. □

På sidorna 51–52 definierade vi fakulteten. Nu definierar vi *binomialkoefficienten* som

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

för $n,k \in \mathbf{N}$ men $k \leq n$. Den verkar kanske 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. Hur ser man detta?

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 $n(n-1)(n+2)\ldots(n-(k-1))$. Men varje val är en ordnad mängd av k element och en mängd av k element kan ordnas på $k!$ olika sätt. Därför har vi, om vi inte vill ta hänsyn till i vilken ordning man väljer dem, räknat $k!$ gånger för många och vi behöver dela $n(n-1)(n+2)\ldots(n-(k-1))$ med $k!$. Därför är antalet sätt att välja ut k element från en mängd av n element

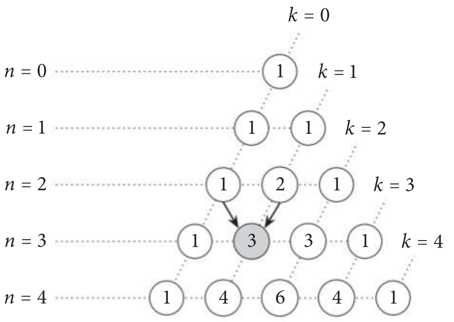
$$\frac{n(n-1)\ldots(n-(k-1))}{k!} = \frac{n!}{k!(n-k)!} = \binom{n}{k}$$

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

Sats 2.12 (Pascals identitet) *För positiva hela tal n och k så att $k \leq n$ har man att*

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

Vi lämnar beviset av sats 2.12 som en övning för läsaren – se uppgift 2.6 – men vi noterar att satsen är precis den regel som ligger bakom Pascals



FIGUR 2.2 Raderna $n = 0$ till 4 av Pascals triangel. Plats k i den n:e raden är $\binom{n}{k}$ där $k = 0, 1, \ldots, n$.

triangel. *Pascals triangel* är tal utformade i en triangel – se figur 2.2 där vi har ritat en cirkel kring varje tal. På de yttersta platserna, både till vänster och höger i varje rad, placerar vi ettor. Vi fyller sedan på de andra platserna enligt regeln att varje tal är summan av de två talen direkt ovanför. I figur 2.2 illustrerar vi hur vi får den gråa trean som summan av ett och två. Genom att fylla platserna med ifyllningsregeln uppifrån och ner kan vi fylla på så många rader vi vill med tal.

För att sedan relatera sats 2.12 till Pascals triangel indexerar vi raderna med ett index n så att på den högsta raden är $n = 0$, på den näst högsta är $n = 1$, och så vidare. På samma sätt indexerar vi platserna längst en rad med indexet k, så att på platsen till vänster i varje rad är $k = 0$, på därnäst är $k = 1$, fram till den yttersta till höger där $k = n$. Med hjälp av sats 2.12 kan vi se att den k:te platsen på den n:te raden är precis $\binom{n}{k}$: Vi kan beräkna att

$$\binom{n}{0} = \frac{n!}{0!(n-0)!} = \frac{n!}{n!} = 1 \quad \text{och} \quad \binom{n}{n} = \frac{n!}{n!(n-n)!} = \frac{n!}{n!} = 1$$

för varje $n \in \mathbf{Z}$, så stämmer det med valet att på både den nollte och n:e platsen på varje rad ska det vara en etta. Och sats 2.12 är helt enkelt ifyllningsregeln för de andra platserna.

Sats 2.13 (Binomialsatsen) *För två icke-noll tal x och y, och $n \in \mathbf{Z}_+$ har man att*

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^{n-k} y^k. \tag{2.7}$$

Bevis. Vi använder induktion. Om $n = 1$ är högerledet av (2.7)

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

som stämmer med vänsterledet. Då antar vi att (2.7) är sant för ett givet $n \in \mathbf{Z}_+$ och sedan räknar vi ut att

$$\begin{aligned}
&(x+y)^{n+1} \\
&= (x+y)(x+y)^n \underset{\substack{\uparrow \\ \text{tack vare antagandet}}}{=} (x+y) \sum_{k=0}^{n} \binom{n}{k} x^{n-k} y^k \\
&= \sum_{k=0}^{n} \binom{n}{k} x^{n+1-k} y^k + \sum_{k=0}^{n} \binom{n}{k} x^{n-k} y^{k+1} \\
&= x^{n+1} + \sum_{k=1}^{n} \binom{n}{k} x^{n+1-k} y^k + y^{n+1} + \sum_{k=0}^{n-1} \binom{n}{k} x^{n-k} y^{k+1} \\
&= x^{n+1} + y^{n+1} + \sum_{k=1}^{n} \binom{n}{k} x^{n+1-k} y^k + \sum_{k=1}^{n} \binom{n}{k-1} x^{n+1-k} y^k \\
&= x^{n+1} + y^{n+1} + \sum_{k=1}^{n} \left(\binom{n}{k} + \binom{n}{k-1} \right) x^{n+1-k} y^k \\
&\underset{\substack{\uparrow \\ \text{enligt lemma 2.12}}}{=} x^{n+1} + y^{n+1} + \sum_{k=1}^{n} \binom{n+1}{k} x^{n+1-k} y^k \\
&= \sum_{k=0}^{n+1} \binom{n+1}{k} x^{n+1-k} y^k .
\end{aligned}$$

Enligt induktion är satsen bevisad. □

2.2 Reella tal

Trots att vi har varit så noggranna med de logiska argumenten vi får genomföra och de operationerna vi får göra med tal, har vi inte pratat så mycket om vad vi menar med tal. Vår utgångspunkt har varit aritmetik där vi är vana vid att beräkna med heltal och även rationella tal, men snart (i kapitel 4) kommer vi se att det inte täcker alla våra behov av tal. Därför försöker vi här precisera vad vi menar med tal och vilka egenskaper de har. Det tar lite tid att införa alla begrepp men redan nu kan vi ange notationen **R** för talmängden vi hittills har hänvisat till med namnet tal, det vill säga för talmängden av *reella tal*.

EGENSKAPER HOS DELMÄNGDER AV REELLA TAL

Vi börjar avsnittet med en diskussion om ytterligare egenskaper hos delmängder av tal. Diskussionen kommer att underlätta nästa avsnitt där vi ger en definition av reella tal. Förebilden för reella tal är en *kontinuerligt rakt linje*, vad det nu kan vara för någonting. Mängden av reella tal kallas ibland även för *reella linjen*.

Största element

Alla par av reella tal kan jämföras med olikheten $\leq$. Det vill säga: För två givna tal a och b är $a \leq b$ eller $b \leq a$. Vi får även $a \leq b$ *och* $b \leq a$ när de två talen är lika. Med hjälp av olikheten kan man betrakta egenskaper hos delmängder av tal. Exempelvis, kan vi definiera begreppet *största elementet*.

Definition 2.14 Ett *största element* i en mängd $A \subseteq \mathbf{R}$ är ett element $x \in A$ sådant att $a \leq x$ för alla $a \in A$.

Definitionen leder till vissa rimliga frågor. Kan en mängd har ett största element över huvud taget? Kan en mängd har fler än ett största element? Har alla mängder största element? Vi kan enkelt svara på första frågan: För mängden $M = \{5, 6, 7, 10\}$ kan man kontrollera direkt att 10 är ett största element. Eftersom mängden innehåller bara fyra element kan man jämföra 10 med alla element i M. Vi ser att

$$5 \leq 10, 6 \leq 10, 7 \leq 10 \quad \text{och} \quad 10 \leq 10.$$

Vi betonar att ett största element i en mängd måste vara ett element i själva mängden, så här behöver vi även dubbelkolla att $10 \in M$. Alla villkor av definitionen är uppfyllda och därmed är 10 ett största element av M.

Man kan även se att mängden inte har något annat element som är ett största element. Det är faktiskt någonting som gäller för alla mängder, vilket följande lemma visar.

Lemma 2.15 *En mängd kan ha högst ett största element. När mängden A har ett största element betecknar vi det* $\max A$.

Bevis. Om vi betraktar två största element x och y av samma mängd så vet vi att $x \leq y$ ty y är ett största element och $y \leq x$ ty x är ett största element. Därför får vi att $x \leq y \leq x$ och därför att $x = y$. □

Motsägelsebevis

Däremot är det inte nödvändigt att en mängd har ett största element. För att bevisa det inför vi en ny bevisföringsmetod. Den kallas för ett *motsägelsebevis* och fungerar så här: För att härleda ett påstående P antar man först att det är falskt, det vill säga man antar $\neg P$. Sedan bevisar man att $\neg P$ tillsammans med de givna förutsättningarna leder till en motsägelse. Därifrån kan man dra slutsatsen att P är giltigt.[4]

Det är en indirekt metod att bevisa någonting, det vill säga att man inte drar P som en direkt slutsats, utan snarare bevisar man att $\neg P$ leder till en motsägelse. Vi illustrerar metoden med ett exempel.

Exempel 2.16 Ge ett motsägelsebevis av det faktum att intervallet $]1,4[$ inte har något största element.

Lösning. Vi börjar med att anta att slutsatsen är falsk, det vill säga vi antar att $]1,4[$ har ett största element – kalla det för x. Per definition är $x \in]1,4[$, det vill säga $1 < x < 4$. Betrakta talet

$$y := (x+4)/2.$$

Eftersom $1 < x$ är $1 < (1+4)/2 < (x+4)/2 = y$ och eftersom $x < 4$ är $y = (x+4)/2 < (4+4)/2 = 4$, är y ett element av intervallet $]1,4[$. Dessutom är

$$x = (x+x)/2 < (x+4)/2 = y,$$

så y är större än x. Det är en motsägelse till antagandet att x är intervallets största element eftersom definition 2.14 kräver att $y \leq x$ för alla $y \in A$. Därför stämmer inte antagandet att x finns och vi drar slutsatsen att $]1,4[$ har inget största element. $\square$

Analogt med största element kan vi definiera begreppet *minsta element* till en delmängd av $\mathbf{R}$.

Definition 2.17 Ett *minsta element* i en mängd $A \subseteq \mathbf{R}$ är ett element $x \in A$ sådant att $x \leq a$ för alla $a \in A$. När ett sådant tal x finns betecknar vi det $x = \min A$.

Precis som i vår undersökning av största element kan man visa att en delmängd av $\mathbf{R}$ har högst ett minsta element och vissa delmängder har inget minsta element alls.

4 Det finns vissa filosofiska frågor kring metoden, men det ligger utanför bokens syften.

Uppåt och nedåt begränsade mängder

Ett svagare men något mer användbart begrepp än när en mängd har en största eller minsta element är när en mängd är begränsad uppåt respektive nedåt. Det är ett svagare begrepp eftersom en övre eller undre begränsning av en mängd inte måste tillhöra själva mängden.

Definition 2.18 Låt $A \subseteq \mathbf{R}$ vara en icke-tom 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.

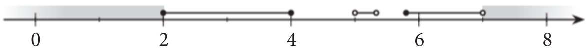
Exempel 2.19 Visa att 4 är en undre begränsning till $B = \{y \in \mathbf{R} \mid y = x^2 - 4x + 9, x \in \mathbf{R}\}$.

Lösning. 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 $y \in B$. □

Omskrivningen $x^2 - 4x + 9 = (x-2)^2 + 5$ är ett exempel på *kvadratkomplettering*. Den är särskilt användbar eftersom den förvandlar ett uttryck som innehåller två kopior av x till ett som innehåller en kopia. I allmänhet är kvadratkomplettering att sätta likhetstecken mellan $x^2 + px + q$ och $(x+a)^2 + b$. Genom att utveckla andra uttrycket ser vi att vi måste ta $a = p/2$ och $b = q - p^2/4$. (Jämför med sats 3.14)

Om man har hittat en undre begränsning c till en mängd då är alla tal mindre än c också undre begränsningar till samma mängd (se figur 2.3). Så i exempel 2.19 är alla tal mindre än 4 också undre begränsningar till mängden B. Om vi tittar noggrant på exempel 2.19 ser vi att vi även bevisade att 5 var en undre begränsning till B och därifrån följde det att 4 var en undre begränsning. Samma argument fungerar för vilket tal som helst mindre än 5.

Man kan säga något liknande om övre begränsningar: Om C är en övre begränsning till en mängd A då är alla tal större än C övre begränsningar till A. Så om vi hittar en övre (eller undre) begränsning till en mängd A finns det en hel mängd av andra övre (respektive undre) begränsningar till A. Tydligen kan mängden av övre (undre) begränsningar inte ha ett största (respektive minsta) element, men kan den ha ett minsta (respektive största) element? Och i så fall under vilka förutsättningar?



FIGUR 2.3 Delmängden $[2,4] \cup]5, \frac{16}{3}[\cup [\frac{29}{5}, 7[$ av reella tal tillsammans med sina (gråa) mängder $[7,\infty[$ av övre begränsningar respektive $]\infty,2]$ av undre begränsningar. Mängden av övre begränsningar fortsätter ut ur bilden åt höger och undre begränsningar fortsätter ut ur bilden åt vänster. Observera att ändpunkterna 2 och 7 är undre respektive övre begränsningar oavsett om de tillhör till mängden eller inte.

Definition 2.20 Låt $A \subseteq \mathbf{R}$ vara en icke-tom mängd. Om det finns ett tal $u \in \mathbf{R}$ sådant att

$$u \text{ är en övre begränsning av } A \tag{2.8}$$

och

$$u \leq \tilde{u} \text{ för alla andra övre begränsningar } \tilde{u} \text{ av } A \tag{2.9}$$

kallas u för en *minsta övre begränsning* (eller *supremum*) till A. Om u finns skriver man

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

Om det finns ett tal $\ell \in \mathbf{R}$ sådant att

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

och

$$\tilde{\ell} \leq \ell \text{ för alla andra undre begränsningar } \tilde{\ell} \text{ av } A$$

kallas ℓ för en *största undre begränsning* (eller *infimum*) till A. Om ℓ finns skriver man

$$\ell = \inf A \quad \text{eller} \quad \ell = \inf_{a \in A} a.$$

Enligt definitionen 2.18 är (2.8) ekvivalent med påståendet

$$a \leq u \text{ för alla } a \in A. \tag{2.10}$$

Villkor (2.9) kan omformuleras som sin kontraposition: Varje tal strängt mindre än u är inte en övre begränsning. Man kan skriva ett godtyckligt tal mindre än u i formen $u-\varepsilon$ där ε (uttalas: epsilon; se även bilaga B) är ett godtyckligt positivt tal. Med sådan notation skriver vi kontrapositionen av (2.9) som påståendet

$$\text{till varje } \varepsilon > 0 \text{ finns det ett } a \in A \text{ sådant att } u - \varepsilon < a. \tag{2.11}$$

Därför är talet u en minsta övre begränsning till A om och endast om både (2.10) och (2.11) är uppfyllda. Med ett liknande resonemang är ℓ en största undre begränsning till A om och endast om

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

och

$$\text{till varje } \varepsilon > 0 \text{ finns det ett } a \in A \text{ sådant att } a < \ell + \varepsilon. \tag{2.12}$$

Observera att valet av a i (2.11) och (2.12) kan (och i de flesta fall kommer att) bero på ε.

Nu utnyttjar vi (2.10) och (2.11) i ett par exempel.

Exempel 2.21 Visa att $\sup A = 0$ för $A = \{x \in \mathbf{R} \mid x < 0\}$.

Lösning. Ett tal a tillhör A om villkoret $a < 0$ är uppfyllt. Därför vet vi även att $a \leq 0$ för alla $a \in A$ och därmed är (2.10) uppfyllt. För att visa (2.11) betraktar vi ett godtyckligt $\varepsilon > 0$. Om vi väljer $a_\varepsilon := -\varepsilon/2$ då är $a_\varepsilon \in A$ (ty $a_\varepsilon < 0$) och $0 - \varepsilon < -\varepsilon/2 = a_\varepsilon$ så för varje ε har vi hittat ett $a_\varepsilon \in A$ sådant att $0 - \varepsilon < a_\varepsilon$ och (2.11) är uppfylld med $u = 0$ och $a = a_\varepsilon$.□

Exempel 2.22 Visa att $\sup A = 10$ då $A = \{7 - 6n - 3n^2 \mid n \in \mathbf{Z}\}$.

Lösning. Först vill vi bevisa (2.10) med $u = 10$. Eftersom elementen av A indexeras med n måste vi visa att

$$7 - 6n - 3n^2 \leq 10 \tag{2.13}$$

för *alla* $n \in \mathbf{Z}$. Men

$$7 - 6n - 3n^2 = -3(n+1)^2 + 10 \leq 10$$

så vi har visat (2.13) för alla $n \in \mathbf{Z}$.

Nu vill vi visa (2.11) med $u = 10$. Det innebär att vi för *varje* $\varepsilon > 0$ måste hitta *något* $n \in \mathbf{Z}$ så att

$$10 - \varepsilon < 7 - 6n - 3n^2. \tag{2.14}$$

Genom att skriva om $7 - 6n - 3n^2 = -3(n+1)^2 + 10$ ser vi att den är ekvivalent med

$$10 - \varepsilon < -3(n+1)^2 + 10$$

och i sin tur är ekvivalent med

$$3(n+1)^2 < \varepsilon. \tag{2.15}$$

För varje givet $\varepsilon > 0$ är det möjligt att välja ett $n \in \mathbf{Z}$ så att (2.15) gäller – det räcker att välja n så att $3(n+1)^2 = 0$, det vill säga $n = -1$. Därför har vi visat att för varje $\varepsilon > 0$ finns det något $n \in \mathbf{Z}$ så att (2.14) gäller.

Med (2.13) och (2.14) har vi bevisat att $\sup A = 10$. □

Så i det här fallet och till skillnad från exempel 2.21 har vi kunnat välja samma n för varje $\varepsilon > 0$. Valet av a i (2.11) (och därför n här) beror i de flesta fall på ε, men inte alltid! Och just här har vi lyckats hitta ett enkelt val som funkar oavsett vad ε är.

Nu utreder vi några allmänna egenskaper hos infimum och supremum av delmängder av $\mathbf{R}$.

Sats 2.23 *En mängd har högst en supremum. En mängd har högst en infimum.*

Bevis. Satsen följer från lemma 2.15 eftersom supremum (eller infimum) är det minsta (respektive största) elementet till mängden av övre (respektive undre) begränsningar.

Men vi kan även bevisa satsen med hjälp av ett motsägelsebevis: Vi bevisar bara det första påståendet och lämnar det andra som en övning för läsaren.

Anta att det finns två olika supremum u_1 och u_2 till en mängd A. Eftersom de inte är lika vet vi att en kommer att vara större än den andra: säg $u_1 < u_2$. Eftersom u_2 är en supremum till mängden A kan vi tillämpa (2.11) med $\varepsilon = (u_2 - u_1)/2$ och dra slutsatsen att det finns $a \in A$ sådant att

$$u_1 \underset{\substack{\uparrow \\ u_1 < u_2}}{<} (u_2 + u_1)/2 = u_2 - (u_2 - u_1)/2 \underset{\substack{\uparrow \\ (2.11)}}{<} a.$$

Men (2.10) medför att $u_1 \geq a$ så vi har hittat en motsägelse till hypotesen att det finns två supremum till A. Därför drar vi slutsatsen att det inte kan finnas mer än en supremum till en mängd. □

I följande sats visar vi att den minsta övre begränsningen och största undre begränsningen av ett intervall är alltid en av intervallets ändpunkter.

Sats 2.24 *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 visar vi att $\sup I = b$. Eftersom intervallet I är lika med mängden av alla x sådant 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 $b_- = b - \varepsilon$, där ε är ett godtyckligt positivt tal, och betraktar två fall: $0 < \varepsilon < 2(b-a)$ och $\varepsilon \geq 2(b-a)$.

Kort uttryckt betraktar vi två fall eftersom b_- kan hamna antingen inom eller utom intervallet.

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

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

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 om intervallet. Därför är det lätt att visa att 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 visar vi att $\inf I = a$. Eftersom intervallet I är lika med mängden av alla x sådant 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_+ inte är 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 om intervallet. Därför är det lätt att visa att 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 begränsning och varje tal större än a inte är en undre begränsning. Därför är a den största undre begränsningen av I. $\square$

Den arkimediska egenskapen

Genom följande exempel finner vi en grundläggande egenskap av reella tal som kallas för *den arkimediska egenskapen*.

Exempel 2.25 Visa att $\sup_n a_n = 3/4$ där följden $(a_n)_n$ definieras enligt formeln

$$a_n = \frac{3n+2}{4n+5}$$

för alla $n \in \mathbf{Z}_+$.

Här skriver vi $\sup_n a_n$ som en alternativ notation till $\sup A$ i fallet mängden A är en följd $(a_n)_n$. Notationen $\inf_n a_n$ används för infimum av följden.

Lösning. Först vill vi bevisa (2.10) med $A = (a_n)_n$ och $u = 3/4$. Eftersom elementen i A indexeras med n måste vi visa att

$$\frac{3n+2}{4n+5} \leq \frac{3}{4} \tag{2.16}$$

för *alla* $n \in \mathbf{Z}_+$. Genom att multiplicera med $4(4n+5)$ ser vi att den är ekvivalent med

$$12n + 8 = 4(3n+2) \leq 3(4n+5) = 12n + 15$$

och den är sant för alla n eftersom $8 \leq 15$. Eftersom (2.16) är ekvivalent med $8 \leq 15$ har vi bevisat (2.16) för alla $n \in \mathbf{Z}_+$.

Nu vill vi bevisa (2.11) med $u = 3/4$. Som sagt indexeras elementen i A med n, så för att bevisa (2.11) måste vi för *varje* $\varepsilon > 0$ hitta *något* $n \in \mathbf{Z}_+$ så att

$$\frac{3}{4} - \varepsilon < \frac{3n+2}{4n+5}. \tag{2.17}$$

Här beror valet av n förstås på ε. Genom att multiplicera med $4(4n+5)$ ser vi att den är ekvivalent med

$$12n+15-4\varepsilon(4n+5) = 3(4n+5)-4\varepsilon(4n+5) < 4(3n+2) = 12n+8$$

och i sin tur är ekvivalent med $7 < 4\varepsilon(4n+5)$ och sedan ekvivalent med

$$n > \frac{7}{16\varepsilon} - \frac{5}{4}. \tag{2.18}$$

För varje givet $\varepsilon > 0$ är det möjligt att välja ett $n \in \mathbf{Z}_+$ så att (2.18) gäller, eller hur? Det är någonting som känns helt rimligt, men ändå noterar vi det som något vi behöver fundera på. Under antagandet att vi kan hitta $n \in \mathbf{Z}_+$ så att (2.18) gäller, har vi för varje $\varepsilon > 0$ visat att det finns $n \in \mathbf{Z}_+$ sådant att (2.17) gäller. Detta tillsammans med att vi har visat (2.16) för alla $n \in \mathbf{Z}_+$ betyder att vi har visat $\sup_n a_n = 3/4$. □

Vad ligger bakom vårt förtroende att det är helt rimligt att hitta n sådant att (2.18) gäller är ett förtroende att det inte finns någon övre begränsning för naturliga tal. Elementen i mängden naturliga tal dyker upp med jämna mellanrum oavsett hur långt längs reella linjen man går. Vi kommer så småningom att kunna bevisa det med hjälp av andra grundläggande egenskaper av reella tal, men än så länge anger vi egenskapen som en sats utan bevis.

Sats 2.26 (Den arkimediska egenskapen) *Mängden av reella tal har följande egenskaper:*

(a) Om $0 \leq a < 1/n$ för alla $n \in \mathbf{Z}_+$ är $a = 0$;

(b) För varje tal $b \in \mathbf{R}$ finns det ett naturliga tal n sådant att $n > b$.

Del (a) och (b) av sats 2.26 är nära förknippade med varandra. Vi kan faktiskt visa att det ena följer av den andra och vice versa och därför är det rimligt att benämna satsen *den* arkimediska egenskapen.

För att visa att del (a) följer av (b) ger vi ett motsägelsebevis. Vi antar att del (a) är falsk: Därför har vi ett tal a som lyder $0 \leq a < 1/n$ för alla $n \in \mathbf{Z}_+$ och även $a \neq 0$. Därför måste $a > 0$ och vi kan sätta $b = 1/a$ och tillämpa del (b) för att få ett n sådant att $n > b$. Däremot, eftersom $1/b = a < 1/n$ för alla $n \in \mathbf{Z}_+$ är $n < b$ för alla $n \in \mathbf{Z}_+$. Det är en motsägelse eftersom vi inte kan ha $n < b$ för alla $n \in \mathbf{Z}_+$ och $n > b$ för något $n \in \mathbf{Z}_+$.

För att visa att del (b) följer av (a) betraktar vi ett godtyckligt $b \in \mathbf{R}$. Om $b \leq 0$ har vi att det naturliga talet 1 är större än b och (b) är bevisat. För $b > 0$ ger vi ett motsägelsebevis. Vi sätter $a = 1/b$ och antar att det inte finns ett naturligt tal n sådant att $n > b$. Därför är $n \leq b = 1/a$ för alla $n \in \mathbf{Z}_+$ så $a \leq n$ för alla $n \in \mathbf{Z}_+$. Så vi har att $0 \leq a \leq 1/n < 1/(n-1)$ för alla $n \in \mathbf{Z}_+ \setminus \{1\}$. Det följer av (a) att $a = 0$, och det är motsägelsefullt till $a = 1/b > 0$.

REELLA TAL SOM OÄNDLIGA DECIMALUTVECKLINGAR

Nu vänder vi tillbaka till hur vi kan definiera de tal vi redan implicit har jobbat med, de så kallade reella talen. Rent praktiskt är den största talmängden vi brukar utföra aritmetiska operationer på de rationella talen – även en dator kan bara hantera ett ändligt antal rationella tal. Reella tal är i någon mening en mycket större mängd av tal än rationella tal, men för att motivera en bra definition av reella tal går vi först till mängden av ändliga decimalutvecklingar, som enligt exempel 1.21 verkar vara en ännu mindre mängd än rationella tal.

Kom ihåg att den ändliga decimalutvecklingen $d_n \dots d_0, d_{-1} \dots d_{-m}$ där varje $d_j \in \{0, 1, \dots, 9\}$ representerade summan (1.22). Med förebilden att de reella talen är punkterna på en rak linje, väljer vi ut en punkt på vår imaginära linje som representerar 0 – den kallas även för *origo* – sedan markerar vi med jämna mellanrum negativa heltal åt ena hållet och positiva heltal åt det andra. Motsvarigheten från punkterna på linjen till tal ges genom att associera till varje punkt dess tecknade avstånd från origo. Ändliga decimalutvecklingar kan representera de punkter som vi har döpt till heltal samt punkter som sitter ett antal tiondelar emellan heltal samt hundradelar emellan tiondelar, och så vidare. Men vi påstår att det finns massor av punkter emellan som inte representeras av ändliga summor (1.22).[5] För att försöka nå alla punkterna tillåter vi decimalutvecklingar att fortsätta i evighet. Men vad menas med detta? Notationsmässigt menar vi, till exempel för ett positivt tal, en *oändlig decimalutveckling* som har formen

$$d_n \dots d_0, d_{-1} \dots \tag{2.19}$$

där varje d_j $(j = n, n-1, \dots)$ är en siffra från mängden $\{0, 1, \dots, 9\}$. Och för att ge (2.19) matematisk mening betraktar vi följden av *trunkerade decimalutvecklingar*

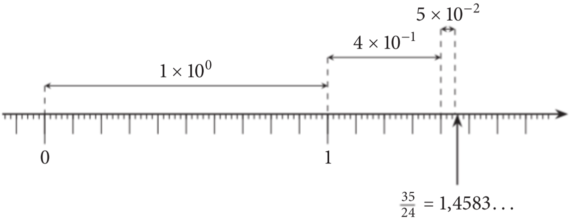
$$\alpha_m = d_n \dots d_0, d_{-1} \dots d_{-m} = \sum_{j=-m}^{n} d_j \times 10^j.$$

Den trunkerade decimalutvecklingen α_m är en underskattning av avståndet från punkten till origo med en noggrannhet av 10^{-m}. Med *noggrannhet* menar vi här att avståndet från origo till punkten minus α_m inte kan överstiga 10^{-m}. Tydligen är $(\alpha_m)_m$ en växande följd. Vår intuition säger oss att den minsta övre begränsningen av $(\alpha_m)_{m \in \mathbf{Z}_+}$ borde vara avståndet från punkten till origo. Se figur 2.4. Vi bestämmer oss för att indexera följden med naturliga tal, så första elementet i följden är alltid ett heltal.

Exempel 2.27 (a) Om $n = 0$ och $d_j = 3$ för alla $j = 0, -1, \dots$ är (2.19) $3{,}333\dots$.

(b) Om $n = 1$ och $d_1 = 7$, $d_0 = 3$, $d_{-1} = 1$, $d_{-2} = 2$, $d_{-3} = 5$ och $d_j = 0$ för alla $j = -4, -5, \dots$ är (2.19) $73{,}125000\dots$.

5 Det är helt rimligt att ifrågasätta intuitionen att det finns punkter som inte nås av ändliga rationella tal. Men utöver vad vi såg i exempel 1.21 kommer vi äntligen i kapitel 4 att bevisa att intuitionen är korrekt.



FIGUR 2.4 De trunkerade decimalutvecklingarna 1, 1,4 och 1,45 är uppskattningar underifrån till 35/24 med noggrannheten 10^0, 10^{-1} respektive 10^{-2}.

(c) Du känner väl till decimalutvecklingen av talet $\pi = 3{,}14159\ldots$ från skolan? Här är $n = 0$, $d_0 = 3$, $d_{-1} = 1$, $d_{-2} = 4$, $d_{-3} = 1$ och så vidare. (Vi kommer definiera talet π i kapitel 5.)

För att gå från intuition till ett konkret matematiskt begrepp behöver vi formulera den föregående föreställningen om en egenskap hos den reella linjen i ett axiom. Axiomet kommer att vara någon slags grundlag för reella tal. Vi vill så gärna att det ska fånga bilden av tal vi har i huvudet, men i grund och botten är det bara ett antagande vi väljer. Vi vill helt enkelt att uppskattningarna α_m alltid når ett reellt tal. Vi vill inte att reella linjen ska ha hål som någon olycklig decimalutveckling kan stöta på. Vi kräver att $(\alpha_m)_{m\in\mathbf{Z}_+}$ alltid har en minsta övre begränsning, men vi formulerar axiomet även för mer allmänna mängder.

Axiom 2.28 (Fullständighetsaxiomet) *Om en icke-tom mängd $A \subseteq \mathbf{R}$ är begränsad uppåt så finns det en minsta övre begränsning $u \in \mathbf{R}$ för A.*

Vi har även de negativa oändliga decimalutvecklingarna

$$-d_n \ldots d_0,d_{-1}\ldots$$

där de trunkerade decimalutvecklingarna

$$\alpha_m = -d_n \ldots d_0,d_{-1}\ldots d_{-m} = \sum_{j=-m}^{n} d_j \times 10^j$$

överskattar avståndet från origo till en punkt till vänster om origo. I det fallet är följden $(\alpha_m)_m$ avtagande. Här vill vi att $\inf_m \alpha_m$ ska existera.

Det är inte svårt att visa att axiom 2.28 är ekvivalent med motsvarande påstående för största undre begränsning: Om en icke-tom mängd $A \subseteq \mathbf{R}$ är begränsad nedåt så finns det en största undre begränsning för A. (Kan du bevisa det?)

Därför, för punkterna både till vänster och till höger om origo, säger axiom 2.28 att vi kan representera dem med oändliga decimalutvecklingar.

Både för punkterna till vänster och till höger om origo (och även origo själv) är följden $(\alpha_m)_{m\in\mathbf{Z}_+}$ av trunkerade decimalutvecklingar en icke-tom mängd av reella tal (även rationella tal). Enligt (1.23) ser vi att varje sådan följd $(\alpha_m)_{m\in\mathbf{Z}_+}$ är begränsad nedåt respektive uppåt. Därför säger axiom 2.28 att vi kan representera punkterna på reella linjen med oändliga decimalutvecklingar. Detta är så gott som en definition av reella tal, men vi fortsätter diskussionen innan vi skriver en formell definition.

Aritmetik med oändliga decimalutvecklingar

Nu har vi en preliminär definition av reella tal, men det är inte klart hur vi kan utföra aritmetiska operationer på dem. Hur lämpliga är algoritmerna vi lärde oss i skolan för oändliga decimalutvecklingar? Prova till exempel att addera två oändliga decimala tal. Man kan förstås addera hyfsat långa trunkerade decimalutvecklingar av talen och nöja sig med svaret, men ingen tror att det är det riktiga svaret utan förhoppningsvis bara inte så långt ifrån svaret. Om vi nu vill vara noggranna med hur vi tänker, hur går vi då till väga?

Man kan hoppas att om talen α och β har de trunkerade decimalutvecklingarna $(\alpha_m)_m$ respektive $(\beta_m)_m$ är följden $(\alpha_m + \beta_m)_m$ en följd av de trunkerade decimalutvecklingarna för summan $\alpha + \beta$. Ibland stämmer det. Till exempel för $\alpha = 5{,}222\ldots$ och $\beta = 1{,}555\ldots$ är

$$\alpha_m = 5{,}\underbrace{2\ldots2}_{m \text{ gånger}} \qquad \text{och} \qquad \beta_m = 1{,}\underbrace{5\ldots5}_{m \text{ gånger}}$$

så

$$\alpha_m + \beta_m = 6{,}\underbrace{7\ldots7}_{m \text{ gånger}}$$

och här är $\sup_m \alpha_m + \sup_m \beta_m = \sup_m(\alpha_m + \beta_m)$. Men ibland fungerar det inte. Det ser vi enklast genom att subtrahera det positiva talet $\beta =$

$0{,}000111\dots$ från $\alpha = 0{,}123000\dots$. Vi har att:

$$\begin{array}{lll}
\alpha_1 = 0{,}1 & \beta_1 = 0{,}0 & \alpha_1 - \beta_1 = 0{,}1 \\
\alpha_2 = 0{,}12 & \beta_2 = 0{,}00 & \alpha_2 - \beta_2 = 0{,}12 \\
\alpha_3 = 0{,}123 & \beta_3 = 0{,}000 & \alpha_3 - \beta_3 = 0{,}123 \\
\alpha_4 = 0{,}123 & \beta_4 = 0{,}0001 & \alpha_4 - \beta_4 = 0{,}1229 \\
\alpha_5 = 0{,}123 & \beta_5 = 0{,}00011 & \alpha_5 - \beta_5 = 0{,}12289 \\
\alpha_6 = 0{,}123 & \beta_6 = 0{,}000111 \text{ och } & \alpha_6 - \beta_6 = 0{,}122889.
\end{array}$$

Observera att följden $(\alpha_m - \beta_m)_m$ inte är en följd av trunkerade decimalutvecklingar. Den är varken växande eller avtagande och man kan beräkna att $\sup_m(\alpha_m - \beta_m) = 0{,}123$ och $\inf_m(\alpha_m - \beta_m) = 0{,}1$ så i allmänt kan varken

$$\sup_m(\alpha_m - \beta_m) = \sup_m \alpha_m - \sup_m \beta_m$$

eller

$$\inf_m(\alpha_m - \beta_m) = \sup_m \alpha_m - \sup_m \beta_m$$

vara giltiga formler. Supremum och infimum är helt enkelt dåligt anpassade som begrepp för att smidigt hjälpa oss att gå från aritmetiska operationer på trunkerade decimalutvecklingar till reella tal. Därför pausar vi diskussionen om aritmetik på reella tal för att ta upp ett nytt begrepp.

Gränsvärden av följder

Begreppet vi behöver för att avsluta diskussionen om aritmetik med reella tal är gränsvärde. Gränsvärdet av en följd $(\alpha_m)_m$ fångar idén att följden kan närma sig ett tal när m är mycket stort. Först inför vi lite ny notation. *Absolutbeloppet* av ett reellt tal x definieras vara

$$|x| = \begin{cases} x & \text{om } x \geq 0; \\ -x & \text{om } x < 0. \end{cases} \tag{2.20}$$

Observera att $-x$ är positivt om x är negativt!

Absolutbeloppet av ett tal är talets avstånd från origo utan tecken.

Anmärkning 2.29 (a) För två reella tal x och y är

$$|xy| = |x||y|.$$

Det kan man bevisa genom att betrakta de möjliga olika tecknen av x och y. Om till exempel $x \geq 0$ och $y < 0$ är $xy \leq 0$ så $|xy| = -xy = (x)(-y) = |x||y|$.

(b) På liknande sätt kan man visa att

$$\left|\frac{x}{y}\right| = \frac{|x|}{|y|}$$

så länge $y \neq 0$.

(c) Däremot kan man inte byta ut multiplikation eller division mot addition. Det vill säga att $|x + y| = |x| + |y|$ är ingen räkneregel! Det är lätt att visa att den är falsk. Om $x = 3$ och $y = -10/3$ är till exempel $|x + y| = 1/3$ men $|x| + |y| = 19/3$. I allmänhet får man olikhet och det visar vi i följande sats.

Definitionen av absolutbeloppet $|x|$ skiljer mellan $x \geq 0$ och $x < 0$, men definitionen är ekvivalent med att säga $|x| = x$ om $x > 0$ och $|x| = -x$ om $x \leq 0$ eftersom $0 = -0$.

Sats 2.30 (Triangelolikheten) *Olikheten*

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

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

Bevis. Om både x och y är icke-negativa har vi att

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

Om både x och y är negativa har vi att

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

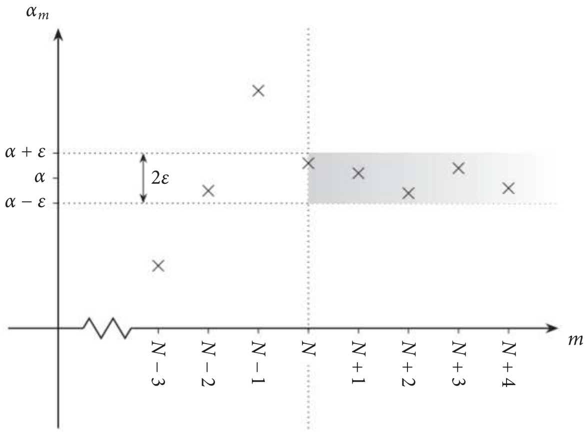
Om x och y har olika tecken får vi anta att $|x| > |y|$ eftersom om det inte stämmer får vi byta roller på x och y. Observera att $x \leq |x|$ för vilket x som helst. Om x är positivt och y är negativt medför det att

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

Om y är positivt och x är negativt har vi att

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

Nu ger vi definitionen av gränsvärde. Definitionen är lite teknisk så vi försöker förklara den efter att vi har läst den.



FIGUR 2.5 Om följden $(\alpha_m)_m$ konvergerar mot α måste det för varje $\varepsilon > 0$ finnas ett $N \in \mathbf{Z}_+$ sådant att element α_m ligger inom det grå området för alla $m \geq N$.

Observera att valet av N kan (och i de flesta fall kommer) bero på ε.

Definition 2.31 En följd $(\alpha_m)_{m\in\mathbf{Z}_+}$ sägs vara *konvergent* om det finns ett tal $\alpha \in \mathbf{R}$ med följande egenskap: För varje $\epsilon > 0$ finns det ett $N \in \mathbf{Z}_+$ sådant att $|\alpha_m - \alpha| < \varepsilon$ för alla $m \geq N$.

I fallet då ett sådant α finns kallas α för *gränsvärdet* av $(\alpha_m)_m$, vi skriver

$$\lim_{m\to\infty} \alpha_m = \alpha$$

eller ”$\alpha_m \to \alpha$ då $m \to \infty$” och säger ”α_m går mot α (då m går mot oändligheten)” eller ”α_m konvergerar mot α (då m går mot oändligheten)”. Ibland säger vi ”$\lim_{m\to\infty} \alpha_m$ existerar” och menar att ”$(\alpha_m)_m$ är konvergent.” En följd som inte är konvergent sägs vara *divergent*.

Idén bakom definitionen är att $(\alpha_m)_m$ sägs konvergera mot α om man för varje mått av exakthet $\varepsilon > 0$ (som man kan förställa sig vara litet) har att elementen α_m stängs in mellan $\alpha - \varepsilon$ och $\alpha + \varepsilon$ under förutsättning att α_m ligger tillräckligt långt ner i följden (uttryckt med hjälp av N som $m \geq N$). Se figur 2.5. Att visa en följd är konvergent innebär att man för varje möjligt val av ε ska kunna förse ett motsvarande N.

Exempel 2.32 Visa att följden $(1/m)_m$ konvergerar mot noll då $m \to \infty$. Med andra ord bevisa att

$$\lim_{m\to\infty} \frac{1}{m} = 0.$$

Lösning. Gränsvärdet kan intuitivt anses vara noll eftersom för jättestora m är $1/m$ pyttelitet. Men för att ordentligt bevisa det utifrån definition 2.31 betraktar vi ett godtyckligt $\varepsilon > 0$. I detta fall är $\alpha_m = 1/m$ och $\alpha = 0$ så definitionen säger att vi behöver hitta ett $N \in \mathbf{Z}_+$ sådant att

$$\left|\frac{1}{m} - 0\right| < \varepsilon \tag{2.21}$$

så länge $m \geq N$. Vi beräknar att

$$\left|\frac{1}{m} - 0\right| = \frac{1}{m}$$

eftersom $m > 0$. Men om $m \geq N$ är

$$\frac{1}{m} \leq \frac{1}{N}.$$

Därför räcker det att för varje $\varepsilon > 0$ hitta $N \in \mathbf{Z}_+$ sådant att $1/N < \varepsilon$. Men vi kan tillämpa den arkimediska egenskapen (sats 2.26(b)) med $b = 1/\varepsilon$ för att få tag i ett $N \in \mathbf{Z}_+$ sådant att $N > 1/\varepsilon$. Därför är

$$\left|\frac{1}{m} - 0\right| = \frac{1}{m} \leq \frac{1}{N} < \varepsilon$$

för detta val av N och alla $m \geq N$. Därmed är (2.21) bevisat och vi har visat att $\lim_{m\to\infty} 1/m = 0$. □

Observera att vårt val av N beror på ε.

En följd av den efterföljande satsen är att vi kan omdefiniera ett reellt tal som ett gränsvärde av sin trunkerade decimalutveckling i stället för supremum eller infimum av sin trunkerade decimalutveckling.

Sats 2.33 *Låt $(\alpha_m)_m$ vara en följd av reella tal.*

(a) *Om $(\alpha_m)_{m\in\mathbf{Z}_+}$ är växande och uppåt begränsad är den även konvergent och*

$$\lim_{m\to\infty} \alpha_m = \sup_{m\in\mathbf{Z}_+} \alpha_m.$$

(b) *Om $(\alpha_m)_{m\in\mathbf{Z}_+}$ är avtagande och nedåt begränsad är den även konvergent och*

$$\lim_{m\to\infty} \alpha_m = \inf_{m\in\mathbf{Z}_+} \alpha_m.$$

Bevis. Vi bevisar bara del (a) eftersom (b) kan bevisas med samma argument.

Vi vet därför att (2.10) och (2.11) gäller med $u = \alpha$ och $A = (\alpha_m)_m$.

Eftersom $(\alpha_m)_m$ är begränsad uppåt vet vi enligt axiom 2.28 att den har en minsta övre begränsning $\sup_{m\in\mathbf{Z}_+} \alpha_m$. Sätt $\alpha = \sup_{m\in\mathbf{Z}_+} \alpha_m$ och låt ε vara ett godtyckligt positivt tal. Enligt definition 2.31 behöver vi visa att det finns ett $N \in \mathbf{Z}_+$ sådant att

$$|\alpha_m - \alpha| < \varepsilon \tag{2.22}$$

för alla $m \geq N$. Men eftersom α är den minsta övre begränsningen av $(\alpha_m)_m$ vet vi från (2.11) att det finns $N \in \mathbf{Z}_+$ sådant att

$$\alpha - \varepsilon < \alpha_N.$$

Eftersom $(\alpha_m)_m$ är växande kan vi även säga att $\alpha_N \leq \alpha_m$ för $m \geq N$ så

$$\alpha - \varepsilon < \alpha_m$$

för $m \geq N$. Från (2.10) har vi även att

$$\alpha_m \leq \alpha.$$

för alla $m \in \mathbf{Z}_+$. Så tillsammans får vi att

$$0 \leq \alpha - \alpha_m < \varepsilon$$

för alla $m \geq N$. Detta medför (2.22) med samma val av N och vi har bevisat del (a) av satsen. □

Att alla monotona begränsade följder konvergerar (som är slutsatsen av sats 2.33) är inte bara en följd av axiom 2.28. Man kan visa att det är ett ekvivalent påstående till axiom 2.28, men vi bevisar inte det här. En konvergent följd behöver förstås inte vara monoton men här visar vi att alla konvergenta följder är begränsade.

Sats 2.34 *En konvergent följd av reella tal är nödvändigtvis både uppåt och nedåt begränsad.*

Bevis. Låt $(\alpha_m)_m$ vara en konvergent följd av reella tal som konvergerar mot $\alpha \in \mathbf{R}$. Vi tar $\varepsilon = 1$ i definition 2.31 och får tag i ett $N \in \mathbf{Z}_+$ sådant att $|\alpha_m - \alpha| \leq 1$ för alla $m \geq N$ så

$$\alpha - 1 \leq \alpha_m \leq \alpha + 1$$

för alla $m \geq N$. Sedan definierar vi

$$C_1 = \max\{\alpha_1, \alpha_2, \ldots, \alpha_{N-1}\} \quad \text{och} \quad C_2 = \max\{C_1, \alpha + 1\}$$

så $\alpha_m \leq C_2$ för alla $m \in \mathbf{Z}_+$ och $(\alpha_m)_m$ är uppåt begränsad.[6] Genom att definiera

$$C_3 = \min\{\alpha_1, \alpha_2, \ldots, \alpha_{N-1}\} \quad \text{och} \quad C_4 = \min\{C_3, \alpha - 1\}$$

ser vi att $\alpha_m \geq C_4$ för alla $m \in \mathbf{Z}_+$ och därmed är $(\alpha_m)_m$ är nedåt begränsad.[7] □

Räkneregler för gränsvärden

Innan vi återkommer till diskussionen om reella tal vill vi etablera vissa räkneregler för gränsvärden. Om vi har två följder $(\alpha_m)_m$ och $(\beta_m)_m$ kan vi bygga nya följder genom att kombinera dem elementvis. Till exempel kan vi bygga följden $(\gamma_m)_m$ där det m:e elementet i den nya följden är summan $\gamma_m = \alpha_m + \beta_m$ av de två m:e elementen i de gamla följderna. På ett liknande sätt kan man bygga differensen $(\alpha_m - \beta_m)_m$, produkten $(\alpha_m \beta_m)_m$ och kvoten $(\alpha_m/\beta_m)_m$.

Sats 2.35 *Låt $(\alpha_m)_m$ och $(\beta_m)_m$ vara två konvergenta följder. Då har vi att:*

(a) Följden $(\alpha_m + \beta_m)_m$ är konvergent och

$$\lim_{m\to\infty} (\alpha_m + \beta_m) = \lim_{m\to\infty} \alpha_m + \lim_{m\to\infty} \beta_m;$$

6 Kom ihåg från lemma 2.15 att notationen $\max\{\alpha_1, \alpha_2, \ldots, \alpha_{N-1}\}$ betyder det största elementet i mängden $\{\alpha_1, \alpha_2, \ldots, \alpha_{N-1}\}$.

7 Kom ihåg från definition 2.17 att notationen $\min\{\alpha_1, \alpha_2, \ldots, \alpha_{N-1}\}$ betyder det minsta elementet i mängden $\{\alpha_1, \alpha_2, \ldots, \alpha_{N-1}\}$.

(b) *Följden* $(\alpha_m - \beta_m)_m$ *är konvergent och*

$$\lim_{m\to\infty}(\alpha_m - \beta_m) = \lim_{m\to\infty}\alpha_m - \lim_{m\to\infty}\beta_m;$$

(c) *Följden* $(\alpha_m\beta_m)_m$ *är konvergent och*

$$\lim_{m\to\infty}(\alpha_m\beta_m) = \left(\lim_{m\to\infty}\alpha_m\right)\left(\lim_{m\to\infty}\beta_m\right); \quad \textit{och}$$

(d) *Om* $\beta_m \neq 0$ *för alla* $m \in \mathbf{Z}_+$ *och* $\lim_{m\to\infty}\beta_m \neq 0$ *är följden* $(\alpha_m/\beta_m)_m$ *konvergent och*

$$\lim_{m\to\infty}\left(\frac{\alpha_m}{\beta_m}\right) = \left(\frac{\lim_{m\to\infty}\alpha_m}{\lim_{m\to\infty}\beta_m}\right).$$

Som special fall av reglerna ovan får vi för en konstant c *att:*

(e) $\lim_{m\to\infty}(\alpha_m + c) = (\lim_{m\to\infty}\alpha_m) + c$*; och*

(f) $\lim_{m\to\infty}(c\alpha_m) = c\lim_{m\to\infty}\alpha_m$*;*

där högerleden existerar om och endast om vänsterleden existerar.

Bevis. För att bevisa del (a) benämner vi $\alpha = \lim_{m\to\infty}\alpha_m$ och $\beta = \lim_{m\to\infty}\beta_m$. Enligt definition 2.31 behöver vi visa att det för varje godtyckligt $\varepsilon > 0$ finns ett $N \in \mathbf{Z}_+$ sådant att

$$|(\alpha_m + \beta_m) - (\alpha + \beta)| < \varepsilon \tag{2.23}$$

för alla $m \geq N$. Men eftersom vi vet att både $(\alpha_m)_m$ och $(\beta_m)_m$ är konvergenta kan vi använda oss av vad definition 2.31 säger angående dem: Vi utnyttjar definition 2.31 med $\varepsilon/2$ i stället för ε (vilket är tillåtet eftersom ε kan vara vilket positivt tal som helst). Vi får att det finns $N_1 \in \mathbf{Z}_+$ sådant att

$$|\alpha_m - \alpha| < \frac{\varepsilon}{2} \tag{2.24}$$

för alla $m \geq N_1$ och $N_2 \in \mathbf{Z}_+$ sådant att

$$|\beta_m - \beta| < \frac{\varepsilon}{2}. \tag{2.25}$$

för alla $m \geq N_2$.

Vi hävdar att valet $N = \max\{N_1, N_2\}$ fungerar bra för att visa (2.23). För att visa det beräknar vi med hjälp av triangelolikheten (sats 2.30) att

$$\begin{aligned} &|(\alpha_m + \beta_m) - (\alpha + \beta)| = |(\alpha_m - \alpha) + (\beta_m - \beta)| \\ &\underset{\substack{\uparrow \\ \text{sats 2.30}}}{\leq} |\alpha_m - \alpha| + |\beta_m - \beta| \underset{\substack{\uparrow \\ (2.24)\ \&\ (2.25)}}{<} \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon \end{aligned}$$

så länge $m \geq N$, eftersom $N \geq N_1$ och $N \geq N_2$, och därmed har vi visat (2.23).

Vi lämnar beviset av de andra delarna som en övning för läsaren och ger bara några tips. Del (b) kan bevisas på ett liknande sätt till del (a). För att visa del (c) kan det vara nyttigt att skriva

$$|\alpha_m \beta_m - \alpha\beta| = |\alpha_m \beta_m - \alpha_m \beta + \alpha_m \beta - \alpha\beta|$$

och sedan använda sats 2.30 och sats 2.34. Del (d) är mer komplicerad att bevisa än (c) men beviset kan ändå genomföras på likartat sätt. Del (e) och (f) kan visas med hjälp av (a) och (c) om man sätter $\beta_m = c$ för alla $m \in \mathbf{Z}_+$ (och i så fall kallas $(\beta_m)_m$ för en konstant följd). □

Men hjälp av exempel 2.32 och sats 2.35 kan man räkna ut många gränsvärden.

Exempel 2.36 Bestäm om följden $(\alpha_m)_{m \in \mathbf{Z}_+}$ är konvergent eller inte där

$$\alpha_m = \frac{5m^2 - 7}{6m^2 + 3m}$$

för alla $m \in \mathbf{Z}_+$. Om den är konvergent avgör gränsvärdet.

Lösning. Vi vill utnyttja räkneregler från sats 2.35 för att bryta ner följden $(\alpha_m)_{m \in \mathbf{Z}_+}$ till en sammansättning av enklare följder. Vi beräknar

$$\alpha_m = \frac{5m^2 - 7}{6m^2 + 3m} = \frac{m^2\left(5 - 7\frac{1}{m^2}\right)}{m^2\left(6 + 3\frac{1}{m}\right)} = \frac{5 - 7\frac{1}{m^2}}{6 + 3\frac{1}{m}}.$$

Observera att $6 + 3\frac{1}{m}$ aldrig är noll för $m \in \mathbf{Z}_+$. Därför säger del (d) av sats 2.35 att $(\alpha_m)_{m \in \mathbf{Z}_+}$ är konvergent och

$$\lim_{m \to \infty} \alpha_m = \frac{\lim_{m \to \infty}\left(5 - 7\frac{1}{m^2}\right)}{\lim_{m \to \infty}\left(6 + 3\frac{1}{m}\right)} \tag{2.26}$$

under förutsättning att både $\lim_{m\to\infty}\left(5-7\frac{1}{m^2}\right)$ och $\lim_{m\to\infty}\left(6+3\frac{1}{m}\right)$ existerar, och $\lim_{m\to\infty}\left(6+3\frac{1}{m}\right)\neq 0$.

Enligt exempel 2.32 existerar $\lim_{m\to\infty} 1/m$ och därför säger del (e) och (f) av sats 2.35 att

$$\begin{aligned}\lim_{m\to\infty}\left(6+3\frac{1}{m}\right) &= 6+\lim_{m\to\infty}\left(3\frac{1}{m}\right) = 6+3\lim_{m\to\infty}\left(\frac{1}{m}\right)\\ &= 6+3\times 0 = 6\neq 0.\end{aligned}$$

Enligt exempel 2.32 och del (c) är

$$\lim_{m\to\infty}\left(\frac{1}{m^2}\right) = \lim_{m\to\infty}\left(\frac{1}{m}\right)\lim_{m\to\infty}\left(\frac{1}{m}\right) = 0\times 0 = 0$$

och därmed kan vi se med hjälp av del (e) och (f) att $\lim_{m\to\infty}\left(5-7\frac{1}{m^2}\right)$ existerar och

$$\begin{aligned}\lim_{m\to\infty}\left(5-7\frac{1}{m^2}\right) &= 5+\lim_{m\to\infty}\left((-7)\frac{1}{m^2}\right) = 5-7\lim_{m\to\infty}\left(\frac{1}{m^2}\right)\\ &= 5-7\times 0 = 5.\end{aligned}$$

Så vi har visat att förutsättningarna för (2.26) är uppfyllda, och därför är $(\alpha_m)_m$ konvergent och $\lim_{m\to\infty}\alpha_m = 5/6$. □

Återkomst till aritmetik med oändliga decimalutvecklingar

På sidan 69 tog vi mängden av oändliga decimalutvecklingar – tolkade som minsta övre eller största undre begränsningar av trunkerade positiva respektive negativa decimalutvecklingar – som en preliminär definition för reella tal. Men i följande avsnitt uppstod det frågor om hur vi kunde enkelt beräkna med dem. Begreppet gränsvärde ger oss ett bättre sätt att tolka decimalutveckling än minsta övre eller största undre begränsningar. Sats 2.33 säger att vi får byta ut vår ursprungliga tolkning av oändliga decimalutvecklingar mot gränsvärden. Utöver det säger sats 2.35 att aritmetik på oändliga decimalutvecklingar kan utföras genom att ta gränsvärden av uträkningar på trunkerade decimalutvecklingar. Därför formulerar vi följande definition.

Definition 2.37 Ett reellt tal är ett tal som kan representeras som en oändlig decimalutveckling

$$\pm d_n\ldots d_0,d_{-1}\ldots$$

tolkad som gränsvärdet $\lim_{m\to\infty} \alpha_m$ av följden

$$\alpha_m = \pm d_n \ldots d_0, d_{-1} \ldots d_{-m} = \pm \sum_{j=-m}^{n} d_j \times 10^j$$

av sina trunkerade decimalutvecklingar. Med tecknet ± menas + eller −. Mängden av alla reella tal betecknas **R** och begreppet av likhet mellan reella tal är det som följs enligt räkneregler för aritmetik samt axiom 2.28.

Vi betonar att varje ändlig decimalutveckling kan skrivas som en oändlig genom att tillägga oändligt många nollor. Därför ingår heltal och även alla ändliga decimalutvecklingar i mängden av reella tal. Sådana decimalutvecklingar har vi redan beräknat med eftersom de kan skrivas som bråk. Nu visar vi ett par exempel där decimalutvecklingar vars siffror upprepas periodiskt kan skrivas som bråk.

Exempel 2.38 Skriv decimalutvecklingen $\alpha = 0{,}6666\ldots$ (det vill säga $d_k = 6$ för alla negativa heltal k) som ett bråk.

Lösning. Om vi skriver

$$\alpha_m = 0{,}\underbrace{6\ldots6}_{m \text{ gånger}}$$

för α:s trunkerade decimalutvecklingar har vi att

$$10\alpha_m = 6{,}\underbrace{6\ldots6}_{m-1 \text{ gånger}} = 6 + \alpha_{m-1}. \tag{2.27}$$

Om vi tar gränsvärden av vänster- och högerledet i (2.27) får vi med hjälp av sats 2.35 att

För att ta gränsvärden här behöver vi veta att $(\alpha_m)_m$ är konvergent. Med det följs av axiom 2.28 eftersom trunkerade decimalutveckling är begränsade följder.

$$10\alpha = 6 + \alpha$$

som i sin tur säger att $9\alpha = 6$. Om vi delar båda sidor med 9 kommer vi fram till att

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

□

Exempel 2.39 Skriv decimalutvecklingen $\alpha = 0{,}125125125\ldots$ (det vill säga $d_{-1-3k} = 1$, $d_{-2-3k} = 2$, $d_{-3-3k} = 5$ för alla icke-negativa heltal k) som ett bråk.

Lösning. I det här fallet är det tre siffror som upprepas i evighet och därför vill vi betrakta

$$1000\alpha_m = 125{,}\underbrace{125\ldots125}_{m-3\text{ siffror}} = 125 + \alpha_{m-3}$$

i stället för (2.27), där återigen $(\alpha_m)_m$ är följden av α:s trunkerade decimalutvecklingar. Om vi tar gränsvärdet av vänster- och högerledet får vi med hjälp av sats 2.35 att

$$1000\alpha = 125 + \alpha$$

som i sin tur säger att $999\alpha = 125$. Om vi delar båda sidor med 999 kommer vi fram till att $\alpha = \frac{125}{999}$. □

Ett par lösa ändar

I detta avsnitt svarar vi på två naturliga frågor som kvarstår. Man kan undra, givet att alla reella tal per definition kan skrivas som oändliga decimalutvecklingar: Finns det bara ett unikt sätt att representera varje reellt tal som en oändlig decimalutveckling? Och du möjligen har märkt att vi inte har bevisat den arkimediska egenskapen (sats 2.26) än, så hur gör man det? De två frågorna svarar vi på nu.

Vissa tycker att att det är ett märkligt faktum att vissa tal faktiskt har två olika decimalutvecklingar. Det är någonting vi är vana vid för rationella tal, men mindre för oändliga decimalutvecklingar. Nu visar vi att talet 1 har två olika decimalutvecklingar: $1{,}000\ldots$ och $0{,}999\ldots$. Vi definierar två följder $(\alpha_m)_m$ och $(\beta_m)_m$ av trunkerade decimalutvecklingar:

$$\alpha_m = 1{,}\underbrace{0\ldots0}_{m\text{ gånger}} \quad \text{och} \quad \beta_m = 0{,}\underbrace{9\ldots9}_{m\text{ gånger}}.$$

Eftersom $\alpha_m = 1$ för alla $m \in \mathbf{Z}_+$ är det en konstant följd och det är inte svårt att visa $\lim_{m\to\infty} \alpha_m = 1$. För $(\beta_m)_m$ har vi att

$$10\beta_m = 9{,}\underbrace{9\ldots9}_{m-1\text{ gånger}} = 9 + \beta_{m-1}$$

så när vi tar gränsvärdet är $10\beta = 9 + \beta$ där $\beta = \lim_{m\to\infty} \beta_m$. Därför är $9\beta = 9$ och därmed är även $\beta = 1$. Så vi har visat att

$$1{,}000\ldots = 1 = 0{,}999\ldots$$

och talet 1 har två decimalutvecklingar.

Nu visar vi att den arkimediska egenskapen är ett följd av fullständighetsaxiomet (axiom 2.28).

Bevis (sats 2.26). Vi visade direkt efter sats 2.26 att del (a) och (b) är ekvivalenta så det räcker av bevisa (b). Vi ger ett motsägelsebevis och antar därför att $n \le b$ för alla $n \in \mathbf{N}$. Därför är b en övre begränsning för $\mathbf{N}$ och enligt axiom 2.28 finns det en minsta övre begränsning u av mängden $\mathbf{N}$. Eftersom u är den *minsta* övre begränsningen är $u - 1$ inte en övre begränsning och det måste finnas $m \in \mathbf{N}$ sådant att $u - 1 < m$. Det betyder att $u < m + 1$ och $m + 1 \in \mathbf{N}$ som är en motsägelse till att u var en övre begränsning av $\mathbf{N}$. □

Det är nog självklart att $m + 1$ måste vara ett naturligt tal om m är det, men det ingår även i induktionsprincipen (axiom 2.6). Som sagt är den ett antagande om strukturen av naturliga tal.

2.3 Sammanfattning och uppgifter

I kapitel 2 har vi:

- pratat om lika slags mängder såsom följder och intervall samt operationer som snittet och unionen man kan utföra på mängder;
- formulerat vår första bevismetod – matematisk induktion – och använt den för att bevisa formler för geometriska och aritmetiska följder samt binomialsatsen;
- utrett begrepp såsom största element och minsta övre begränsningar för delmängder av tal med hjälp av bland annat motsägelsebevis och den arkimediska egenskapen;
- diskuterat vad vi kan mena med reella tal och sett att begreppet gränsvärde ger en bättre verktyg för att beräkna med oändliga decimalutvecklingar än supremum eller infimum;
- visat att vissa tal har två decimalutvecklingar och att den arkimediska egenskapen är en följd av fullständighetsaxiomet.

UPPGIFTER

2.1 Vilka av följande följder kan vara aritmetiska eller geometriska följder?

(a) $(7, 10, 13, 16, \dots)$

(b) $(4, 12, 36, 108, \dots)$

(c) $(0, 0, 0, 0, \dots)$

(d) $(1, 1, 2, 3, 5, 8, 13, \dots)$

(e) $(2, 3, 5, 7, 11, 13, \dots)$

2.2 Bevisa följande likheter:

(a) $\prod_{k=1}^{n} 7 = 7^n$

(b) $\prod_{k=1}^{n} 5^k = 5^{n(n+1)/2}$

(c) $\prod_{k=1}^{n} 3k = 3^n n!$

2.3 Ge ett induktionsbevis av följande likheter som gäller för alla $n \in \mathbf{Z}_+$.

(a)

$$\sum_{k=1}^{n} \frac{2k-1}{2} = \frac{n^2}{2}$$

(b)

$$\sum_{k=1}^{n} k^3 = \left(\frac{n(n+1)}{2}\right)^2$$

2.4 Ge ett induktionsbevis av följande olikheter.

(a) $4n \leq 2^n$ för alla heltal $n \geq 4$.

(b) $2n + 1 \leq 2^n$ för alla heltal $n \geq 3$.

2.5 Hitta vad som är *fel* med följande induktionsbevis.

(a) Bevisa

$$k^2 \leq k \tag{2.28$_k$}$$

för varje $k \in \mathbf{Z}_+$.

Lösning. Vi kan lätt kontrollera att basfallet $(2.28)_1$ är sant: Om vi tar $k = 1$ så är $k^2 = 1^2 \leq 1 = k$.

Nu antar vi att $(2.28)_n$ är sant för något givet $n \in \mathbf{Z}_+$ och betraktar fallet $(2.28)_{n+1}$. I så fall har vi att

$$(n+1)^2 \underset{(2.28)_{n+1}}{\leq} (n+1) \leq 3n+1 \implies n^2+2n+1 \leq 3n+1 \iff n^2 \leq n.$$

Sista olikheten är precis $(2.28)_n$. Enligt induktionsprincipen är satsen bevisad. □

(b) Bevisa

$$2k = 0 \tag{$(2.29)_k$}$$

för varje $k \in \mathbf{N}$.

Lösning. Vi bevisar först basfallet: Om $k = 0$ så är $2k = 2 \times 0 = 0$ och därför är $(2.29)_0$ sant.

Nu antar vi $(2.29)_k$ för alla $k \leq n$ och något givet $n \in \mathbf{Z}_+$ och betraktar $(2.29)_{n+1}$. Observera att man kan skriva $n + 1 = i + j$ där i och j är två icke-negativa heltal mindre än eller lika med n. Enligt induktionsantagandet är både $(2.29)_i$ och $(2.29)_j$ sanna. Då får vi beräkna att

$$2(n+1) = 2(i+j) = 2i + 2j \underset{(2.29)_i \,\&\, (2.29)_j}{=} 0 + 0 = 0.$$

Enligt induktionsprincipen är satsen bevisad. □

(c) Bevisa

$$k + 1 < k \tag{$(2.30)_k$}$$

för varje $k \in \mathbf{Z}_+$.

Lösning. Som vanligt antar vi $(2.30)_n$ för något givet $n \in \mathbf{Z}_+$ och betraktar $(2.30)_{n+1}$: Vi kan beräkna att

$$(n+1) + 1 \underset{(2.30)_n}{<} (n) + 1 = (n+1)$$

som är $(2.30)_{n+1}$. Enligt induktionsaxiomet är satsen bevisad. □

2.6 Bevisa sats 2.12.

2.7 Visa att

$$\binom{n}{k} \leq \frac{n^k}{k!}$$

för hela tal $k \leq n$.

2.8 Hitta alla möjliga par av reella tal (a,b) som uppfyller

$$|a+b| = |a| + |b|.$$

2.9 Visa att följden $(n)_{n\in\mathbf{Z}_+}$ är divergent. [Tips: Man kan till exempel använda sig av sats 2.26 och 2.34.]

2.10 Med hjälp av sats 2.35 och exempel 2.32 visa att följande följder $(\alpha_m)_{m\in\mathbf{Z}_+}$ är konvergenta. Avgör i varje fall vad gränsvärdet $\lim_{m\to\infty} \alpha_m$ är för tal. [Tips: Titta på exempel 2.36.]

(a) $\alpha_m = \frac{2m^2+1}{m^2+4m}$

(b) $\alpha_m = \frac{m-3}{2m+1}$

(c) $\alpha_m = \frac{4-5m}{m^2+6}$

(d) $\alpha_m = \frac{2m^3+m}{9m^3+3}$

(e) $\alpha_m = \frac{m^4-8}{m+3m^4}$

(f) $\alpha_m = \frac{5m^6+1}{2+3m^6}$

2.11 För att visa att en siffra eller några siffror i en decimal utveckling upprepas i evighet skriver vi en linje ovanför siffrorna som upprepas. Till exempel $7/3 = 2.333\ldots$ skrivs som $2.\overline{3}$ och $25/99 = 0.252525\ldots$ skrivs som $0.\overline{25}$.

(a) Skriv de decimala utvecklingarna $0.\overline{27}$, $6.\overline{15}$ och $4.\overline{118}$ som bråk.

(b) Räkna ut de första fyra siffrorna i en decimal utveckling för $1/8$, $1/3$, $1/2$ och $4/7$.

2.12 Vilken eller vilka av de följande mängderna är lika med intervallet $[0,4]$? Motivera ditt svar.

(a) $\{x \in \mathbf{R} \,|\, 4x - 7 \geq 2 \text{ och } x \leq 4\}$

(b) $\{x \in \mathbf{R} \,|\, 2x^2 + 4 < 12 - 8x\}$

(c) $\{x \in \mathbf{R} \,|\, x^2 - 2x + 1 \leq 1 \text{ eller } 8 \leq 6x - x^2\}$

(d) $\{x \in \mathbf{R} \,|\, x^2 - 4x \leq 0 \text{ och } 8 \leq 6x - x^2\}$

(e) $\{x \in \mathbf{R} \,|\, x^2 - 4x < 0,\ x^2 - 4x + 3 = 0 \text{ eller } x^2 + 8 = 6x\}$

2.13 Bevisa att följden $(a_n)_{n\in\mathbf{Z}_+}$, definierad genom uttrycket

$$a_n = \frac{n^n}{(2n)!}$$

för varje $n \in \mathbf{Z}_+$, är uppåt begränsad. [Tips: Försök jämföra $(2n)!$ med n^n.]

2.14 Undersök om mängden

$$\{y \in \mathbf{R} \mid y = (x-1)(x+5) \text{ och } x \in [-6,2]\}$$

har en infimum eller supremum. Motivera ditt svar och i de fall de existerar bestäm deras värde.

2.15 Syftet med den här uppgiften är att bevisa följande påstående:

$$a,b \in \mathbf{Z} \implies a^2 - 4b \neq -2. \tag{2.31}$$

Det intressanta är hur man kan bevisa satsen. Såvitt jag vet finns det inget direkt sätt att bevisa påståendet. I uppgiften ska vi ge ett motsägelsebevis.

(a) Formulera antagandena vi vill adoptera för att genomföra ett motsägelsebevis av påståendet (2.31).

(b) Utifrån antagandena från (a), bevisa att a^2 är jämnt.

(c) Kom ihåg svaret från uppgift 1.6 och dra en slutsats om talet a under antagandena från (a).

(d) Sedan bevisa att 1 måste vara jämnt.

(e) Visa påståendet (2.31) utifrån (d).

Funktioner och former

3.1 Funktioner: begrepp och egenskaper

Begreppet *funktion* är centralt i matematisk analys. I det här avsnittet preciserar vi vad vi menar med en funktion och inför relaterade begrepp såsom definitionsmängd och värdemängd. Vi pratar om hur vi kan representera funktioner med hjälp av en graf samt hur man på olika sätt kan lägga ihop funktioner. Vi tittar även närmare på polynom som är särskilt viktiga exempel på funktioner.

FUNKTIONER, DEFINITIONSMÄNGDER 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 lika med alfabetet, sedan kan vi definiera en funktion som för en bil anger första bokstaven 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*. Många blandar ihop f och $f(x)$ när de skriver, men i strikt mening är f tecknet för funktionen medan $f(x)$ är elementet i B som f skickar ett $x \in A$ till.

Exempel 3.1 (a) Ta $A = B = \mathbf{R}$. Sedan kan man definiera en funktion $f\colon \mathbf{R} \to \mathbf{R}$ enligt formeln $f(x) = x^2 + 2x + 1$ för alla $x \in \mathbf{R}$.

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

(c) Ta $A = \mathbf{R}$ och $B = \mathbf{Z}$. Sedan kan man definiera en funktion $t\colon \mathbf{R} \to \mathbf{Z}$ enligt regeln

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

Observera att en funktion $f\colon A \to B$ inte nödvändigtvis är definierad med hjälp av ett enstaka aritmetiskt uttryck: så länge man för varje $x \in A$ kan räkna fram $f(x) \in B$ från x via någon regel är den en funktion. Ibland skriver man $x \mapsto f(x)$. På så sätt kan man slippa använda ett namn för funktionen: 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 $f : A \to B$ och $g : C \to D$ sägs vara *lika* om $A = C$ och $f(x) = g(x)$ för alla $x \in A$.

Exempel 3.2 (a) Definiera $f : \mathbf{R} \setminus \{4\} \to \mathbf{R}$ enligt formeln

$$f(x) = \frac{x^2 + 4x - 32}{(x-4)} \quad \text{för alla } x \in \mathbf{R} \setminus \{4\}$$

och $g : \mathbf{R} \setminus \{4\} \to \mathbf{R}$ enligt formeln

$$g(x) = x + 8 \quad \text{för alla } x \in \mathbf{R} \setminus \{4\}.$$

Funktionerna f och g är lika eftersom de har samma definitionsmängd och

$$f(x) = \frac{x^2 + 4x - 32}{(x-4)} = \frac{(x-4)(x+8)}{(x-4)} = x + 8 = g(x)$$

för alla $x \in \mathbf{R} \setminus \{4\}$.

(b) Om vi definierar $h : \mathbf{R} \to \mathbf{R}$ enligt formeln

$$h(x) = x + 8 \quad \text{för alla } x \in \mathbf{R}$$

är h inte lika med f eftersom de har olika definitionsmängder.

(c) Om vi definierar $k : \mathbf{R} \to \mathbf{R}$ enligt formeln

$$k(x) = \begin{cases} \frac{x^2+4x-32}{(x-4)} & \text{om } x \neq 4; \\ 13 & \text{om } x = 4. \end{cases}$$

är k inte lika med h eftersom $k(4) = 13 \neq 12 = h(4)$.

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

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

för funktionens *värdemängd*. Observera att värdemängden är en delmängd av målmängden B: För varje $x \in A$ måste $f(x)$ vara ett element i B.

Däremot finns det inget krav att alla $y \in B$ måste kunna skrivas i formen $y = f(x)$ för något $x \in A$ så värdemängden kan vara en äkta delmängd av målmängden. Till exempel, till f i exempel 3.2(a) har vi valt målmängden $B = \mathbf{R}$ – ett rimligt val eftersom uttrycket $(x^2 + 4x - 32)/(x - 4)$ är säkert ett reellt tal för varje $x \in \mathbf{R} \setminus \{4\}$ – men det finns till exempel inget $x \in \mathbf{R} \setminus \{4\}$ som löser $12 = f(x)$, så $y = 12$ är ett element i f:s målmängd B men inte i f:s värdemängd.

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

$$\begin{aligned}(f+g)(x) &= f(x) + g(x),\\ (f-g)(x) &= f(x) - g(x) \quad \text{och}\\ (fg)(x) &= f(x)g(x)\end{aligned} \tag{3.1}$$

för $x \in A$. Kvoten f/g definieras ha definitionsmängden $\{x \in A \mid g(x) \neq 0\}$ och ges av

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

för varje $x \in A$ sådant att $g(x) \neq 0$. Alltså, till skillnad från $f + g$, $f - g$ och fg, kan f/g:s definitionsmängd skilja sig från f och g:s definitionsmängd.

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 3.3 *Låt $f\colon A \to B$, $g\colon B \to C$ och $h\colon C \to D$. Vi har då att $h \circ (g \circ f) = (h \circ g) \circ f$.*

Bevis. Både $h \circ (g \circ f)$ och $(h \circ g) \circ f$ har definitionsmängden A. Vi har också att

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

och

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

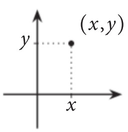
för alla $x \in A$. □

Exempel 3.4 (a) Låt $g: \mathbf{R} \to \mathbf{R}$ vara definierad enligt formeln $g(x) = x^2 - 2x + 1$ för alla $x \in \mathbf{R}$ och $h: \mathbf{R} \to \mathbf{R}$ vara definierad enligt formeln $h(x) = x + 5$ för alla $x \in \mathbf{R}$. Då är

(i) $(g + h)(x) = (x^2 - 2x + 1) + (x + 5) = x^2 - x + 6$,
(ii) $(g - h)(x) = (x^2 - 2x + 1) - (x + 5) = x^2 - 3x - 4$,
(iii) $(gh)(x) = (x^2 - 2x + 1)(x + 5) = x^3 + 3x^2 - 9x + 5$,
(iv) $(g \circ h)(x) = (x + 5)^2 - 2(x + 5) + 1 = x^2 + 8x + 16$ och
(v) $(h \circ g)(x) = (x^2 - 2x + 1) + 5 = x^2 - 2x + 6$

(b) Låt $n \in \mathbf{Z}_+$ och definiera $F_n: \mathbf{R} \to \mathbf{R}$ enligt formeln $F_n(x) = x^n$ för alla $x \in \mathbf{R}$. Då är

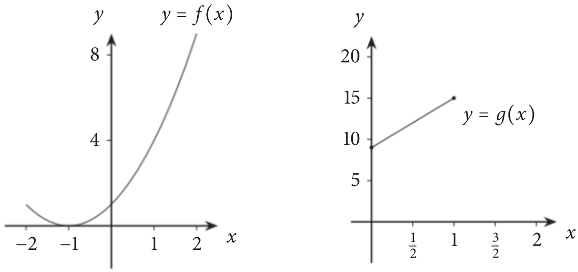
(i) $F_n F_m = F_{n+m}$
(ii) $F_n \circ F_m = F_{nm}$



FIGUR 3.1 Koordinataxlarna och en punkt med koordinaterna (x,y). Variablerna x och y är så pass vanliga att den vågräta axeln även kallas för x-axeln och den lodräta axeln för y-axeln.

KOORDINATSYSTEM OCH GRAFER

I ett *plan*, som man kan identifiera med ett idealt slätt papper med oändlig utsträckning, ritar vi två kopior av den reella linjen, den ena vinkelrät mot den andra så att de skär varandra i sina respektive origo. Se figur 3.1. Skärningspunkten kallas även för *origo* i planet och vi döper den ena linjen till den *lodräta axeln* och den andra till den *vågräta axeln*. En godtycklig punkt i planet kan identifieras genom två tal: Det första är avståndet med tecken x från punkten till den lodräta axeln. Det andra är avståndet med tecken y från punkten till den vågräta axeln. Talen kallas för *(kartesiska) koordinater* och skrivs (x,y). Mängden av alla par (x,y) där $x,y \in \mathbf{R}$ noteras $\mathbf{R}^2$. 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. Vi har redan använt oss av koordinatplanet i figur 2.5 utan att nämna det explicit!



(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 3.2 Här har vi ritat graferna av de första två funktionerna från exempel 3.1.

Man kan använda koordinatplanet 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

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

Sedan 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.[1] Se figur 3.2.

Grafer är förstås inte de enda delmängderna av $\mathbf{R}^2$ vi kan representera på koordinatplanet: den begränsade noggrannheten av pennan, storleken av pappret och tålamodet hos ritaren är de enda begränsningarna.

MONOTONICITET

För att förstå hur en funktion beter sig är det användbart att definiera begreppet *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

$$x_1,x_2 \in E \text{ och } x_1 < x_2 \text{ medför det att } f(x_1) \le f(x_2)$$

1 Kom ihåg att ett negativt avstånd till höger (ovanför) betyder till vänster (under).

och *avtagande* på E om

$$x_1,x_2 \in E \text{ och } x_1 < x_2 \text{ medför det att } f(x_1) \geq f(x_2).$$

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

$$x_1,x_2 \in E \text{ och } x_1 < x_2 \text{ medför det att } f(x_1) < f(x_2)$$

och *strängt avtagande* på E om

$$x_1,x_2 \in E \text{ och } x_1 < x_2 \text{ medför det att } f(x_1) > f(x_2).$$

Om en funktion är (strängt) växande, respektive (strängt) avtagande på hela definitionsmängden D säger man bara att funktionen är (strängt) växande, respektive (strängt) avtagande. En funktion som är avtagande eller växande (på en mängd) kallas för *monoton* (på mängden).

Exempel 3.5 (a) Funktionen $f\colon \mathbf{R} \to \mathbf{R}$ definierad som $f(x) = x^2$ för alla $x \in \mathbf{R}$ är strängt växande på $[0,\infty)$. För att visa det beräknar vi

$$x_2^2 - x_1^2 = (x_2 - x_1)(x_2 + x_1).$$

För $x_1,x_2 \in [0,\infty)$ sådana att $x_1 < x_2$ är $0 \leq x_1 < x_2$ så $x_1 + x_2 > 0$. Vi har även $x_2 - x_1 > 0$ ty $x_1 < x_2$. Därför är $(x_2 - x_1)(x_2 + x_1) > 0$ och därmed är $f(x_2) - f(x_1) \leq 0$ och $f(x_1) \leq f(x_2)$.

(b) På $(-\infty,0]$ är funktionen f strängt avtagande eftersom $x_1 < x_2 \leq 0$ medför att $f(x_1) = x_1^2 > x_2^2 = f(x_2)$. Beviset är detsamma som det i del (a) förutom att vi nu i stället kan argumentera att $x_2 + x_1 < 0$.

(c) Delar (a) och (b) medför att f varken är växande eller avtagande (på $\mathbf{R}$).

För att underlätta undersökningen av mer komplicerade funktioner kan vi bygga upp vissa allmänna regler vad gäller olika kombinationer av monotona funktioner.

Sats 3.6 *Låt $A \subseteq \mathbf{R}$, $f\colon A \to \mathbf{R}$ och $g\colon A \to \mathbf{R}$, och anta att $E \subseteq A$.*

(a) Om f är växande på E och g är strängt växande på E då är $f + g$ strängt växande på E.

(b) Om f är avtagande på E och g är strängt avtagande på E då är $f + g$ strängt avtagande på E.

Påståendena gäller även om man stryker "strängt" från både hypoteserna och slutsatserna.

Bevis. Här visar vi bara (a) och lämnar de andra som en övning för läsaren – se uppgift 3.2.

Då antar vi att f är växande på E och g är strängt växande på E. Det innebär att $x_1 < x_2$ medför $f(x_1) \leq f(x_2)$ och $g(x_1) < g(x_2)$ för $x_1, x_2 \in E$. Därför är $f(x_1) + g(x_1) \leq f(x_2) + g(x_1) < f(x_2) + g(x_2)$ så $(f + g)(x_1) < (f + g)(x_2)$ och $f + g$ är strängt växande på E. □

Exempel 3.7 Tydligen är funktionen $x \mapsto 5x + 2$ strängt växande på $\mathbf{R}$ och vi såg i del (a) av exempel 3.5 att $x \mapsto x^2$ är strängt växande på $[0,\infty)$. Därför säger sats 3.6 (a) att summan $x \mapsto x^2 + 5x + 2$ är strängt växande på $[0,\infty)$.

För att testa vår förståelse av begreppet monotonicitet kan man utreda om vi kan bevisa varianter av påståenden i sats 3.6. Till exempel kan man ersätta $f + g$ med $f - g$ eller fg, och vad behöver man ändra för att få en giltigt sats? Sådana frågor tar vi upp i uppgift 3.1 och 3.3 i avsnitt 3.3.

Vi kan också visa några räkneregler för sammansättningar av funktioner.

Sats 3.8 *Låt $A, B, C \subseteq \mathbf{R}$, $f\colon A \to B$ och $g\colon B \to C$.*

(a) *Om både f och g är strängt växande då är $(g \circ f)$ strängt växande.*
(b) *Om både f och g är strängt avtagande då är $(g \circ f)$ strängt växande.*
(c) *Om f är strängt växande och g är strängt avtagande då är $(g \circ f)$ strängt avtagande.*
(d) *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 bevisar vi bara (b) och lämnar de andra som en övning för läsaren – se uppgift 3.4.

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

Det är förstås slarvigt att skriva x^0 som faktiskt står som första termen av Σ-summan i (3.2). Vi vet att x^0 är odefinierad när $x = 0$, men vi gör det bara i det här avsnittet eftersom notationen underlättar vårt arbete här.

POLYNOM

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

$$p(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n = \sum_{k=0}^{n} a_kx^k \tag{3.2}$$

för något $n \in \mathbf{N}$ och $a_k \in \mathbf{R}$ $(k = 0, 1, \ldots, n)$ kallas för *polynom*. Talen a_k kallas för polynomets *koefficienter*. Om $a_n \neq 0$ kallas n polynomets *grad*, vilket skrivs som $n = \deg(p)$.[2] Speciellt om $n = 1$ kallas polynomet ett *förstagradspolynom* och 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 $a_k \in \mathbf{R}$.[3]

Precis som vi har gjort för andra begrepp kan vi direkt från definitionen utreda vissa egenskaper av begreppet grad. Nedan samlar vi några egenskaper i en sats vars bevis vi lämnar som en övning för läsaren – se uppgift 3.12.

Sats 3.9 *Om p och q är polynom så är $p + q$, pq och $p \circ q$ polynom. Dessutom*

(a) $\deg(p + q) \leq \max\{\deg(p), \deg(q)\}$,[4]

(b) $\deg(p + q) = \max\{\deg(p), \deg(q)\}$ *om* $\deg(p) \neq \deg(q)$,

(c) $\deg(pq) = \deg(p) + \deg(q)$ *och*

(d) $\deg(p \circ q) = \deg(p)\deg(q)$

för alla icke-triviala p och q.

Följande sats är nyckeln till att förstå många egenskaper hos polynom.

Sats 3.10 *Till varje par av polynom p och q varav q är icke-trivialt finns det polynom k och r så att*

$$\begin{aligned} &p(x) = q(x)k(x) + r(x) \text{ för alla } x \in \mathbf{R}, \text{ och} \\ &r \text{ är antingen trivialt eller uppfyller } \deg(r) < \deg(q). \end{aligned} \tag{3.3}$$

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

2 Notationen $\deg(p)$ härstammar från det engelska ordet *degree*.

3 I kapitel 7 betraktar vi polynom med koefficienter som tillhör en annan mängd än $\mathbf{R}$.

4 Kom ihåg att notationen $\max\{a, b\}$ betyder den största talet av a och b.

Observera att (3.3) har samma form som ekvation (1.13) för heltal. Sats 3.10 säger att division även är möjligt för polynom. Som namnet antyder är resten r liten jämfört med kvoten q, men för polynom behöver man mäta storleken av ett polynom med hjälp av polynomets grad eftersom vi har, till skillnad från heltal, ingen vanlig olikhet mellan polynom. Innan vi bevisar satsen är det bra att gå genom ett exempel.

Exempel 3.11 Dela $p(x) = 2x^4 - 5x^2 - x + 3$ med $q(x) = 2x^2 + 3$.

Lösning. Vi letar efter polynom k och r sådana att $p(x) = q(x)k(x) + r(x)$ och $\deg(r) < \deg(q) = 2$. Eftersom $\deg(p) = 4$ måste $\deg(k) = 2$ för att $\deg(p) = \deg(qk+r)$. Därför har k formen $k(x) = a_2x^2+a_1x+a_0$. Uppgiften är därför att hitta $a_0, a_1, a_2 \in \mathbf{R}$ sådana att

$$r(x) = p(x) - q(x)k(x) = p(x) - q(x)\left(a_2x^2 + a_1x + a_0\right)$$

har grad strängt mindre än två. Vi beräknar

$$\begin{aligned} r(x) &= p(x) - q(x)\left(a_2x^2 + a_1x + a_0\right) \\ &= (2x^4 - 5x^2 - x + 3) - (2x^2 + 3)(a_2x^2 + a_1x + a_0). \end{aligned}$$

Härifrån kan vi se att högsta graden av termerna är fyra och de enda termer som innehåller x^4 är $2x^4$ från första parentesen och $-2a_2x^4$ från utvecklingen av de sista två parenteserna. För att minimera graden av r väljer vi därför $a_2 = 1$ så de två termerna tar ut varandra. Sedan är

$$\begin{aligned} r(x) &= (2x^4 - 5x^2 - x + 3) - (2x^2 + 3)(x^2 + a_1x + a_0) \\ &= (2x^4 - 5x^2 - x + 3) - (2x^4 + 3x^2) - (2x^2 + 3)(a_1x + a_0) \\ &= (-8x^2 - x + 3) - (2x^2 + 3)(a_1x + a_0). \end{aligned}$$

Vi ser att högsta graden av termerna nu är tre och den enda termen som innehåller x^3 är $-2a_1x^2$ från utvecklingen av de sista två parenteserna. För att minimera graden av r väljer vi därför $a_1 = 0$. Så

$$\begin{aligned} r(x) &= (-8x^2 - x + 3) - (2x^2 + 3)(0 + a_0) \\ &= (-8x^2 - x + 3) - (2a_0x^2 + 3a_0). \end{aligned}$$

Till sist vill vi bli av med termerna av grad två och väljer därför $a_0 = -4$ så vi kommer fram till att

$$r(x) = -x + 15.$$

Vi har räknat ut att $k(x) = x^2 - 4$ och p delat med q är därför $x^2 - 4$ med rest $-x + 15$:

$$p(x) = q(x)(x^2 - 4) + (-x + 15).$$

Vi kan även skriva uppräkningar med hjälp av den liggande stolen, precis som man gör i lång division:

$$
\begin{array}{l|l}
x^2 - 4 & \\
\hline
2x^4 - 5x^2 - x + 3 & 2x^2 + 3 \\
\cline{2-2}
2x^4 + 3x^2 & \\
\cline{1-1}
\quad -8x^2 - x + 3 & \\
\quad -8x^2 - 12 & \\
\cline{1-1}
\quad -x + 15 &
\end{array}
$$

Här syns inte valet $a_1 = 0$ eftersom räkningen hoppar direkt från ett fjärdegrads- till ett andragradspolynom. $\square$

Precis samma metod ger ett bevis av sats 3.10.

Bevis (av sats 3.10). Först visar vi att sådana k och r finns. Vi delar upp beviset beroende på graden av p.

Fall 1: p är trivialt

I så fall är $p(x) = 0$ och tydligen kan man ta $k(x) = r(x) = 0$ för alla $x \in \mathbf{R}$.

Fall 2: $\deg(p) < \deg(q)$

Då kan vi ta $k(x) = 0$ och $r(x) = p(x)$ för alla $x \in \mathbf{R}$.

Fall 3: $\deg(p) \geq \deg(q)$

Skriv

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

där $n = \deg(p)$ och $m = \deg(q)$. Definitionen av polynomens grad innebär att $a_n \neq 0$ och $b_m \neq 0$. Därför, om vi definierar $k_1(x) :=$

$(a_n/b_m)x^{n-m}$ är

$$\begin{aligned}
p_1(x) &:= p(x) - q(x)k_1(x) \\
&= \sum_{k=0}^{n} a_k x^k - \left(\sum_{k=0}^{m} b_k x^k\right)\left(\frac{a_n}{b_m}\right)x^{n-m} \\
&= \sum_{k=0}^{n} a_k x^k - \sum_{k=0}^{m} \frac{a_n b_k}{b_m} x^{n-m+k} \\
&= \sum_{k=0}^{n} a_k x^k - \sum_{k=n-m}^{n} \frac{a_n b_{k-n+m}}{b_m} x^k \\
&= \sum_{k=0}^{n-1} a_k x^k - \sum_{k=n-m}^{n-1} \frac{a_n b_{k-n+m}}{b_m} x^k
\end{aligned} \tag{3.4}$$

där vi har använt oss av en variabelbyte i andra summan i den fjärde likheten och sista likheten gäller eftersom $a_n b_{k-n+m}/b_m = a_k$ om $k = n$. Därmed har p_1 grad strängt mindre än $\deg(p)$. Om $\deg(p_1) < \deg(q)$ har vi bevisat satsen med $k = k_1$ och $r = p_1$. Annars upprepar vi argumentet för att konstruera polynom p_2 och k_2, sedan p_3 och k_3, fram till p_ℓ och k_ℓ där $\deg(p_\ell) < \deg(q)$.

För tydlighets skull beskriver vi det s:e steget där vi har redan konstruerat p_{s-1} och k_{s-1} och $\deg(p_{s-1}) \geq \deg(q) = m$. Omdefiniera $n = \deg(p_{s-1})$ och a_k till p_{s-1}:s koefficienter. Därför är

$$p_{s-1}(x) = \sum_{k=0}^{n} a_k x^k$$

där $a_n \neq 0$, och $b_m \neq 0$ precis som förut. Låt

$$k_s(x) := (a_n/b_m)x^{n-m}$$

och

$$p_s(x) := p_{s-1}(x) - q(x)k_s(x). \tag{3.5}$$

Vi kan beräkna som vi gjorde i (3.4) att p_s har grad strängt mindre än n.

Eftersom $\deg(p_s)$ är minst 1 mindre än $\deg(p_{s-1})$ vet vi att det finns ett $\ell \leq \deg(p) - \deg(q) + 1$ sådan att $\deg(p_\ell) < \deg(q)$. Utifrån (3.4) och (3.5) ser vi att $k := k_1 + k_2 \cdots + k_\ell$ och $r := p_\ell$ uppfyller (3.3).

Nu visar vi att k och r är entydiga. Anta att det finns två par av polynom k_1 och r_1, och k_2 och r_2 som uppfyller (3.3):[5] Vi vill visa att de

5 Att vi återanvänder notationen k_1 och k_2 borde inte vara något problem.

faktiskt är samma par. Utifrån (3.3) är

$$0 = p(x) - p(x) = q(x)(k_2(x) - k_1(x)) + (r_2(x) - r_1(x))$$

för alla $x \in \mathbf{R}$. Vi definierar $\tilde{k} = k_2 - k_1$ och $\tilde{r} = r_2 - r_1$, så

$$0 = q(x)\tilde{k}(x) + \tilde{r}(x) \tag{3.6}$$

för alla $x \in \mathbf{R}$.

Först ger vi ett motsägelsebevis på att $\tilde{k}$ måste vara trivialt. Om $\tilde{k}$ är icke-trivialt medför del (b) och (c) från sats 3.9 att

$$\deg(q\tilde{k} + \tilde{r}) = \deg(q) + \deg(\tilde{k}) > 0,$$

eftersom $\tilde{r}$ antingen är trivialt eller $\deg(\tilde{r}) < \deg(q)$ och q är icke-trivialt. Därför måste $q\tilde{k} + \tilde{r}$ vara icke-trivialt, vilket är en motsägelse till (3.6). Slutsatsen är då att $\tilde{k} = 0$ och därför är $k_2 = k_1$.

Nu vet vi att (3.6) har formen $0 = \tilde{r}(x)$ så det följer direkt att $r_2 = r_1$ och vi har visat att det finns ett entydigt par av polynom k och r som uppfyller (3.3). □

Vi säger att ett polynom p är *delbart* med ett annat polynom q om (3.3) gäller med $r(x) = 0$ för alla $x \in \mathbf{R}$. Ett tal $c \in \mathbf{R}$ kallas för ett *nollställe* till polynomet p om det löser ekvationen $p(c) = 0$. Följande sats ger ett samband mellan nollställen och delbarhet med förstagradspolynom.

Sats 3.12 *Talet $c \in \mathbf{R}$ är ett nollställe till polynomet p om och endast om p är delbart med $x \mapsto x - c$.*

Bevis. Om c är ett nollställe till p så är $p(c) = 0$. Om vi tillämpar sats 3.10 med $q(x) = x - c$ får vi att

$$p(x) = k(x)(x - c) + r(x) \tag{3.7}$$

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 $x = c$ i (3.7) 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 ett polynom k och för alla $x \in \mathbf{R}$. Därför är $p(c) = k(c)(c - c) = 0$ och c är ett nollställe till p. $\square$

Följande sats säger att ett polynom av grad n kan ha högst n nollställen.

Sats 3.13 *För ett polynom av formen $p(x) = \sum_{k=0}^{n} a_k x^k$ anta att det finns tal $x_0 < x_1 < \cdots < x_n$ så att $p(x_j) = 0$ för alla $j = 0, 1, \ldots, n$. Vi kan då dra slutsatsen att $p(x) = 0$ för alla $x \in \mathbf{R}$.*

Bevis. Vi bevisar satsen med hjälp av induktion över n.

Först betraktar vi $n = 0$. Då är

$$p(x) = a_0 \quad \text{för alla } x \in \mathbf{R} \tag{3.8}$$

och

$$p(x_0) = 0 \quad \text{för ett givet } x_0 \in \mathbf{R}. \tag{3.9}$$

Om vi tar $x = x_0$ i (3.8) medför det tillsammans med (3.9) att $a_0 = p(x_0) = 0$. Därför är, enligt (3.8), $p(x) = a_0 = 0$ 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 högst $m + 1$ 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 3.12 att

$$p(x) = k(x)(x - x_{m+1}) \tag{3.10}$$

där k är ett polynom av grad högst m (enligt del (c) av sats 3.9). Men eftersom $x_j \neq x_{m+1}$ för alla $j = 0, 1, \ldots, m$ vet vi från (3.10) 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 på k och dra slutsatsen att $k(x) = 0$ för alla $x \in \mathbf{R}$. Enligt (3.10) är då $p(x) = 0$ för alla $x \in \mathbf{R}$ som är satsens slutsats där $n = m + 1$.

Enligt induktionsprincipen (axiom 2.6) har vi bevisat satsen för alla $n \in \mathbf{N}$. $\square$

Följande sats säger att olika termer i ett polynom inte kan ta ut varandra – ett polynoms värden bestämmer dess koefficienter. Man kan även kombinera sats 3.13 och 3.14 för att säga att värden av ett polynom av grad n i n punkter preciserar polynomets koefficienter. Se uppgift 3.7.

Sats 3.14 *För ett polynom av formen* $p(x) = \sum_{k=0}^{n} a_k x^k$ *och sådan att* $p(x) = 0$ *för alla* $x \in \mathbf{R}$ *måste* $a_k = 0$ *för alla* $k = 0, 1, \ldots, n$.

Bevis. Som ovan 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 $a_0 = 0$ 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 $p(x) = \sum_{k=0}^{m+1} a_k x^k$ av grad högst $m + 1$ sådant 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

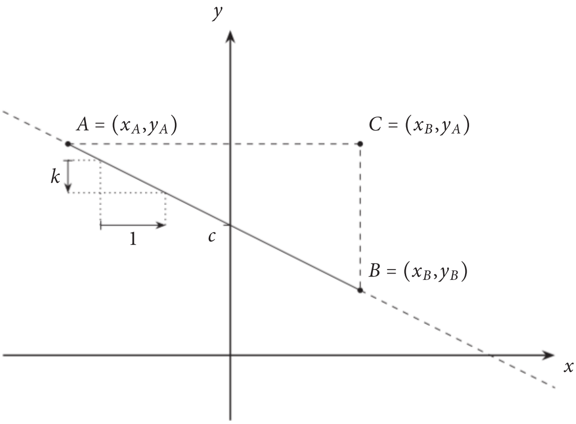
$$p(x) = \sum_{k=1}^{m+1} a_k x^k = \left(\sum_{k=1}^{m+1} a_k x^{k-1} \right) x. \tag{3.11}$$

Polynomet $q(x) := \sum_{k=1}^{m+1} a_k x^{k-1}$ har grad högst m och eftersom $p(x) = 0$ för alla $x \in \mathbf{R}$ är $q(x) = 0$ för alla $x \in \mathbf{R} \setminus \{0\}$ enligt (3.11). Men i så fall medför sats 3.13 att $q(x) = 0$ även för $x = 0$ (eftersom vi kan välja vilka $x_j \neq 0$ som helst för att tillämpa sats 3.13). Enligt induktionsantagandet är alla q:s koefficienter lika med noll, det vill säga $a_k = 0$ för alla $k = 1, 2, \ldots, m + 1$.

Enligt induktionsprincipen (axiom 2.6) har vi bevisat satsen för alla $n \in \mathbf{N}$. □

3.2 Former, längd och area

I det här avsnittet koncentrerar vi oss på geometriska begrepp såsom linjer, bågar och former, och area, längd och vinkel. Vi kommer även att prata om den välkända *Pythagoras sats*.



FIGUR 3.3 Strecket mellan A och B är värdemängden av funktionen $\gamma\colon [0,1] \to \mathbf{R}$ definierad i (3.12). Strecket kan förlängas till en hel linje (streckad). Ekvation (3.14) ger den kartesiska formen av linjen. Den skär y-axeln i c och har lutningen k, det vill säga linjens y-höjd ändrar sig med k enheter för varje enhetsförflyttning längs x-riktningen. Bilden visar en negativ lutning k och därför sjunker linjen från vänster till höger.

LINJER, BÅGAR OCH VINKLAR

I figur 3.3 har vi ritat ett *streck* mellan två punkter A och B i koordinatplanet: Om vi säger att A har koordinator (x_A, y_A) och B har koordinator (x_B, y_B), har funktionen $\gamma\colon [0,1] \to \mathbf{R}^2$ som ges av

$$\gamma(t) = (x_A + t(x_B - x_A), y_A + t(y_B - y_A)) \quad \text{för } t \in [0,1] \tag{3.12}$$

värdemängd som är strecket mellan A och B. Funktionen γ kallas för en *parametrisering* av strecket mellan A och B. Ett streck är ett exempel på en *kurva* – se bilaga C.3. Koordinaterna till en godtycklig punkt (x,y) på strecket som ges av γ uppfyller

$$\begin{aligned} x &= x_A + t(x_B - x_A) \quad \text{och} \\ y &= y_A + t(y_B - y_A) \end{aligned} \tag{3.13}$$

för något $t \in [0,1]$. Vi kan skriva om (3.13) till en ekvation i x och y genom att eliminera t: Från den första ekvationen är

$$t = \frac{x - x_A}{x_B - x_A}$$

om $x_B \neq x_A$ och med hjälp av den andra då är

$$\begin{aligned} y &= y_A + \frac{x - x_A}{x_B - x_A}(y_B - y_A) \\ &= \left(\frac{y_B - y_A}{x_B - x_A}\right)x + \left(y_A - x_A\frac{y_B - y_A}{x_B - x_A}\right) \end{aligned} \tag{3.14}$$

som är i så kallad *kartesisk form* $y = kx + c$ med *lutningen*

$$k = \frac{y_B - y_A}{x_B - x_A}$$

och skär y-axeln i

$$c = y_A - x_A\frac{y_B - y_A}{x_B - x_A}.$$

Observera att ett lodrätt streck (det vill säga när $x_B = x_A$) inte kan skrivas i kartesisk form. Alla punkter (x,y) på ett icke-lodrätt streck mellan A och B uppfyller (3.14) men en punkt (x,y) som uppfyller (3.14) måste inte ligga mellan A och B: den kan hamna på förlängningen av strecket mellan A och B till en hel rak linje och skulle kunna motsvara en förlängning av funktionen γ med samma formel (3.12) fast för $t \in \mathbf{R}$ i stället för bara $t \in [0,1]$.

Vi vill utvidga begreppet längd från intervall (se sidan 45) till allmänna streck i planet. Det kommer till viss del att vara en fortsättning på diskussionen om reella tal från avsnitt 2.2. Vi betecknar längden av strecket mellan punkterna A och B eller avståndet mellan A och B som

$$\ell(A,B). \tag{3.15}$$

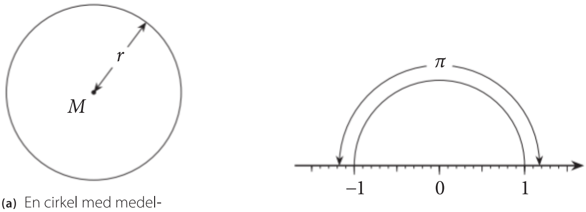
Ett syfte med det här avsnittet är att hitta ett samband mellan punkternas koordinator och avståndet $\ell(A,B)$. Men för tillfället fortsätter vi utan formel och litar på vår föraning om avstånd.

Med hjälp av avståndet (3.15) kan vi definiera vår första form: Mängden

$$\{P \in \mathbf{R}^2 : \ell(M,P) = r\} \tag{3.16}$$

av alla punkter som ligger på ett avstånd $r > 0$ från en given punkt M kallas för en *cirkel* med *medelpunkten* M och *radien* r. Se figur 3.4(a).

Vi kan betrakta en cirkels *omkrets* eller även en *båge*, det vill säga en del av omkretsen, på liknande sätt som längden av ett rakt streck. I



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

(b) Talet π definieras som längden av en halv enhetscirkel.

FIGUR 3.4 En cirkel och längden π.

bilaga C.3 ger vi en mer allmän definition av längden av en kurva, fast det är ganska svårhanterat, så i de flesta fall undviker vi att använda den direkt och i stället lita på symmetri och våra geometriska intuition. Vi gör det genom att först skapa en referenslängd som vi kan jämföra andra längder med.

Betrakta en *enhetscirkel*, det vill säga en cirkel med radien 1. Längden ändras inte om vi förflyttar cirkeln så det spelar ingen roll var medelpunkten ligger. Vi klipper cirkeln i två lika halvor med hjälp av en rak linje genom medelpunkten. Vi betecknar längden av vardera halvan π (figur 3.4(b)).[6] Enligt symmetri är cirkelns omkrets 2π, en fjärdedel av omkretsen lika med $\pi/4$, och så vidare.

Om vi betraktar en cirkel med radien r istället för 1 säger vår geometriska intuition att alla längder multipliceras med r. Det betyder att en cirkel med radien r har omkretsen

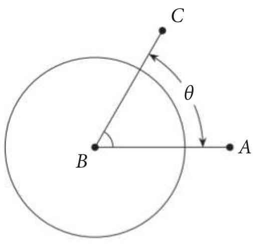
$$2\pi r$$

och en q:e del av omkretsen är förstås

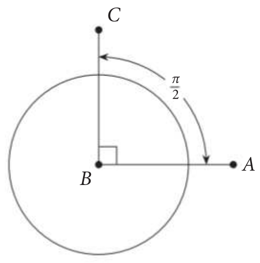
$$\frac{2\pi r}{q}. \tag{3.17}$$

När två streck möter varandra i en punkt kan vi tala om *vinkeln* mellan de två linjerna med hjälp av enhetscirkeln. Betrakta två streck, det ena mellan A och B som vi kallar för AB, och det andra mellan B och C, som vi kallar för BC. Rita en cirkel med radien 1 och medelpunkten

6 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.

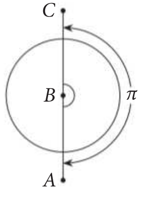


(b) En vinkel mellan AB och BC är en rätvinkel när den är lika med $\pi/2$.

FIGUR 3.5 Vinklar.

B. Vinkeln definieras som längden θ av den delen av cirkeln som ligger mellan AB och BC (eventuellt förlängda vid behov, så att strecken skär cirkeln). Se figur 3.5(a).

I synnerhet om A, B och C ligger på en gemensam rät linje i denna ordningsföljd är vinkeln π – se figur 3.6. Med den här definitionen kallas mätenheten för vinklar *radianer*. Det är tvetydigt åt vilket håll man ska mäta vinkeln: om man mäter åt ett håll är vinkeln högst π och åt andra hållet är den minst π.[7] Den första mätenheten för vinklar de flesta lär sig om är *grader*, då ett helt varv är 360° (360 grader) i stället för 2π radianer, men radianer är en mer naturlig mätenhet som förenklar formler i mer avancerad teori. Vinkeln $\pi/2$ dyker upp ganska ofta och därför betecknar vi den med en liten fyrkant i hörnet i stället för en båge och kallar den för en *rätvinkel*. Se figur 3.5(b) och jämför med figur 3.5(a).

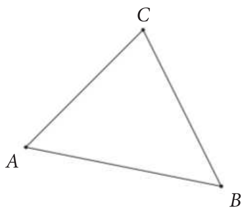


FIGUR 3.6 När A, B och C ligger på en gemensam rät linje är vinkeln mellan AB och BC π.

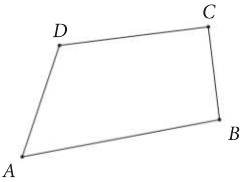
TRIANGLAR, FYRHÖRNINGAR OCH AREA

En *triangel* skapas av tre punkter A, B och C i ett plan som inte ligger på en gemensam rät linje tillsammans med tre streck som börjar och slutar i punkterna A, B och C. Se figur 3.7(a). Observera att det bara finns ett sätt att dra tre olika linjer mellan tre punkter. Punkterna A, B och C kallas för *hörnpunkter* och linjerna kallas för *sidor* (eller *kanter*). En triangel som har en rätvinkel mellan två sidor kallas för en *rätvinklig*

7 Senare i kapitel 5 lägger vi till ett tecken till vinkeln och tillåter även flera varv runt enhetscirkeln, så en vinkel kan vara vilket positivt eller icke-positivt reellt tal som helst.



(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 fyra 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 3.7 Två former.

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

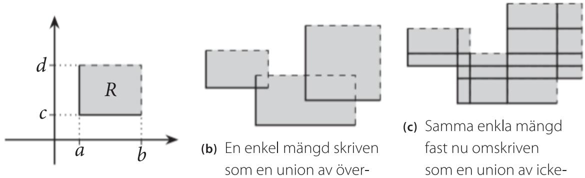
En *fyrhörning* skapas av fyra punkter A, B, C och D med fyra linjer som börjar och slutar i punkterna A, B, C och D på så sätt att linjerna inte skär varandra och varje punkt är början eller slutet på två linjer. Se figur 3.7(b). Precis som för en triangel kallas punkterna A, B, C och D för *hörnpunkter* 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.

Varje form delar planet i två disjunkta mängder: formens *inre* och *yttre*. Ibland smälter vi samman en form med dess inre. Vi kan mäta storleken av en mängd genom att mäta *arean* av den (eller arean av det inre när det gäller en form). Arean av en mängd är analog med längden av en kurva, men hur definierar man area? Det finns faktiskt flera sätt, men definitionen vi beskriver här kallas för ett *Jordanmått*.

Vi börjar genom att definiera vad arean av en rektangulär mängd är. Mer precist sagt definierar vi en *rektangulär mängd* R som den *kartesiska produkten* av två intervall av formen $[a,b)$ och $[c,d)$, det vill säga

$$R = [a,b) \times [c,d) := \{(x,y) \mid x \in [a,b) \text{ och } y \in [c,d)\}.$$



(a) En rektangulär mängd $R = [a,b) \times [c,d)$.

(b) En enkel mängd skriven som en union av överlappande rektangulära mängder.

(c) Samma enkla mängd fast nu omskriven som en union av icke-överlappande rektangulära mängder.

FIGUR 3.8 Rektangulära mängder. Vi markerar om en sida tillhör mängden eller inte med en heldragen, respektive streckad linje.

för några $a,b,c,d \in \mathbf{R}$, $a < b$ och $c < d$. Sedan definieras arean av R som

$$|R| := (b-a)(d-c).$$

Se figur 3.8(a). Extremt kan vi även betrakta $R = \varnothing$ som en rektangulär mängd – här kan man tänka att vi har tagit $a = b$ eller $c = d$, men i strikt mening har vi inte definierat intervall med lika ändpunkter – och vi definierar $|\varnothing| := 0$. För en union av icke-överlappande rektangulära mängder $R_j := [a_j,b_j) \times [c_j,d_j)$ $(a_j \leq b_j, c_j \leq d_j, j = 1,\ldots,n)$ definieras arean som summan av areorna av varsin rektangulär mängd:

En samling av mängder $(S_j)_j$ kallas för *icke-överlappande* eller *parvis disjunkta* om snittet mellan varje par $S_{j_1} \cap S_{j_2}$ $(j_1 \neq j_2)$ är den tomma mängden $\varnothing$.

$$\left|\bigcup_{j=1}^{n} R_j\right| := \sum_{j=1}^{n} |R_j| = \sum_{j=1}^{n} (b_j - a_j)(d_j - c_j). \tag{3.18}$$

För en union S av rektangulära mängder som är överlappande, kan man skriva om den som en union av icke-överlappande rektangulära mängder (se figur 3.8(b) och 3.8(c)) och sedan definiera arean enligt (3.18). Utifrån den definitionen är det möjligt att visa att arean $|S|$ av S är oberoende av hur man skriver om S som en union av icke-överlappande rektangulära mängder. Vi kallar mängder som är unioner av rektangulära mängder för *enkla mängder*.

Om vi skulle definiera arean av mer komplicerade mängder $M \subset \mathbf{R}^2$ skulle man förvänta sig att om $S \subseteq M \subseteq T$ skulle $|S| \leq |M| \leq |T|$. Eftersom vi redan har definierat arean av alla enkla mängder försöker vi uppskatta mängden M både inifrån och utifrån med enkla mängder S och T. Allmänt vet vi att

$$\sup_{S \subseteq M} |S| \leq \inf_{T \supseteq M} |T| \tag{3.19}$$

där supremum tas över alla möjliga enkla mängder $S \subseteq M$ och infimum tas över alla möjliga enkla mängder $T \supseteq M$. Men i fallet

$$\sup_{S \subseteq M} |S| = \inf_{T \supseteq M} |T|$$

definierar vi $|M| = \sup_{S \subseteq M} |S|$. På så sätt har vi definierat arean av delmängder av $\mathbf{R}^2$ som är väl approximerade med enkla mängder – sådana mängder M kallas för *(Jordan)mätbara*. Man kan fråga om alla delmängder av $\mathbf{R}^2$ är mätbara, det vill säga: Är alla delmängder av $\mathbf{R}^2$ väl approximerade med enkla mängder? Svaret är nej men ett bevis av detta är utanför bokens ram. Emellertid är många mängder mätbara och vår definition är nyttig i många sammanhang.

Vår intuition om area säger att den borde vara oberoende av en mängds position och orientering i planet. Det stämmer för vår definition här: Eftersom arean av en rektangulär mängd är oförändrad under en parallellförflyttning, gäller detsamma för parallellförflyttningar av alla mängder. När det gäller rotationer av mängder är det inte så uppenbart, men det kan också visas genom att visa att arean av en rektangulär mängd är oförändrad under en rotation.

Vi förväntar oss också att arean av unionen $M_1 \cup M_2$ av två icke-överlappande mätbara mängder M_1 och M_2 är mätbar och

$$|M_1 \cup M_2| = |M_1| + |M_2|. \tag{3.20}$$

I fallet $M_1 \cap M_2 \neq \varnothing$ bör

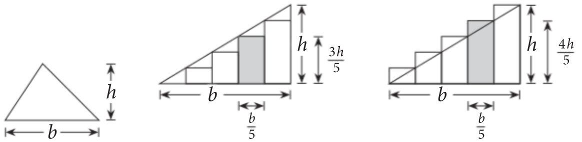
$$|M_1 \cup M_2| = |M_1| + |M_2| - |M_1 \cap M_2|.$$

Vi bevisar inte sanningen av dessa förväntningar här eftersom det är lite klurigt trots att själva resultaten är troliga. I stället vänder vi oss mot ett enklare resultat.

Sats 3.15 *Arean av en triangel med bas b och höjd h (se figur 3.9(a)) är $bh/2$.*

Bevis. Eftersom arean är oberoende av en mängds position och orientering i planet, kan vi anta att triangelns bas ligger på x-axeln.

Vi börjar med att betrakta en rätvinklig triangel med kateterna av längden b och h. Man kan förstås argumentera med hjälp av symmetri att



(a) En triangel med bas b och höjd h.

(b) En enkel mängd som är en delmängd av en rätvinklig triangel.

(c) En enkel mängd som har triangeln som en delmängd.

FIGUR 3.9 En triangel är väl approximerad av enkla mängder.

arean måste vara hälften av arean av en rektangel med sidolängderna b och h och därför är arean $bh/2$, men för att träna på de här nya begreppen ger vi också ett bevis utifrån definitionen.

Vi döper triangeln (eller snarare triangelns inre) till M och betraktar den enkla mängden $S_n \subseteq M$ som är unionen av $(n-1)$ rektangulära mängder av bredden b/n där den j:e har höjden jh/n – se figur 3.9(b). Vi beräknar

$$|S_n| = \sum_{j=1}^{n-1} \left(\frac{b}{n}\right)\left(\frac{jh}{n}\right) = \frac{bh}{n^2}\sum_{j=1}^{n-1} j \underset{\substack{\uparrow \\ \text{sats 1.13}}}{=} \frac{bh(n-1)n}{2n^2} = \frac{bh}{2} - \frac{bh}{2n}.$$

Supremum över alla möjliga enkla mängder $S \subseteq M$ måste vara minst lika stort som supremum över S_n för $n \in \mathbf{Z}_+$. Därför är

$$\sup_{S\subseteq M} |S| \geq \sup_{n\in\mathbf{Z}_+} |S_n|.$$

Utöver det är $(|S_n|)_n$ en växande och uppåt begränsad följd så

$$\sup_{n\in\mathbf{Z}_+} |S_n| \underset{\substack{\uparrow \\ \text{sats 2.33(a)}}}{=} \lim_{n\to\infty} |S_n| \underset{\substack{\uparrow \\ \text{sats 2.35}}}{=} \frac{bh}{2} - \frac{bh}{2}\lim_{n\to\infty}\frac{1}{n} \underset{\substack{\uparrow \\ \text{ex. 2.32}}}{=} \frac{bh}{2}.$$

Å andra sidan kan vi betrakta den enkla mängden $T_n \supseteq M$ som är unionen av n rektangulära mängder av bredden b/n där den j:e har höjden jh/n – se figur 3.9(c). Vi beräknar

$$|T_n| = \sum_{j=1}^{n} \left(\frac{b}{n}\right)\left(\frac{jh}{n}\right) = \frac{bh}{n^2}\sum_{j=1}^{n} j \underset{\substack{\uparrow \\ \text{sats 1.13}}}{=} \frac{bhn(n+1)}{2n^2} = \frac{bh}{2} + \frac{bh}{2n}.$$

Infimum över alla möjliga enkla mängder $T \supseteq M$ kan inte vara större än infimum över T_n ($n \in \mathbf{Z}_+$). Därför är

$$\inf_{T \supseteq M} |T| \leq \inf_{n \in \mathbf{Z}_+} |T_n|.$$

Följden $(|T_n|)_n$ är avtagande och nedåt begränsad så

$$\inf_{n \in \mathbf{Z}_+} |T_n| \underset{\text{sats 2.33(b)}}{=} \lim_{n \to \infty} |T_n| \underset{\text{sats 2.35}}{=} \frac{bh}{2} + \frac{bh}{2} \lim_{n \to \infty} \frac{1}{n} \underset{\text{ex. 2.32}}{=} \frac{bh}{2}.$$

Tillsammans är då

$$\inf_{T \supseteq M} |T| \leq \frac{bh}{2} \leq \sup_{S \subseteq M} |S|$$

och, eftersom vi alltid har (3.19), är $\inf_{T \supseteq M} |T| = \sup_{S \subseteq M} |S| = bh/2$. Därför är M mätbar och $|M| = bh/2$.

Ifall triangeln inte är rätvinklig efter en eventuell rotation och parallellförflyttning är triangelns hörn koordinaterna $(0,0)$, $(b,0)$ och (a,h) där $0 < a < b$. Då kan vi dela upp $M = M_1 \cup M_2$ där

$$\begin{aligned} M_1 &= \{(x,y) \mid 0 < y < hx/a < h\} \quad \text{och} \\ M_2 &= \{(x,y) \mid 0 < y < h(b-x)/(b-a) \leq h\}. \end{aligned}$$

Vi kan kontrollera att $M_1 \cap M_2 = \varnothing$ och enligt argumentet ovan är $|M_1| = ah/2$ och $|M_2| = (b-a)h/2$ så

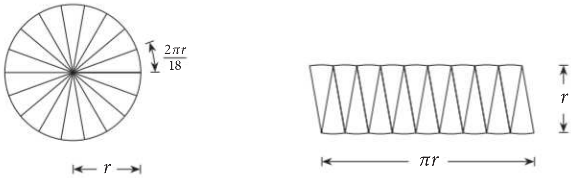
$$|M| = |M_1 \cup M_2| \underset{(3.20)}{=} |M_1| + |M_2| = \frac{ah}{2} + \frac{(b-a)h}{2} = \frac{bh}{2}. \qquad \square$$

Med lite mer arbete (och trigonometri från kapitel 5 gör det betydligt enklare) kan man visa följande sats.

Sats 3.16 *Arean av en cirkel med radien r är πr^2.*

Här bevisar vi inte satsen, men det är relativt enkelt att åtminstone motivera den. Betrakta en cirkel med radie r och dela upp den i n lika stora pajbitar – se figur 3.10(a). Varje pajbit är *ungefär* en likbent triangel[8]

8 En likbent triangel är en triangel som har två sidor som är lika långa.



(a) En cirkel uppdelad i 18 lika stora pajbitar.

(b) En förflyttning och omorientering av pajbitarna.

FIGUR 3.10 En motivering till formeln för en cirkels area.

med höjd r och bas $2\pi r/n$. Vi kan flytta om pajbitarna så att de *ungefär* formar en rektangel med längderna r och πr – se figur 3.10(b). Båda formerna ska ha samma areor eftersom den ena är förflyttningar och rotationer av delmängder av den andra. Arean av rektangeln är πr^2 så cirkelns area ska också vara πr^2.

Motivationen ovan låter övertygande men är inget fullständigt bevis av sats 3.16 eftersom pajbitarna inte bildar en exakt rektangel. Man behöver ta gränsvärdet $n \to \infty$ för att få en riktig rektangel men sedan dyker det upp subtila frågor om hur väl omflyttningen av pajbitarna närmar sig en rektangel.

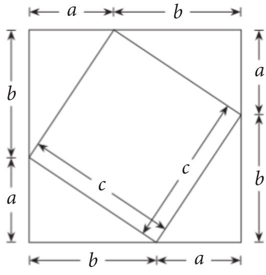
PYTHAGORAS SATS

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[9] redan tusen år innan Pythagoras föddes men vi vet inte om de kände till sambandet med trianglar. Det finns även en kinesisk avhandling, samtida med Pythagoras, som innehåller ett bevis av satsen. [11, s. 26–31]

Sats 3.17 (Pythagoras sats) *Betrakta en rätvinklig triangel som har sidor med längden a, b och c. Anta att hypotenusan har längden c. Vi har då att*

$$a^2 + b^2 = c^2.$$

9 Till exempel, 3, 4 och 5, och 5, 12 och 13.



FIGUR 3.11 En kvadrat med sidor av längden c inom en kvadrat med sidor av längden $a + b$.

Bevis. Betrakta en kvadrat med sidorna av längden $a + b$. Markera en punkt på toppsidan som har längden a från vänsterkanten. Gör detsamma på de tre andra sidorna och rita en kvadrat som har de fyra punkterna som hörnpunkter. Se figur 3.11.

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 största 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 $a^2 + b^2 = c^2$. □

Med hjälp av Pythagoras sats kan vi infria målet vi skrev på sidan 102: Att hitta ett samband mellan längden $\ell(A,B)$ mellan två punkterna A och B och punkternas koordinater. Anta att A och B har koordinaterna (x_A, y_A) respektive (x_B, y_B) (se figur 3.3). Vi kan bilda en triangel med hörnen A, B och en tredje punkt C som definieras ha koordinaterna (x_B, y_A). Sedan har triangeln hypotenusan AB och kateterna AC och BC. Längderna av kateterna är lätta att räkna ut: $\ell(A,C) = |x_B - x_A|$ och $\ell(B,C) = |y_B - y_A|$ så sats 3.17 säger att

$$\ell(A,B)^2 = (x_B - x_A)^2 + (y_B - y_A)^2. \tag{3.21}$$

En följd av detta är att alla punkter (x, y) som ligger på ett avstånd r från en punkt $M = (a, b)$ uppfyller ekvationen

$$r^2 = (x - a)^2 + (y - b)^2. \tag{3.22}$$

Det vill säga att mängden (3.16) är en delmängd av mängden av alla lösningarna $P = (x, y)$ till (3.22). Mängden av alla lösningarna $P = (x, y)$ till (3.22) är faktiskt precis lika med (3.16) – det visar vi i exempel 4.9.

3.3 Sammanfattning och uppgifter

I kapitel 3 har vi:

- diskuterat funktioner och deras förknippade mängder;
- definierat vad vi menar med en monoton funktion och utrett vad vi kan säga när vi till exempel adderar eller sammansätter monotona funktioner;
- diskuterat koordinatsystemet och hur vi kan använda det för att rita grafer av funktioner;
- utrett egenskaper såsom delbarhet hos polynom och sett att vi kan dela polynom på liknande sätt som man kan dela heltal;
- diskuterat geometriska former, linjer och bågar;
- använt suprema och infima för att ge en konkret definition av area;
- bevisat Pythagoras sats och använt den för att ge ett samband mellan två punkters koordinater och avståndet mellan punkterna.

UPPGIFTER

3.1 Man kan inte ersätta summan $f + g$ i sats 3.6 med produkten fg. Ge ett exempel som visar att det inte går. Förklara med dina egna ord varför det inte går.

3.2 Bevisa de kvarstående delarna av sats 3.6.

3.3 Bevisa en variant av sats 3.6 med $f - g$ i stället för $f + g$.

3.4 Bevisa de kvarstående delarna av sats 3.8.

3.5 Skissa graferna av följande funktioner $f\colon \mathbf{R} \to \mathbf{R}$.

(a) $f(x) = 2x - 2$.

(b) $f(x) = \begin{cases} 2x - 2 & \text{om } x \leq 2, \\ 2 + 2x - x^2 & \text{om } x > 2. \end{cases}$

(c) $f(x) = 2x^2 + 8x + 16$.

3.6 Bevisa direkt från definitionen att följande funktioner är strängt växande.

(a) $f\colon \mathbf{R} \to \mathbf{R}$ definierad enligt formeln $f(x) = 2x - 2$.

(b) $f\colon [0,\infty) \to \mathbf{R}$ definierad enligt formeln $f(x) = x^2$.

(c) $f\colon \mathbf{R} \to \mathbf{R}$ definierad enligt formeln $f(x) = x^3$.

3.7 Bevisa den följande satsen genom att använda sats 3.13 och 3.14.

Sats *Anta att p är ett polynom av formen $p(x) = \sum_{k=0}^{n} a_k x^k$ och det finns tal $x_0 < x_1 < \cdots < x_n$ så att $p(x_j) = 0$ för alla $j = 0, 1, \ldots, n$. Vi kan då dra slutsatsen att $a_k = 0$ för alla $k = 0, 1, 2, \ldots, n$.*

3.8 Hitta $a, b, c \in \mathbf{R}$ så att ekvationen

$$(2a - 5)x^2 + (5b + c)x + (c - a) = 0$$

gäller för alla $x \in \mathbf{R}$. Motivera ditt svar.

3.9 Skissa mängderna

(a) $\{(x,y) \in \mathbf{R}^2 \mid x^2 + y^2 = 1\}$,

(b) $\{(x,y) \in \mathbf{R}^2 \mid x^2 + y^2 = r^2\}$ och

(c) $\{(x,y) \in \mathbf{R}^2 \mid (x-a)^2 + (y-b)^2 = r^2\}$,

för givna $(a,b) \in \mathbf{R}^2$ och $r > 0$. Motivera dina skisser med hjälp av Pythagoras sats (sats 3.17).

3.10 Skissa mängderna

(a) $\{(x,y) \in \mathbf{R}^2 \mid |x| + |y| = 1\}$,

(b) $\{(x,y) \in \mathbf{R}^2 \mid |x+y| + |x-y| = 2\}$ och

(c) $\{(x,y) \in \mathbf{R}^2 \mid x \geq 0,\ y \geq 0 \text{ och } x + y \leq 1\}$.

Motivera dina skisser.

3.11 Bevisa genom lämpliga uppskattningar med enkla mängder att mängden

$$P = \{(x,y) \in \mathbf{R}^2 \mid 0 \leq y < x^2 \text{ och } 0 \leq x < h\}$$

har arean $h^3/3$. Formeln (2.6) kan vara till nytta här.

3.12 Bevisa del (b) och (c) från sats 3.9. Hitta ett exempel av polynom p och q som visar att man inte nödvändigtvis får likhet i del (a).

DEL II

Elementära funktioner, trigonometri och komplexa tal

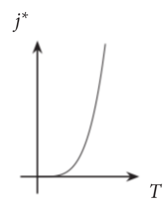
Kvadratrötter och andra inversa funktioner

4.1 Inversa funktioner och irrationella tal

I det här avsnittet utforskar vi när det på något sätt är möjligt att göra ogjort vad en funktion har gjort. Vi utforskar närmare bestämt när man utifrån ett givet element i funktionens värdemängd, kan avgöra vilket element i funktionens definitionsmängd som skickas av funktionen till detta element. Senare i avsnittet upptäcker vi även ytterligare egenskaper hos reella tal, nämligen att det finns reella tal som inte ens kan skrivas som ett bråk mellan heltal!

INVERSA FUNKTIONER

Det finns väldigt många funktioner som dyker upp i tillämpningar av matematik i olika ämnen. Inom fysik finns det till exempel Stefan-Boltzmanns lag som säger att man kan beskriva värmestrålningen j^* för en kropp – det vill säga elektromagnetisk strålning som utsänds av en kropp – som en funktion av kroppens absoluta temperatur T:



FIGUR 4.1 Grafen av en kropps värmestrålning j^* som en funktion av temperatur T.

$$j^* = \sigma T^4 \tag{4.1}$$

för alla $T \geq 0$. Man kan därför tänka att j^* är en funktion av T där $\sigma > 0$ är en given konstant. Både definitions- och målmängden av funktionen är $[0,\infty)$ eftersom absolut temperatur och värmestrålning (energi per tid per areaenhet) aldrig är negativa. Det är naturligt att undra: Kan man vända på problemet och entydigt avgöra vad en kropps temperatur är om man vet kroppens värmestrålning? Eller finns det två temperaturer som leder till samma värmestrålning? Finns det även värmestrålning som aldrig skulle uppkomma oavsett kroppens temperatur? Från grafen (se figur 4.1) skulle man svara direkt att det absolut är möjligt att avgöra temperaturen från en kropps värmestrålning men vi vill svara på frågan på ett matematiskt konkret sätt. Intressant nog avslöjar svaret mer om de reella talens natur än fysik.

Vi omformulerar frågan för en funktion med en mer allmän definitions- och målmängd: För en funktion $f : A \to B$ med definitionsmängden A och målmängden B, vad behövs för att få en unik lösning $x \in A$ till ekvationen

$$f(x) = y \tag{4.2}$$

för varje givet $y \in B$?

Vi kan dela upp frågan i två delar. Först vill vi för varje $y \in B$ hitta (minst) en lösning $x \in A$ till (4.2). Om det är möjligt kallas f för en *surjektiv* funktion. Sedan vill vi till varje $y \in B$ inte hitta mer än ett $x \in A$ så att (4.2) gäller. Om det stämmer kallas f för en *injektiv* funktion. En surjektiv och injektiv funktion kallas för *bijektiv* (eller *inverterbar*).

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

$$f^{-1}(y) = x$$

för varje $y \in B$, där x är den unika lösningen till (4.2) i A.

Exempel 4.1 Visa att funktionen $f : \mathbf{R} \setminus \{1\} \to \mathbf{R} \setminus \{6\}$, definierad enligt formeln

$$f(x) = 5 + \frac{x-2}{x-1},$$

för alla $x \neq 1$, är inverterbar.

Lösning. Vi vill bevisa att f är både surjektiv och injektiv för då är den inverterbar.

För att visa f är surjektiv måste man för varje $y \in \mathbf{R} \setminus \{6\}$ hitta ett $x \in \mathbf{R} \setminus \{1\}$ så att $f(x) = y$. Om man av någon anledning har något förslag på vad x kan vara som en funktion av y räcker det att verifiera att $f(x) = y$ genom att sätta förslaget på x i f, samt kontrollera att förslaget x tillhör definitionsmängden $\mathbf{R} \setminus \{1\}$ så länge $y \in \mathbf{R} \setminus \{6\}$. I det här fallet provar vi förslaget

$$x = 1 + \frac{1}{6-y}. \tag{4.3}$$

Hur vi kom fram till det här förslaget är en intressant fråga, men vi måste inte svara på den för att bevisa att f är surjektiv, så vi lägger undan frågan till nästa avsnitt. Vi börjar med att kontrollera att förslaget (4.3) tillhör funktionens definitionsmängd $\mathbf{R} \setminus \{1\}$: Högerledet av (4.3) är tydligen definierat och ger ett reellt tal för varje $y \in \mathbf{R} \setminus \{6\}$. Vi beräknar även att

$$1 + \frac{1}{6-y} = 1 \implies \frac{1}{6-y} = 0 \implies 1 = 0.$$

Eftersom $1 \neq 0$ är x i (4.3) aldrig lika med 1, därför tillhör det $\mathbf{R} \setminus \{1\}$. Nu kan vi sätta in x:et från (4.3) i funktionen:

$$\begin{aligned} f\left(1 + \frac{1}{6-y}\right) &= 5 + \frac{\left(1 + \frac{1}{6-y}\right) - 2}{\left(1 + \frac{1}{6-y}\right) - 1} \\ &= 5 + \frac{((6-y)+1) - 2(6-y)}{((6-y)+1) - (6-y)} \\ &= 5 + (y-5) = y \end{aligned}$$

så genom att visa att gissningen (4.3) fungerar har vi visat att det för varje $y \neq 6$ finns minst ett x sådant att $f(x) = y$. Därmed är f surjektiv.

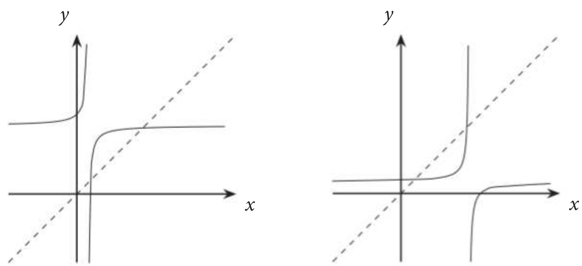
För att visa att den är injektiv vill vi verifiera att det för varje $y \neq 6$ finns högst en lösning x till ekvationen $f(x) = y$. Ett annat sätt att formulera det är att säga vi måste bevisa $f(x_1) = f(x_2) \implies x_1 = x_2$.[1] Vi beräknar

$$\begin{aligned} f(x_1) = f(x_2) &\implies 5 + \frac{x_1 - 2}{x_1 - 1} = 5 + \frac{x_2 - 2}{x_2 - 1} \\ &\implies \frac{x_1 - 2}{x_1 - 1} = \frac{x_2 - 2}{x_2 - 1} \\ &\implies (x_1 - 2)(x_2 - 1) = (x_2 - 2)(x_1 - 1) \\ &\implies x_1x_2 - x_1 - 2x_2 + 2 = x_1x_2 - x_2 - 2x_1 + 2 \\ &\implies -x_1 - 2x_2 = -x_2 - 2x_1 \\ &\implies x_1 = x_2 \end{aligned}$$

så f är faktiskt injektiv. □

Vår lösning innehåller även uttrycket för den inversa funktionen till f:

1 Med andra ord vill vi visa, om två ”x-värden” ger samma ”y-värde” då måste de där x-värdena faktiskt vara desamma – x-värdena var faktiskt ett värde från början.



(a) Grafen av funktionen f från exempel 4.1.

(b) Grafen av inversen f^{-1} av funktionen f från exempel 4.1.

FIGUR 4.2 Grafen av den inversa funktionen f^{-1} till en funktion f är alltid spegelbilden av grafen av f i linjen $x = y$.

Det är förstås förslaget (4.3), så f^{-1} är en funktion från definitionsmängden $\mathbf{R} \setminus \{6\}$ till målmängden $\mathbf{R} \setminus \{1\}$ som ges av

$$f^{-1}(x) = 1 + \frac{1}{6 - x}$$

för alla $x \neq 6$.

Det är instruktivt att rita både grafen av f och grafen av f^{-1}. I figur 4.2 ser vi att den ena är spegelbilden av den andra i linjen $y = x$. Det är förstås alltid relationen mellan graferna av en funktion och dess invers.

När man utreder om en funktion har en invers är det viktigt att tänka noga på funktionens definitionsmängd och målmängd. Nu funderar vi på en fortsättning på exempel 4.1.

Exempel 4.2 Visa att om vi ändrar målmängden i exempel 4.1 till $\mathbf{R}$ så $f : \mathbf{R} \setminus \{1\} \to \mathbf{R}$ är f inte längre surjektiv.

Lösning. Vi ger ett motsägelsebevis till påståendet ”Det finns inget x sådant att $f(x) = 6$”: Vi antar att påståendet är falskt, det vill säga att det finns x så att $f(x) = 6$. Det medför att

$$5 + \frac{x-2}{x-1} = 6 \implies \frac{x-2}{x-1} = 1 \implies x - 2 = x - 1 \implies -2 = -1.$$

Eftersom $-2 \neq -1$ måste antagandet att det finns sådant x vara falskt. □

IRRATIONELLA TAL OCH EN FÖLJD FÖR POTENSFUNKTIONER

Man skulle kunna bli besviken på beviset i exempel 4.1 att f är surjektiv. Vi utnyttjade (4.3) utan att kunna motivera varifrån den kommer. Rent logiskt är det inget problem för själva beviset, men rent praktiskt är det svårt att tillämpa argumentet i andra sammanhang. Men vi kan motivera (4.3) på följande sätt.

Exempel 4.3 Lös direkt ekvationen $f(x) = y$ för x där f är funktionen från exempel 4.1 och $y \in \mathbf{R} \setminus \{6\}$.

Lösning. Vi beräknar

$$\begin{aligned} y = 5 + \frac{x-2}{x-1} &\implies y(x-1) = 5(x-1) + (x-2) \\ &\implies yx - y = 6x - 7 \\ &\implies 7 - y = x(6-y) \\ &\implies (6-y) + 1 = x(6-y) \\ &\underset{y \neq 6}{\implies} x = 1 + \frac{1}{6-y}. \end{aligned} \tag{4.4}$$

Sista implikationen gäller endast om $y \neq 6$, men det är inget problem eftersom det är precis det vi vet: y tillhör f:s målmängd, det vill säga tillhör $\mathbf{R} \setminus \{6\}$, så $y \neq 6$. Vi har visat att om ekvationen $f(x) = y$ har en lösning x måste den vara (4.3). Men (4.4) visar inte att detta x *är* en lösning, bara att det är den enda möjligheten om det finns en lösning alls (jämför med hur vi löste (1.11)). Vad vi har faktiskt visat är att f är injektiv: för varje y från målmängden finns det högst en möjlig lösning x till $f(x) = y$.

För att visa att x i (4.3) är en lösning måste vi kontrollera två saker: detta x tillhör f:s definitionsmängd $\mathbf{R} \setminus \{1\}$ och att $f(x) = y$. Vi kontrollerade att x tillhörde definitionsmängden i exempel 4.1 så vi upprepar inte argumentet här. För att visa $f(x) = y$ kan vi sätta detta x in i ekvationen precis som vi också gjorde i exempel 4.1. Men det går lika bra att kontrollera att implikationspilarna i (4.4) gäller åt andra hållet: Igen beräknar vi

$$
\begin{aligned}
y = 5 + \frac{x-2}{x-1} \quad \underset{x \neq 1}{\overset{\uparrow}{\Longleftarrow}} \quad & y(x-1) = 5(x-1) + (x-2) \\
\Longleftarrow \quad & yx - y = 6x - 7 \\
\Longleftarrow \quad & 7 - y = x(6-y) \\
\Longleftarrow \quad & (6-y) + 1 = x(6-y) \\
\Longleftarrow \quad & x = 1 + \frac{1}{6-y}.
\end{aligned}
$$

Kravet för den första implikationen är uppfyllt eftersom definitionsmängden inte innehåller 1, så vi har visat att x är en lösning.

Här ser vi direkt från uträkningarna att inversa funktionen är

$$f^{-1}(y) = 1 + \frac{1}{6-y}$$

för alla $y \neq 6$. □

Det vore bra om vi kunde räkna ut alla inversa funktioner på liknande sätt – som i exempel 4.3 – men det är kanske ännu mer intressant att även för vissa inverterbara funktioner är det inte möjligt, åtminstone inte bara med de aritmetiska operationer vi har betraktat hittills.

Nu betraktar vi potensfunktioner – som har nästan samma form som Stefan-Boltzmanns lag (4.1). För givet $n \in \mathbf{Z}_+$ definierar vi

$$f_n \colon [0,\infty) \to [0,\infty)$$

enligt formeln

$$f_n(x) = x^n \tag{4.5}$$

för $x \geq 0$. Det är ett märkligt faktum att vi inte kan räkna ut en invers till f_n på samma sätt som vi gjorde i exempel 4.3 om $n \geq 2$. Det bevisar vi – specifikt för f_2 (se sidan 128) – genom att först tänka mer på egenskaper hos reella tal.

I avsnitt 2.2 beskrev vi ganska noggrant mängden av reella tal som oändliga decimalutvecklingar. Men en hel del decimalutvecklingar kan skrivas som rationella tal: Metoden från exempel 1.20 visar att ändliga decimalutvecklingar är rationella tal och metoden från exempel 2.38 och

2.39 visar att även oändliga decimalutvecklingar vars siffror upprepar sig är rationella tal. Dessutom visar uppgift 1.9 att aritmetiska operationer (det vill säga addition, subtraktion, multiplikation och division) med rationella tal alltid ger rationella tal som svar. Man kan då fråga: Finns det fler reella tal än rationella tal? Den överraskande svaret är ja! Nyckeln för att bevisa detta är följande sats.

Sats 4.4 *Om det finns en lösning x till $x^2 = 2$ då kan x inte vara rationellt.*

Satsen är ungefär lika gammal som Pythagoras sats och är lika vacker idag som den dagen då den upptäcktes. G.H. Hardy [8, 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.

Innan vi bevisar satsen bevisar vi följande hjälpsats. Vi har egentligen sett lemma 4.5 tidigare i uppgift 1.6.

Lemma 4.5 *Om man vet att ett heltal i kvadrat n^2 är jämnt så måste n också vara jämnt.*

Bevis. Vi bevisar lemmat genom att bevisa dess kontraposition. Därför antar vi att n är ett udda heltal, det vill säga det kan skrivas som $n = 2k+1$ för något heltal k. Vi beräknar att

$$n^2 = (2k+1)^2 = 4k^2 + 4k + 1 = 2(2k^2 + 2k) + 1.$$

Eftersom $2k^2 + 2k$ är ett heltal ser vi att även n^2 är udda. Så vi har visat att

$$n \text{ är udda} \implies n^2 \text{ är udda}$$

och genom att ta kontrapositionen av detta får vi att

$$n^2 \text{ är jämnt} \implies n \text{ är jämnt} \qquad \square$$

Lemma 4.5 kommer att vara till nytta i beviset av sats 4.4.

Bevis (av sats 4.4). Vi ger ett motsägelsebevis och antar därför att x är rationellt, och kan därför skrivas som ett bråk av heltal. Vi får även anta att nämnaren och täljaren inte har någon gemensam delare, eftersom eventuella gemensamma delare kan strykas. Därför kan vi skriva $x = n/m$ för $n, m \in \mathbf{Z}$ där n och m inte har någon gemensam delare.

Därav får vi att $2 = x^2 = n^2/m^2$ som medför att $2m^2 = n^2$. Så n^2 är jämnt och enligt lemma 4.5 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 lemma 4.5 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, och därför är x inte rationellt. □

Sats 4.4 ger oss ett sätt att gå till väga för att svara på frågan från sidan 123 – finns det reella tal som inte är rationella? Om det inte finns något rationellt tal x som löser $x^2 = 2$ har vi två alternativ kvar: Antingen finns det inte heller någon reell lösning eller så finns det reella tal som inte är rationella. Formulerat med hjälp av funktionen f_2: Antingen är f_2 inte surjektiv eller så finns det reella tal som inte är rationella. Vi kallar reella tal som inte är rationella för *irrationella*. I nästa avsnitt visar vi utifrån våra antaganden om reella tal från avsnitt 2.2 att f_2 faktiskt är surjektiv och som följd finns det irrationella tal!

Om det finns ett irrationellt tal finns det flera: Man kan till exempel visa med hjälp av uppgift 1.9 att om α är irrationellt måste även $r + \alpha$ vara irrationellt för varje rationellt tal r.

4.2 En ny algebraisk operation: rot av tal

Nu fortsätter vi vår undersökning av funktionen f_2 definierad i (4.5). Med motivation grundad i Stefan-Boltzmanns lag (4.1) tror vi att det finns en invers till f_2 samt alla f_n, men vi kommer också se att det inte finns något enkelt aritmetiskt uttryck för den.

KVADRATRÖTTER

Vi börjar genom att visa att f_2 är surjektiv. Vi måste bevisa att det för varje $a \in [0,\infty)$ finns en lösning $\ell \in [0,\infty)$ till $f_2(\ell) = a$. Vi skriver ut ekvationen explicit:

$$\ell^2 = a \tag{4.6}$$

Återigen säger vår intuition att det är möjligt, men beviset som vi presenterar här bygger på den arkimediska egenskapen (sats 2.26) och fullständighetsaxiomet (axiom 2.28).

Vi inför notationen $\mathbf{R}_0^+ := [0,\infty) = \{x \mid x \geq 0\}$ för mängden av icke-negativa reella tal och delar upp den i två delar – vi definierar

$$\begin{aligned} M_a^- &= \{x \in \mathbf{R}_0^+ \mid x^2 \leq a\} \quad \text{och} \\ M_a^+ &= \mathbf{R} \setminus M_a^- = \{x \in \mathbf{R}_0^+ \mid x^2 > a\}. \end{aligned} \tag{4.7}$$

Kom ihåg att det inte är samma sak att säga att ett tal är icke-negativt som att säga att det är positivt. Ett tal x är icke-negativt om $x \geq 0$ medan det är positivt om $x > 0$. Noll är det enda reella talet som varken är positivt eller negativt.

Alla tal $x \in \mathbf{R}_0^+$ tillhör antingen M_- eller M_+: Man kan tänka att vi testar varje $x \in \mathbf{R}_0^+$ och varje x hamnar antingen i M_a^- eller M_a^+ och på så sätt delar vi upp $\mathbf{R}_0^+$. Vår intuition av vad fullständighetsaxiomet innebär antyder att det även borde finnas ett x som sitter precis på gränsen mellan M_a^- och M_a^+, och uppfyller $x^2 = 2$. Men vi behöver ett bevis för att bekräfta vår intuition och därför försöker vi använda axiomet för att bevisa följande sats.

Sats 4.6 *Låt $a \in [0,\infty)$ och $\ell := \sup M_a^-$ där M_a^- definieras i* (4.7). *Då är ℓ en lösning till* (4.6).

Bevis. Först vill vi kontrollera att ℓ är väldefinierat, det vill säga att mängden M^- har ett supremum. Ett sätt att göra detta är att kontrollera att alla förutsättningarna för axiom 2.28 är uppfyllda.

Vi noterar att M_a^- inte är tomt: 0 tillhör M_a^- oavsett a:s värde ty $0^2 \leq a$. Dessutom är M_a^- uppåt begränsad: Om $a \geq 1$ och $x > a$ är $x^2 > a^2 \geq a$ så $x \notin M_a^-$, och därför drar vi slutsatsen att a är en övre begränsning för M_a^- om $a \geq 1$. I fallet $a < 1$ kan vi säga $M_a^- \subseteq M_1^-$ och vi har bevisat att M_1^- har 1 som en övre begränsning. Därmed är 1 även en övre begränsning till M_a^-. Vi har då bevisat

$$x \leq \max\{a, 1\} \text{ för alla } x \in M_a^- \tag{4.8}$$

så enligt axiom 2.28 finns det en minsta övre begränsning till M_-. Det vill säga, ℓ är väldefinierat.

Nu vill vi visa att ℓ faktiskt är en lösning till (4.6). Vi betraktar två fall.

Fall 1: $a = 0$

I så fall kan vi se $M_a^- = \{0\}$, så tydligen är $\ell = 0$.

Fall 2: $a > 0$

Här ska vi använda oss av egenskaperna (2.9) och (2.11) hos en minsta övre begränsning för att visa

$$0 \leq |\ell^2 - a| < 1/n \tag{4.9}$$

för godtyckliga $n \in \mathbf{Z}_+$. Om vi lyckas med det följer det direkt av sats 2.26(a) att $|\ell^2 - a| = 0$ och därför $\ell^2 - a = 0$, så ℓ löser (4.6).

Utifrån (4.8) vet vi att $\max\{a, 1\}$ är en övre begränsning till M_a^- så (2.9) medför att

$$\ell \leq \max\{a, 1\}. \tag{4.10}$$

Men vi kan även uppskatta ℓ underifrån: Om $0 < a < 1$ är $a^2 < a$ så $a \in M_a^-$ och därför är $a \leq \ell$. Om $a \geq 1$ är $1 \in M_a^-$ eftersom $1^2 = 1 \leq a$ och därför är $1 \leq \ell$. Detta medför att

$$\min\{a, 1\} \leq \ell. \tag{4.11}$$

Nu bevisar vi (4.9). Enligt (2.11) finns det för varje $\varepsilon > 0$ ett $x \in M_a^-$ sådant att $\ell - \varepsilon < x$. För ε som uppfyller $0 < \varepsilon \leq \min\{a, 1\}$ medför (4.11) att $\ell - \varepsilon \geq 0$ och i så fall är $(\ell - \varepsilon)^2 < x^2$. Eftersom $x \in M_a^-$ är $x^2 \leq a$. Vi vet därför att

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

Med hjälp av (4.10) dra vi slutsatsen att

$$\ell^2 - a < (2\ell - \varepsilon)\varepsilon < (2\max\{a, 1\} - 0)\varepsilon = 2\max\{a, 1\}\varepsilon \tag{4.12}$$

för alla $\varepsilon \in (0, \min\{a, 1\}]$.

Eftersom ℓ är den minsta övre begränsningen till M_a^- och $\ell < \ell + \varepsilon$ måste $\ell + \varepsilon \in M_a^+$. Vi kan beräkna att

$$\begin{aligned}
\ell + \varepsilon \in M_a^+ \implies & (\ell + \varepsilon)^2 \geq a \\
\iff & \ell^2 + 2\ell\varepsilon + \varepsilon^2 \geq a \\
\iff & \ell^2 - a \geq -2\ell\varepsilon - \varepsilon^2 = -(2\ell + \varepsilon)\varepsilon.
\end{aligned}$$

Vi använder uppskattningen (4.10) för att beräkna

$$\begin{aligned} \ell^2 - a &\geq -(2\ell + \varepsilon)\varepsilon \\ &> -(2\max\{a,1\} + \min\{a,1\})\varepsilon \\ &= -3\max\{a,1\}\varepsilon \end{aligned} \tag{4.13}$$

för alla $\varepsilon \in (0, \min\{a,1\}]$. Uppskattningarna (4.12) och (4.13) tillsammans medför att

$$0 \leq |\ell^2 - a| < 3\max\{a,1\}\varepsilon$$

för alla $\varepsilon \in (0, \min\{a,1\}]$. Nu får vi välja $\varepsilon = 1/(3\max\{a,1\}n)$ för att se att (4.9) gäller för tillräckliga större $n \in \mathbf{Z}_+$, det vill säga för n så stort att valet av ε ligger i intervallet $(0, \min\{a,1\}]$. Men som följd får vi att (4.9) gäller för *alla* $n \in \mathbf{Z}_+$. (Varför?) Som sagt, vi kan nu när (4.9) är bevisad dra slutsatsen att $\ell^2 = a$. □

Så nu vet vi att det för varje $a \geq 0$ finns minst en lösning ℓ till (4.6). En naturlig följdfråga är: Hur många lösningar finns det? Satsen nedan ger ett delsvar på frågan.

Sats 4.7 *För ett givet $a \in \mathbf{R}_0^+$ är talet ℓ definierat i sats 4.6 det enda icke-negativa talet som löser* (4.6).

Bevis. Anta att det finns två icke-negativa lösningar ℓ_1 och ℓ_2 till (4.6), så $\ell_1^2 = \ell_2^2 = a$. Vi får också anta att

$$0 \leq \ell_1 < \ell_2. \tag{4.14}$$

(Annars byter vi bara plats på ℓ_1 och ℓ_2.) Vi kan beräkna

$$0 = a - a = \ell_2^2 - \ell_1^2 = (\ell_2 - \ell_1)(\ell_2 + \ell_1) \tag{4.15}$$

Utifrån (4.14) är $(\ell_2 + \ell_1) > 0$ och $(\ell_2 - \ell_1) > 0$, så

$$(\ell_1 - \ell_2)(\ell_1 + \ell_2) > 0$$

som är en motsägelse till (4.15) och därför kan det inte finnas två lösningar till (4.6). □

Vi kan sammanfatta vad vi har gjort så här: Sats 4.6 säger att $f_2 \colon \mathbf{R}_0^+ \to \mathbf{R}_0^+$ är surjektiv och sats 4.7 säger att den är injektiv, därför är f_2 inverterbar. Det gäller trots att inversen inte kan skrivas som en sammansättning av aritmetiska operationer på sin variabel och rationella konstanter, och inget mer – om inversen kunde skrivas som en sådan sammansättning skulle uppgift 1.9 säga att $f_2^{-1}(2)$ är rationellt vilket vore motsägelsefullt med sats 4.4. För varje $a \in \mathbf{R}_0^+$ är

$$\ell = f_2^{-1}(a) \tag{4.16}$$

den unika icke-negativa lösningen till (4.6). Vi använder en särskild notation som kallas *kvadratrot*:

$$\sqrt{x} := f_2^{-1}(x)$$

för alla $x \in \mathbf{R}_0^+$. Vi har genomfört ett indirekt bevis på att vi kan ta *roten ur* ett icke-negativt reellt tal och har dessutom lärt oss att det inte ens var möjligt att skriva en formel för kvadratroten som bara innehåller rationella konstanter och aritmetiska operationer.

ANDRAGRADSPOLYNOM

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

$$ax^2 + bx + c = 0 \tag{4.17}$$

för $a \neq 0$ och $b, c \in \mathbf{R}$. Det vill säga vi vill hitta en formel för nollställen till andragradspolynom. Först måste vi tänka på vad som händer när vi utvidgar funktionen f_2 från en funktion med definitionsmängden $\mathbf{R}_0^+$ till en funktion med definitionsmängden $\mathbf{R}$: Vi definierar $F_2 \colon \mathbf{R} \to \mathbf{R}$ enligt formeln $F_2(x) = x^2$ för alla $x \in \mathbf{R}$. Funktionen F_2 sägs vara en *utvidgning* av f_2 eftersom $F_2(x) = f_2(x)$ för x från f_2:s definitionsmängd och f_2:s definitionsmängd är en äkta delmängd av F_2:s.

Orden *lösning*, *rot* och *nollställe* är nära förknippade. En rot är ett annat ord för en lösning x till ekvationen $f(x) = 0$ men det är funktionen f – och inte ekvationen $f(x) = 0$ – som sägs ha ett nollställe i x.

Vi vet därför redan från exempel 3.5(a), respektive 3.5(b), att F_2 är strängt växande på $\{x \mid x \geq 0\}$, respektive strängt avtagande på $\{x \mid x \leq 0\}$. Argumentet i beviset av sats 4.7 kan användas för att visa att det inte kan finnas två eller flera icke-positiva lösningar x till $x^2 = d$ och vi räknar ut direkt att $(-\sqrt{d})^2 = d$ så $-\sqrt{d}$ är den icke-positiva lösning 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 (4.17) observera att

Kom ihåg att uträkningen i (4.18) ibland kallas för *kvadratkomplettering*. Här har vi också en koefficient framför x^2-termen som gör den lite mer komplicerad. Se också exempel 2.19 och 2.22.

$$ax^2 + bx + c = a\left(x + \frac{b}{2a}\right)^2 - \frac{b^2}{4a} + c \tag{4.18}$$

så (4.17) är ekvivalent med

$$\left(x + \frac{b}{2a}\right)^2 = \frac{b^2 - 4ac}{4a^2}. \tag{4.19}$$

Det är därför uttrycket $b^2 - 4ac$ – ibland kallat för polynomets *diskriminant* – som avgör hur många lösningar (4.17) har. Det finns lösningar till (4.19) om och endast om $b^2 - 4ac \geq 0$ och i så fall är (4.19) ekvivalent med

$$\left(x + \frac{b}{2a}\right) = \pm\sqrt{\frac{b^2 - 4ac}{4a^2}} \tag{4.20}$$

enligt argumentet ovan. Alltså får vi att (4.17) har precis två lösningar

$$x = -\frac{b}{2a} \pm \sqrt{\frac{b^2 - 4ac}{4a^2}} \quad \text{om } b^2 - 4ac > 0, \tag{4.21}$$

precis en lösning

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

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

Anmärkning 4.8 Här passar vi på att betona skillnaden i hur tecknet ± ska tolkas i olika sammanhang. Det kan betyda

”+ och −” eller ”+ eller −”

beroende på kontexten. Om vi vill till exempel prata om en ekvivalent omskrivning av (4.17) i fallet $b^2 - 4ac > 0$ måste ± betyda ”+ eller −”: (4.17) gäller om och endast om antingen (4.21) gäller med ett plustecken *eller* den gäller med ett minustecken. Men om vi vill prata om alla möjliga lösningar x till (4.17) i fallet $b^2 - 4ac > 0$ måste ± betyda ”+ och −”: Både (4.21) med ett plustecken *och* (4.21) med ett minustecken löser (4.17).

Tecknet ± är ett bra exempel där det är alldeles för lätt att förkorta bort meningen av vad vi vill kommunicera med notation. Ord som ”och” och ”eller” är förstås grundläggande till logiken av matematiska argument, men dessvärre är det oftast lämnat till läsaren att lista ut vad författaren har förkortat bort med smidig notation.

Exempel 4.9 I Pythagoras sats (sats 3.17) hittade vi ett samband mellan en rätvinklig triangels sidolängder a, b och c. Diskussionen i det här avsnittet visar att det endast finns två möjliga värden för triangelns hypotenusa (så länge som $a,b > 0$):

$$c = \sqrt{a^2 + b^2} \quad \text{eller} \quad c = -\sqrt{a^2 + b^2}.$$

Eftersom c är en längd kan vi utesluta den andra lösning så

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

Det innebär att vi får ett uttryck för avståndet mellan två punkter $A = (x_A, y_B)$ och $B = (x_B, y_B)$: Utifrån (3.21) måste

$$\ell(A,B) = \sqrt{(x_B - x_A)^2 + (y_B - y_A)^2}$$

eftersom $\ell(A,B)$ inte kan vara negativ.

Slutligen kan vi även säga att en punkt $P = (x,y)$ tillhör cirkeln (3.16) med medelpunkten $M = (a,b)$ och radien $r > 0$ om och endast om

$$\begin{aligned}
&\ell(M,P) = r \\
&\iff \sqrt{(x-a)^2 + (y-b)^2} = r \\
&\iff (x-a)^2 + (y-b)^2 = r^2,
\end{aligned}$$

där sista ekvivalensen gäller eftersom $r > 0$. Det innebär, som vi lovade i slutet av avsnitt 3.2, att mängden (3.16) är inte bara en delmängd av mängden av alla lösningarna $P = (x,y)$ till (3.22), utan att de är lika.

RATIONELLA POTENSER

Med bara ett fåtal ändringar får vi generalisera det föregående argumentet från f_2 till godtycklig f_n definierad i (4.5). För ett $n \geq \mathbf{Z}_+$ och $a > 0$

ersätter vi mängderna M_a^- och M_a^+ i (4.7) med

$$M_{a,n}^- = \{x \in \mathbf{R}_0^+ \mid x^n \le a\} \quad \text{och}$$
$$M_{a,n}^+ = \mathbf{R}_0^+ \setminus M_a^{-,n} = \{x \in \mathbf{R}_0^+ \mid x^n > a\}.$$

Ungefär samma argument som ovan leder till följande generalisering av sats 4.6.

Sats 4.10 *Låt $n \in \mathbf{Z}_+$, $a \in [0,\infty)$ och $\ell := \sup M_{a,n}^-$. Då är ℓ en lösning till $\ell^n = a$.*

Fallet $n = 1$ är förstås mycket enklare än när $n > 1$. Då har vi även en explicit och enkel formel för ℓ: Det är förstås $\ell = a$. Men kraften av sats 4.10 och 4.11 ligger i fallen $n > 1$ där en explicit formel för ℓ saknas.

Vi lämnar beviset som en övning för läsaren. Vi har också en generaliseringen av sats 4.7.

Sats 4.11 *Låt $n \in \mathbf{Z}_+$. Funktionen $f_n\colon [0,\infty) \to [0,\infty)$ definierad i (4.5) är strängt växande och därmed injektiv.*

Bevis. Vi antar att

$$0 \le x_1 < x_2. \tag{4.22}$$

I fallet $n = 1$ medför (4.22) direkt att $f_1(x_1) < f_1(x_2)$ så vi har visat att f_1 är strängt växande. För andra värden av n kan vi beräkna[2]

$$f_n(x_2) - f_n(x_1) = x_2^n - x_1^n = (x_2 - x_1)\left(\sum_{j=0}^{n-1} x_2^j x_1^{n-1-j}\right).$$

Utifrån (4.22) är $(x_2 - x_1) > 0$ och

$$\sum_{j=0}^{n-1} x_2^j x_1^{n-1-j} = \underbrace{x_2^{n-1}}_{>0} + \underbrace{\sum_{j=0}^{n-2} x_2^j x_1^{n-1-j}}_{\ge 0} > 0,$$

så

$$f_n(x_2) - f_n(x_1) > 0 \implies f_n(x_2) > f_n(x_1)$$

och vi har bevisat att f_n är strängt växande för $n \ge 2$. □

2 När $x_1 = 0$ begår vi samma notationsbrott här som vi gjorde i (3.2) (se marginalanteckningen bredvid (3.2)), men det bör inte leda till något missförstånd.

Anmärkning 4.12 En strängt monoton funktion f är alltid injektiv eftersom den inte kan anta samma värde i två olika x-värden. Om vi betraktar två olika värden x_1 och x_2 är det ena mindre än det andra, så är till exempel $x_1 < x_2$, och därför är $f(x_1) < f(x_2)$ om f är strängt växande och $f(x_1) > f(x_2)$ om f är strängt avtagande. Vi kan därför aldrig ha $f(x_1) = f(x_2)$ så f måste vara injektiv.

Satserna 4.10 och 4.11 visar att funktionen $f_n\colon \mathbf{R}_0^+ \to \mathbf{R}_0^+$ definierad enligt (4.5), är bijektiv. Vi definierar då den *n:e roten* till y som

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

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

Exempel 4.13 Vi återgår till Stefan-Boltzmanns lag (4.1). Med hjälp av allt arbete vi har gjort i det här avsnittet är det lätt att räkna ut temperaturen $T > 0$ från en kropps värmestrålning $j^* > 0$. Med hjälp den 4:e roten kan vi beräkna utifrån (4.1) att

$$j^* = \sigma T^4 \iff \frac{j^*}{\sigma} = T^4 \iff T = \sqrt[4]{\frac{j^*}{\sigma}}$$

där vi har använt oss av det faktum att T måste vara positivt. Därför kan vi svara jakande på frågan från sidan 117 att vår intuition var helt korrekt och man kan avgöra en kropps temperatur från dess värmestrålning.

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

$$a^{m/n} := \sqrt[n]{a^m} \tag{4.23}$$

Observera att definitionen av en rationell potens kräver att potensen först skrivs med en positiv nämnare.

för alla $a \geq 0$, $m \in \mathbf{Z}$ och $n \in \mathbf{Z}_+$. I synnerhet säger (4.23) att vi får skriva $\sqrt[n]{a}$ som $a^{1/n}$. Observera att här har vi inte utökat definitionen av rationella potenser för icke-positiva a fast definitionen för heltalspotenser (1.15) även gäller för $a < 0$. Även om (4.23) har någon mening för vissa n och m när $a < 0$ slipper vi definiera $a^{n/m}$ för negativa a eftersom då skulle a^r för $r \in \mathbf{Q}$ bero på representationen av r som kvoten n/m. Till exempel är $1/3 = 2/6$ men den rimligaste tolkningen av $\sqrt[3]{(-1)^1}$ är -1 medan $\sqrt[6]{(-1)^2} = 1$.

Man får utvidga sats 1.8 från heltalspotenser till rationella potenser åtminstone när $a,b > 0$. Vi skriver den nya versionen av satsen nedan.

Sats 4.14 *För givna godtyckliga* $a, b > 0$ *och* $r, s \in \mathbf{Q}$ *har vi att*

(a) $a^r a^s = a^{r+s}$,

(b) $a^r a^{-s} = a^{r-s}$,

(c) $a^r b^r = (ab)^r$, *och*

(d) $(a^r)^s = a^{rs}$.

Bevis. Här bevisar vi (c) och lämnar de andra som en övning för läsaren.

Först ska vi bevisa att

$$\sqrt[n]{ab} = \sqrt[n]{a}\sqrt[n]{b} \tag{4.24}$$

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

$$x^n = ab. \tag{4.25}$$

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

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

Vi vill bevisa att $x = yz$. Men enligt (c) från sats 1.8 är

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

så yz är också en lösning till (4.25). Som en produkt av två icke-negativa tal är yz också icke-negativ. Eftersom det finns högst en icke-negativ lösning till (4.25) så måste $x = yz$, alltså är (4.24) bevisad.

Vi får skriva $r \in \mathbf{Q}$ som $r = m/n$ för $n \in \mathbf{Z}_+$ och $m \in \mathbf{Z}$. Enligt definitionen (4.23) och (4.24) 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 (c) från sats 1.8 och återigen definitionen (4.23). Därmed är (c) bevisad. □

4.3 Sammanfattning och uppgifter

I kapitel 4 har vi:

- identifierat ett behov av att kunna göra ogjort vad en funktion har gjort och sett att en funktion måste vara både injektiv och surjektiv för att kunna definiera en invers;
- sett att en funktions injektivitet och surjektivitet hänger på funktionens definitions- och målmängd;
- accepterat att om vi över huvud taget vill hitta lösningar till vissa ekvationer måste vi betrakta tal som inte är rationella och operationer (såsom suprema) som går utöver de fyra fundamentala aritmetiska operationerna;
- definierat kvadratrötter som inversen av en funktion och visat att den är väldefinierad;
- utvidgat begreppet potens från heltalspotenser till rationella potenser.

UPPGIFTER

4.1 (a) Betrakta ett heltal m. Bevisa att om m^2 är delbart med 3 då är m delbart med 3.

(b) Bevisa att om $c^2 = 3$ då är c irrationellt.

4.2 (a) Betrakta ett heltal m. Bevisa att om m^2 är delbart med 6 då är m delbart med 6.

(b) Bevisa att om $c^2 = 6$ då är c irrationellt.

4.3 Hitta ett heltal m så att m^2 är jämnt delbart med 9 fast m inte är det.

4.4 Utifrån uppgifterna 4.1, 4.2 och 4.3 formulera en generalisering av sats 4.4.

4.5 Betrakta funktionen $f\colon \mathbf{R} \to \mathbf{R}$ definierad enligt

$$f(x) = \begin{cases} 2x - 3 & \text{om } x < 2; \\ -4x + 13 & \text{om } 2 \leq x \leq 3; \\ x + 2 & \text{om } x > 3. \end{cases}$$

(a) Rita grafen av f.

(b) Bevisa att f är inverterbar och hitta den inversa funktionen f^{-1}.

4.6 Betrakta en funktion $f\colon D \to M$ som ges av uttrycket

$$f(x) = \sqrt{\frac{x+1}{x-1}} \quad \text{för } x \in D.$$

Syftet med den här uppgiften är att utreda definitionsmängden $D \subseteq \mathbf{R}$ och målmängden $M \subseteq \mathbf{R}$ så att f är bijektiv.

(a) Bestäm den största mängden av x så att $f(x)$ är definierat. Ta den mängden som funktionens definitionsmängd D.

(b) Undersök vad man måste välja för M så att f blir surjektiv. Motivera ditt svar.

(c) Med M som i (b) är $f\colon D \to M$ injektiv? Motivera ditt svar.

(d) Ge ett explicit uttryck för inversen till f.

4.7 (a) Betrakta två positiva tal a och b. Visa att
$$(a+b)^{1/n} \leq a^{1/n} + b^{1/n}$$
för alla $n \in \mathbf{Z}_+$.

(b) Hitta ett exempel på a och b där vi får sträng olikhet i olikheten ovan för varje $n > 1$. Det vill säga, ge ett exempel där $(a+b)^{1/n} < a^{1/n} + b^{1/n}$ för alla $n \in \mathbf{Z}_+ \setminus \{1\}$.

Trigonometri

5.1 Definitioner och formler

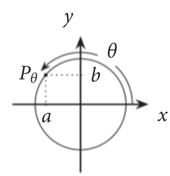
Ordet ”trigonometri” kommer ifrån det grekiska ordet för ”triangel” (trígōnon) och ”mått” (métron) och handlar om relationer mellan längder och vinklar i trianglar. Trigonometri är ett kraftfullt verktyg som både löser många praktiska problem och avslöjar många mönster och relationer inom geometrin.

DEFINITIONER

Trots namnet 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 hela $\mathbf{R}$ 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 enligt regeln att för givet $\theta \in \mathbf{R}$ mäter man en båge med längden θ från punkten $(1, 0)$ kring enhetscirkeln – moturs om $\theta \geq 0$ och medurs om $\theta < 0$ – och sedan tar

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

där a och b är koordinaterna av bågens ändpunkt P_θ som visas i figur 5.1.



FIGUR 5.1 $\cos\theta$ och $\sin\theta$ definieras som den x-respektive den y-koordinat av punkten man hamnar på när man går runt enhetscirkeln på ett avstånd θ.

Längden θ vi pratar om här är förstås en vinkel, men den är egentligen ett slags utvidgning av begreppet vinkel som vi definierade på sidan 104. Där var en vinkel aldrig mer än omkretsen av enhetscirkeln med medelpunkt i skärningspunkten av två streck, och var därför aldrig mer än 2π. Men definitionen här innefattar även en riktning (från vinkelns tecken) och eventuella varv kring enhetscirkeln (ty vinklar kan vara vilka reella tal som helst) och är därför i någon mening en mer dynamisk definition.

Där ger vi en ganska statisk beskrivning av vinkeln mellan två streck som skär varandra, men genom att tillåta reella vinklar kan vi ge vinkeln en riktning och även notera antalet varv vi tar runt en punkt.

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

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

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

$$\tan\theta = \frac{\sin\theta}{\cos\theta}. \tag{5.1}$$

TRIGONOMETRISKA FORMLER

På sidan 103 definierade vi talet π som längden runt hälften av enhetscirkelns omkrets, och därför har hela omkretsen längden 2π. Om vi adderar ett helt varv – det vill säga om vi adderar 2π – till cosinus eller sinus variabel hamnar vi på samma punkt i figur 5.1. Så funktionernas värden är oförändrade om vi adderar 2π till variabeln. Därför kallas cosinus och sinus för *periodiska funktioner* med period 2π:

$$\begin{aligned} \cos(\theta + 2\pi) &= \cos\theta \quad \text{och} \\ \sin(\theta + 2\pi) &= \sin\theta \end{aligned} \tag{5.2}$$

för alla $\theta \in \mathbf{R}$. Avståndet från en godtycklig punkt $P_\theta = (\cos\theta, \sin\theta)$ på enhetscirkeln till origo $(0,0)$ är förstås 1 och därför säger (3.21) (som är en följd av Pythagoras sats – sats 3.17) att

$$\cos^2\theta + \sin^2\theta = 1 \tag{5.3}$$

för alla $\theta \in \mathbf{R}$.

Av tradition skriver man $(\cos\theta)^n$ som $\cos^n\theta$, samt $\sin^n\theta = (\sin\theta)^n$ för positiva heltal n.

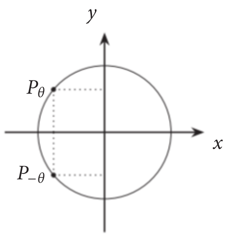
Figur 5.2(a) visar att punkterna $P_\theta = (\cos\theta, \sin\theta)$ och $P_{-\theta} = (\cos(-\theta), \sin(-\theta))$ är spegelbilder av varandra i x-axeln. Därför är deras x-koordinater lika och deras y-koordinater skiljer sig åt bara med tecknet, det vill säga den ena är minus den andra. Så

$$\cos(-\theta) = \cos\theta \tag{5.4}$$

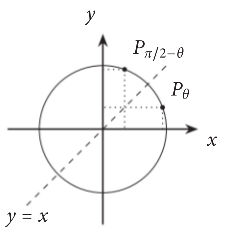
och

$$\sin(-\theta) = -\sin\theta \tag{5.5}$$

för alla $\theta \in \mathbf{R}$. Eftersom cosinus lyder (5.4) kallar vi cosinus för en *jämn funktion* och eftersom sinus lyder (5.5) kallar vi sinus för en *udda funktion*.



(a) Punkterna $P_{-\theta}$ och P_θ är spegelbilder av varandra i x-axeln. Därför är deras x-koordinater lika och den ena punktens y-koordinat är minus den andras.



(b) Punkterna $P_{\pi/2-\theta}$ och P_θ är spegelbilder av varandra i linjen $y = x$. Därför är den enas x-koordinat den andras y-koordinat.

FIGUR 5.2 Spegelsymmetri ledar till trigonometriska likheter (5.4)–(5.7).

En motsvarade spegelsymmetri (se figur 5.2(b)) visar att punkterna $P_\theta = (\cos\theta, \sin\theta)$ och $P_{\pi/2-\theta} = (\cos(\pi/2-\theta), \sin(\pi/2-\theta))$ är spegelbilder av varandra i linjen $y = x$. Därför är den enas x-koordinat den andras y-koordinat så

$$\cos(\pi/2-\theta) = \sin\theta \quad \text{och} \tag{5.6}$$
$$\sin(\pi/2-\theta) = \cos\theta \tag{5.7}$$

för alla $\theta \in \mathbf{R}$.

Additions- och subtraktionsformler

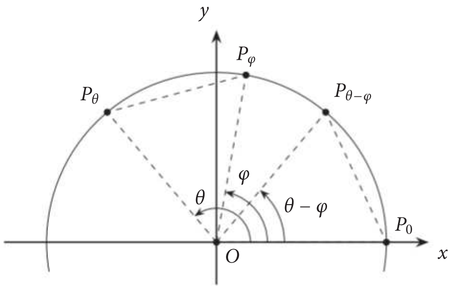
Det finns även formler för att uttycka $\cos(\theta \pm \varphi)$ och $\sin(\theta \pm \varphi)$ som en funktion av $\cos\theta$, $\cos\varphi$, $\sin\theta$ och $\sin\varphi$ för godtyckliga θ och φ.

Notera att man *inte* kan förvänta sig att till exempel $\cos(\theta+\varphi) = \cos\theta + \cos\varphi$. Definitionen av cosinus innebär mätning av avstånd både runt en cirkel och längs en axel. Därför är en så enkel formel helt orimlig!

Vi börjar genom att visa att

$$\cos(\theta-\varphi) = \cos\theta\cos\varphi + \sin\theta\sin\varphi. \tag{5.8}$$

Vi använder oss av den rotationssymmetri som visas i figur 5.3. Betrakta två punkter $P_\theta = (\cos\theta, \sin\theta)$ och $P_\varphi = (\cos\varphi, \sin\varphi)$ på enhetscirkeln tillsammans med origo O. De tre punkterna bildar en triangel och triangelns sidolängder är oförändrade om vi roterar triangeln $-\varphi$ runt O. Rotationen ger triangeln med hörnen O, $P_{\theta-\varphi}$ och P_0, där $P_{\theta-\varphi}$ =



FIGUR 5.3 Triangeln med hörnen O, $P_{\theta-\varphi}$ och P_0 är en rotation runt O av triangeln med hörnen O, P_θ och P_φ.

$(\cos(\theta-\varphi), \sin(\theta-\varphi))$ och $P_0 = (1,0)$. I synnerhet är avståndet mellan P_θ och P_φ lika med avståndet mellan $P_{\theta-\varphi}$ och P_0. Med hjälp av (3.21) och (5.3) beräknar vi

$$\begin{aligned}
\ell(P_\theta,P_\varphi)^2 &= (\cos\varphi - \cos\theta)^2 + (\sin\varphi - \sin\theta)^2 \\
&= (\cos^2\varphi - 2\cos\varphi\cos\theta + \cos^2\theta) \\
&\quad + (\sin^2\varphi - 2\sin\varphi\sin\theta + \sin^2\theta) \\
&= 2 - 2\cos\varphi\cos\theta - 2\sin\varphi\sin\theta
\end{aligned}$$

och

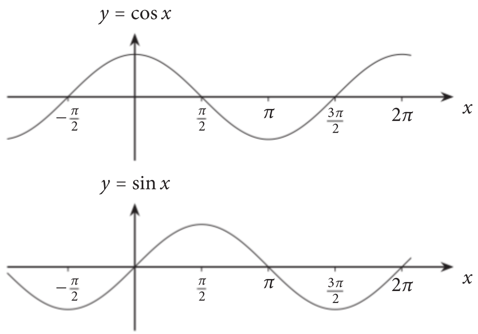
$$\begin{aligned}
\ell(P_{\theta-\varphi},P_0)^2 &= (1-\cos(\theta-\varphi))^2 + \sin^2(\theta-\varphi) \\
&= (1 - 2\cos(\theta-\varphi) + \cos^2(\theta-\varphi)) + \sin^2(\theta-\varphi) \\
&= 2 - 2\cos(\theta-\varphi).
\end{aligned}$$

Eftersom $\ell(P_\theta,P_\varphi)^2 = \ell(P_{\theta-\varphi},P_0)^2$ får vi sätta högerleden lika med varandra och sedan får vi (5.8) direkt.

Eftersom (5.8) gäller för vilka θ och φ som helst får vi ersätta φ med $-\varphi$ för att få

$$\begin{aligned}
\cos(\theta+\varphi) &= \cos(\theta-(-\varphi)) \\
&= \cos\theta\cos(-\varphi) + \sin\theta\sin(-\varphi) \qquad (5.9) \\
&= \cos\theta\cos\varphi - \sin\theta\sin\varphi
\end{aligned}$$

där vi har använt (5.4) och (5.5) i sista likheten.



FIGUR 5.4 Grafen av funktionerna cosinus och sinus.

Genom att använda (5.6), (5.7), (5.8) kan vi beräkna

$$\begin{aligned}\sin(\theta + \varphi) &= \sin(\pi/2 - ((\pi/2 - \theta) - \varphi)) \\ &= \cos((\pi/2 - \theta) - \varphi) \\ &= \cos(\pi/2 - \theta)\cos\varphi + \sin(\pi/2 - \theta)\sin\varphi \\ &= \sin\theta\cos\varphi + \cos\theta\sin\varphi.\end{aligned} \tag{5.10}$$

Det finns ett helt menageri av trigonometriska likheter så det kan vara svårt att både komma ihåg och bevisa dem. Vi har sett att det ofta är symmetrier – till exempel rotations- eller spegelsymmetri – som ligger bakom relationer mellan olika uttryck av trigonometriska funktioner. Som ett minnesknep för att snabbt kontrollera eller komma ihåg trigonometriska likheter kan det vara värt att ha en bild av funktionernas grafer. Vi skissar dem i figur 5.4. För att underlätta uträkningar är det värt mödan att lära sig (5.3), (5.9) och (5.10) utantill. Många andra trigonometriska likheter följer direkt av dem.

Till exempel kan vi snabbt härleda en additionsformel för tangens. Enligt definitionen (5.1) av tangens, (5.9) och (5.10) är

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

så länge som varken θ, φ eller $\theta + \varphi$ är lika med $\pi/2 + n\pi$ för något $n \in \mathbf{Z}$ och vi delar därför aldrig med noll.

Dubbla vinklar

Om vi sätter $\varphi = \theta$ i (5.9) och (5.10) får vi att

$$\begin{aligned}\cos(2\theta) &= \cos(\theta + \theta) = \cos\theta\cos\theta - \sin\theta\sin\theta \\ &= \cos^2\theta - \sin^2\theta\end{aligned} \tag{5.11}$$

respektive

$$\begin{aligned}\sin(2\theta) &= \sin(\theta + \theta) = \sin\theta\cos\theta + \cos\theta\sin\theta \\ &= 2\sin\theta\cos\theta.\end{aligned}$$

Från (5.11) och (5.3) får man

$$\cos(2\theta) = 2\cos^2\theta - 1 \iff \cos^2\theta = \frac{1 + \cos(2\theta)}{2} \tag{5.12}$$

och

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

Nog så intressant här är att vi har lyckats skriva cosinus och sinus i kvadrat som en uttryck av bara $\cos(2\theta)$. Det är ingen slump att det är just 2θ som dyker upp i samband med potensen 2. Det kommer vi utreda vidare i kapitel 7.

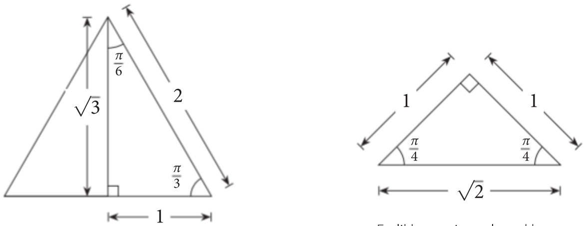
Exakta värden

Återigen kan vi använda oss av symmetri för att få reda på egenskaper hos trigonometriska funktioner. Den här gången räknar vi ut exakta värden av de trigonometriska funktionerna $\cos\theta$ och $\sin\theta$ för vissa θ.

Betrakta en liksidig triangel med sidlängd 2 (se figur 5.5(a)).[1] Rita ett streck från ett hörn till mittpunkten av den motsatta sidan. Strecket delar även hörnets vinkel i två lika delar och vi har fått två trianglar med vinklarna $\pi/2$, $\pi/3$ och $\pi/6$ och sidolängderna 2, 1 och – enligt sats 3.17 – $\sqrt{2^2 - 1^2} = \sqrt{3}$. Därför är

$$\cos(\pi/3) = \sin(\pi/6) = \frac{1}{2} \quad \text{och} \quad \cos(\pi/6) = \sin(\pi/3) = \frac{\sqrt{3}}{2}.$$

1 En liksidig triangel är en triangel vars alla sidor är lika långa.



(a) En liksidig triangel med sidolängd 2 delat i två.

(b) En likbent triangel med kateter av längden 1 och en rät vinkel mittemot hypotenusan.

FIGUR 5.5 Man kan räkna ut exakta värden av de trigonometriska funktionerna på $\pi/2$, $\pi/3$ och $\pi/6$.

Vi kan också betrakta en likbent rätvinklig triangel med kateter av längden 1 och vinkeln mittemot hypotenusan lika med $\pi/2$ (se figur 5.5(b)). Då är hypotenusan av längden $\sqrt{2}$ (återigen tack vare sats 3.17) och de två andra vinklarna är båda lika med $\pi/4$. Man räknar direkt ut att

$$\cos(\pi/4) = \sin(\pi/4) = \frac{1}{\sqrt{2}}.$$

5.2 Olikheter och arcusfunktioner

Nu tillämpar vi en del av den teori vi har utvecklat till trigonometriska funktioner. Men hjälp av begreppet area kan vi visa ett par olikheter som gäller för trigonometriska funktioner. Sedan genom att begränsa definitionsmängderna av de trigonometriska funktionerna kan vi definiera några slags inversa funktioner till dem.

TRIGONOMETRISKA OLIKHETER

Olikheter är förvånansvärt kraftfulla verktyg, särskilt när man kommer till mer avancerad matematik. Här visar vi två relationer mellan en trigonometrisk funktion och dess variabel.

Sats 5.1 *För* $\theta \in (0, \pi/2)$ *är*

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

Bevis. I figur 5.6 avbildas en fjärdedel av en enhetscirkel och vissa streck. Låt T_1 vara triangeln med hörnpunkterna O, A och D som ritas i figur 5.6(a), T_2 vara triangeln med hörnpunkterna A, B och D och T_3 vara triangeln med hörnpunkterna O, B och C. Låt dessutom W vara den delen av enhetscirkeln mellan linjen från O till B och linjen från O till D. Enligt figur 5.6(a) ser vi att

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

så genom att ta arean av mängderna får vi olikheterna

$$|T_1 \cup T_2| < |W| < |T_3|. \tag{5.13}$$

För att vi verkligen får stränga olikheter i (5.13) och inte $\leq$, krävs att mängderna är lite mer än bara delmängder. Om skillnaden mellan mängderna vore bara några punkter skulle areorna vara lika, men vi ser i figur 5.6(a) att man alltid kan klämma in en rektangulär mängd med positiv area mellan mängderna. Då måste vi få stränga olikheter när vi tar arean av mängderna.

Precis som vi argumenterade att (3.17) var en q:e del av en cirkels omkrets, måste $|W|$ vara en $\theta/(2\pi)$:e del av enhetscirkelns area. Vi vet från sats 3.16 att enhetscirkelns area är π så $|W| = \theta/2$. Från (3.20) ser vi att $|T_1 \cup T_2| = |T_1| + |T_2|$ och med hjälp av figur 5.6(b) och sats 3.15 räknar vi ut att

$$\begin{aligned} |T_1| + |T_2| &= \frac{\cos\theta \sin\theta}{2} + \frac{(1-\cos\theta)\sin\theta}{2} \qquad \text{och} \\ |T_3| &= \frac{\tan\theta}{2}. \end{aligned}$$

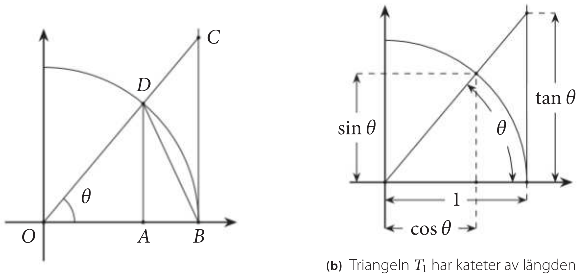
Därför är (5.13)

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

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

ARCUSFUNKTIONER

De trigonometriska funktionerna sinus, cosinus och tangens är inte bijektiva. De är i synnerhet inte injektiva som syns i fallen av cosinus



(a) Triangeln med hörnpunkterna O, B och C är en förstoring av triangeln T_1 med hörnpunkterna O, A och D. Därför är deras respektive sidolängder proportionella.

(b) Triangeln T_1 har kateter av längden $\sin\theta$ och $\cos\theta$. Kateten av längden $\cos\theta$ förlängs till 1 i triangeln med hörnpunkterna O, B och C, så den andra har längden $\sin\theta/\cos\theta = \tan\theta$.

FIGUR 5.6 En fjärdedel av en enhetscirkel och en vinkel $\theta \in (0,\pi/2)$.

och sinus från (5.2). Tangens är inte heller injektiv eftersom

$$\tan\theta = \tan(\theta+\pi)$$

så länge som θ tillhör tangens definitionsmängd. Därför kan vi inte rakt av hitta inversa funktioner till de funktionerna. Däremot är det möjligt att på ett lämpligt sätt ersätta funktionernas definitionsmängder med mindre mängder så att funktionerna blir injektiva och ersätta deras målmängder med mindre mängder så att de blir surjektiva. Med *mindre mängder* menar vi äkta delmängder av de ursprungliga definitions- och målmängderna. Då finns det hopp om att åtminstone de modifierade funktionerna kan vara inverterbara.

För mängderna $E \subset D$ och M, och en funktion $F\colon D \to M$ kan vi definiera en ny funktion $f\colon E \to M$ genom att sätta $f(x) = F(x)$ för alla $x \in E$. Då kallas f för *restriktionen* av F från D till E. Vår uppgift är då att hitta restriktioner av trigonometriska funktioner samt nya målmängder så att de är bijektiva. Det finns förstås olika val av definitionsmängder som fungerar lika bra, men här gör vi det vanligaste valet.

I avsnitt 4.2 såg vi att det innebar mycket arbete att bevisa att $f_2\colon \mathbf{R}_0^+ \to \mathbf{R}_0^+$ definierad enligt formeln (4.5) var bijektiv. Här ska vi vara lite slarviga och delvis förlita oss på vår geometriska intuition. Vi kommer även att vara lite slarviga med notationen och använder samma namn för både en funktion och dess restriktion.

Eftersom cosinus är definierad som x-koordinater av punkter på enhetscirkeln känns det rimligt att cosinus värdemängd är intervallet $[-1,1]$: Från figur 5.1 ser vi att för ett godtyckligt $a \in [-1,1]$ kan vi hitta ett $\theta \in \mathbf{R}$ sådant att $\cos\theta = a$. Därför tror vi att cosinus skulle vara surjektiv om vi begränsar målmängden till $[-1,1]$.

Vi kan ge ett mer rigoröst bevis för injektivitet utifrån enkela geometriska egenskaper hos cosinus. Vi börjar genom att visa att cosinus är strängt avtagande på $[0,\pi/2]$. Det följer av (5.9) att

$$\begin{aligned}\cos(x+h)-\cos x &= \cos x\cos h - \sin x\sin h - \cos x \\ &= \cos x(\cos h - 1) - \sin x \sin h.\end{aligned} \tag{5.14}$$

Utöver att $\cos\theta \leq 1$ för alla θ kan man se av definitionen av cosinus och sinus att

$$\begin{aligned}\cos x > 0 &\quad \text{för } x \in (-\pi/2, \pi/2),\\ \cos h < 1 &\quad \text{för } h \in (0,\pi/2], \text{ och}\\ \sin\theta \geq 0 &\quad \text{för } \theta \in [0,\pi/2].\end{aligned}$$

Därför i fallet $x \in [0,\pi/2)$ och $h \in (0,\pi/2]$ är

$$\cos x(\cos h - 1) < 0 \quad \text{och} \quad \sin x \sin h \geq 0. \tag{5.15}$$

Om $0 \leq x < y \leq \pi/2$ kan man skriva $y = x + h$ med $0 < h \leq \pi/2$ och det följer av (5.14) och (5.15) att

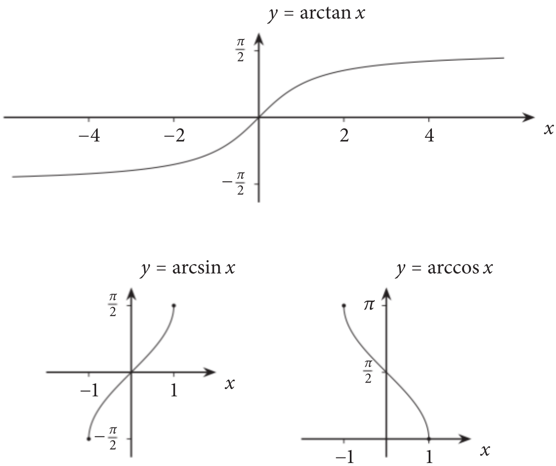
$$\begin{aligned}\cos y - \cos x &= \cos(x+h) - \cos x\\ &= \cos x(\cos h - 1) - \sin x \sin h\\ &< 0\end{aligned}$$

och cosinus är därför strängt avtagande på $[0,\pi/2]$.

Följaktligen är $\theta \mapsto -\cos(\pi - \theta)$ strängt avtagande på $[\pi/2,\pi]$ och eftersom

$$-\cos(\pi-\theta) = -\cos(\pi/2 - (\theta - \pi/2)) \underset{\substack{\uparrow\\(5.6)}}{=} -\sin(\theta - \pi/2)$$

$$\underset{\substack{\uparrow\\(5.5)}}{=} \sin(\pi/2 - \theta) \underset{\substack{\uparrow\\(5.7)}}{=} \cos(\theta)$$



FIGUR 5.7 Grafer av arcusfunktioner.

är cosinus också strängt avtagande på $[\pi/2, \pi]$.

Vi har därför visat att cosinus är strängt avtagande på $[0, \pi/2] \cup [\pi/2, \pi] = [0, \pi]$ och enligt anmärkning 4.12 är den injektiv om vi begränsar dess definitionsmängd till $[0, \pi]$. Tillsammans med dess surjektivitet på $[-1, 1]$ innebär det att

$$\cos\colon [0, \pi] \to [-1, 1]$$

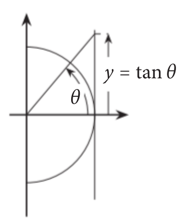
är bijektiv. Det följer med hjälp av (5.6) att

$$\sin\colon [-\pi/2, \pi/2] \to [-1, 1]$$

är bijektiv (och även strängt växande).

Med stöd av figur 5.8 ser man att till varje reellt tal y finns det en vinkel $\theta \in (-\pi/2, \pi/2)$ sådan att $\tan\theta = y$. Därför är tangens surjektiv med sin ursprungliga målmängd $\mathbf{R}$. Utifrån monoticiteten av cosinus och sinus kan vi visa att restriktionen av tangens till $(-\pi/2, \pi/2)$ är injektiv – se uppgift 5.9. Som följd är

$$\tan\colon (-\pi/2, \pi/2) \to \mathbf{R}$$



FIGUR 5.8 För varje reellt tal y finns det en vinkel $\theta \in (-\pi/2, \pi/2)$ sådan att $\tan\theta = y$.

bijektiv. Slutsatsen är att vi får definiera

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

som inversa funktioner till cosinus, sinus och tangens. Vi ritar graferna av de inversa trigonometriska funktionerna i figur 5.7.

5.3 Sammanfattning och uppgifter

I kapitel 5 har vi:

- definierat trigonometriska funktioner med hjälp av enhetscirkeln;
- utifrån spegel- och rotationssymmetri bevisat många trigonometriska formler;
- räknat ut vissa exakta värden av trigonometriska funktioner;
- bevisat olikheter mellan trigonometriska funktioner och deras argument;
- hittat restriktioner av cosinus, sinus och tangens som är bijektiva och definierat arcusfunktioner.

UPPGIFTER

5.1 Bevisa den motsvarande formeln till (5.10) för $\sin(\theta - \varphi)$.

5.2 Bevisa med hjälp av uppgift 5.1, (5.8), (5.9) och (5.10) att

(a) $2\cos\theta\cos\varphi = \cos(\theta - \varphi) + \cos(\theta + \varphi)$, och

(b) $2\sin\theta\cos\varphi = \sin(\theta - \varphi) + \sin(\theta + \varphi)$.

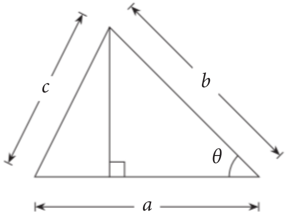
5.3 Hitta alla lösningar $\theta \in \mathbf{R}$ till ekvationen

$$0 = \cos(2\theta) - \cos\theta + 1.$$

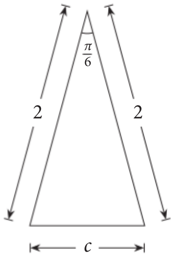
5.4 (a) Betrakta triangeln i figur 5.9(a) med sidlängderna a, b och c och en vinkel θ mitt emot sidan av längden c. Använd bland annat Pythagoras sats (sats 3.17) för att bevisa

$$c^2 = a^2 + b^2 - 2ab\cos\theta.$$

Likheten kallas för *cosinussatsen*.



(a) En triangel delad i två stödjer beviset av cosinussatsen.



(b) En likbent triangel med två sidor av längd 2.

FIGUR 5.9 Cosinussatsen kan användas för att räkna ut exakta värden av trigonometriska funktioner.

(b) Använd bland annat cosinussatsen och figur 5.9(b) för att visa att

$$\sin\left(\frac{\pi}{12}\right) = \frac{\sqrt{3}-1}{2\sqrt{2}} \quad \text{och} \quad \cos\left(\frac{\pi}{12}\right) = \frac{\sqrt{3}+1}{2\sqrt{2}}.$$

Det kan vara nyttigt att komma ihåg att $4 \pm 2\sqrt{3} = 3 \pm 2\sqrt{3} + 1 = (\sqrt{3} \pm 1)^2$.

5.5 Beräkna följande uttryck.

(a) $\arcsin\left(\frac{1}{2}\right)$

(b) $\arccos\left(-\frac{\sqrt{2}}{2}\right)$

(c) $\arccos\left(-\frac{\sqrt{3}}{2}\right)$

(d) $\arccos\left(\cos\left(\frac{17\pi}{5}\right)\right)$

(e) $\arctan\left(-\sqrt{3}\right)$

(f) $\sin\left(\arctan\left(\frac{12}{5}\right)\right)$

5.6 Ge ett motsägelsebevis till påståendet att

$$\cos\theta + \sin\theta \geq 1$$

för alla $\theta \in [0, \pi/2]$.

5.7 Använd (5.3) och sats 5.1 för att bevisa

$$\cos^2\theta \geq 1 - \theta^2$$

och

$$\cos\theta \geq 1 - \theta$$

för $\theta \in [0, \pi/2]$.

5.8 Skriv om

$$34\sqrt{3}\cos\theta + 34\sin\theta$$

som ett uttryck i $\theta \in \mathbf{R}$ som innehåller högst en trigonometrisk funktion.

5.9 En produkt eller en kvot av monotona funktioner är inte nödvändigtvist monoton – se uppgift 3.1. Men i den här uppgiften visar vi att

$$x \mapsto \tan x = \frac{\sin x}{\cos x}$$

är monoton på $(-\pi/2, \pi/2)$. Från sidor 146–147 vet vi att:

(a) $\cos\theta > 0$ för $\theta \in (-\pi/2, \pi/2)$;

(b) cosinus är strängt avtagande på $[0, \pi]$; och

(c) sinus är strängt växande på $[-\pi/2, \pi/2]$.

Använd (a)–(c) för att visa att tangens är strängt växande på $[0, \pi/2)$. Visa sedan att tangens är strängt växande på $(-\pi/2, \pi/2)$.

Den naturliga exponentialfunktionen och dess invers

6.1 Exponentialfunktion

Med ränta på ränta som ett motiv utreder vi egenskaperna hos uttrycket $(1 + x/n)^n$. Även rent matematiskt är uttrycket intressant när vi tänker på gränsvärdet då $n \to \infty$. Betrakta fallet $x = 1$: Å ena sidan är $(1 + 1/n)$ alltid större än ett, så om vi höjer upp det till en hög potens n skulle man kunna tro det ger ett jättestort tal som blir större ju större n är och därför borde följden inte ha någon övre begränsning; å andra sidan är $\lim_{n\to\infty}(1 + 1/n) = 1$ (enligt exempel 2.32 och sats 2.35) och $1^n = 1$, så gränsvärdet av $(1 + x/n)^n$ då $n \to \infty$ borde vara 1.

Ingetdera argumentet gäller som bevis eftersom vi inte kan räkna ut gränsvärdet av en sammansättning genom att bryta upp uttrycket där båda delar beror på n och sedan ta gränsvärdet av varsin del. En noggrann undersökning visar att verkligheten är precis mitt emellan de två felaktiga slutsatserna.

Vår efterforskning leder till en ny funktion – som vi kommer kalla för den naturliga exponentialfunktionen – och senare i slutet av avsnitt 6.2 även en utvidgning av begreppet potens – den här gången till irrationella potenser.

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 att så man kan bli ännu rikare! Banken har råd att betala ränta eftersom den investerar de pengar 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.

(a) Är det bättre för kunden att banken gör flera delbetalningar?
(b) Finns det någon övre begränsning 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. Det är ett bra exempel där en olikhet kan var ett kraftfullt verktyg för att utreda matematiska frågor.

Lemma 6.1 (Bernoullis olikhet) *För alla reella tal $x \geq -1$ och alla $n \in \mathbf{Z}_+$ får man att*

$$(1+x)^n \geq 1+nx. \tag{6.1}$$

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

Bevis. Vi ger ett induktionsbevis. Först ser vi vad som händer då $n = 1$: $(1+x)^n = (1+x) \geq (1+x) = 1+nx$ så (6.1) är sant om $n = 1$. Sedan antar vi att (6.1) är sant för $n = m$ där $m \in \mathbf{Z}_+$ och betraktar fallet $n = m+1$:

$$\begin{aligned}(1+x)^{m+1} &= (1+x)^m(1+x) \geq (1+mx)(1+x)\\ &= 1+mx+x+mx^2 \geq 1+(m+1)x\end{aligned}$$

så (6.1) med $n = m+1$ följer från (6.1) med $n = m$. Enligt axiom 2.6 är (6.1) bevisad för alla $n \in \mathbf{Z}_+$.

I vilket steg använder vi hypotesen $x \geq -1$?

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

Med hjälp av lemma 6.1 kan vi svara jakande på fråga (a) ovan. Ju fler delbetalningar banken gör, desto mer ränta får kunden.

Hypotesen av sats 6.2 är alltid uppfylld om $x \geq 0$. Så följden $((1+\frac{x}{n})^n)_n$ är växande för icke-negativa räntor.

Sats 6.2 *För alla $x \in \mathbf{R}$ och $n \in \mathbf{Z}_+$ så att $n \geq -x$ är*

$$\left(1+\frac{x}{n}\right)^n \leq \left(1+\frac{x}{n+1}\right)^{n+1}. \tag{6.2}$$

Bevis. Vi betraktar fallet $n = -x$ som ett separat fall. I så fall är $\left(1+\frac{x}{n}\right)^n = 0$ och $\left(1+\frac{x}{n+1}\right)^{n+1} > 0$, så (6.2) gäller tydligen.

Om $n \neq -x$ måste $n > -x$ på grund av hypotesen $n \geq -x$, så $1+\frac{x}{n} > 0$ och det räcker att bevisa

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

Enligt lemma 6.1 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 $1+\frac{x}{n+1} \geq 0$ under antagandet att $n > -x$. □

Vi kan också svara ja på fråga (b) på sidan 152.

Sats 6.3 *Det finns en funktion $B\colon \mathbf{R} \to \mathbf{R}$ sådan att för alla $x \in \mathbf{R}$ och $n \in \mathbf{Z}_+$ är*

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

Bevis. Vi betraktar två fall: $n > x$; och $n \geq -x$. För varje $x \in \mathbf{R}$ täcker de två fallen alla positiva heltal n.

Fall 1: $n > x$.

Vi beräknar att

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

Eftersom $1-\frac{x}{n} > 0$ medför det att

$$\left(1+\frac{x}{n}\right)^n \leq \left(1-\frac{x}{n}\right)^{-n}. \tag{6.3}$$

Enligt sats 6.2 (tillämpad med $-x$ i stället för x) 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 definierar vi $n_0(x)$ lika med det minsta positiva heltalet strängt större än x och får att

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

och då är enligt (6.3)

$$\left(1+\frac{x}{n}\right)^n \leq \left(1-\frac{x}{n_0(x)}\right)^{-n_0(x)} =: B(x). \tag{6.4}$$

Fall 2: $n \geq -x$.

Det finns (tillräckligt stort) m sådant att $m > n \geq -x$ och $m > x$ så enligt sats 6.2 (nu tillämpad med x inte $-x$) och fall 1 (tillämpad med m i stället för n) är

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

EXPONENTIALFUNKTIONEN

Vi definierar till varje $x \in \mathbf{R}$ en följd $(\exp_n(x))_n$ enligt formeln

$$\exp_n(x) = \begin{cases} 0 & \text{om } n < -x, \\ \left(1 + \frac{x}{n}\right)^n & \text{om } n \geq -x. \end{cases} \tag{6.5}$$

Vi har i föregående avsnitt sett hur (6.5) är kopplad till ränta på ett bankkonto. I sats 6.2 och 6.3 har vi visat att $(\exp_n(x))_n$ är en växande och uppåt begränsad följd och därför enligt sats 2.33 är följden även konvergent. Gränsvärdet av (6.5) då $n \to \infty$ har många intressanta egenskaper och det är värt att studera som ett självständigt begrepp. Därför definierar vi en funktion $\exp\colon \mathbf{R} \to \mathbf{R}$ som kallas för *(den naturliga) exponentialfunktionen* enligt formeln

$$\exp(x) := \lim_{n\to\infty} \exp_n(x) = \lim_{n\to\infty} \left(1 + \frac{x}{n}\right)^n . \tag{6.6}$$

Speciellt benämner vi

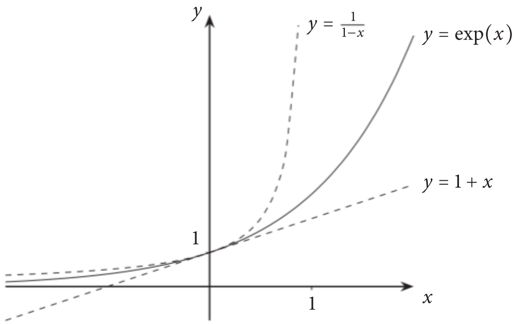
$$e := \exp(1) = \lim_{n\to\infty} \left(1 + \frac{1}{n}\right)^n \tag{6.7}$$

som ibland kallas för *Eulers* eller *Napiers tal* efter matematikerna Leonhard Euler och John Napier.

Anmärkning 6.4 (a) Observera att det är $\exp_n(x)$ och inte $(1 + \frac{x}{n})^n$ som måste vara växande i n. I fallet $x < 0$ växlar $(1 + \frac{x}{n})^n$ mellan positiva och negativa tal för $1 \leq n < -x$. Därför kunde vi enligt sats 2.33 likaledes definiera

$$\exp(x) = \sup_{n\in\mathbf{Z}_+} \exp_n(x). \tag{6.8}$$

(b) De två gränsvärdena i (6.6) är alltid lika trots att följderna ibland är definierade annorlunda (när heltal $n < -x$ uppstår). Det är eftersom ett gränsvärde av en följd $(a_n)_n$ handlar om vad som händer med följdens *svans* – det vill säga vad som händer med a_n för tillräckligt större n. Att följderna inte nödvändigtvis är lika för de första tio, hundra eller miljon elementen spelar ingen roll. Man kan inte säga samma sak för den minsta övre begränsningen av en följd och därför går det inte att ersätta $\exp_n(x)$ i (6.8) med $(1 + \frac{x}{n})^n$.



FIGUR 6.1 Exponentialfunktionens graf ligger under grafen av $x \mapsto 1/(1-x)$ för $x < 1$ och över både grafen av funktionen $x \mapsto x + 1$ och x-axeln.

Enkla egenskaper hos exponentialfunktionen

Nu börjar vi utreda egenskaper hos exponentialfunktionen. Vår första sats visar oss bland annat att grafen av exponentialfunktionen befinner sig mellan grafen av två olika funktioner. Se figur 6.1.

Sats 6.5 *Exponentialfunktionen* $\exp\colon \mathbf{R} \to \mathbf{R}$ *har följande egenskaper:*

(a) $\exp(x) > 0$ *för alla* $x \in \mathbf{R}$*;*

(b) $\exp(x) \geq 1 + x$ *för alla* $x \in \mathbf{R}$*;*

(c) $\exp(x)\exp(-x) = 1$ *för alla* $x \in \mathbf{R}$*; och*

(d) $\exp(x) \leq 1/(1-x)$ *för alla* $x < 1$.

Bevis. Observera först att

$$\exp_n(x) > 0$$

för $n > -x$ och eftersom $\exp(x)$ är en övre begränsning av $(\exp_n(x))_n$ (se (6.8)) måste även $\exp(x) > 0$ och (a) är bevisad.

Enligt lemma 6.1 är

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

för $n \geq -x$, så eftersom $\exp(x)$ är en övre begränsning av $(\exp_n(x))_n$ (se (6.8) igen) är $\exp(x) \geq 1 + x$ och (b) är bevisad.

För att bevisa (c) använder vi (6.6) i stället för (6.8). Vi beräknar för $n > -x$ å ena sidan att

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

och å andra sidan att

$$\exp_n(x)\exp_n(-x) = \left(1-\frac{x^2}{n^2}\right)^n \underset{\substack{\uparrow\\ \text{lemma 6.1}}}{\geq} 1-\frac{x^2}{n}.$$

Därför får vi att

$$1-\frac{x^2}{n} \leq \exp_n(x)\exp_n(-x) \leq 1. \tag{6.9}$$

Olikheterna i (6.9) tillåter oss att räkna ut gränsvärdet av $\exp_n(x)\exp_n(-x)$ direkt från definition 2.31: Tydligen medför (6.9) att

$$|1-\exp_n(x)\exp_n(-x)| \leq \frac{x^2}{n}$$

och därför för varje $\varepsilon > 0$ kan vi (enligt sats 2.26(b)) välja $N > x^2/\varepsilon$ så att om $n \geq N$ är

$$|1-\exp_n(x)\exp_n(-x)| \leq \frac{x^2}{n} \leq \frac{x^2}{N} < \varepsilon.$$

Därför säger definition 2.31 att $\lim_{n\to\infty}\exp_n(x)\exp_n(x) = 1$ och eftersom $(\exp_n(x))_n$ konvergerar mot $\exp(x)$ säger sats 2.35(c) att

$$\begin{aligned}1 &= \lim_{n\to\infty}\exp_n(x)\exp_n(x)\\ &= \left(\lim_{n\to\infty}\exp_n(x)\right)\left(\lim_{n\to\infty}\exp_n(x)\right) = \exp(x)\exp(-x)\end{aligned}$$

och (c) är bevisad.

Vi ger två olika bevis av (d). Det första förlitar sig på (b) och (c):

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

om $x < 1$. Men som ett alternativ till detta kan vi i stället titta närmare på beviset av sats 6.3. För $x < 1$ är det minsta positiva heltalet strängt större än x lika med $n_0(x) = 1$. Därför säger (6.4) att $B(x) = (1 - x)^{-1}$ och eftersom $B(x)$ är en övre begränsning av $(\exp_n(x))_n$ enligt (6.8) och (2.9) är det även större än eller lika med den minsta övre begränsningen $\exp(x)$. □

Sats 6.5(c) är ett specialfall av en mer allmän räkneregel som exponentialfunktionen lyder. Det utreder vi vidare i följande avsnitt (se sats 6.9).

Innan vi går vidare vänder vi tillbaka till föregående bevis. Vi såg i (6.9) att vi kunde beräkna gränsvärdet av en följd genom att stänga in den mellan två andra följder som har samma gränsvärde. Metoden fungerar i en mer allmän kontext än just (6.9) och vi kommer att använda den framöver så vi formulerar den som en sats. Den kallas för instängningssatsen.

Sats 6.6 (Instängningssatsen) *Betrakta tre följder $(\alpha_n)_{n\in\mathbf{Z}_+}$, $(\beta_n)_{n\in\mathbf{Z}_+}$ och $(\gamma_n)_{n\in\mathbf{Z}_+}$. Anta att $(\alpha_n)_n$ och $(\gamma_n)_n$ konvergerar mot ett gemensamt gränsvärde A (det vill säga $\lim_{n\to\infty}\alpha_n = A = \lim_{n\to\infty}\gamma_n$) och för något $M \in \mathbf{Z}_+$ är*

$$\alpha_n \le \beta_n \le \gamma_n \tag{6.10}$$

för alla $n \ge M$. Då är även $(\beta_n)_n$ konvergent och

$$\lim_{n\to\infty}\beta_n = A.$$

Att (6.10) bara antas gälla för $n \ge M$ är en annan manifestation av att det är vad som händer längs följdens svans som är viktigt när det gäller gränsvärdet av en följd. (Jämför med anmärkning 6.4(b).)

Bevis. Vi betraktar ett godtyckligt $\varepsilon > 0$. Eftersom både $(\alpha_n)_n$ och $(\gamma_n)_n$ konvergerar mot A finns det $N_1, N_2 \in \mathbf{Z}_+$ sådana att

$$|\alpha_n - A| < \varepsilon \text{ för } n > N_1, \text{ och } |\gamma_n - A| < \varepsilon \text{ för } n > N_2. \tag{6.11}$$

Om vi sätter $N = \max\{N_1, N_2, M\}$ får vi från (6.10) och (6.11) att

$$-\varepsilon < \alpha_n - A \le \beta_n - A \le \gamma_n - A < \varepsilon$$

om $n \ge N$. Därför är $|\beta_n - A| < \varepsilon$ om $n \ge N$. Enligt definition 2.31 är $(\beta_n)_n$ konvergent och $\lim_{n\to\infty}\beta_n = A$. □

För att visa kraften av sats 6.6 tar vi ett exempel från trigonometri.

Exempel 6.7 I signalbehandling behöver man utreda funktionen $\operatorname{sinc}(x) := \sin(x)/x$. Här använder vi instängningssatsen för att visa

$$\lim_{n\to\infty} \operatorname{sinc}\left(\frac{1}{n}\right) = \lim_{n\to\infty} n \sin\left(\frac{1}{n}\right) = 1. \tag{6.12}$$

Lösning. Från sats 5.1 vet vi att

$$\sin\left(\frac{1}{n}\right) < \frac{1}{n} < \tan\left(\frac{1}{n}\right).$$

Vi kan skriva om den första olikheten till

$$n \sin\left(\frac{1}{n}\right) < 1$$

och – med hjälp av (5.1) och uppgift 5.7 – från den andra får vi att

$$1 - \frac{1}{n^2} \leq \cos\left(\frac{1}{n}\right) < n \sin\left(\frac{1}{n}\right).$$

Därför gäller (6.10) med valen $\alpha_n = 1 - 1/n^2$, $\beta_n = n\sin(1/n)$ och $\gamma_n = 1$ för alla $n \in \mathbf{Z}_+$. Tydligen är $\lim_{n\to\infty} \gamma_n = \lim_{n\to\infty} 1 = 1$ och från exempel 2.32 och sats 2.35 är

$$\lim_{n\to\infty} \alpha_n = \lim_{n\to\infty} \left(1 - \frac{1}{n^2}\right) = 1 - \left(\lim_{n\to\infty} \frac{1}{n}\right)^2 = 1.$$

Så vi kan tillämpa sats 6.6 med $A = 1$ och dra slutsatsen att $\beta_n \to 1$ då $n \to \infty$. Därmed har vi bevisat likhet (6.12). □

Här lyckades vi till och med bevisa (6.10) i instängningssatsen med stränga olikheter. Det innebär förstås inte att man får stränga olikheter mellan följdernas gränsvärden. Man får inte nödvändigtvis mer än vanlig olikhet mellan gränsvärdena av två följder när en sträng olikhet mellan följdernas element finns.

Att $(\operatorname{sinc}(1/n))_n$ konvergerar mot 1 (det vill säga att (6.12) gäller) är intressant eftersom både $\sin(1/n)$ och $1/n$ är små för större n. Vad det säger är att båda kvantiteterna är ungefär lika små för stort n.

Motivationen bakom exponentialfunktionens namn

Nu återvänder vi till exponentialfunktionen. Följande hjälpsats är ett tekniskt hjälpmedel för att bevisa sats 6.9 nedan. Man kan med fördel återkomma till dess bevis efter att man har läst beviset av sats 6.9.

Lemma 6.8 *För $x \geq 0$ och positiva heltal n är*

$$\exp(x) \geq \sum_{k=1}^{n} \frac{x^k}{n^{k-1}k!}.$$

Bevis. Olikheten är självklar om $x = 0$. När $x > 0$ kan vi tillämpa Binomialsatsen (sats 2.13). Vi får att

$$\begin{aligned}\exp(x) \geq \exp_n(x) &= \left(1 + \frac{x}{n}\right)^n = \sum_{k=0}^{n} \binom{n}{k} \frac{x^k}{n^k} \\ &\geq \sum_{k=1}^{n} \binom{n}{k} \frac{x^k}{n^k} = \sum_{k=1}^{n} \frac{n!x^k}{(n-k)!k!n^k},\end{aligned}$$

men

$$\frac{n!}{(n-k)!n^k} = \frac{(n-1)!}{(n-k)!n^{k-1}} \geq \frac{1}{n^{k-1}}$$

eftersom $(n-1)! \geq (n-k)!$ om $1 \leq k \leq n$. Därför är

$$\sum_{k=1}^{n} \frac{n!x^k}{(n-k)!k!n^k} \geq \sum_{k=1}^{n} \frac{x^k}{n^{k-1}k!}$$

och satsen är bevisad. □

Följande sats ger oss en aning om varför vi valde namnet exponentialfunktion. Den lyder precis samma räkneregel som den som gäller för potenser (jämför med sats 1.8(a) och sats 4.14(a)).

Sats 6.9 *För alla $x, y \in \mathbf{R}$ är*

$$\exp(x + y) = \exp(x)\exp(y).$$

Bevis. För ett godtyckligt positivt heltal $n > \max\{-x, -y, -(x + y)\}$ beräknar vi att

$$\begin{aligned}\exp_n(x)\exp_n(y) &= \left(1 + \frac{x}{n}\right)^n \left(1 + \frac{y}{n}\right)^n \\ &= \left(1 + \frac{x + y}{n} + \frac{xy}{n^2}\right)^n.\end{aligned}$$

Och enligt binomialsatsen (sats 2.13)[1] är

$$\begin{aligned}
&\left(1+\frac{x+y}{n}+\frac{xy}{n^2}\right)^n \\
&=\left(\left(1+\frac{x+y}{n}\right)+\frac{xy}{n^2}\right)^n \\
&=\sum_{k=0}^{n}\binom{n}{k}\left(1+\frac{x+y}{n}\right)^{n-k}\left(\frac{xy}{n^2}\right)^k \\
&=\left(1+\frac{x+y}{n}\right)^n+\sum_{k=1}^{n}\binom{n}{k}\left(1+\frac{x+y}{n}\right)^{n-k}\left(\frac{xy}{n^2}\right)^k \\
&=\exp_n(x+y)+\sum_{k=1}^{n}\binom{n}{k}\left(1+\frac{x+y}{n}\right)^{n-k}\left(\frac{xy}{n^2}\right)^k.
\end{aligned}$$

Därför är

$$\begin{aligned}
&\exp_n(x)\exp_n(y) \\
&=\exp_n(x+y)+\sum_{k=1}^{n}\binom{n}{k}\left(1+\frac{x+y}{n}\right)^{n-k}\left(\frac{xy}{n^2}\right)^k
\end{aligned} \tag{6.13}$$

för alla $n > \max\{-x, -y, -(x+y)\}$ och genom att ta gränsvärdena av båda sidor skulle satsen följa om vi kunde visa att sista termen går mot noll, det vill säga

$$\lim_{n\to\infty}\sum_{k=1}^{n}\binom{n}{k}\left(1+\frac{x+y}{n}\right)^{n-k}\left(\frac{xy}{n^2}\right)^k=0. \tag{6.14}$$

För att bevisa (6.14) räcker det enligt instängningssatsen (sats 6.6) att bevisa

$$\left|\sum_{k=1}^{n}\binom{n}{k}\left(1+\frac{x+y}{n}\right)^{n-k}\left(\frac{xy}{n^2}\right)^k\right| \le \frac{C(x,y)}{n} \tag{6.15}$$

för något $C(x,y)$ som inte beror på n. Med hjälp av triangelolikheten (sats 2.30) beräknar vi att

$$\begin{aligned}
&\left|\sum_{k=1}^{n}\binom{n}{k}\left(1+\frac{x+y}{n}\right)^{n-k}\left(\frac{xy}{n^2}\right)^k\right| \\
&\le\sum_{k=1}^{n}\binom{n}{k}\left|1+\frac{x+y}{n}\right|^{n-k}\left(\frac{|xy|^k}{n^{2k}}\right) \\
&\le\sum_{k=1}^{n}\binom{n}{k}\left(1+\frac{|x|+|y|}{n}\right)^{n-k}\left(\frac{|xy|}{n^2}\right)^k.
\end{aligned} \tag{6.16}$$

1 Om x eller y är noll (och vi inte får tillämpa sats 2.13) kan vi direkt härleda (6.13), där Σ-summan är 0. Då är (6.14) självklar.

Enligt sats 6.3 är

$$\left(1+\frac{|x|+|y|}{n}\right)^{n-k} \leq \left(1+\frac{|x|+|y|}{n}\right)^{n} \leq B(|x|+|y|)$$

där funktionen B är funktionen från sats 6.3. Kom även ihåg från uppgift 2.7 att

$$\binom{n}{k} \leq \frac{n^k}{k!}.$$

Därför fortsätter vi från (6.16) och uppskattar

$$\begin{aligned}
&\sum_{k=1}^{n}\binom{n}{k}\left(1+\frac{|x|+|y|}{n}\right)^{n-k}\left(\frac{|xy|}{n^2}\right)^{k} \\
&\leq \sum_{k=1}^{n}\frac{n^k}{k!}B(|x|+|y|)\left(\frac{|xy|^k}{n^{2k}}\right) \\
&= \sum_{k=1}^{n}\frac{|xy|^k B(|x|+|y|)}{n^k k!} \\
&= \frac{B(|x|+|y|)}{n}\sum_{k=1}^{n}\frac{|xy|^k}{n^{k-1}k!} \\
&\leq \frac{B(|x|+|y|)\exp(|xy|)}{n}
\end{aligned}$$

där sista olikheten följs av lemma 6.8, och (6.15) är bevisat med

$$C(x,y) = B(|x|+|y|)\exp(|xy|)$$

så länge $n \geq M$ där $M := \max\{-x, -y, -(x+y)\}$.

Som sagt kan vi nu avsluta beviset: Vi tillämpar instängningssatsen (sats 6.6) med M som ovan, $\alpha_n = -C(x,y)/n$, β_n lika med summan i (6.15) och $\gamma_n = C(x,y)/n$ för att dra (6.14) som en slutsats. Genom att ta gränsvärden av båda leden i (6.13) får vi att $\exp(x)\exp(y) = \exp(x+y) + 0$ och satsen är bevisad. □

Men hjälp av sats 6.9 kan vi visa att exponentialfunktionen är strängt växande. Det stämmer överens med bilden vi skissade i figur 6.1.

Sats 6.10 *Exponentialfunktionen* exp: $\mathbf{R} \to \mathbf{R}$ *är strängt växande.*

Bevis. Vi betraktar två reella tal $x < y$. Vi kan förstås skriva $y = x + h$ där $h > 0$ och sedan räknar vi ut att

$$\begin{aligned} \exp(y) = \exp(x+h) &\underset{\text{sats 6.9}}{\uparrow}{=} \exp(x)\exp(h) \\ &\underset{\text{sats 6.5(a) och (b)}}{\overset{}{\geq}} \exp(x)(1+h) \underset{\text{sats 6.5(a) och } h>0}{>} \exp(x). \end{aligned}$$

□

Nu tittar vi lite närmare på funktioner som uppfyller räkneregeln som nämndes i sats 6.9. Vi betraktar en funktion $f\colon \mathbf{Q} \to \mathbf{R}$ som lyder regeln

$$f(x+y) = f(x)f(y) \tag{6.17}$$

för alla $x, y \in \mathbf{Q}$ och

$$f(1) = a \tag{6.18}$$

för något tal $a > 0$. Kom ihåg att vi har bara definierat rationella potenser så det inte är så konstigt att vi tar $\mathbf{Q}$ som f:s definitionsmängd (och inte till exempel $\mathbf{R}$). Observera att funktionen $x \mapsto a^x$ lyder (6.17) och (6.18), men i följande sats visar vi att $x \mapsto a^x$ är den unika funktionen som gör det. För givet $a > 0$ kallas även $x \mapsto a^x$ för *en exponentialfunktion med bas a*, så kan vi redan här gissa varför den naturliga exponentialfunktionen fick sitt namn. Men just nu ska vi ha i åtanke att den naturliga exponentialfunktionen exp och $x \mapsto a^x$ har två olika definitioner.

Vi betonar att när vi säger exponentialfunktionen eller den naturliga exponentialfunktionen menar vi $\exp$ som definieras i (6.6). I fallet vi vill hänvisa till $x \mapsto a^x$ säger vi i stället en exponentialfunktion med bas a.

Sats 6.11 *Om $f\colon \mathbf{Q} \to \mathbf{R}$ uppfyller* (6.17) *och* (6.18) *då är*

$$f(r) = a^r \tag{6.19}$$

för alla $r \in \mathbf{Q}$.

Bevis. Vi vet att $x = x/2 + x/2$ och därför är

$$f(x) = f(x/2 + x/2) \underset{(6.17)}{=} f(x/2)f(x/2) = f(x/2)^2 \geq 0 \tag{6.20}$$

för alla $x \in \mathbf{Q}$. Vi kan också beräkna att

$$a \underset{(6.18)}{=} f(1) = f(1+0) \underset{(6.17)}{=} f(1)f(0) \underset{(6.18)}{=} af(0),$$

så

$$f(0) = 1 \tag{6.21}$$

ty vi kan dela med $a \neq 0$. Det innebär att vi kan säga

$$f(x)f(-x) \underset{\substack{\uparrow \\ (6.17)}}{=} f(x + (-x)) = f(x - x) = f(0) \underset{\substack{\uparrow \\ (6.21)}}{=} 1 \tag{6.22}$$

för alla $x \in \mathbf{Q}$. Slutligen kan vi visa att

$$f(nx) = f(x)^n \tag{6.23}$$

för alla $x \in \mathbf{Q}$ och $n \in \mathbf{Z}_+$ med ett induktionsbevis. Alltså om $n = 1$ är $f(nx) = f(x) = f(x)^n$ och om $f(nx) = f(x)^n$ för något $n \in \mathbf{Z}_+$ är

$$f((n+1)x) = f(nx+x) \underset{\substack{\uparrow \\ (6.17)}}{=} f(nx)f(x) \underset{\substack{\uparrow \\ \text{antagandet}}}{=} f(x)^n f(x) = f(x)^{n+1}.$$

Därför enligt axiom 2.6 är (6.23) bevisad.

Om man skriver x/n i stället för x i (6.23) är

$$f(x) = f(x/n)^n$$

så detta och (6.20) medför att $y = f(x/n)$ är den icke-negativa lösningen till ekvationen

$$f(x) = y^n$$

och därmed är

$$\sqrt[n]{f(x)} = f(x/n) \tag{6.24}$$

för alla $x \in \mathbf{Q}$ och $n \in \mathbf{Z}_+$. Sedan kan vi beräkna att

$$f(n/m) \underset{\substack{\uparrow \\ (6.24)}}{=} \sqrt[m]{f(n)} \underset{\substack{\uparrow \\ (6.23)}}{=} \sqrt[m]{f(1)^n} = f(1)^{n/m} \underset{\substack{\uparrow \\ (6.18)}}{=} a^{n/m}$$

för alla $n,m \in \mathbf{Z}_+$ och (6.19) är bevisad för $r = n/m > 0$.

För $r < 0$ kan vi skriva $r = -n/m$ där $m, n \in \mathbf{Z}_+$ och beräkna

$$f(-n/m) \underset{(6.22)}{\uparrow}{=} \frac{1}{f(n/m)} \underset{(6.19)\text{ för } r>0}{\uparrow}{=} \frac{1}{\sqrt[m]{(f(1))^n}} \underset{(4.24)}{\uparrow}{=} \sqrt[m]{\frac{1}{f(1)^n}}$$

$$\underset{(1.16)}{\uparrow}{=} \sqrt[m]{f(1)^{-n}} \underset{(4.23)}{\uparrow}{=} f(1)^{-n/m} \underset{(6.18)}{\uparrow}{=} a^{-n/m},$$

så (6.19) är bevisad för $r = -n/m < 0$.

För $r = 0$ följer (6.19) från (6.21) och faktumet $a^0 = 1$. □

Sats 6.9 och 6.11 tillåter oss att bevisa en anmärkningsvärd sak: Exponentialfunktionen vars definition hade mer att göra med ränta på ränta än potenser är faktiskt en exponentialfunktion med bas e.

Sats 6.12 *För alla $r \in \mathbf{Q}$ är*

$$\exp(r) = e^r.$$

Bevis. Sats 6.9 och (6.7) visar att exponentialfunktionen med definitionsmängden begränsad till $\mathbf{Q}$ uppfyller (6.17) och (6.18) med $a = e$. Därför är satsen en föjd av sats 6.11 med $a = e$. □

När man ser sats 6.12 förstår man direkt varför vi döpte exponentialfunktionen så. Men sats 6.12 har även konsekvenser för hur och för vilka tal vi kan definiera potenser. Denna punkt återkommer vi till i avsnitt 6.2.

Exponentialfunktionen som en serie

Innan vi går vidare visar vi en annan representation av exponentialfunktionen. Funktionen är inte ett polynom men vi kan visa att den är ett gränsvärde av polynom. Eftersom polynom har formen (3.2) behövs det att vi tar gränsvärden av summor och därför inför vi notationen

$$\sum_{k=m}^{\infty} \alpha_k := \lim_{n\to\infty} \sum_{k=m}^{n} \alpha_k \tag{6.25}$$

för gränsvärdet då $n \to \infty$ av summan från m till n av en följd $(\alpha_k)_k$. En sådan oändlig summa (6.25) kallas för en *serie* och summorna

$$\sum_{k=m}^{n} \alpha_k$$

(för $n \geq m$) kallas för *delsummorna* eller *partialsummorna* av serien. Vi skriver även

$$\sum_{k=m}^{\infty} \alpha_k = \alpha_m + \alpha_{m+1} + \alpha_{m+2} + \cdots$$

med en ellipsis för att indikera att serien är mer än vad vi explicit har skrivit. En serie måste inte vara mer än en formell oändlig summa men när gränsvärdet i (6.25) existerar, identifierar vi serien med gränsvärdet – och då säger vi att serien är *konvergent* – annars är den inget mer är en formell summa – och serien sägs vara *divergent*.

Exempel 6.13 Vi har faktiskt redan sett exempel på serier. I definition 2.37 definierade vi ett reellt tal som gränsvärdet av en följd av trunkerade decimalutvecklingar. Varje trunkerad decimalutveckling var i sin tur en delsumma. Där tog vi gränsvärdet i begynnelsevärdet för summationsindexet i stället för slutvärdet, men det är väsentligen samma begrepp – med ett variabelbyte kan man skriva det som i (6.25).

Exempel 6.14 Bevisa att serien

$$\sum_{k=0}^{\infty} \left(\frac{1}{2}\right)^k = 1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \cdots$$

är konvergent.

Lösning. Först beräknar vi med hjälp av sats 1.14 att delsummorna är

$$\sum_{k=0}^{n} \left(\frac{1}{2}\right)^k = \sum_{k=1}^{n+1} \left(\frac{1}{2}\right)^{k-1} = \frac{1-\left(\frac{1}{2}\right)^{n+1}}{1-\frac{1}{2}} = 2\left(1 - \frac{1}{2^{n+1}}\right).$$

Men enligt uppgift 2.4 är $0 \leq \frac{1}{2^{n+1}} \leq \frac{1}{4(n+1)}$ för $n \geq 4$ så instängningssatsen (sats 6.6) och exempel 2.32 säger att $\lim_{n\to\infty} \frac{1}{2^{n+1}} = 0$. Därför är enligt sats 2.35 serien konvergent och vi kan även beräkna att

$$\sum_{k=0}^{\infty} \left(\frac{1}{2}\right)^k = \lim_{n\to\infty} \sum_{k=0}^{n} \frac{1}{2^k} = \lim_{n\to\infty} 2\left(1 - \frac{1}{2^{n+1}}\right) = 2\,(1-0) = 2. \qquad \square$$

Nu skriver vi ner serierepresentationen av exponentialfunktionen.

Sats 6.15 *För varje $x \neq 0$ är*[2]

$$\exp(x) = \sum_{k=0}^{\infty} \frac{x^k}{k!}.$$

Serien i sats 6.15 är ett exempel på en *potensserie* och är även exponentialfunktionens *taylorutveckling* kring noll. Tyvärr ligger den teorin utanför bokens ram, men det intressanta är att vi ändå kan bevisa satsen med mer elementära verktyg.

Innan vi ger ett fullständigt bevis av sats 6.15 motiverar vi satsen med en uträkning: Sats 2.13 tillåter oss att skriva

$$\left(1 + \frac{x}{n}\right)^n = \sum_{k=0}^{n} \binom{n}{k} \left(\frac{x}{n}\right)^k = \sum_{k=0}^{n} \frac{n!}{(n-k)!n^k} \frac{x^k}{k!}. \tag{6.26}$$

Gränsvärdet av vänsterledet är förstås (enligt (6.6)) $\exp(x)$ så satsen skulle vara bevisad om vi kunde visa

$$\frac{n!}{(n-k)!n^k} \to 1 \tag{6.27}$$

då $n \to \infty$. Men det är subtilt eftersom uttrycket i (6.27) beror på både n och k. Det innebär att termer i delsummorna kommer att konvergera i olika takt i k till sina gränsvärden i n. Vi behöver därför en hjälpsats.

Lemma 6.16 *Vi har uppskattningarna*

$$0 \leq 1 - \frac{n!}{(n-k)!n^k} \leq \frac{(k-1)^2}{n}$$

för heltal $n \geq k \geq 0$.

Bevis. Vi kan beräkna att

$$\begin{aligned} \frac{n!}{(n-k)!n^k} &= \frac{n(n-1)\ldots(n-k+1)}{n^k} \\ &= \prod_{j=0}^{k-1} \frac{(n-j)}{n} \leq \prod_{j=1}^{k} 1 = 1. \end{aligned}$$

samt

$$\begin{aligned} \frac{n!}{(n-k)!n^k} &= \prod_{j=0}^{k-1} \frac{(n-j)}{n} \geq \prod_{j=1}^{k-1} \frac{n-(k-1)}{n} \\ &= \left(1 - \frac{k-1}{n}\right)^{k-1} \underset{\text{sats 6.1}}{\geq} 1 - \frac{(k-1)^2}{n}. \end{aligned}$$

De två olikheterna ger lemmat. □

2 I fallet $x = 0$ blir första termen i serien 0^0 och är därför odefinierad. Jämför med anmärkning 1.15(b) och även marginalanteckningen på sidan 94.

Innan vi kan bevisa sats 6.15 behöver vi visa två hjälpsatser till. Först visar vi att serien i sats 6.15 är konvergent om $x \geq 0$ och sedan visar vi att om en serie är konvergent så kan vi hitta en *svans* av serien som är godtyckligt liten.

Lemma 6.17 *För varje $x > 0$ är serien $\sum_{k=0}^{\infty} \frac{x^k}{k!}$ konvergent.*

Bevis. Följden $(S_n)_{n\in\mathbf{N}}$ av delsummor

$$S_n = \sum_{k=0}^{n} \frac{x^k}{k!}$$

är tydligen växande (i n) om $x > 0$. Så enligt sats 2.33(a) räcker det att visa att delsummorna har en övre begränsning (som förstås kan bero på x, men inte n). För att visa det påstår vi att

$$\frac{x^k}{k!} \leq \left(\frac{1}{2}\right)^{k-m} \frac{x^m}{m!} \tag{6.28}$$

för $k > m \geq 2x$. Om (6.28) är sant kan vi beräkna att

$$\begin{aligned}
S_n = \sum_{k=0}^{n} \frac{x^k}{k!} &= \sum_{k=0}^{m} \frac{x^k}{k!} + \sum_{k=m+1}^{n} \frac{x^k}{k!} \\
&\leq \sum_{k=0}^{m} \frac{x^k}{k!} + \frac{x^m}{m!} \sum_{k=m+1}^{n} \left(\frac{1}{2}\right)^{k-m} \\
&\leq \sum_{k=0}^{m} \frac{x^k}{k!} + \frac{x^m}{2m!} \sum_{k=0}^{\infty} \left(\frac{1}{2}\right)^{k} \\
&= \sum_{k=0}^{m} \frac{x^k}{k!} + \frac{x^m}{m!}
\end{aligned}$$

tack vare exempel 6.14. Eftersom högerledet är en summa av $m+2$ termer som beror endast på m och x och vi kan välja m beroende endast på x, har vi hittat den övre begränsningen vi behöver.

För att avsluta beviset måste vi bevisa (6.28): För $k > m$ är

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

så för $k > m \geq 2x$ är

$$\frac{x^k}{k!} \leq \left(\frac{x}{m+1}\right)^{k-m} \frac{x^m}{m!} \leq \left(\frac{1}{2}\right)^{k-m} \frac{x^m}{m!}$$

så (6.28) är bevisad. □

Lemma 6.18 *Om en serie $\sum_{k=0}^{\infty} \alpha_k$ är konvergent då är*

$$\lim_{m\to\infty} \sum_{k=m}^{\infty} \alpha_k = 0.$$

Bevis. Vi skriver

$$\sum_{k=m}^{n} \alpha_k = \sum_{k=0}^{n} \alpha_k - \sum_{k=0}^{m} \alpha_k$$

och tar med stöd från sats 2.35 gränsvärdet i n för att få

$$\sum_{k=m}^{\infty} \alpha_k = \sum_{k=0}^{\infty} \alpha_k - \sum_{k=0}^{m} \alpha_k.$$

Vi vet att $\sum_{k=0}^{\infty} \alpha_k$ är konvergent och därför döper vi $\ell := \sum_{k=0}^{\infty} \alpha_k$ så

$$\sum_{k=m}^{\infty} \alpha_k = \ell - \sum_{k=0}^{m} \alpha_k.$$

Vi vet även att $\lim_{m\to\infty} \sum_{k=0}^{m} \alpha_k = \ell$ så genom att återigen med stöd från sats 2.35 ta gränsvärdet i m får vi att

$$\lim_{m\to\infty} \sum_{k=m}^{\infty} \alpha_k = \ell - \ell = 0. \qquad \square$$

Äntligen är vi redo att bevisa sats 6.15.

Bevis (av sats 6.15). Med hjälp av (6.26) och triangelolikheten (sats 2.30) ser vi att

$$\begin{aligned}
\left| \sum_{k=0}^{m} \frac{x^k}{k!} - \left(1 + \frac{x}{n}\right)^n \right| &= \left| \sum_{k=0}^{m} \frac{x^k}{k!} - \sum_{k=0}^{n} \frac{n!}{(n-k)!n^k} \frac{x^k}{k!} \right| \\
&= \left| \sum_{k=0}^{m} \left(1 - \frac{n!}{(n-k)!n^k}\right) \frac{x^k}{k!} - \sum_{k=m+1}^{n} \frac{n!}{(n-k)!n^k} \frac{x^k}{k!} \right| \\
&\le \sum_{k=0}^{m} \left| 1 - \frac{n!}{(n-k)!n^k} \right| \frac{|x|^k}{k!} + \sum_{k=m+1}^{n} \left| \frac{n!}{(n-k)!n^k} \right| \frac{|x|^k}{k!}
\end{aligned}$$

för $0 \leq m < n$. Vi använder lemma 6.16 för att vidare beräkna att

$$\begin{aligned}
&\sum_{k=0}^{m}\left|1-\frac{n!}{(n-k)!n^k}\right|\frac{|x|^k}{k!}+\sum_{k=m+1}^{n}\left|\frac{n!}{(n-k)!n^k}\right|\frac{|x|^k}{k!}\\
&\leq \sum_{k=0}^{m}\frac{(k-1)^2}{n}\frac{|x|^k}{k!}+\sum_{k=m+1}^{n}\frac{|x|^k}{k!}\\
&\leq \frac{(m-1)^2}{n}\sum_{k=0}^{m}\frac{|x|^k}{k!}+\sum_{k=m+1}^{n}\frac{|x|^k}{k!}
\end{aligned}$$

och därför vet vi att

$$\begin{aligned}
\left|\sum_{k=0}^{m}\frac{x^k}{k!}-\left(1+\frac{x}{n}\right)^n\right| &\leq \frac{(m-1)^2}{n}\sum_{k=0}^{m}\frac{|x|^k}{k!}+\sum_{k=m+1}^{n}\frac{|x|^k}{k!}\\
&\leq \frac{(m-1)^2}{n}E(x)+\sum_{k=m+1}^{\infty}\frac{|x|^k}{k!}
\end{aligned} \tag{6.29}$$

där vi har skrivit

$$E(x) := \sum_{k=0}^{\infty}\frac{|x|^k}{k!}$$

som vi enligt lemma 6.17 vet är konvergent. Utöver det vet vi från lemma 6.18 att

$$\lim_{m\to\infty}\sum_{k=m+1}^{\infty}\frac{|x|^k}{k!}=\lim_{m\to\infty}\sum_{k=m}^{\infty}\frac{|x|^k}{k!}=0. \tag{6.30}$$

Nu betraktar vi ett godtyckligt $\varepsilon > 0$. Ekvationen (6.30) tillsammans med (6.6) (och definition 2.31 av ett gränsvärde) betyder att vi får välja ett $M \in \mathbf{Z}_+$ sådant att

$$\sum_{k=m+1}^{\infty}\frac{|x|^k}{k!}<\frac{\varepsilon}{3} \tag{6.31}$$

för $m \geq M$ och

$$\left|\left(1+\frac{x}{n}\right)^n-\exp(x)\right|<\frac{\varepsilon}{3} \tag{6.32}$$

för $n \geq M$. För varje $m \geq M$ får vi vid behov välja ett större n så att vi även har att

$$n \geq \max\left\{\frac{3(m-1)^2E(x)}{\varepsilon}, m\right\}. \tag{6.33}$$

Sedan är

$$\begin{aligned}
\left|\sum_{k=0}^{m}\frac{x^k}{k!}-\exp(x)\right| &\underset{\substack{\uparrow\\ \text{sats 2.30}}}{\leq} \left|\sum_{k=0}^{m}\frac{x^k}{k!}-\left(1+\frac{x}{n}\right)^n\right|+\left|\left(1+\frac{x}{n}\right)^n-\exp(x)\right| \\
&\underset{\substack{\uparrow\\ (6.29)\ \&\ (6.32)}}{<} \left(\frac{(m-1)^2}{n}E(x)+\sum_{k=m+1}^{\infty}\frac{|x|^k}{k!}\right)+\frac{\varepsilon}{3} \\
&\underset{\substack{\uparrow\\ (6.31)}}{<} \left(\frac{(m-1)^2}{n}E(x)+\frac{\varepsilon}{3}\right)+\frac{\varepsilon}{3} \underset{\substack{\uparrow\\ (6.33)}}{\leq} \varepsilon
\end{aligned}$$

och sats 6.15 är bevisad. □

6.2 Den naturliga logaritmfunktionen och irrationella potenser

I det här avsnittet visar vi att exponentialfunktionen är inverterbar och vi utreder egenskaper hos inversen – som vi kallar för den naturliga logaritmfunktionen. Med hjälp av båda funktionerna utvidgar vi begreppet potens för andra gången, nu från rationella tal till alla reella tal.

INVERSEN TILL EXPONENTIALFUNKTIONEN

Vi börjar med det följande nödvändiga resultatet för att ta inversen av exponentialfunktionen.

Sats 6.19 *Exponentialfunktionen* $\exp\colon \mathbf{R} \to (0,\infty)$ *är bijektiv.*

Bevis. Vi måste förstås bevisa att exponentialfunktionen är både injektiv och surjektiv.

Utifrån sats 6.10 vet vi att exponentialfunktionen är strängt växande. Därför vet vi direkt att den är injektiv (se anmärkning 4.12).

För att visa att den är surjektiv följer vi ett liknande argument som det i beviset av sats 4.6. Där visade vi att funktionen $x \mapsto x^2$ med målmängden $[0,\infty)$ var surjektiv och nu vill vi förstås göra samma sak för exponentialfunktionen. Vi vill visa att

$$\exp(x) = y \tag{6.34}$$

har för varje $y \in (0, \infty)$ en lösning x. För givet $y \in (0, \infty)$ delar vi upp $\mathbf{R}$ i två disjunkta mängder:

$$E_y^- := \{x \in \mathbf{R} \mid \exp(x) \le y\} \quad \text{och}$$
$$E_y^+ := \mathbf{R} \setminus E_y^- = \{x \in \mathbf{R} \mid \exp(x) > y\}.$$

Om vi betraktar

$$x_0 := \sup E_y^-$$

förväntar vi oss att $x = x_0$ är lösningen till (6.34). Men först måste vi kontrollera att den minsta övre begränsningen av E_y^- över huvud taget finns, det vill säga att x_0 är väldefinierat.

Enligt axiom 2.28 har E_y^- en minsta övre begränsning om E_y^- är icke-tom och uppåt begränsad. Enligt sats 6.5(b) är

$$1 + x \le \exp(x) \le y$$

för alla $x \in E_y^-$, så $x \le y - 1$ och $y - 1$ är en övre begränsning till E_y^-. För att visa att E_y^- är icke-tom betrakta talet

$$\tilde{x} := \frac{y-1}{y}.$$

Eftersom $y \in (0, \infty)$ är $\tilde{x} < 1$ och därför är

$$\exp(\tilde{x}) \le \frac{1}{1-\tilde{x}} = \frac{1}{1 - \left(\frac{y-1}{y}\right)} = y$$

enligt sats 6.5(d). Vi drar slutsatsen att $\tilde{x} \in E_y^-$ så E_y^- är icke-tom.

För att visa att $x = x_0$ är en lösning till (6.34) försöker vi bevisa att

$$|y - \exp(x_0)| < \frac{C(y)}{n} \tag{6.35}$$

för alla $n \in \mathbf{Z}_+$ och något $C(y)$ som inte beror på n. Om vi kan visa (6.35) då är $\exp(x_0) = y$ enligt sats 2.26(a).

Eftersom x_0 är den minsta övre begränsningen av E_y^- finns det $x_n \in E_y^-$ sådant att $x_0 - 1/n < x_n$. Därför är

$$\begin{aligned} y &\underset{\substack{\uparrow \\ x_n \in E_y^-}}{\ge} \exp(x_n) \underset{\substack{\uparrow \\ \text{sats 6.10}}}{>} \exp(x_0 - 1/n) \underset{\substack{\uparrow \\ \text{sats 6.9}}}{=} \exp(x_0)\exp(-1/n) \\ &\underset{\substack{\uparrow \\ \text{sats 6.5(a) \& (b)}}}{\ge} \exp(x_0)\left(1 - \frac{1}{n}\right) \end{aligned}$$

vilket medför att

$$y - \exp(x_0) > -\frac{\exp(x_0)}{n} \tag{6.36}$$

för alla $n \in \mathbf{Z}_+$.

Vi vet även att $x_0 + 1/n \in E_y^+$ eftersom det är större än den övre begränsningen x_0 till E_y^-. Därför är

$$\begin{aligned} y &< \exp(x_0 + 1/n) \underset{\text{sats 6.9}}{\underset{\uparrow}{=}} \exp(x_0)\exp(1/n) \\ &\underset{\text{sats 6.5(a) \& (d)}}{\underset{\uparrow}{\leq}} \frac{\exp(x_0)}{1 - 1/n} = \exp(x_0) + \frac{\exp(x_0)}{n-1} \end{aligned}$$

för $n \geq 2$ som medför att

$$y - \exp(x_0) < \frac{\exp(x_0)}{n-1} \tag{6.37}$$

för alla $n \in \mathbf{Z}_+ \setminus \{1\}$.

Eftersom x_0 bara beror på y kan vi ta $C(y) = \exp(x_0)$ och sedan följer (6.35) av (6.36) och (6.37). $\square$

Med stöd av sats 6.19 kan vi ta inversen av exponentialfunktionen. Vi kallar den inversa funktionen *(den naturliga) logaritmfunktionen* eller den *naturliga logaritmen* och betecknar den

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

Varje egenskap hos exponentialfunktionen kan skrivas om som en egenskap hos den naturliga logaritmen.

Sats 6.20 *Den naturliga logaritmen* $\ln\colon (0,\infty) \to \mathbf{R}$ *har följande egenskaper:*

(a) $x \mapsto \ln(x)$ *är en strängt växande funktion;*
(b) $\ln(1) = 0$ *och* $\ln(e) = 1$*;*
(c) $\ln(a) \leq a - 1$ *för alla* $a \in (0, \infty)$*;*
(d) $\ln(a) \geq (a-1)/a$ *för alla* $a \in (0, \infty)$*;*
(e) $\ln(ab) = \ln(a) + \ln(b)$ *för alla* $a, b \in (0, \infty)$*; och*
(f) $\ln(a^r) = r\ln(a)$ *för alla* $a \in (0, \infty)$ *och* $r \in \mathbf{Q}$.

Bevis. (a)

För godtyckliga $a < b$ vet vi att det finns x och y så att $\exp(x) = a$ och $\exp(y) = b$. För att visa att ln är strängt växande behöver vi visa att $x < y$. Men från sats 6.10 vet vi att $x \geq y \implies a \geq b$. Genom att ta kontrapositionen av detta får vi att $a < b \implies x < y$.

(b)

Eftersom $\exp(0) = 1$ och $\exp(1) = e$ så är $\ln(1) = 0$ och $\ln(e) = 1$.

(c)

Betrakta olikheten i sats 6.5(b) och sätt $\exp(x) = a$ (så $x = \ln a$). Då säger den att

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

som medför att $\ln(a) \leq a - 1$.

(d)

Betrakta olikheten i sats 6.5(d) och sätt $\exp(x) = a$ för $x < 1$, så $a < e$. Den säger att

$$a \leq \frac{1}{1 - \ln(a)}$$

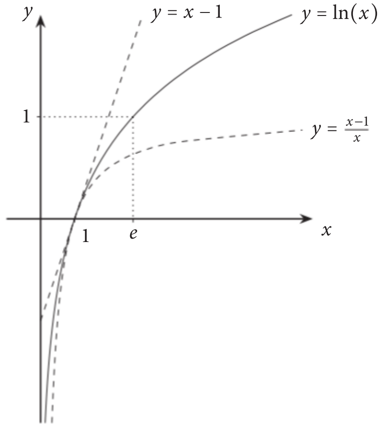
vilket medför att $(a-1)/a \leq \ln(a)$. Om $a \geq e$ kan man kontrollera direkt att olikheten är sann: Vi vet att $\ln(a) \geq \ln(e) = 1$ enligt del (a) och (b), och $(a-1)/a \leq 1$, så $\ln(a) \geq 1 \geq (a-1)/a$.

(e)

Betrakta likheten i sats 6.9 och sätt $\exp(x) = a$ och $\exp(y) = b$. Satsen säger att

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

och genom att ta logaritmfunktionen av båda leden medför det att $\ln(ab) = \ln(a) + \ln(b)$.



FIGUR 6.2 I sats 6.20 bevisar vi några egenskaper hos den naturliga logaritmfunktionen. Dess graf sitter mellan grafen av $x \mapsto x - 1$ och $x \mapsto (x-1)/x$. Vi vet även att $\ln$ är strängt växande, $\ln(1) = 0$ och $\ln(e) = 1$.

(f)

Enligt sats 6.9 lyder funktionen $f(r) := \exp(r\ln(a))$ ($a > 0$ och $r \in \mathbf{Q}$) regeln (6.17). Den uppfyller tydligen även (6.18). Därför, enligt sats 6.11, är

$$\exp(r\ln(a)) = a^r. \tag{6.38}$$

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

I figur 6.2 skissar vi grafen av den naturliga logaritmfunktionen.

IRRATIONELLA POTENSER

Vi kan använda exponentialfunktionen och den naturliga logaritmfunktionen för att skriva om en rationell potens av ett positivt tal: Observera att, medan högerledet av (6.38) är odefinierat om $r \in \mathbf{R} \setminus \mathbf{Q}$, är vänsterledet definierat för alla $r \in \mathbf{R}$! Det innebär att (6.38) ger oss ett smidigt sätt att utöka begreppet a^r till irrationella potenser r, vilket är konsekvent med vår tidigare definition. Alltså definierar vi (om)

$$a^x := \exp(x\ln(a)) \tag{6.39}$$

för $a > 0$ och $x \in \mathbf{R}$. Att vi har härlett (6.38) utifrån (4.23) säkerställer att definitionen (6.39) stämmer överens med definitionen (4.23) om $x \in \mathbf{Q}$.

Ännu en gång får vi upprepa sats 1.8. Först utvidgade vi den från heltalspotenser till rationella potenser i sats 4.14. Och nu utvidgar vi sats 4.14 från rationella potenser till reella (inklusive irrationella) potenser. Men med definitionen (6.39) är beviset mycket enklare än beviset av sats 4.14.

En direkt följd av (6.39) är att sats 6.20(f) gäller inte för bara $r \in \mathbf{Q}$ utan alla reella r. Det ser vi genom att ta logaritmen av (6.39).

Sats 6.21 *Med givna godtyckliga $a, b > 0$ och $x, y \in \mathbf{R}$ har vi att*

(a) $a^x a^y = a^{x+y}$,

(b) $a^x a^{-y} = a^{x-y}$,

(c) $a^x b^x = (ab)^x$, *och*

(d) $(a^x)^y = a^{xy}$.

Bevis. Vi beräknar att

$$\begin{aligned} a^x a^y &\underset{\substack{\uparrow \\ (6.39)}}{=} \exp(x\ln(a))\exp(y\ln(a)) \underset{\substack{\uparrow \\ \text{sats } 6.9}}{=} \exp(x\ln(a) + y\ln(a)) \\ &= \exp((x+y)\ln(a)) \underset{\substack{\uparrow \\ (6.39)}}{=} a^{x+y} \end{aligned}$$

så (a) är bevisat, sedan är

$$\begin{aligned} a^x a^{-y} &\underset{\substack{\uparrow \\ (6.39)}}{=} \exp(x\ln(a))\exp(-y\ln(a)) \underset{\substack{\uparrow \\ \text{sats } 6.9}}{=} \exp(x\ln(a) - y\ln(a)) \\ &= \exp((x-y)\ln(a)) \underset{\substack{\uparrow \\ (6.39)}}{=} a^{x-y} \end{aligned}$$

så (b) är bevisat, och

$$\begin{aligned} a^x b^x &\underset{\substack{\uparrow \\ (6.39)}}{=} \exp(x\ln(a))\exp(x\ln(b)) \underset{\substack{\uparrow \\ \text{sats } 6.9}}{=} \exp(x\ln(a) + x\ln(b)) \\ &= \exp(x(\ln(a) + \ln(b))) \underset{\substack{\uparrow \\ \text{sats } 6.20(\text{e})}}{=} \exp(x\ln(ab)) \underset{\substack{\uparrow \\ (6.39)}}{=} (ab)^x. \end{aligned}$$

så (c) är bevisat, och även

$$
\begin{aligned}
(a^x)^y &\underset{(6.39)}{=} \exp(y\ln(a^x)) \underset{(6.39)}{=} \exp(y\ln(\exp(x\ln(a)))) \\
&= \exp(y(x\ln(a))) = \exp(xy\ln(a)) \underset{(6.39)}{=} a^{xy}.
\end{aligned}
$$

så (d) är bevisat. □

6.3 Sammanfattning och uppgifter

I kapitel 6 har vi:

- funderat på ränta på ränta och hittat en följd vars gränsvärde kallas för den naturliga exponentialfunktionen;
- utrett egenskaper hos exponentialfunktionen med hjälp av de matematiska verktygen vi har byggt i tidigare kapitel;
- hittat en ny sats för gränsvärden – instängningssatsen – och använt den för att visa att exponentialfunktionen lyder samma regel som gäller för potenser;
- visat att den enda funktion som kan uppfylla en potensregel på rationella tal är $x \mapsto a^x$ så $\exp(x) = e^x$ för $x \in \mathbf{Q}$;
- hittat ännu en formel för den naturliga exponentialfunktionen som är en serie;
- visat att exponentialfunktionen är bijektiv och döpt inversen till den naturliga logaritmfunktionen;
- bevisat egenskaper hos den naturliga logaritmfunktionen och utvidgat begreppet potens till irrationella tal.

UPPGIFTER

6.1 Anta att $\exp(x) = 4\sqrt{3}$ och $\exp(y) = \sqrt{3}$. Räkna ut

(a) $\exp(x+y)$; (b) $\exp(2x+y)$; och (c) $\exp(x-3y)$.

6.2 Om $x \in \mathbf{R}$ uppfyller $\exp(x+4) = \exp(x) + \exp(4)$; räkna ut alla möjliga värden för $\exp(x)$.

6.3 Använd (6.7) och (6.8) samt beviset av sats 6.3 för att visa

$$\left(1+\frac{1}{n}\right)^n \le e \le \left(1-\frac{1}{2}\right)^{-2}$$

för alla $n \in \mathbf{Z}_+$ och i synnerhet $9/4 \le e \le 4$.

6.4 Bevisa genom lämpliga uppskattningar med enkla mängder att mängden

$$E = \{(x,y) \in \mathbf{R}^2 \mid 0 \le y < e^x \text{ och } 0 \le x < h\}$$

har arean $e^h - 1$.

6.5 Visa att cosh: $[0,\infty) \to [1,\infty)$, som är definierad enligt formeln

$$\cosh(x) = \frac{e^x + e^{-x}}{2} \quad \text{för } x \in [0,\infty),$$

är en bijektiv funktion. Räkna ut dess invers. Funktionen kallas för *cosinus hyperbolicus*.

6.6 Förenkla följande uttryck:

(a) $\exp(2\ln(\sqrt{x+1}) + \ln(x-1))$ för $x > 1$; och

(b) $3\ln(\exp(1+x)\exp(1-x))$ för $x \in \mathbf{R}$.

6.7 Låt $a > 0$ vara ett givet tal skilt från 1. I (6.39) definieras exponentialfunktionen $x \mapsto a^x$ med bas a för alla reella tal som en sammansättning av exponential- och logaritmfunktionen.

(a) Visa att $x \mapsto a^x$ med definitionsmängd $\mathbf{R}$ och målmängd $(0,\infty)$ är en bijektiv funktion utifrån exp:s och ln:s egenskaper.

(b) Inversen kallas för *logaritmfunktionen i basen a* och skrivs som $\log_a$. Visa att

$$\log_a(x) = K\ln(x)$$

för alla $x > 0$ och något $K \in \mathbf{R}$ som är oberoende av x. Ge en formel för K.

(c) Visa att

$$\log_a(xy) = \log_a(x) + \log_a(y).$$

för alla $x, y \in (0,\infty)$.

Komplexa tal

7.1 Algebraiska operationer

Vår resa inom matematisk analys har huvudsakligen varit en upptäcktsresa genom tal och geometri. Från en enkelt början i aritmetik och logik har vi skapat en axiomatisk grund för reella tal. Med de grundläggande egenskaperna hos reella tal som verktyg kunde vi upptäcka nya algebraiska begrepp såsom rötter och exponentialfunktioner och geometriska begrepp såsom area. I det här sista kapitlet utvidgar vi den reella linjen tillsammans med de aritmetiska operationerna man kan utföra på den till det så kallade komplexa planet. Det är förstås i och för sig själv spännande att det över huvud tagit är möjligt, men utöver detta kommer det att leda till en djupare förståelse av de begrepp vi har träffat på tidigare såsom polynom och trigonometri.

I bilaga C.4 motiverar vi utifrån aritmetik med reella tal hur aritmetik på det komplexa planet måste se ut. Men slutsatsen av diskussionen är ganska enkel: Ett *komplext tal* är ett tal z som skrivs som en summa

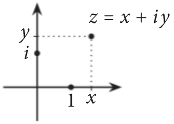
$$z = x + iy \tag{7.1}$$

där $x, y \in \mathbf{R}$. Tecknet i är ett tal som följer de vanliga aritmetiska reglerna tillsammans med regeln att

$$i^2 = -1. \tag{7.2}$$

Regeln säger att i är en rot till polynomet $p(x) = x^2 + 1$ – ett polynom som vi vet inte har några reella rötter. Därför är talet i inte ett reellt tal och kallas för ett *imaginärt tal*. Det betyder inte att i är mindre (eller mer) verkligt än andra tal – det är bara ett namn. Alla multiplar iy av i kallas för *imaginära tal*, så (7.1) säger att ett komplext tal är en summa av ett reellt och ett imaginärt tal. Mängden av alla komplexa tal noteras $\mathbf{C}$ och två komplexa tal $z = x + iy$ och $w = u + iv$ $(x, y, u, v \in \mathbf{R})$ sägs vara lika när $x = u$ och $y = v$. Det vill säga

$$z = w \iff x = u \text{ och } y = v. \tag{7.3}$$



FIGUR 7.1 Det komplexa planet där ett komplext tal $z = x + iy$ representeras som punkten (x,y). Talen 1 och i sitter på varsin axel som kallas för den reella respektive imaginära axeln.

Geometriskt representerar z i (7.1) punkten i planet med koordinaterna (x,y) (se figur 7.1). På så sätt kan vi identifiera **C** med ett plan – *det komplexa planet* – precis som vi identifierade **R** med en linje – den reella linjen.

ARITMETIK MED KOMPLEXA TAL

Att utföra addition, subtraktion och multiplikation med komplexa tal är lika enkelt (eller svårt) som med reella tal, eftersom räknereglerna är nästan identiska. Därför kan vi definiera heltals potenser av komplexa tal precis som vi definierade heltals potenser av reella tal, det vill säga att (1.15), (1.16) och sats 1.8 gäller även för $a, b \in \mathbf{C}$. Den additionella regeln (7.2) innebär att man kan förenkla bort alla potenser av i högre än 1 och även dem som är mindre än noll.

Att definiera rationella potenser av ett komplext tal kräver tyvärr en djupare analys. Se anmärkning 7.8. Å andra sidan kommer vi i avsnitt 7.2 lyckas att utvidga den exponentialfunktionen till komplexa tal.

Exempel 7.1 Med hjälp av (7.2) kan vi förenkla positiva potenser så här:

$$i^3 = i^2 i = -i, \quad i^4 = i^2 i^2 = (-1)(-1) = 1,$$
$$i^5 = i^2 i^2 i = (-1)(-1)i = i, \ldots$$

Och negativa potenser så här:

$$i^{-1} = \frac{1}{i} = \frac{i}{i^2} = \frac{i}{-1} = -i, \quad i^{-2} = \frac{1}{i^2} = \frac{1}{-1} = -1,$$
$$i^{-3} = \frac{1}{i^3} = \frac{i}{i^4} = \frac{i}{(-1)(-1)} = i, \ldots$$

Det innebär att man alltid kan förenkla uttryck som innehåller potenser av i så att bara $i^0 = 1$ och $i^1 = i$ finns kvar.

Exempel 7.2 Vi kan multiplicera två komplexa tal $2 + 3i$ och $7 - i$ så här:

$$\begin{aligned}(2+3i)(7-i) &= 2 \times 7 + 2 \times (-i) + (3i) \times 7 + (3i) \times (-i)\\ &= 14 - 2i + 21i - 3i^2\\ &\underset{\substack{\uparrow\\(7.2)}}{=} 14 - 2i + 21i - 3(-1) = 17 + 19i.\end{aligned}$$

På så sätt har vi lyckats skriva om produkten $(2+3i)(7-i)$ i formen (7.1).

Metoden från exempel 7.2 fungerar för att skriva vilken produkt av komplexa tal som helst i formen (7.1): För $z = x + iy$ och $w = u + iv$ är

$$\begin{aligned} wz &= (u+iv)(x+iy) = ux + iuy + ivx + i^2 vy \\ &\underset{(7.2)}{=} ux + iuy + ivx - vy = (ux - vy) + i(uy + vx). \end{aligned}$$

Addition och subtraktion är ännu enklare. Vi kan alltid skriva om en summa eller differens av två komplexa tal till formen (7.1):

$$\begin{aligned} w + z &= (u+iv) + (x+iy) = (u+x) + i(v+y); \\ w - z &= (u+iv) - (x+iy) = (u-x) + i(v-y). \end{aligned}$$

Exempel 7.3 Till exempel är

$$(2+3i) + (7-i) = (2+7) + (3-1)i = 9 + 2i$$

och

$$(2+3i) - (7-i) = (2-7) + (3-(-1))i = -5 + 4i.$$

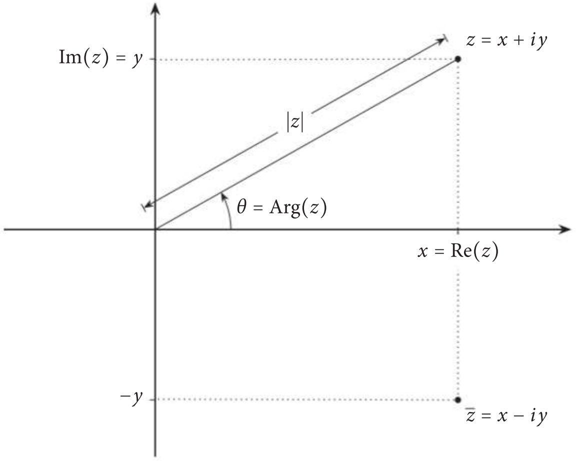
För $z = x + iy \in \mathbf{C}$ definierar vi *(komplex)konjugatet* till z att vara

$$\bar{z} = x - iy$$

och *absolutbeloppet*

$$|z| = \sqrt{x^2 + y^2}. \tag{7.4}$$

Konjugatet $\bar{z}$ är därför en spegling av z i reella axeln (x-axeln) och absolutbeloppet $|z|$ är avståndet mellan z och origo (se figur 7.2). Absolutbeloppet definierat i (7.4) är en utvidgning från reella tal till komplexa tal av begreppet definierat i (2.20) med samma namn. *Realdelen* av z är $\mathrm{Re}(z) := x$ och $\mathrm{Im}(z) := y$ kallas för *imaginärdelen* av z. Så z kallas för imaginärt om $\mathrm{Re}(z) = 0$ och reellt om $\mathrm{Im}(z) = 0$.



FIGUR 7.2 Real- och imaginärdelen av ett komplext tal samt argumentet och absolutbeloppet. Konjugatet $\overline{z}$ är en spegling i reella axeln.

Begreppet konjugat kan användas för att skriva om en kvot av komplexa tal i formen (7.1). Observera att

$$z\overline{z} = (x + iy)(x - iy) = x^2 - i^2 y^2 \underset{(7.2)}{=} x^2 + y^2 \underset{(7.4)}{=} |z|^2 \tag{7.5}$$

för $z = x + iy$ ($x, y \in \mathbf{R}$). Vi kan därför beräkna att

$$\frac{1}{z} = \frac{\overline{z}}{z\overline{z}} \underset{(7.5)}{=} \frac{\overline{z}}{|z|^2} = \frac{x - iy}{x^2 + y^2} = \frac{x}{x^2 + y^2} - i\frac{y}{x^2 + y^2}$$

för $z \neq 0$ och för $w = u + iv$ får vi även beräkna att

$$\begin{aligned}\frac{w}{z} &= \frac{u + iv}{x + iy} = \frac{(u + iv)(x - iy)}{(x + iy)(x - iy)} = \frac{(ux + vy) + i(vx - uy)}{x^2 + y^2} \\ &= \frac{ux + vy}{x^2 + y^2} + i\frac{(vx - uy)}{x^2 + y^2}.\end{aligned}$$

Exempel 7.4 Vi kan beräkna

$$\frac{6 - 2i}{12 + 35i} = \frac{(6 - 2i)(12 - 35i)}{(12 + 35i)(12 - 35i)} = \frac{2 + 234i}{1369} = \frac{2}{1369} + \frac{234}{1369}i.$$

Ett *argument* av z definieras som en vinkel θ mellan den positiva reella axeln och strålen från origo genom punkten z. Vi skriver $\theta = \arg(z)$. I fallet $z = 0$ är argumentet av z odefinierat. För $z \neq 0$ är argumentet $\arg(z)$ inte entydigt bestämt av z utan definieras bara upptill en additiv konstant $2k\pi$ där $k \in \mathbf{Z}$. Argumentet θ av z som ligger i intervallet $(-\pi, \pi]$ kallas för *principalargumentet* och skrivs $\operatorname{Arg}(z)$. Det ges av formeln

$$\operatorname{Arg}(z) = \begin{cases} -\pi + \arctan\left(\frac{y}{x}\right) & \text{om } x < 0 \text{ och } y < 0; \\ -\dfrac{\pi}{2} & \text{om } x = 0 \text{ och } y < 0; \\ \arctan\left(\frac{y}{x}\right) & \text{om } x > 0; \\ \dfrac{\pi}{2} & \text{om } x = 0 \text{ och } y > 0; \\ \pi + \arctan\left(\frac{y}{x}\right) & \text{om } x < 0 \text{ och } y \geq 0; \\ \text{odefinierad} & \text{om } x = 0 \text{ och } y = 0. \end{cases} \tag{7.6}$$

Exempel 7.5 För $z = \sqrt{3} - 3i$ är $\operatorname{Re}(z) := \sqrt{3}$ och $\operatorname{Im}(z) := -3$. Vi kan även beräkna från (7.4) att

$$|z| = \sqrt{\left(\sqrt{3}\right)^2 + (-3)^2} = \sqrt{3+9} = \sqrt{12} = 2\sqrt{3}$$

och eftersom $\operatorname{Re}(z) > 0$ säger (7.6) att

$$\operatorname{Arg}(z) = \arctan\left(\frac{-3}{\sqrt{3}}\right) = \arctan\left(-\sqrt{3}\right) = -\frac{\pi}{3}.$$

Vi har redan sett att real- och imaginärdelar av ett komplext tal ger talets kartesiska koordinator i planet. Men även absolutbeloppet och argumentet av ett komplext tal identifierar punkten i planet där talet ligger. De är helt enkelt koordinaterna till talet i ett annat koordinatsystem: Systemet kallas för *polära koordinatsystemet* och systemet identifierar en punkt i planet genom att specificera punktens avstånd från origo och vinkeln den bildar med x-axeln. Från trigonometri har vi relationerna

$$\operatorname{Re}(z) = |z|\cos(\arg(z)) \quad \text{och} \quad \operatorname{Im}(z) = |z|\sin(\arg(z)). \tag{7.7}$$

Det är en nyttig övning att bevisa följande sats och vi lämnar det som en övning för läsaren – se också uppgift 7.2.

Sats 7.6 *För alla komplexa tal z och w har vi att*

(a) $\overline{zw} = \bar{z}\,\overline{w}$,
(b) $\overline{z+w} = \bar{z} + \overline{w}$,
(c) $|zw| = |z||w|$,
(d) $|z+w| \leq |z| + |w|$,
(e) $\arg(zw) = \arg(z) + \arg(w) + 2k\pi$ *för* $k \in \mathbf{Z}$ *och*
(f) $\arg(z/w) = \arg(z) - \arg(w) + 2k\pi$ *för* $k \in \mathbf{Z}$.

Sats 7.6(d) säger att triangelolikheten (sats 2.30) gäller även för komplexa tal.

ANDRAGRADSPOLYNOM OCH KOMPLEXA TAL

Som sagt följer alla beräkningar som gäller addition, multiplikation, subtraktion och division samma regler som gäller för reella tal. I exempel 7.1 har vi även beräknat med heltalspotenser för komplexa tal som vi definierar på samma sätt som för reella tal.

Men kom ihåg att kvadratrötter är någonting utöver de fyra aritmetiska operationerna så vi måste återvända till frågan i samband med komplexa tal. Men för att komma igång erinrar vi oss vad vi lärde oss i kontexten om reella tal: För varje positivt reellt tal y har ekvationen

$$x^2 = y \tag{7.8}$$

två reella lösningar, en som är positiv $x = \sqrt{y}$ och en som är negativ $x = -\sqrt{y}$. I fallet $y = 0$ har (7.8) en lösning, det vill säga $x = \sqrt{0} = 0$, och i fallet $y < 0$ har (7.8) ingen reell lösning alls. Nu kommer vi att betrakta samma problem fast vi byter ut de reella talen x och y mot de komplexa talen z respektive w: För givet $w \in \mathbf{C}$ vill vi lösa

$$z^2 = w. \tag{7.9}$$

Man kan skriva $z = x + iy$ och $w = u + iv$ för $x, y, u, v \in \mathbf{R}$ och då kommer (7.9) att vara

$$x^2 - y^2 + 2xyi = (x + iy)^2 = u + iv.$$

Därför att enligt (7.3) är real- respektive imaginärdelar lika:

$$x^2 - y^2 = u; \quad \text{och} \tag{7.10}$$

$$2xy = v. \tag{7.11}$$

Man kan förstås lösa de två ekvationerna för x och y men det är lite smidigare att också ta absolutbeloppet av (7.9). Därav får man likheten

$$x^2 + y^2 = \sqrt{u^2 + v^2} \tag{7.12}$$

så om vi tar summan och skillnaden av (7.10) och (7.12) får vi 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 att det finns totalt fyra möjliga lösningar $z = x + iy$ till (7.9), med två val för x och två för y. Men tänk på vad (7.11) 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; Om $v = 0$ är minst en av x och y lika med noll. Det medför att vi har (högst) två möjliga lösningar till (7.9):

Man kan kontrollera att (7.13) ger oss två *olika* kandidatlösningar z till (7.9) för varje $w \neq 0$. Om $w = 0$ visar (7.13) att det inte kan finnas flera lösningar till (7.9) än $z = 0$.

$$z = \begin{cases} \pm\left(\sqrt{\frac{u+\sqrt{u^2+v^2}}{2}} - i\sqrt{\frac{\sqrt{u^2+v^2}-u}{2}}\right) & \text{om } v < 0; \\ \pm\left(\sqrt{\frac{u+\sqrt{u^2+v^2}}{2}} + i\sqrt{\frac{\sqrt{u^2+v^2}-u}{2}}\right) & \text{om } v \geq 0. \end{cases} \tag{7.13}$$

Det kvarstår att verifiera att de två kandidatlösningarna ovan faktiskt ger lösningar till (7.9). Det enklaste sättet att göra detta är att verifiera att (7.13) är lösningar till den ekvivalenta formuleringen (7.10) och (7.11) och vi lämnar det som en övning för läsaren.

Att kunna lösa (7.9) med (7.13) är nyckeln för att hitta komplexa lösningar till andragradspolynom, även dem med komplexa koefficienter.

Exempel 7.7 Lös

$$z^2 - 2iz - 9 - 6i = 0$$

för $z \in \mathbf{C}$.

Lösning. Vi börjar med att kvadratkomplettera:

$$z^2 - 2iz - 9 - 6i = (z - i)^2 - 8 - 6i$$

så vi behöver lösa

$$(z - i)^2 = 8 + 6i$$

för z. Det kan vi jämföra med (7.9) och därför vet vi från (7.13) med $u = 8$ och $v = 6$ att

$$z - i = \pm\left(\sqrt{\frac{8 + \sqrt{8^2 + 6^2}}{2}} + i\sqrt{\frac{\sqrt{8^2 + 6^2} - 8}{2}}\right) = \pm(3 + i)$$

så ekvationen har två lösningar: $z = 3 + 2i$ och $z = -3$. □

Anmärkning 7.8 Vi avstår från att utvidga kvadratrötter till mer än icke-negativa reella tal. Precis som (7.8) har två lösningar för positiva y har vi sett att (7.9) har två lösningar för icke-noll w. Men till skillnad från (7.8) där vi i (4.16) valde den icke-negativa lösningen finns det inget naturligt sätt att välja mellan de två lösningarna i (7.13). Det behöver man på något sätt göra så att f_2 med komplexa definitions- och målmängder blir en bijektiv funktion (kom ihåg f_2 från (4.5)). Man kan förstås göra något val – genom att välja antingen + eller − i ± i (7.13) – men man kan inte göra detta på så sätt att alla räkneregler vi hittills har bevisat bevaras. Till exempel ser vi direkt att $z = i$ och $z = -i$ är lösningar till (7.9) när $w = -1$. Men om vi sa att $\sqrt{-1}$ var i, kan sats 4.14(c) inte gälla eftersom den leder till motsägelsen att

$$1 = \sqrt{1} = \sqrt{(-1)(-1)} = \sqrt{-1}\sqrt{-1} = i \times i = -1.$$

Samma problem uppstår om vi i stället tar $\sqrt{-1}$ lika med $-i$.

Som följd avstår vi från att ge en formel för komplexa nollställen till andragradspolynom såsom det i exempel 7.7. Om man var särskilt envis kunde man säga att $\pm\sqrt{(b^2 - 4ac)/(4a^2)}$ kan tolkas som båda lösningar z till $z^2 = (b^2 - 4ac)/(4a^2)$ och att formeln (4.21) gäller även för komplexa a, b och c. Vi behöver till exempel inte sats 4.14 för att härleda den och uträkningarna gäller även för komplexa a, b och c. Men i så fall döljer formeln arbetet som ligger bakom (7.13) och förser oss inte med en färdig lösning.

7.2 Funktioner och geometri

Samspelet mellan algebra och geometri i komplex analys innebär att det finns mycket struktur att utreda. I det här avsnittet får vi ett smakprov på vad som finns.

KOMPLEXA EXPONENTIALFUNKTIONEN

Vi börjar genom att utreda om vi kan utvidga exponentialfunktionen exp till en funktion på **C**. Den var i (6.6) definierad som ett gränsvärde så det känns rimligt att definiera utvidgningen exp: $\mathbf{C} \to \mathbf{C}$ på samma sätt:

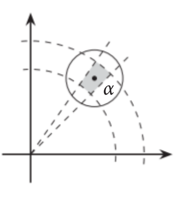
$$\exp(z) := \lim_{n\to\infty} \left(1 + \frac{z}{n}\right)^n \tag{7.14}$$

för $z \in \mathbf{C}$. Men definitionen kräver att vi först svarar på ett par frågor: Vad menar vi med gränsvärdet av en följd av komplexa tal? Och sedan: Existerar gränsvärdet i (7.14)?

För att svara på första frågan tittar vi på definition 2.31. Den enda modifikation som krävs för att betrakta $\alpha \in \mathbf{C}$ i stället för $\alpha \in \mathbf{R}$ är att absolutbeloppet tolkas som (7.4) i stället för (2.20). Med den nya tolkningen är definition 2.31 ett utmärkt val för gränsvärdet, även av komplexa följder. Eftersom triangelolikheten gäller även för komplexa tal (se sats 7.6(d)) kan man kontrollera att beviset av sats 2.35 gäller även för komplexa gränsvärden, så vi har precis samma räkneregler som för reella gränsvärden.

För att svara på andra frågan kan vi inte längre lita på sats 2.33 eftersom följden $(1 + \frac{z}{n})^n$ i allmänhet är en följd av komplexa tal – det finns inga olikheter mellan komplexa tal och därför kan vi inte längre prata om växande följder i det här kontexten. Vi kan i stället räkna ut ett gränsvärde direkt med stöd från vårt arbete i kapitel 5. Det gör vi först i det speciella fallet där z är imaginärt, det vill säga $z = i\theta$ och $\theta \in \mathbf{R}$.

Definition 2.31 säger att $\lim_{n\to\infty} \alpha_n = \alpha$ om man för varje $\varepsilon > 0$ kan hitta $N \in \mathbf{Z}_+$ sådant att $|\alpha_m - \alpha| < \varepsilon$ för alla $m \geq N$. En jämförelse med (3.22) och (7.4) visar oss att det säger att alla α_m $(m \geq N)$ ligger inom en cirkel i det komplexa planet med medelpunkten α med radien ε. Figur 7.3 visar oss att för att hamna i cirkelns inre räcker det att absolutbeloppet och argumentet av α_n konvergerar mot $|\alpha|$ respektive $\arg(\alpha)$ då $n \to \infty$. Därför kan vi visa att en följd $(\alpha_n)_n$ av komplexa tal konvergerar genom att visa att två följder $(|\alpha_n|)_n$ och $(\arg(\alpha_n))_n$ av reella tal konvergerar.



FIGUR 7.3 En punkt i komplexa planet måste ligga inom en cirkel med medelpunkt α om dess absolutbelopp och argument inte skiljer sig mycket från α:s.

Vi sätter därför

$$\alpha_n := \left(1 + \frac{i\theta}{n}\right)^n \tag{7.15}$$

för $n \in \mathbf{Z}_+$. Först beräknar vi att

$$\begin{aligned} |\alpha_n| &= \left|\left(1 + \frac{i\theta}{n}\right)^n\right| \underset{\substack{\uparrow \\ \text{sats 7.6(c)}}}{=} \left|1 + \frac{i\theta}{n}\right|^n \underset{\substack{\uparrow \\ (7.4)}}{=} \left(1 + \frac{\theta^2}{n^2}\right)^{n/2} \\ &\underset{\substack{\uparrow \\ (6.39)}}{=} \exp\left(\frac{n}{2}\ln\left(1 + \frac{\theta^2}{n^2}\right)\right). \end{aligned}$$

Sedan kan vi uppskatta den underifrån:

$$\exp\left(\frac{n}{2}\ln\left(1 + \frac{\theta^2}{n^2}\right)\right) \underset{\substack{\uparrow \\ \text{sats 6.5(b)}}}{\geq} 1 + \frac{n}{2}\ln\left(1 + \frac{\theta^2}{n^2}\right) \underset{\substack{\uparrow \\ \text{sats 6.20(d)}}}{\geq} 1 + \frac{n}{2}\frac{\frac{\theta^2}{n^2}}{1 + \frac{\theta^2}{n^2}} \geq 1.$$

Och även överifrån: Sats 6.20(c) säger att $\frac{n}{2}\ln\left(1 + \frac{\theta^2}{n^2}\right) \leq \frac{\theta^2}{2n}$ så

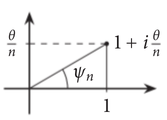
$$\exp\left(\frac{n}{2}\ln\left(1 + \frac{\theta^2}{n^2}\right)\right) \underset{\substack{\uparrow \\ \text{sats 6.5(d)}}}{\leq} \frac{1}{1 - \frac{n}{2}\ln\left(1 + \frac{\theta^2}{n^2}\right)} \underset{\substack{\uparrow \\ \text{sats 6.20(c)}}}{\leq} \frac{1}{1 - \frac{\theta^2}{2n}}$$

för $n > \theta^2/2$. Därför är

$$1 \leq |\alpha_n| \leq \frac{1}{1 - \frac{\theta^2}{2n}}$$

för $n > \theta^2/2$ och det följer från instängningssatsen (sats 6.6), exempel 2.32 och sats 2.35 att

$$\lim_{n\to\infty} |\alpha_n| = 1. \tag{7.16}$$



FIGUR 7.4 Det komplexa talet $1 + i\theta/n$ har absolutbeloppet $\sqrt{1 + \theta^2/n^2}$ och argumentet $\psi_n := \arctan(\theta/n)$.

Nu uppskattar vi argumentet av (7.15). Vi har att

$$\arg(z_n) = \arg\left(\left(1 + \frac{i\theta}{n}\right)^n\right) \underset{\substack{\uparrow \\ \text{sats 7.6(e)}}}{=} n\arg\left(1 + i\frac{\theta}{n}\right) = n\psi_n \tag{7.17}$$

där vinkeln ψ_n visas i figur 7.4. Om $\theta > 0$ är $\psi_n \in (0,\pi/2)$ så medför sats 5.1 och figur 7.4 att

$$\frac{\frac{\theta}{n}}{\sqrt{1+\theta^2/n^2}} = \sin\psi_n < \psi_n < \tan\psi_n = \frac{\theta}{n}$$

så om vi multiplicerar med n får vi med hjälp av (7.17) att

$$\frac{\theta}{\sqrt{1+\theta^2/n^2}} < \arg(z_n) < \theta. \tag{7.18}$$

Vi kan vidare uppskatta

$$\begin{aligned}\sqrt{1+\theta^2/n^2} &\underset{\substack{\uparrow\\(6.39)}}{=} \exp\left(\frac{1}{2}\ln\left(1+\frac{\theta^2}{n^2}\right)\right) \underset{\substack{\uparrow\\\text{sats } 6.5(\text{b})}}{\geq} 1+\frac{1}{2}\ln\left(1+\frac{\theta^2}{n^2}\right)\\ &\underset{\substack{\uparrow\\\text{sats } 6.20(\text{d})}}{\geq} 1+\frac{1}{2}\frac{\frac{\theta^2}{n^2}}{\left(1+\frac{\theta^2}{n^2}\right)}\end{aligned}$$

så om vi sätter det i (7.18) får vi att

$$\frac{\theta\left(1+\frac{\theta^2}{n^2}\right)}{1+\frac{3\theta^2}{2n^2}} < \arg(z_n) < \theta.$$

Det följer från instängningssatsen (sats 6.6), exempel 2.32 och sats 2.35 att

$$\lim_{n\to\infty}\arg(z_n) = \theta \tag{7.19}$$

om $\theta > 0$. Ett liknande argument gäller för $\theta < 0$: Då måste man använda sats 5.1 tillsammans med (5.4) och (5.5) för att visa $\tan\psi < \psi < \sin\psi$ för $\psi \in (-\pi/2,0)$ och sedan ger samma argument som ovan (7.19) för $\theta < 0$. Vi behöver inte visa någonting när $\theta = 0$ för då är $\alpha_n = 1$ för alla $n \in \mathbf{Z}_+$ och (som vi redan vet) konvergerar α_n mot $\exp(0) = 1$.

Från (7.16) och (7.19) vet vi att α_n i (7.15) konvergerar mot en punkt på enhetscirkeln i det komplexa planet med medelpunkt i origo, med argument θ. Det betyder att

$$\exp(i\theta) = \cos\theta + i\sin\theta \tag{7.20}$$

för $\theta \in \mathbf{R}$. I (7.20) ser vi en spännande och oväntad koppling mellan exponentialfunktionen och trigonometriska funktioner!

För att bevisa att gränsvärdet (7.14) existerar för andra komplexa tal behöver vi först observera att den grundläggande räkneregeln (6.17) för exponentialfunktionen gäller även för komplexa argument. Mer precist sagt kan vi visa följande variant av sats 6.9. Satsen ser lite annorlunda ut eftersom vi inte vet i förväg att gränsvärdet (7.14) existerar för alla $z \in \mathbf{C}$.

Sats 7.9 *Om gränsvärdet* (7.14) *existerar för* $z = z_1$ *och* $z = z_2$ *där* $z_1, z_2 \in \mathbf{C}$ *då existerar* (7.14) *även för* $z = z_1 + z_2$ *och*

$$\exp(z_1 + z_2) = \exp(z_1)\exp(z_2).$$

Bevis. Vi kan använda det komplexa absolutbeloppet (7.4) i stället för (2.20) och den komplexa varianten av triangelolikheten (sats 7.6(d)) för att upprepa beviset av (6.15) med z_1 och z_2 i stället för x och y. Därför gäller även (6.14) med z_1 och z_2 i stället för x och y.

Man kan kontrollera att binomialsatsen (sats 2.13) också gäller för komplexa tal och därför får vi även (6.13) med z_1 och z_2 i stället för x och y. Eftersom vänsterledet av (6.13) konvergerar, (6.14) och den komplexa varianten av sats 2.35 säger att gränsvärdet $\exp(z_1 + z_2)$ existerar och $\exp(z_1 + z_2) = \exp(z_1)\exp(z_2)$. □

Nu kan vi visa att (7.14) är väldefinierad för alla $z \in \mathbf{C}$.

Sats 7.10 *För alla* $z \in \mathbf{C}$ *existerar gränsvärdet* (7.14) *och*

$$\exp(z) = e^x(\cos y + i \sin y)$$

för $z = x + iy$ *(*$x, y \in \mathbf{R}$*).*

Bevis. Vi vet att gränsvärdet (7.14) existerar i fallen $z = x$ för $x \in \mathbf{R}$ och $z = iy$ för $y \in \mathbf{R}$. Därför enligt sats 7.9 existerar även (7.14) om $z = x + iy$ och

$$\exp(z) = \exp(x + iy) = \exp(x)\exp(iy) \underset{\substack{\uparrow \\ (6.39)\text{ med } a = e}}{=} e^x \exp(iy)$$

$$\underset{\substack{\uparrow \\ (7.20)}}{=} e^x(\cos y + i \sin y)$$ □

Analogt med det reella fallet använder vi även notationen $e^z := \exp(z)$ för $z \in \mathbf{C}$. En vanligt sätt att skriva ett komplext tal är i så kallad *polär form*: Enligt (7.20) är $e^{i\theta} = \cos\theta + i\sin\theta$ så från (7.7) får vi skriva

$$z = \mathrm{Re}(z) + i\,\mathrm{Im}(z) = r\cos\theta + ir\sin\theta = re^{i\theta}$$

där $r = |z|$ och $\theta = \arg(z)$ för $z \neq 0$. I fallet $z = 0$ är $\arg(z)$ odefinierad, men eftersom $r = |z| = 0$ kan vi ändå skriva $z = re^{i\theta}$ för vilket θ som helst. Observera att även om argumentet av ett komplext tal inte är entydigt bestämt är $z = re^{i\theta}$ ett och samma tal oavsett vilket av z:s argument man tar för θ: Eftersom två val θ_1 och θ_2 av argument skiljer sig med en heltalsmultipel av 2π och $\theta \mapsto e^{i\theta}$ är en periodisk funktion med perioden 2π, är $e^{i\theta_1} = e^{i\theta_2}$.

Med de här definitionerna är följande sats enkel att bevisa.

Sats 7.11 *För alla $\theta \in \mathbf{R}$ är*

(a) $e^{i\pi} + 1 = 0$ *(Eulers identitet),*

(b) $(\cos\theta + i\sin\theta)^n = \cos(n\theta) + i\sin(n\theta)$ *för $n \in \mathbf{Z}$ (de Moivres formel),*

(c) $\cos\theta = (e^{i\theta} + e^{-i\theta})/2$*, och*

(d) $\sin\theta = (e^{i\theta} - e^{-i\theta})/2i$*.*

Bevis. Vi bevisar de Moivres formel och lämnar de andra delarna som en övning för läsaren – se uppgift 7.4. Kom ihåg att definitionerna (1.15) och (1.16) gäller även för komplexa a.

Först antar vi att $n > 0$. Vi kan tillämpa sats 7.9 $n - 1$ gånger för att se

$$\exp(ni\theta) = (\exp(i\theta))^n$$

och med hjälp av (7.20) är

$$\begin{aligned} \exp(ni\theta) &= \cos(n\theta) + i\sin(n\theta) \quad \text{och} \\ (\exp(i\theta))^n &= (\cos\theta + i\sin\theta)^n. \end{aligned}$$

Vi kan lägga ihop de tre likheterna för att bevisa (b) när $n > 0$.

Nu antar vi att $n < 0$, sätter $n = -m$ och beräknar med hjälp av (b) för positiva heltal att

$$
\begin{aligned}
&(\cos\theta + i\sin\theta)^n \\
&\underset{\substack{\uparrow\\(1.16)}}{=} \frac{1}{(\cos\theta + i\sin\theta)^m} = \frac{1}{\cos(m\theta) + i\sin(m\theta)} \\
&= \frac{\cos(m\theta) - i\sin(m\theta)}{(\cos(m\theta) + i\sin(m\theta))(\cos(m\theta) - i\sin(m\theta))} \\
&= \frac{\cos(m\theta) - i\sin(m\theta)}{\cos^2(m\theta) + \sin^2(m\theta)} \underset{\substack{\uparrow\\(5.3)}}{=} \cos(m\theta) - i\sin(m\theta) \\
&\underset{\substack{\uparrow\\(5.4)\ \&\ (5.5)}}{=} \cos(n\theta) + i\sin(n\theta)
\end{aligned}
$$

så (b) är bevisad när $n < 0$.

I fallet $n = 0$ har vi inte så mycket att bevisa: Vi kan beräkna direkt från (1.16) att $(\cos\theta + i\sin\theta)^0 = 1$ och vi vet även att $\cos(0) + i\sin(0) = 1$.□

Eulers identitet (sats 7.11(a)) är särdeles vacker eftersom den innehåller tre operationer som var väsentliga för vår resa – addition, multiplikation och exponentiering – samt fem tal med mycket olika ursprung – det additionsneutrala elementet 0, det multiplikationsneutrala elementet 1, hälften π av en enhetscirkels omkrets, den naturliga basen e och den imaginära enheten i.

Vi kan vidare utnyttja kopplingen (7.20) mellan exponentialfunktionen och trigonometriska funktioner för att härleda en serierepresentation för cosinus och sinus. Först noterar vi att sats 6.15 gäller även för komplexa argument – det kan vi se tack vare samma bevis med absolutbeloppet tolkat som (7.4) i stället för (2.20).

Sats 7.12 *För alla $\theta \neq 0$ är*

$$
\cos\theta = \sum_{k=0}^{\infty} \frac{(-1)^k \theta^{2k}}{(2k)!} = 1 - \frac{\theta^2}{2!} + \frac{\theta^4}{4!} - \cdots \tag{7.21}
$$

och

$$
\sin\theta = \sum_{k=0}^{\infty} \frac{(-1)^k \theta^{2k+1}}{(2k+1)!} = \theta - \frac{\theta^3}{3!} + \frac{\theta^5}{5!} - \cdots . \tag{7.22}
$$

Bevis. För att bevisa (7.21) använder vi sats 7.11(c) och den komplexa varianten av sats 6.15:

$$\cos\theta = \frac{e^{i\theta} + e^{-i\theta}}{2} = \frac{1}{2}\left(\sum_{k=0}^{\infty}\frac{(i\theta)^k}{k!} + \sum_{k=0}^{\infty}\frac{(-i\theta)^k}{k!}\right).$$

Eftersom en serie bara är gränsvärdet av sina delsummor kan vi använda den komplexa varianten av sats 2.35 för att skriva om de två serierna som en:

$$\sum_{k=0}^{\infty}\frac{(i\theta)^k}{k!} + \sum_{k=0}^{\infty}\frac{(-i\theta)^k}{k!} = \sum_{k=0}^{\infty}\frac{\left((i\theta)^k + (-i\theta)^k\right)}{k!}.$$

Nu tittar vi på termerna i sista serien: Om k är udda är $(i\theta)^k + (-i\theta)^k = (i\theta)^k - (i\theta)^k = 0$. Om $k = 2\ell$ för $\ell \in \mathbf{N}$ är $(i\theta)^k + (-i\theta)^k = 2i^{2\ell}(\theta)^{2\ell} = 2(-1)^{\ell}(\theta)^{2\ell}$ eftersom $i^2 = -1$. Därför är

$$\sum_{k=0}^{\infty}\frac{\left((i\theta)^k + (-i\theta)^k\right)}{k!} = 2\sum_{\ell=0}^{\infty}\frac{(-1)^{\ell}\theta^{2\ell}}{(2\ell)!}$$

och (7.21) är bevisad.

Beviset av (7.22) är likadant fast man använder del (d) av sats 7.11 i stället för (c). □

KOMPLEX GEOMETRI

Genom att utnyttja sambanden mellan geometri i planet och komplexa tal kan vi omformulera vissa problem av geometrisk natur till problem i komplex aritmetik.

Exempel 7.13 Skriv om $2\arctan(4/7) + \arctan(1/4)$ som ett uttryck som innehåller en arcusfunktion.

Lösning. Vi kan formulera om problemet till ett problem om komplexa tal. Vi vet att $\operatorname{Arg}(7 + 4i) = \arctan(4/7)$ och $\operatorname{Arg}(4 + i) = \arctan(1/4)$. Därför enligt sats 7.6(e) är

$$\begin{aligned}
&2\arctan(4/7) + \arctan(1/4)\\
&= 2\operatorname{Arg}(7 + 4i) + \operatorname{Arg}(4 + i) = \operatorname{Arg}\left((7 + 4i)^2(4 + i)\right)\\
&= \operatorname{Arg}(76 + 257i) = \arctan(257/76).
\end{aligned}$$

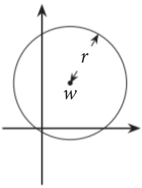
□

Eftersom både $\operatorname{Arg}(7 + 4i)$ och $\operatorname{Arg}(4 + i)$ är mellan 0 och $\pi/4$ måste $2\operatorname{Arg}(7+4i) + \operatorname{Arg}(4+i)$ vara principalargumentet av $(7 + 4i)^2(4 + i)$.

Vi kan skriva ekvationer vars lösningsmängder ger geometriska former som vi har sett i avsnitt 3.2.

Till exempel har vi redan definierat en cirkel som mängden av alla punkter (x,y) i planet som ligger på ett visst avstånd $r > 0$ från cirkelns medelpunkt (a,b). Om vi identifierar punkterna (a,b) och (x,y) med de komplexa talen $w = a + ib$ respektive $z = x + iy$, uppfyller en punkt $z \in \mathbf{C}$ ekvationen

$$|z - w| = r \tag{7.23}$$



FIGUR 7.5 En cirkel i det komplexa planet med medelpunkten w och radien r kan beskrivas med ekvationen $|z - w| = r$.

precis när z ligger på cirkeln (se figur 7.5). Vi kan enkelt kontrollera att ekvationen är ekvivalent med (3.22) eftersom

$$|z - w| = \sqrt{(x-a)^2 + (y-b)^2}$$

och r var positivt.

En linje kan beskrivas som alla punkter (x,y) som ligger lika långt från två givna punkter (u_1,v_1) och (u_2,v_2) i planet (se figur 7.6(a)). Om vi identifierar (x,y), (u_1,v_1) och (u_2,v_2) med $z = x + iy$, $w_1 = u_1 + iv_1$ respektive $w_2 = u_2 + iv_2$ kan vi omformulera den föregående meningen som påståendet att mängden av alla lösningar till ekvationen

$$|z - w_1| = |z - w_2| \tag{7.24}$$

är en linje. Man kan verifiera att ekvation (7.24) har samma lösningar som (3.14) så länge $v_1 \neq v_2$: Vi beräknar

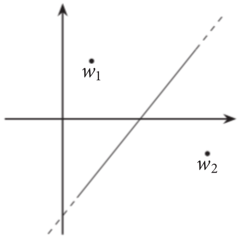
$$\begin{aligned}
&|z - w_1| = |z - w_2| \\
&\iff |z - w_1|^2 = |z - w_2|^2 \\
&\iff (x - u_1)^2 + (y - v_1)^2 = (x - u_2)^2 + (y - v_2)^2 \\
&\iff 2(u_2 - u_1)x + 2(v_2 - v_1)y = u_2^2 - u_1^2 + v_2^2 - v_1^2 \\
&\iff y = -\frac{(u_2 - u_1)}{(v_2 - v_1)}x + \frac{u_2^2 - u_1^2 + v_2^2 - v_1^2}{2(v_2 - v_1)}.
\end{aligned}$$

Så i det här fallet är linjens lutning

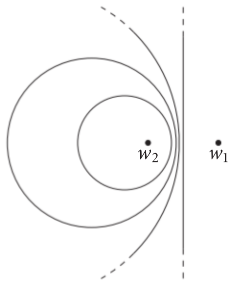
$$k = -\frac{(u_2 - u_1)}{(v_2 - v_1)}$$

och den skär y-axeln i

$$c = \frac{u_2^2 - u_1^2 + v_2^2 - v_1^2}{2(v_2 - v_1)}.$$



(a) En linje i komplexa planet är alla punkter som ligger lika långt från två komplexa tal $w_1 = u_1 + iv_1$ och $w_2 = u_2 + iv_2$.



(b) Lösningar till (7.25) med $w_1 = 1$, $w_2 = 0$ och $\alpha = 2, 3/2, 5/4$ respektive 1. Den raka linjen ($\alpha = 1$) kan ses som något slags gränsvärde av cirklarna då $\alpha \to 1$.

FIGUR 7.6 Cirklar och linjer kan beskrivas med hjälp av det relativa avståndet från två referenspunkter w_1 och w_2.

I fallet $v_1 = v_2$ beskriver (7.24) en lodrät linje eftersom i så fall är

$$\begin{aligned} &|z - w_1| = |z - w_2| \\ &\iff 2(u_2 - u_1)x = u_2^2 - u_1^2 \\ &\iff x = \frac{u_2^2 - u_1^2}{2(u_2 - u_1)} = \frac{u_2 + u_1}{2} \end{aligned}$$

så länge w_1 och w_2 inte är samma punkt (så $u_1 \neq u_2$).

Genom komplex geometri ser vi även ett närmare samband mellan cirklar och linjer. Vi kan nämligen se ekvationen (7.24) som ett särskilt fall av ekvationen

$$|z - w_1| = \alpha|z - w_2| \tag{7.25}$$

som innehåller en parameter $\alpha > 0$. Det är inte konstigt att (7.24) (eller (7.25) med $\alpha = 1$) är en ekvation för en linje, men mindre intuitivt är det att (7.25) är en ekvation för en cirkel i det komplexa planet när $\alpha \neq 1$ (se figur 7.6(b)). Vi noterar resultatet som en sats.

Sats 7.14 *För $\alpha \in (0,1) \cup (1,\infty)$, $w_1 = u_1 + iv_1$ och $w_2 = u_2 + iv_2$ ($u_1, v_1, u_2, v_2 \in \mathbf{R}$, $w_1 \neq w_2$) är mängden av lösningar till ekvation (7.25) en* *cirkel med medelpunkt*

$$\left(u_2 + \frac{u_2 - u_1}{\alpha^2 - 1}, v_2 + \frac{v_2 - v_1}{\alpha^2 - 1}\right)$$

och radie

$$\frac{\alpha\sqrt{(u_2 - u_1)^2 + (v_2 - v_1)^2}}{|\alpha^2 - 1|}.$$

Bevis. För att förenkla uträkningarna betraktar vi först det särskilda fallet av (7.25) där $w_2 = 0$ och $w_1 = u + iv$ ($u, v \in \mathbf{R}$):

$$\begin{aligned}
&|z - u - iv| = \alpha|z| \\
&\iff (x - u)^2 + (y - v)^2 = \alpha^2(x^2 + y^2) \\
&\iff (\alpha^2 - 1)x^2 + 2ux - u^2 + (\alpha^2 - 1)y^2 + 2vy - v^2 = 0 \\
&\iff (\alpha^2 - 1)\left(x + \frac{u}{\alpha^2 - 1}\right)^2 - \frac{u^2}{\alpha^2 - 1} - u^2 + \qquad (7.26) \\
&\quad (\alpha^2 - 1)\left(y + \frac{v}{\alpha^2 - 1}\right)^2 - \frac{v^2}{\alpha^2 - 1} - v^2 = 0 \\
&\iff \left(x + \frac{u}{\alpha^2 - 1}\right)^2 + \left(y + \frac{v}{\alpha^2 - 1}\right)^2 = \frac{\alpha^2\left(u^2 + v^2\right)}{(\alpha^2 - 1)^2}
\end{aligned}$$

som är (jämför med (3.22)) ekvationen av en cirkel med medelpunkt

$$\left(-\frac{u}{\alpha^2 - 1}, -\frac{v}{\alpha^2 - 1}\right)$$

och radie

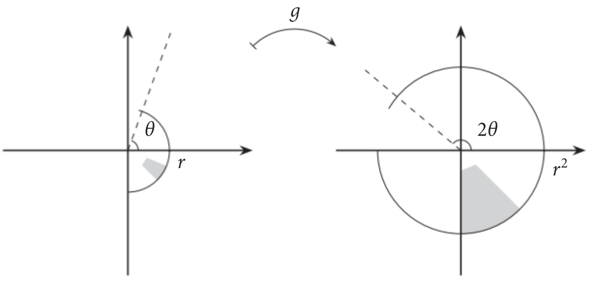
$$\frac{\alpha\sqrt{u^2 + v^2}}{|\alpha^2 - 1|}.$$

Genom att byta ut z mot $z - w_2$ och sätta $u = u_1 - u_2$ och $v = v_1 - v_2$ adderar vi w_2 till medelpunkten och får (7.25) från (7.26). □

Det är inte bara vissa cirklar som kan skrivas i formen (7.25) utan för vilken cirkel som helst kan vi hitta lämpliga val av w_1, w_2 och α så att mängden av lösningarna till (7.25) är den utvalda cirkeln.

Sats 7.15 *En cirkel i det komplexa planet med medelpunkt $a + ib$ ($a, b \in \mathbf{R}$) och radie $r > 0$ är lösningarna till* (7.25) *med valet*

$$w_1 = (2r + a) + ib \quad w_2 = \left(\frac{r}{2} + a\right) + ib \quad \textit{och} \quad \alpha = 2.$$



FIGUR 7.7 Exempel på hur punkter i det komplexa planet avbildas av funktionen $g(z) = z^2$.

Bevis. Först observerar vi att valen $w_1 = 2r$, $w_2 = r/2$ och $\alpha = 2$ i (7.25) är enligt sats 7.14 en cirkel med medelpunkten i origo och radien r. Men en cirkel med samma radie och medelpunkten $a + ib$ är bara en förflyttning av den med medelpunkten i origo. Därför är en cirkel med medelpunkten $a + ib$ och radien r lösningarna till (7.25) med $w_1 = (2r + a) + ib$, $w_2 = (r/2 + a) + ib$ och $\alpha = 2$. □

Sats 7.14 och 7.15 visar oss att (7.25) är ett annat sätt att representera cirklar precis som (3.22) och (7.23). Men valen av w_1, w_2 och α i sats 7.15 är inte entydliga: Till exempel för en given cirkel är varje rotation av w_1 och w_1 kring cirkelns medelpunkt också ett lämpligt val. Därför, till skillnad från (3.22) och (7.23), är representationen inte entydigt bestämd för varje cirkel.

AVBILDNINGAR AV KOMPLEXA FUNKTIONER

I det här avsnittet funderar vi på hur vi kan representera funktioner av en komplex variabel geometriskt. För en funktion $f : \mathbf{C} \to \mathbf{C}$ går det inte att rita en graf eftersom **C** representeras som ett plan – ett tvådimensionellt objekt – och därför skulle vi behöver fyra dimensioner för att rita grafen (två för målmängden **C** och två för definitionsmängden **C**). Det finns ingen ideal lösning till problemet att rymden vi lever i bara är tredimensionell, men ett sätt att förstå vad en funktion gör är att rita hur vissa mängder i definitionsmängden avbildas i målmängden. Vi betraktar ett par exempel.

Exempel 7.16 Betrakta funktionen $g\colon \mathbf{C} \to \mathbf{C}$ som definieras enligt formeln $g(z) = z^2$. För att få en bättre känsla för funktionen skriver vi z i polär form, det vill säga $z = re^{i\theta}$ där $r > 0$ och $\theta \in \mathbf{R}$. Då är

$$g(re^{i\theta}) = (re^{i\theta})^2 = r^2 e^{i2\theta}.$$

Då ser vi direkt att g skickar en punkt z till en punkt $g(z)$ där:

- $g(z)$:s argument är dubbelt det av z (2θ för $g(z)$ kontra θ för z); och
- $g(z)$:s absolutbelopp är z:s absolutbelopp i kvadrat (r^2 för $g(z)$ kontra r för z).

Det innebär att om vi betraktar en stråle från origo, skickar g den till en annan stråle vars vinkel mellan den reella axeln är dubbelt så stor. En cirkel med medelpunkten i origo och radie r skickas av g till en annan cirkel med samma medelpunkt och radie r^2. I figur 7.7 har vi till vänster ritat en streckad stråle som bildar vinkeln θ med reella axeln och sedan till höger dess avbildning som g bildar. Till vänster ser vi en båge (en del av cirkeln med radie r) och till höger avbildningen (av radie r^2). Även ett område inom cirkeln tillsammans med avbildningen visas.

Exempel 7.17 Betrakta funktionen $h\colon \mathbf{C} \to \mathbf{C}$ som definieras enligt formeln $h(z) = e^z$. Den här gången skriver vi z i kartesisk form, det vill säga $z = x + iy$. Men om vi skriver ut $h(z)$ ser vi att

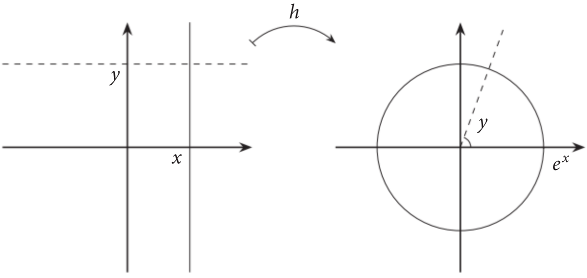
$$h(x + iy) = e^{x+iy} = e^x e^{iy} = e^x(\cos y + i \sin y)$$

och därför kan vi förstå avbildningar genom att tänka i polära koordinater. Vi ser att:

- z:s realdel bestämmer $h(z)$:s absolutbelopp genom sambandet $|h(z)| = e^x$. Därför avbildas en vågrätt linje som en stråle – se de streckade linjerna i figur 7.8; och
- z:s imaginärdel är $h(z)$:s argument (det vill säga $\arg(h(z)) = y$). Därför avbildas en lodrät linje som en cirkel – se figur 7.8.

Exempel 7.18 (Möbiusavbildningar) Låt $a, b, c, d \in \mathbf{C}$ vara sådana att $c \neq 0$ och $ad - bc \neq 0$, och betrakta funktionen $f\colon \mathbf{C} \setminus \{-d/c\} \to \mathbf{C} \setminus \{a/c\}$ som ges av formeln

$$f(z) = \frac{az + b}{cz + d}.$$



FIGUR 7.8 Exempel på hur punkter i det komplexa planet avbildas av funktionen $h(z) = e^z$.

Sådana funktioner kallas för *Möbiusavbildningar*. Tack vare villkoret $ad - bc \neq 0$ kan vi beräkna som i exempel 4.3 och visa att f är bijektiv – se uppgift 7.8. Definitionen av en bijektiv funktion och dess invers på sidan 118 gäller likväl när funktionens definitions- och målmängd är delmängder av $\mathbf{C}$ som när de är delmängder av $\mathbf{R}$. Utfallet av uppgift 7.8 är att inversen $f^{-1}\colon \mathbf{C} \setminus \{a/c\} \to \mathbf{C} \setminus \{-d/c\}$ ges av formeln

$$f^{-1}(z) = \frac{dz - b}{-cz + a}.$$

Observera att inversen av en Möbiusavbildning också är en Möbiusavbildning. Möbiusavbildningar har många intressanta egenskaper och den vi vill visa här är att avbildningar av cirklar och linjer är sig själva cirklar och linjer.

I föregående avsnitt har vi visat att cirklar och linjer alltid kan skrivas som lösningar till (7.25). (Se diskussionen efter (7.24) samt sats 7.14 och 7.15.) Därför betraktar vi en punkt $f(z)$ i f:s målmängd som lyder (7.25) för något val av w_1, w_2 och α:

$$\begin{aligned}
&|f(z) - w_1| = \alpha|f(z) - w_2| \\
\iff\ & \left|\frac{az+b}{cz+d} - w_1\right| = \alpha\left|\frac{az+b}{cz+d} - w_2\right| \\
\iff\ & \left|\frac{az+b-w_1(cz+d)}{cz+d}\right| = \alpha\left|\frac{az+b-w_2(cz+d)}{cz+d}\right| \\
\iff\ & |(a - w_1c)z + (b - w_1d)| = \alpha\,|(a - w_2c)z + (b - w_2d)|\,.
\end{aligned}$$

Nu betraktar vi tre fall: Antingen $a - w_1c = 0$ eller $a - w_2c = 0$ eller båda uttrycken skilda från noll. Minst ett måste vara skilt från noll eftersom $w_1 \neq w_2$.

Om $a - w_1 c = 0$ är

$$\begin{aligned} &|(a - w_1 c)z + (b - w_1 d)| = \alpha\,|(a - w_2 c)z + (b - w_2 d)| \\ \iff & \left|\frac{b - w_1 d}{a - w_2 c}\right| = \alpha \left|z + \left(\frac{b - w_2 d}{a - w_2 c}\right)\right| \end{aligned}$$

som har formen (7.23) med $r = \left|\frac{b-w_1 d}{a-w_2 c}\right|$ och $w = \left(\frac{b-w_2 d}{a-w_2 c}\right)$.

Ett liknande argument gäller när $a - w_2 c = 0$.

Om båda uttrycken är skilda från noll kan vi räkna ut att

$$\begin{aligned} &|(a - w_1 c)z + (b - w_1 d)| = \alpha\,|(a - w_2 c)z + (b - w_2 d)| \\ \iff & \left|z + \frac{b - w_1 d}{a - w_1 c}\right| = \frac{\alpha|a - w_2 c|}{|a - w_1 c|} \left|z + \frac{b - w_2 d}{a - w_2 c}\right| \end{aligned}$$

som har formen (7.25) med $w_1 = \frac{b-w_1 d}{a-w_1 c}$, $w_2 = \frac{b-w_2 d}{a-w_2 c}$ och $\alpha = \frac{\alpha|a-w_2 c|}{|a-w_1 c|}$.

Därför har vi visat att en cirkel eller linje i f:s målmängd alltid är avbildningen av en cirkel eller linje i f:s definitionsmängd. Vilken cirkel eller linje i definitionsmängden avgörs bara av koefficienterna i uttrycket för f samt ekvationen för cirkeln i målmängden. Att alla cirklar och linjer i definitionsmängden ger upphov till cirklar eller linjer i målmängden följer av samma argument eftersom f är bijektiv och inversen också är en Möbiusavbildning.

7.3 Sammanfattning och uppgifter

I kapitel 7 har vi:

- definierat vad vi menar med komplexa tal och hur vi kan representera dem som punkter i planet;
- utrett hur vi kan förenkla potenser av i samt andra uttyck vi får när vi utför aritmetiska operationer på komplexa tal;
- utvidgat begreppet absolutbelopp till komplexa tal och definierat konjugatet och argumentet av ett komplext tal samt utrett deras räkneregler;
- hittat nollställen till andragradspolynom även i det fall själva polynomet har komplexa koefficienter;
- utvidgat exponentialfunktionen till komplexa tal genom att visa att gränsvärdet (7.14), som definierar funktionen, konver-

gerar även för komplexa argument och kopplat exponentialfunktionen till trigonometriska funktioner;
- använt sambandet mellan exponentialfunktionen och de trigonometriska funktionerna för att skriva cosinus och sinus som serier;
- använt sambanden mellan geometri och komplexa tal för att lösa geometriska problem med aritmetiska verktyg samt hitta komplexa varianter av ekvationer för linjer och cirklar;
- jobbat för att förstå komplexa funktioner geometriskt trots att vi inte kan rita grafer på samma sätt som för reella funktioner;
- visat att Möbiusavbildning avbildar cirklar och linjer på cirklar och linjer.

UPPGIFTER

7.1 Skriv om följande uttryck i formen $a + ib$ för $a, b \in \mathbf{R}$:

(a) $(2+3i)+(6-2i)$; (c) $(2-8i)(1+3i)$; (e) $6i(1+3i)$; och
(b) $(-2+5i)-(1-7i)$; (d) $(-4+6i)/(1-2i)$; (f) $(1-3i)/(1+3i)$.

7.2 Visa att $\overline{wz} = \overline{w}\,\overline{z}$ och $\overline{w+z} = \overline{w} + \overline{z}$ för alla $w, z \in \mathbf{C}$. (Det vill säga bevisa del (a) och (b) av sats 7.6.)

7.3 Visa att om P är ett polynom med reella koefficienter och $P(z) = 0$ för något $z \in \mathbf{C}$, då är även $P(\overline{z}) = 0$ för samma $z \in \mathbf{C}$.

7.4 Använd (7.20) för att bevisa del (a), (c) och (d) av sats 7.11.

7.5 Skriv om följande uttryck i formen $re^{i\theta}$ för $r > 0$ och $\theta \in \mathbf{R}$:

(a) $\sqrt{3} + i$; (c) $(-12 + 5i) + (3 + 4i)$; och
(b) $(1 + i)^4$; (d) $3e^{i8\pi/3} \times 7e^{i\pi/3}$.

7.6 Vad är medelpunkten och radien av cirkeln som ges av lösningarna z till ekvationen

$$|z - (1 + i)| = 3|z - (4 - 3i)|?$$

7.7 (a) Visa att om vi begränsar den komplexa exponentialfunktionen definitionsmängd till $A = \{x + iy : x \in \mathbf{R}, -\pi \leq y < \pi\}$ och ta målmängden lika med $B = \mathbf{C} \setminus \{0\}$ är $\exp : A \to B$ inverterbar.

(b) Hitta en formel för inversen till $\exp : A \to B$.

7.8 Betrakta Möbiusavbildningen

$$f(z) = \frac{az + b}{cz + d}$$

från exempel 7.18 med definitionsmängden $\mathbf{C} \setminus \{-d/c\}$ och målmängden $\mathbf{C} \setminus \{a/c\}$. Här antar vi att $a,b,c,d \in \mathbf{C}$ är sådana att $c \neq 0$ och $ad - bc \neq 0$.

Visa att f är bijektiv och beräkna en formel för dess invers.

Lösningsförslag

1.1 Påstående (b) är falskt. För att bevisa det räcker det att hitta *ett* tal $x \geq -5$ sådant att $x^2 < 25$ ty påståendet säger att $x^2 \geq 25$ för *alla* $x \geq -5$. Vi kan till exempel ta $x = -3$: $-3 \geq -5$ men $(-3)^2 < 25$.

Vi kan bevisa att (a) är sant genom att multiplicera bägge sidor av ”$x \geq 5$” med positiva tal – kom ihåg vår diskussion om enkla regler i avsnitt 1.1 och särskilt den om (1.10).[1] Vi beräknar

$$x \geq 5 \implies 5x \geq 5 \times 5 = 25$$

och

$$x \geq 5 \implies x^2 \geq 5x$$

eftersom x och 5 är icke-negativa. Vi kan lägga ihop $5x \geq 25$ och $x^2 \geq 5x$ och får $x^2 \geq 5x \geq 25$. Därför $x \geq 5 \implies x^2 \geq 25$.

Samma argument fungerar inte för (b) eftersom x inte nödvändigtvis är icke-negativt.

1.2 Först är det bra att fråga: Vad menas med varje implikation här? Vad som menas är att så länge ett x uppfyller påståendet som sitter på en pils bakre ände så får vi att x också uppfyller påståendet på pilens spets.

(a) Vi använder oss av samma metod som i uppgift 1.1. Från olikheten $x \geq 3$ får vi både $x^2 \geq 3x$ och $3x \geq 9$ eftersom både x och 3 är positiva. Därför lägger vi ihop dem och får $x^2 \geq 3x \geq 9$. Vi har därför visat att $x \geq 3 \implies x^2 \geq 9$.

(b) Samma metod visar att $x \geq 5$ medför både $x^2 \geq 5x$ och $5x \geq 25$, och därför $x^2 \geq 5x \geq 25 \geq 20$. Det innebär att $x \geq 5 \implies x^2 \geq 20$.

1 Man kan även jämföra den med axiom C.2(d) i bilaga C.1.

(c) Om vi adderar samma tal till båda sidor av en olikhet behåller vi olikheten.[2] Därför $x > 5 \implies x - 2 > 5 - 2 = 3$. Eftersom x måste vara positivt så kan vi säga $x - 2 > 3 \implies x(x-2) > 3x$, och enligt samma argument $x > 5 \implies 3x > 3 \times 5 = 15$ ty 3 är positivt. Allihop då

$$x > 5 \implies x(x-2) > 3x > 15.$$

1.3 (c) Kontrapositionen av påståendet ”$A \implies B$” är ”$\neg B \implies \neg A$” där A och B också är påståenden. Därför tar vi A lika med påståendet ”$x > 5$” och B lika med ”$x(x-2) > 15$” så $\neg A$ är ”$x \leq 5$” och $\neg B$ är ”$x(x-2) \leq 15$”. Kontrapositionen är då

$$x(x-2) \leq 15 \implies x \leq 5.$$

1.4 (b) Implikationen i uppgiften är

$$x(x-2) > 15 \implies x > 5.$$

Med detta menas: Om x är ett tal som uppfyller olikheten $x(x-2) > 15$ så uppfyller det även olikheten $x > 5$. Därför, för att visa att den är felaktig räcker det att hitta *ett* x som uppfyller $x(x-2) > 15$ men inte $x > 5$.

För att hitta ett sådant x använder vi oss av lite intuition. Ni vet förmodligen att det finns två enkla sätt då ett uttryck som $x(x-2)$, som är ett andragradspolynom i x, kan vara stort: antingen x är väldigt positivt, eller x är väldigt negativt. Vi vill hitta ett x som uppfyller $x \not> 5$, så vi provar ett negativt x.

Till exempel, provar vi $x = -10$. Då är $x(x-2) = 120 > 15$ men samtidigt är $x = -10 \not> 5$. Därför har vi bevisat att påståendet ”$x > 5 \impliedby x(x-2) > 15$” är felaktigt.

1.5 (a) För alla heltal n är $n^2 - 3n + 1 \geq 0$.

(b) Det finns minst ett par tal x och y sådant att $x + y \neq y + x$.

(c) Det finns ett x sådant att $x > 8$ och $x^2 - 14x + 48 \leq 0$.

1.6 (a) $2^2 = 4$ och $4^2 = 16$ är jämna.

(b) Vi undersöker villkoret ”n är jämnt”. Om n är jämnt då kan det skrivas som $n = 2k$ för ett heltal k. Sedan är även $n^2 = (2k)^2 = 2(2k^2)$ jämnt. Därför kan vi säga att ett tillräckligt villkor för att n^2 ska vara jämnt är att n är jämnt.

2 Jämför med axiom C.2(c) i bilaga C.1.

(c) Man kan visa att samma villkor är också ett nödvändigt villkor för att n^2 ska vara jämnt. Se beviset av lemma 4.5. Därför är n^2 jämnt om och endast om n är jämnt.

1.7 Vi kan bevisa det på samma sätt som sats 1.5. Eftersom n_1 delat med 82 har rest 49 vet vi från (1.13) att det finns ett heltal q_1 så att

$$n_1 = 82q_1 + 49.$$

Eftersom n_2 delat med 82 har rest 9 vet vi också från (1.13) att det finns ett heltal q_2 så att

$$n_2 = 82q_2 + 9.$$

Därför kan vi dra slutsatsen att

$$\begin{aligned} n_1n_2 &= (82q_1 + 49)(82q_2 + 9) \\ &= (82q_1q_2 + 9q_1 + 49q_2)82 + 441 \\ &= (82q_1q_2 + 9q_1 + 49q_2 + 5)82 + 31, \end{aligned}$$

så (1.13) i definition 1.4 uppfylls med $n = n_1n_2$, $q = (q_1q_2m + 9q_1 + 49q_2 + 5)$ och $r = 31$ som förstås alla är hela tal. Dessutom är $0 \leq 31 < 82$. Därför är 31 resten av n_1n_2 delat med 82.

1.8 Observera först att slutsatsen inte gäller i allmänhet utan fler villkor än $a < b$. Till exempel, om $n = 2$, $a = -5$ och $b = 3$ är $a < b$ men $a^n = 25 \nless 9 = b^n$.

Det är möjligt att hitta många olika svar till uppgiften. Här går vi igenom ett par exempel.

Först betraktar vi villkoret ”$a > 0$ och n positivt”. Vi härleder[3] att

$$a < b \text{ och } a > 0 \implies a^2 < ab.$$

Under samma antagande är även $b > 0$, så vi kan också dra slutsatsen att $ab < b^2$. Därmed

$$a < b \text{ och } a > 0 \implies a^2 < ab < b^2 \implies a^2 < b^2.$$

Samma bevis fungerar för att bevisa slutsatsen med ett positivt heltal n i stället för potensen 2. Vi kan till exempel beräkna med samma logik att $a < b$ och $a > 0$ medför att

$$a^4 < a^3b < a^2b^2 < ab^3 < b^4$$

3 Jämför med axiom C.2(d).

som visar slutsatsen i fallet $n = 4$. Man kan förstås argumentera att ett induktionsbevis skulle vara tydligare för att behandla allmänna n, men det kommer vi till i kapitel 2.

Ett andra villkor som fungerar är "$n > 0$ är udda". Här använder vi ett knep för att bevisa det i fallet $n = 3$: Vi kan skriva

$$a^3 - b^3 = \frac{(a-b)}{2}\left(a^2 + (a+b)^2 + b^2\right) \tag{A.1}$$

och olikheten $a < b$ medför att $a - b < 0$ och $\left(a^2 + (a+b)^2 + b^2\right) > 0$ eftersom alla termer är icke-negativa och högst en term kan vara noll. Därför är högerledet i (A.1) negativt för $a < b$ och vi har visat att

$$a < b \implies a^3 < b^3.$$

Beviset kan generaliseras till udda $n > 0$.

1.9 Eftersom r och s är rationella finns det hela tal p, q, m och n så att $p \neq 0$, $q \neq 0$, $r = m/p$ och $s = n/q$.

(a) $r + s = m/p + n/q = (mq)/(pq) + (np)/(pq) = (mq + np)/(pq)$ som medför att $r + s$ är rationellt ty $mq + np$ och pq är hela tal och $pq \neq 0$.

(b) $rs = (m/p)(n/q) = (mn)/(pq)$ som medför att rs är rationellt ty mn och pq är hela tal och $pq \neq 0$.

1.10 (a) 111, 10 001, 1 100 respektive 100 000.

(b) 21, 122, 110 respektive 1 012.

1.11 (a) Vi beräknar

$$\sum_{j=-m}^{n} d_j 10^j \leq \sum_{j=-m}^{n} 9 \times 10^j \underset{\substack{\uparrow \\ \text{sats 1.14}}}{=} 9 \times 10^{-m}\left(\frac{1 - 10^{n+m+1}}{1 - 10}\right)$$
$$= 10^{-m}\left(10^{n+m+1} - 1\right) < 10^{n+1}.$$

(b) Talet $t = \sum_{j=-m}^{n} d_j 10^j$ har en decimalutveckling med högst $n + 1$ siffror till vänster om decimalkommat och m siffror till höger. Däremot har 10^{n+1} en decimalutveckling som har en icke-noll siffra i $(n + 2)$:e platsen till vänster om decimalkommat och är därför större än $\sum_{j=-m}^{n} d_j 10^j$.

Siffrorna till vänster om decimalkommat bidrar högst $10^{n+1} - 1$ till t. De andra siffrorna, det vill säga de till höger om decimalkommat, bidrar med ett tal strängt mindre än 1 till t – här kunde vi ha fått ett beroendeförhållande med m, men vi har lyckats med en uppskattning som är oberoende av m! Därmed har vi återigen visat att summan t mindre än 10^{n+1}.

2.1 Följderna (a) och (c) är aritmetiska följder. Följderna (b) och (c) är geometriska följder.

2.2 (a) Här är faktorn 7 oberoende av k så talet bara upprepas n gånger:

$$\prod_{k=1}^{n} 7 = \underbrace{7 \times \cdots \times 7}_{n \text{ gånger}} = 7^n.$$

(b) Enligt sats 1.8(a) kan potenserna av 5 adderas:

$$\prod_{k=1}^{n} 5^k = 5^{\sum_{k=1}^{n} k} \underset{\substack{\uparrow \\ \text{sats 1.13}}}{=} 5^{n(n+1)/2}.$$

(c) Vi beräknar

$$\begin{aligned}\prod_{k=1}^{n} 3k = 3(3 \times 2)(3 \times 3) \ldots (3n) &= 3^n (1 \times 2 \times 3 \times \cdots \times n) \\ &= 3^n n!\end{aligned}$$

2.3 (b) Vi vill bevisa

$$\sum_{k=1}^{n} k^3 = \left(\frac{n(n+1)}{2}\right)^2 \qquad (\text{A.2})_n$$

för alla positiva heltal n.

Först kontrollerar vi fallet $(\text{A.2})_1$. Vi beräknar att $\sum_{k=1}^{1} k^3 = 1$ och $\left(\frac{1(1+1)}{2}\right)^2 = 1$, så $(\text{A.2})_1$ är bevisat.

Nu antar vi att $(\text{A.2})_m$ är sant för något positivt heltal m och betraktar $(\text{A.2})_{m+1}$. Vänsterledet av $(\text{A.2})_{m+1}$ är

$$\sum_{k=1}^{m+1} k^3 = \sum_{k=1}^{m} k^3 + (m+1)^3 = \left(\frac{m(m+1)}{2}\right)^2 + (m+1)^3$$

enligt antagandet $(\text{A.2})_m$. Vi kan beräkna

$$\begin{aligned}\left(\frac{m(m+1)}{2}\right)^2 + (m+1)^3 &= (m+1)^2 \left(\frac{m^2}{4} + (m+1)\right) \\ &= \frac{(m+1)^2((m+1)+1)^2}{4}\end{aligned}$$

som är högerledet i $(A.2)_{m+1}$. Därför är implikationen $(A.2)_m \implies (A.2)_{m+1}$ bevisad.

Sanningen av $(A.2)_1$ och implikationen $(A.2)_m \implies (A.2)_{m+1}$ för $m \in \mathbf{Z}_+$ medför enligt induktionsprincipen att $(A.2)_n$ är sant för alla positiva heltal n.

2.4 (a) Vi vill bevisa

$$4n \le 2^n \tag{A.3$_n$}$$

för alla heltal $n \ge 4$.

Först kontrollerar vi fallet $(A.3)_4$. Vi räknar ut att $4n = 4 \times 4 = 16$ och $2^n = 2^4 = 16$. Eftersom $16 \le 16$ är $(A.3)_4$ bevisat.

Nu antar vi att $(A.3)_m$ är sant för något heltal $m \ge 4$ och betraktar $(A.3)_{m+1}$. Vänsterledet av $(A.3)_{m+1}$ är

$$4(m+1) = 4m + 4 \le 2^m + 4 = 2^m + 2^2$$

enligt antagandet $(A.3)_m$. Eftersom $m \ge 2$ är

$$2^m + 2^2 \le 2^m + 2^m = 2^m(1+1) = 2^{m+1}$$

som är högerledet i $(A.3)_{m+1}$. Därför är implikationen $(A.3)_m \implies (A.3)_{m+1}$ bevisad.

Enligt induktionsprincipen är $(A.3)_n$ därför sant för alla heltal $n \ge 4$.

Här har vi tillämpat axiom 2.6 där basfallet är $n = 4$ i stället för $n = 1$, men det gör inget förutom att beviset ger $(A.3)_n$ endast för $n \ge 4$. Metoden funkar lika bra eftersom man kan tänka att vi tar påståendet $P(n)$ i axiom 2.6 lika med $(A.3)_{n+3}$.

Observera att beviset av induktionssteget $(A.3)_m \implies (A.3)_{m+1}$ gäller för $m \ge 2$ trots att olikheten $(A.3)_n$ är falsk för $n = 2,3$ (och även $n = 1$). Det är ingen motsägelse eftersom man behöver både basfallet och induktionssteget för att induktionsbeviset ska vara giltigt. Eftersom basfallet inte kan bevisas för $n = 2,3$ är beviset fullständigt endast för $n \ge 4$.

2.5 (a) Problemet ligger i induktionssteget. Trots att vi säger att vi antar $(2.28)_n$ i själva räkningen är det $(2.28)_{n+1}$ vi antar och sedan bevisar vi $(2.28)_{n+1} \implies (2.28)_n$. Så implikationen vi har bevisat är baklänges från den vi behöver för att genomföra induktionssteget i ett induktionsbevis av $(2.28)_k$.

(b) Vi kan endast skriva $n+1 = i + j$ där både i och j är icke-negativa heltal mindre än eller lika med n i fallet $n \geq 1$. Om vi försöker göra det i fallet $n = 0$ får vi motsägelsen att $0 + 1 = 0 + 0$. Därför har vi endast ett giltigt bevis för

$$(2.29)_k \text{ för alla } 0 \leq k \leq n \implies (2.29)_{n+1}$$

för $n \geq 1$. Men basfallet var $(2.29)_0$ så vi saknar implikationen $(2.29)_0 \implies (2.29)_1$.

Observera att påståendet $P(n)$ vi försöker bevisa med ett induktionsbevis skiljer sig lite från dem vi har sett tidigare: Här är $P(n)$ ”$(2.29)_k$ för alla $0 \leq k \leq n$” i stället för att helt enkelt vara $(2.29)_n$. Det är ett helt rimligt val och inte där problemet ligger – jämför till exempel med antagandet om polynomets grad i induktionsbeviset av sats 3.13.

(c) Vi har inte bevisat basfallet $(2.30)_1$.

2.6 Vi beräknar

$$\begin{aligned}
&\binom{n}{k} + \binom{n}{k-1} \\
&= \frac{n!}{k!(n-k)!} + \frac{n!}{(k-1)!(n-k+1)!} \\
&= \frac{n!}{k(k-1)!(n-k)!} + \frac{n!}{(k-1)!(n-k+1)(n-k)!} \\
&= \frac{n!}{(k-1)!(n-k)!}\left(\frac{1}{k} + \frac{1}{(n-k+1)}\right) \\
&= \frac{n!}{(k-1)!(n-k)!}\left(\frac{(n-k+1)+k}{k(n-k+1)}\right) \\
&= \frac{n!}{(k-1)!(n-k)!}\left(\frac{n+1}{k(n-k+1)}\right) \\
&= \frac{(n+1)!}{k!(n+1-k)!} = \binom{n+1}{k}.
\end{aligned}$$

2.7 Vi kan beräkna att

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{n(n-1)\ldots(n-k+1)}{k!} \leq \frac{n^k}{k!}$$

eftersom $n - \ell \leq n$ för $\ell = 0, 1, \ldots, k-1$.

2.8 Vi beräknar att

$$\begin{aligned} |a+b| = |a|+|b| &\implies |a+b|^2 = (|a|+|b|)^2 \\ &\implies a^2+b^2+2ab = a^2+b^2+2|a||b| \\ &\implies ab = |a||b|. \end{aligned}$$

Om $a > 0$ är $|a| = a$ och därför $ab = |a||b| \implies b = |b|$, så är $b \geq 0$. Om $a < 0$ är $|a| = -a$ och därför $ab = |a||b| \implies b = -|b|$, så är b icke-positivt. Om $a = 0$ då kan b vara vilket tal som helst. Därmed har vi räknat ut alla kandidater till lösningar: $|a+b| = |a|+|b|$ medför att $a > 0$ och $b \geq 0$, $a < 0$ och $b \leq 0$, eller $a = 0$ och vilket b som helst.

Nu måste vi kontrollera att kandidaterna verkligen är lösningar. Och det är inte så svårt. Till exempel: Om $a = |a|$ och $b = |b|$ är

$$|a+b| = ||a|+|b|| = |a|+|b|.$$

På liknande sätt visar man att de andra kandidaterna också är lösningar.

2.9 För att genomföra ett motsägelsebevis antar vi att $(n)_{n\in\mathbf{Z}_+}$ är konvergent. I så fall säger sats 2.34 att $(n)_{n\in\mathbf{Z}_+}$ är uppåt begränsad. Men sats 2.26(b) säger att inget $b \in \mathbf{R}$ kan vara en övre begränsning av $(n)_{n\in\mathbf{Z}_+}$. Det är en motsägelse till att $(n)_{n\in\mathbf{Z}_+}$ är uppåt begränsad och därmed kan $(n)_{n\in\mathbf{Z}_+}$ inte vara konvergent.

2.10 (a) Beräkningar ger

$$\frac{2m^2+1}{m^2+4m} = \frac{2+\frac{1}{m^2}}{1+4\frac{1}{m}}.$$

Från sats 2.35(c), (e) och (f), och exempel 2.32 vet vi att $\lim_{m\to\infty}(1+\frac{4}{m}) = 1+0 \neq 0$ och $\lim_{m\to\infty}(2+\frac{1}{m^2}) = 2+0$, så från sats 2.35(d) vet vi att $(\alpha_m)_{m\in\mathbf{Z}_+}$ är konvergent och $\lim_{m\to\infty}\alpha_m = 2/1 = 2$.

(b) $(\alpha_m)_{m\in\mathbf{Z}_+}$ är konvergent och $\lim_{m\to\infty}\alpha_m = 1/2$.

(c) $(\alpha_m)_{m\in\mathbf{Z}_+}$ är konvergent och $\lim_{m\to\infty}\alpha_m = 0$.

(d) $(\alpha_m)_{m\in\mathbf{Z}_+}$ är konvergent och $\lim_{m\to\infty}\alpha_m = 2/9$.

(e) $(\alpha_m)_{m\in\mathbf{Z}_+}$ är konvergent och $\lim_{m\to\infty}\alpha_m = 1/3$.

(f) Beräkningar ger

$$\frac{5m^6+1}{2+3m^6} = \frac{5+\frac{1}{m^6}}{\frac{2}{m^6}+3}.$$

Från sats 2.35(c), (e) och (f), och exempel 2.32 vet vi att $\lim_{m\to\infty}(\frac{2}{m^6}+3) = 0+3 \neq 0$ och $\lim_{m\to\infty}(5+\frac{1}{m^6}) = 5+0$, så från sats 2.35(d) vet vi att $(\alpha_m)_{m\in\mathbf{Z}_+}$ är konvergent och $\lim_{m\to\infty}\alpha_m = 5/3$.

2.11 (a) $0.\overline{27} = 27/99 = 3/11$, $6.\overline{15} = 6 + 15/99 = 203/33$, och $4.\overline{118} = 4 + 118/999 = 4114/999$.

(b) De första fyra siffrorna är $1/8 = 0{,}125\ldots$ alternativt $1/8 = 0{,}124\ldots$, $1/3 = 0{,}333\ldots$, $1/2 = 0{,}500\ldots$ alternativt $1/2 = 0{,}499\ldots$, och $4/7 = 0{,}5714\ldots$.

Vi har faktiskt att $1/8 = 0{,}125$, $1/3 = 0{,}\overline{3}$, $1/2 = 0.5$ och $4/7 = 0{,}\overline{5714285}$. Observera att $1/8$ och $1/2$ båda har två oändliga decimalutvecklingar: $1/8 = 0{,}125\overline{0} = 0{,}124\overline{9}$, respektive $1/2 = 0{,}5\overline{0} = 0{,}4\overline{9}$.

2.12 (a) Eftersom $4x-7 \geq 2 \iff 4x \geq 9 \iff 9/4 \leq x$, så är

$$\{x\in\mathbf{R} \mid 4x-7\geq 2 \text{ och } x\leq 4\} = \{x\in\mathbf{R} \mid 9/4 \leq x \leq 4\}$$

som är intervallet $[9/4, 4]$ och inte $[0,4]$.

(b) Mängden $M = \{x\in\mathbf{R} \mid 2x^2+4 < 12-8x\}$ kan inte vara intervallet $[0,4]$ eftersom $2(-1)^2+4 = 6 < 20 = 12-8(-1)$ så $-1\in M$ trots att $-1\notin[0,4]$.

(c) Vi har

$$x^2-2x+1\leq 1 \iff x(x-2)\leq 0.$$

Sista olikheten gäller i fallet en av x och $x-2$ är icke-positivt och den andra är icke-negativt. Det innebär att antingen $x\geq 0$ och $x\leq 2$, eller $x\leq 0$ och $x\geq 2$. Eftersom $x\leq 0$ och $x\geq 2$ inte kan gälla samtidigt är

$$x^2-2x+1\leq 1 \iff x(x-2)\leq 0 \iff 0\leq x\leq 2.$$

På liknande sätt ser vi att

$$\begin{aligned} 8\leq 6x-x^2 &\iff x^2-6x+8\leq 0\\ &\iff (x-4)(x-2)\leq 0\\ &\iff 2\leq x\leq 4 \end{aligned}$$

så

$$\begin{aligned} &\{x\in\mathbf{R} \mid x^2-2x+1\leq 1 \text{ eller } 8\leq 6x-x^2\}\\ &= \{x\in\mathbf{R} \mid 0\leq x\leq 2 \text{ eller } 2\leq x\leq 4\}\\ &= [0,2]\cup[2,4] = [0,4]. \end{aligned}$$

(d) Ett liknande argument till det i (c) ger oss att

$$\begin{aligned}&\{x \in \mathbf{R} \mid x^2 - 4x \le 0 \text{ och } 8 \le 6x - x^2\}\\&= \{x \in \mathbf{R} \mid 0 \le x \le 4 \text{ och } 2 \le x \le 4\}\\&= [0,4] \cap [2,4] = [2,4] \ne [0,4].\end{aligned}$$

(e) Återigen kan vi beräkna

$$\begin{aligned}x^2 - 4x < 0 &\iff x(x-4) < 0 \iff 0 < x < 4,\\x^2 - 4x + 3 = 0 &\iff (x-3)(x-1) = 0\\&\iff x = 3 \text{ eller } x = 1, \quad \text{och}\\x^2 + 8 = 6x &\iff (x-4)(x-2) = 0\\&\iff x = 4 \text{ eller } x = 2.\end{aligned}$$

Därför är

$$\begin{aligned}&\{x \in \mathbf{R} \mid x^2 - 4x < 0, x^2 - 4x + 3 = 0 \text{ eller } x^2 + 8 = 6x\}\\&= (0,4) \cup \{3,1\} \cup \{4,2\} = (0,4] \ne [0,4].\end{aligned}$$

2.13 Vi kan skriva

$$(2n)! = 2n(2n-1)(2n-2)\dots(n+1)n(n-1)\dots 2 \times 1.$$

För att jämföra $(2n)!$ med n^n jämför vi faktorerna i $(2n)!$ med faktorerna i n^n. Vi vet att $(2n-k) \ge n$ så länge $0 \le k \le n$, därför är

$$2n(2n-1)(2n-2)\dots(n+1) \ge n^n$$

eftersom vänsterledet innehåller n faktorer. Vi får också uppskatta

$$n(n-1)\dots 2 \times 1 \ge 1.$$

Vi drar slutsatsen att

$$(2n)! = 2n(2n-1)(2n-2)\dots(n+1)n(n-1)\dots 2\times 1 \ge n^n \times 1 = n^n$$

och därför är

$$a_n = \frac{n^n}{(2n)!} \le 1$$

för varje $n \in \mathbf{Z}_+$.

2.14 Ett godtyckligt element i mängden

$$M = \{y \in \mathbf{R} \mid y = (x-1)(x+5) \text{ och } x \in [-6, 2]\}$$

har formen $y = (x-1)(x+5) = (x+2)^2 - 9$ och eftersom $(x+2)^2 \geq 0$ är -9 en undre begränsning till M. Utöver det tillhör -9 mängden eftersom $-9 = (x-1)(x+5)$ för $x = -2$ och $-2 \in [-6, 2]$. Därför kan varje tal större än -9 inte vara en undre begränsning. Vi dra slutsatsen att $\inf M = -9$.

För att visa att M har en minsta övre begränsning kontrollerar vi tre olika fall: $x \in [-6, -5]$; $x \in (-5, 1)$; och $x \in [1, 2]$. Anledningen till att välja de där fallen är att när x förflyttar sig från ett intervall till ett annat byter minst en faktor i uttrycket $y = (x-1)(x+5)$ tecken.

Fall 1: $x \in [-6, -5]$.

Då är $-7 \leq x - 1 \leq -6$ och $-1 \leq x + 5 \leq 0$ så $0 = (-6)(0) \leq y = (x-1)(x+5) \leq (-7)(-1) = 7$. Och y kan även anta sin övre begränsning 7 genom att ta $x = -6 \in [-6, -5]$.

Fall 2: $x \in (-5, 1)$.

Då är $-6 < x - 1 < 0$ och $0 < x + 5 < 6$ så $y = (x-1)(x+5)$ blir negativt och därmed mindre än y-värden från fall 1.

Fall 3: $x \in [1, 2]$.

Då är $0 \leq x-1 \leq 1$ och $6 \leq x+5 \leq 7$ så $0 = (0)(6) \leq y = (x-1)(x+5) \leq (1)(7) = 7$. Återigen kan y anta värdet 7 genom att vi sätter in $x = 2 \in [1, 2]$.

Från fall 1–3 drar vi slutsatsen att 7 är en övre begränsning till M och eftersom $7 \in M$ kan ett tal mindre än 7 inte vara en övre begränsning. Därför är $\sup M = 7$. (Om en mängd har en minsta övre begränsning som tillhör mängden är det faktiskt mängdens största element, så vi har även visat att $\max M = 7$.)

2.15 (a) För att genomföra ett motsägelsebevis antar vi att:

$$\text{det finns } a, b \in \mathbf{Z} \text{ sådana att } a^2 - 4b = -2. \tag{A.4}$$

(b) Under antagandet (A.4) är

$$a^2 = 4b - 2 = 2(2b-1)$$

och eftersom $2b - 1$ är ett heltal skulle a^2 därför vara jämnt.

(c) Enligt uppgift 1.6 är a^2 jämnt (om och) endast om a är jämnt. Så enligt (b) medför (A.4) att a är jämnt.

(d) Antagandet a är jämnt och (A.4) innebär att a kan skrivas som $a = 2k$ för något $k \in \mathbf{Z}$ och att

$$2 = 4b - a^2 = 4b - (2k)^2 = 4(b - a^2).$$

Det leder till att $1 = 2(b - a^2)$ och att 1 är jämnt.

(e) Antagandet (A.4) medför att a är jämnt och därför säger (d) att 1 är jämnt. Talet 1 är förstås inte jämnt, så det är en motsägelse. Antagandet (A.4) måste därför vara falskt och dess negation (2.31) sant.

3.1 Vi kan till exempel ta $f(x) = g(x) = x$ och $E = [-2, 2]$. Då är f och g växande på E men fg inte är växande på $E = [-2, 2]$.

Olikheten man får när man multiplicerar båda leden av en olikhet med någonting beror på tecknet av det man multiplicerar med så produkten lyder inte nödvändigtvis samma olikhet. (Se diskussionen om enkla regler i avsnitt 1.1 och särskilt den om (1.10).[4])

3.2 (b) Vi antar att f är avtagande på E och g är strängt avtagande på E. Det innebär att $x_1 < x_2$ medför $f(x_1) \geq f(x_2)$ och $g(x_1) > g(x_2)$ för $x_1, x_2 \in E$. Därför är $f(x_1) + g(x_1) \geq f(x_2) + g(x_1) > f(x_2) + g(x_2)$ så $x_1 < x_2$ medför att $(f + g)(x_1) > (f + g)(x_2)$ och $f + g$ är strängt avtagande på E.

Om vi stryker "strängt" från både (a) och (b) gäller samma argument förutom att den andra olikheten inte är sträng och vi får bara monotonicitet och inte sträng monotonicitet.

3.3 Vi kan till exempel bevisa att:

(a) Om f är växande på E och g är avtagande på E då är $f - g$ växande på E; och

(b) Om f är avtagande på E och g är strängt växande på E då är $f - g$ strängt avtagande på E.

Bevis. (a) Om $x_1 < x_2$ är $f(x_1) \leq f(x_2)$ och $g(x_1) \geq g(x_2)$ för $x_1, x_2 \in E$. Därför är $f(x_1) - g(x_1) \leq f(x_2) - g(x_1) \leq f(x_2) + g(x_2)$ så $(f + g)(x_1) \leq (f + g)(x_2)$. Därmed är $f - g$ växande på E.

4 Man kan även jämför den med axiom C.2(d) i bilaga C.1.

(b) Om $x_1 < x_2$ är $f(x_1) \geq f(x_2)$ och $g(x_1) < g(x_2)$ för $x_1, x_2 \in E$. Därför är $f(x_1) - g(x_1) \geq f(x_2) - g(x_1) > f(x_2) - g(x_2)$ så $(f - g)(x_1) > (f - g)(x_2)$. Därmed är $f - g$ strängt avtagande på E.

3.4 (a) Då antar vi att både f och g är strängt växande. Det innebär att $x_1 < x_2$ ($x_1, x_2 \in A$) medför $f(x_1) < f(x_2)$ och det i sin tur medför att $(g \circ f)(x_1) = g(f(x_1)) < g(f(x_2)) = (g \circ f)(x_2)$.

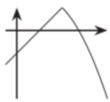
(c) Då antar vi att f är strängt växande och g är strängt avtagande. Det innebär att $x_1 < x_2$ ($x_1, x_2 \in A$) medför $f(x_1) < f(x_2)$ och det i sin tur medför att $(g \circ f)(x_1) = g(f(x_1)) > g(f(x_2)) = (g \circ f)(x_2)$.

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

Ett liknande argument gäller om man stryker ”strängt” från hypoteserna och i så fall får man vanliga olikheter i stället för stränga.

3.5 (a)

(b)



(c)

3.6 (a) Om $x < y$ är $2x < 2y$ och $f(x) = 2x - 2 < 2y - 2 = f(y)$ så är f strängt växande.

(b) Om $0 \leq x < y$ är $f(y) - f(x) = y^2 - x^2 = (y - x)(y + x) > 0$ ty $y - x > 0$ och $y + x > 0$.

(c) Om $x < y$ är $f(y) - f(x) = y^3 - x^3 = (y - x)(y^2 + xy + x^2) > 0$ ty $y - x > 0$ och

$$y^2 + xy + x^2 = \frac{1}{2}\left((x + y)^2 + x^2 + y^2\right)$$

som är positivt för att högst ett av x och y kan vara noll.

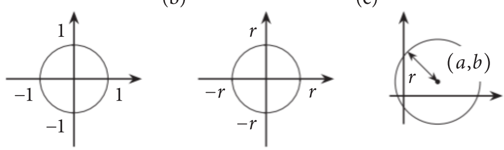
3.7 Om $p(x) = \sum_{k=0}^{n} a_k x^k$ är ett polynom och det finns tal $x_0 < x_1 < \cdots < x_n$ så att $p(x_j) = 0$ för alla $j = 0, 1, \ldots, n$, då är $p(x) = 0$ för alla $x \in \mathbf{R}$ enligt sats 3.13. Därför tillämpar vi sats 3.14 och drar slutsatsen $a_k = 0$ för alla $k = 1, 2, \ldots, n$.

3.8 Det enda sättet för ett polynom att vara identiskt noll är enligt sats 3.14 att alla dess koefficienter är noll. Därför måste vi lösa systemet

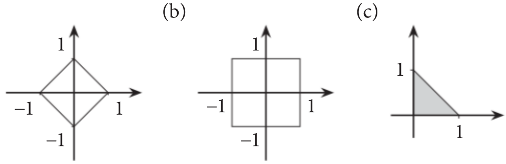
$$\begin{aligned} 2a - 5 &= 0 \\ 5b + c &= 0 \quad \text{och} \\ c - a &= 0. \end{aligned}$$

Vi räknar ut att $a = 5/2$, $c = 5/2$ och $b = -1/2$.

3.9 (a) (b) (c)



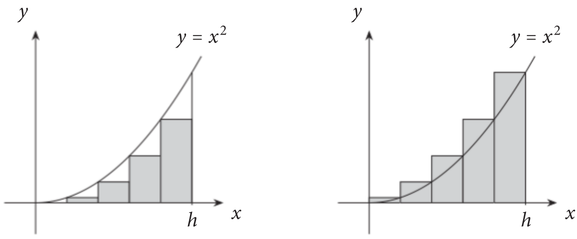
3.10 (a) (b) (c)



3.11 Vi kan underskatta arean av P med hjälp av mängderna E_n som är unionen av $n-1$ icke-överlappande rektangulära mängder $[hj/n, h(j+1)/n) \times [0, (hj/n)^2)$ $(j = 1, \dots, n-1)$. Se figur A.1(a). Eftersom $E_n \subseteq P$ är $|E_n| \le |P|$. Vi kan beräkna

$$\begin{aligned} |E_n| &= \sum_{j=1}^{n-1} \frac{h}{n} \left(\frac{hj}{n}\right)^2 = \frac{h^3}{n^3} \sum_{j=1}^{n-1} j^2 \underset{(2.6)}{\overset{}{\underset{\uparrow}{=}}} \frac{h^3}{n^3} \frac{(n-1)n(2n-1)}{6} \\ &= \frac{h^3}{3} - \frac{h^3}{2n}\left(1 - \frac{1}{3n}\right). \end{aligned}$$

Vi kan överskatta arean av P med hjälp av mängderna F_n som är unionen av n icke-överlappande rektangulära mängder $[h(j-1)/n, hj/n) \times [0, (hj/n)^2)$ $(j = 1, \dots, n)$. Se figur A.1(b). Eftersom



(a) En underskattning av arean av P med hjälp av fyra rektangulära mängder.

(b) En överskattning av arean av P med hjälp av fem rektangulära mängder.

FIGUR A.1 Två uppskattningar av mängden P med enkla mängder.

$F_n \supseteq P$ är $|F_n| \geq |P|$. Vi kan beräkna

$$|F_n| = \sum_{j=1}^{n} \frac{h}{n}\left(\frac{hj}{n}\right)^2 = \frac{h^3}{n^3}\sum_{j=1}^{n} j^2 \underset{\substack{\uparrow \\ (2.6)}}{=} \frac{h^3}{n^3}\frac{n(n+1)(2n+1)}{6}$$

$$= \frac{h^3}{3} + \frac{h^3}{2n}\left(1 + \frac{1}{3n}\right).$$

Vi vet att $\sup_n |E_n| \leq \sup_{E\subseteq P} |E| \leq \inf_{F\supseteq P} |F| \leq \inf_n |F_n|$ där $\sup_{E\subseteq P} |E|$ är supremum taget över alla möjliga enkla mängder $E \subseteq P$ och $\sup_{F\supseteq P} |F|$ är infimum taget över alla möjliga enkla mängder $F \supseteq P$. Vi ser från beräkningar ovan att $\sup_n |E_n| = h^3/3$ och $\inf_n |F_n| = h^3/3$ så $h^3/3 \leq \sup_{E\subseteq P} |E| \leq \inf_{F\supseteq P} |F| \leq h^3/3$ och därför är $\sup_{E\subseteq P} |E| = \inf_{F\supseteq P} |F| = h^3/3$ och därmed är $|P| = h^3/3$.

3.12 Vi skriver

$$p(x) = \sum_{k=0}^{n} a_k x^k,$$

där $a_n \neq 0$, och

$$q(x) = \sum_{j=0}^{m} b_j x^j,$$

där $b_m \neq 0$. Därför är $\deg p = n$ och $\deg q = m$.

Först bevisar vi (b). Enligt antagandet i satsen är $m \neq n$ men genom att eventuellt byta plats på p och q kan vi anta att $n > m$. Sedan beräknar vi att

$$p(x)+q(x) = \sum_{k=0}^{n} a_k x^k + \sum_{j=0}^{m} b_j x^j = \sum_{k=m+1}^{n} a_k x^k + \sum_{j=0}^{m} (a_j + b_j) x^j.$$

Vi ser att högsta potensen av x i $p(x)+q(x)$ är n och eftersom $a_n \neq 0$ är $\deg(p+q) = n = \max\{n,m\}$.

För att bevisa (c) räknar vi att

$$p(x)q(x) = \left(\sum_{k=0}^{n} a_k x^k\right)\left(\sum_{j=0}^{m} b_j x^j\right) = \sum_{k=0}^{n}\sum_{j=0}^{m}(a_k b_j) x^{k+j}.$$

Högsta potensen av x i högerledet är $n+m$ och $a_n b_m \neq 0$ eftersom vi vet att både $a_n \neq 0$ och $b_m \neq 0$. Därför är $\deg(pq) = n + m = \deg p + \deg q$.

För att visa att man kan få en sträng olikhet i (a) kan vi till exempel ta $p(x) = 3x^2 + 2x - 6$ och $q(x) = -3x^2 + 4x$. Sedan är $\deg(p+q) = \deg(6x-6) = 1 < 2 = \max\{2,2\} = \max\{\deg p, \deg q\}$.

4.1 (a) Det är enklare att bevisa kontrapositionen: Om m inte är delbart med 3 då är m^2 inte delbart med 3.

Om m inte är delbart med 3 kan det skrivas antingen som $m = 3k+1$ eller $m = 3k-1$ för något heltal k. Då är

$$m^2 = (3k \pm 1)^2 = 3(3k^2 \pm 2k) + 1$$

så m^2 är inte heller delbart med 3.

(b) Vi ger ett motsägelsebevis: Anta att c är rationellt och därför kan skrivas som $c = n/m$ där n och m är hela tal utan gemensamma delare. Vi beräknar

$$3 = c^2 = \frac{n^2}{m^2} \implies 3m^2 = n^2.$$

Därför är n^2 delbart med 3 och enligt del (a) är n delbart med 3, så $n = 3k$ för något heltal k. Återigen beräknar vi

$$3m^2 = n^2 \text{ och } n = 3k \implies 3m^2 = 9k^2 \implies m^2 = 3k^2.$$

Därför är m^2 delbart med 3 och enligt del (a) är m delbart med 3.

Att både n och m är delbara med 3 är en motsägelse till antagandet att c kan skrivas som $c = n/m$ där n och m är hela tal utan gemensamma delare. Därför är c irrationellt.

4.2 (a) Det är enklare att bevisa kontrapositionen: Om m inte är delbart med 6 då är m^2 inte delbart med 6.

Om m är inte delbart med 6 kan det skrivas som $m = 3k + r$ för något heltal k och $r = 1,2,3,4$ eller 5. Då är

$$m^2 = (3k + r)^2 = 3(3k^2 + 2kr) + r^2.$$

Vi vet att $r^2 = 1$, 4, 9, 16 eller 25 så r^2 är aldrig delbart med 6, och därför är m^2 inte delbart med 6

(b) Vi ger ett motsägelsebevis: Anta att c är rationellt och därför kan skrivas som $c = n/m$ där n och m är hela tal utan gemensamma delare. Vi beräknar

$$6 = c^2 = \frac{n^2}{m^2} \implies 6m^2 = n^2.$$

Därför är n^2 delbart med 6 och enligt del (a) är n delbart med 6, så $n = 6k$ för något heltal k. Återigen beräknar vi

$$6m^2 = n^2 \text{ och } n = 6k \implies 3m^2 = 36k^2 \implies m^2 = 6k^2.$$

Därför är m^2 delbart med 6 och enligt del (a) är m delbart med 6.

Att både n och m är delbara med 6 är ett motsägelse till antagandet att c kan skrivas som $c = n/m$ där n och m är hela tal utan gemensamma delare. Därför är c irrationellt.

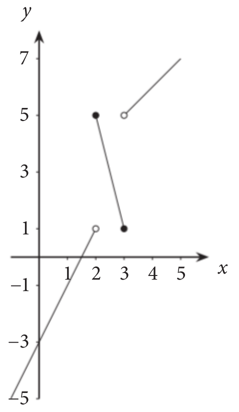
4.3 Till exempel är $m = 6$ inte jämnt delbart med 9 men $m^2 = 6^2 = 36$ är delbart med 9.

4.4 Det finns flera möjliga generaliseringar av sats 4.4. Här är två möjliga svar.

Sats *Låt p vara ett primtal. Då är lösningarna c till $c^2 = p$ irrationella.*

Sats *För olika primtal $p_1 < p_2 < \cdots < p_k$ definiera $y = \prod_{j=1}^{k} p_j$. Då är lösningarna c till $c^2 = y$ irrationella.*

4.5 (a)



(b) Vi betraktar ekvationen

$$f(x) = y \tag{A.5}$$

för de tre relevanta x-intervallen $(-\infty, 2)$, $[2, 3]$ och $(3, \infty)$. I första intervallet får vi att

$$y = 2x - 3 \quad \text{och} \quad x < 2$$

som är ekvivalent med

$$x = \frac{y+3}{2} \quad \text{och} \quad y < 1$$

så för $y < 1$ har ekvationen en unik lösning $x = (y+3)/2$ i $(-\infty, 2)$.

I intervallet $[2, 3]$ får vi att

$$y = -4x + 13 \quad \text{och} \quad 2 \le x \le 3$$

som är ekvivalent med

$$x = \frac{13-y}{4} \quad \text{och} \quad 1 \le y \le 5$$

så för $1 \le y \le 5$ har ekvationen (A.5) en unik lösning $x = (13-y)/4$ i $[2, 3]$.

I intervallet $(3, \infty)$ får vi att

$$y = x + 2 \quad \text{och} \quad x > 3$$

som är ekvivalent med

$$x = y - 2 \quad \text{och} \quad y > 5$$

så för $y > 5$ har ekvationen (A.5) en unik lösning $x = y - 2$ i $(3, \infty)$.

Sammanfattningsvis för varje y i intervallet $(-\infty, 1)$, $[1, 5]$ eller $(5, \infty)$ har (A.5) en unik lösning x i intervallet $(-\infty, 2)$, $[2, 3]$ respektive $(3, \infty)$. Utöver det är både y-intervallen $(-\infty, 1)$, $[1, 5]$ och $(5, \infty)$, och x-intervallen $(-\infty, 2)$, $[2, 3]$ respektive $(3, \infty)$ parvis disjunkta och har en union som är $\mathbf{R}$. Därför för varje $y \in \mathbf{R}$ har (A.5) en unik lösning $x \in \mathbf{R}$ och därmed är f är inverterbar. Enligt våra uträkningar ges inversen av

$$f^{-1}(y) = \begin{cases} (y+3)/2 & \text{om } y < 1; \\ (13-y)/4 & \text{om } 1 \leq y \leq 5; \\ y - 2 & \text{om } y > 5. \end{cases}$$

4.6 (a) Uttrycket $\sqrt{\frac{x+1}{x-1}}$ är definierat om och endast om nämnaren i kvoten är nollskild och hela kvoten är icke-negativ. Nämnaren i kvoten är nollskild precis då $x \neq 1$. Och hela kvoten är icke-negativ när antingen $x+1 \geq 0$ och $x-1 > 0$, eller $x+1 \leq 0$ och $x-1 < 0$. Kraven kan omskrivas till kravet $x > 1$ eller $x \leq -1$. Därför definierar vi $D = (-\infty, -1] \cup (1, \infty)$.

(b) Vi vill hitta alla möjliga värden för y som kan uppstå genom uttrycket

$$y = f(x) := \sqrt{\frac{x+1}{x-1}},$$

där $x \leq -1$ eller $x > 1$. Därför, för ett givet y måste vi hitta en lämplig lösning x som ger detta y: Först beräknar vi att

$$y = f(x) := \sqrt{\frac{x+1}{x-1}} \implies y^2 = \frac{x+1}{x-1}$$
$$\implies (x-1)y^2 = (x+1) \iff x(y^2 - 1) = 1 + y^2.$$

Om $y = \pm 1$ säger den sista likheten att $0 = 2$ som är en motsägelse. Vi har därför visat att det inte finns något x så att $f(x) = \pm 1$. Därför kan varken 1 eller -1 tillhöra målmängden M om f ska vara surjektiv.

Om y varken är -1 eller 1 hittar vi en lämplig kandidat för x genom att dela båda leden med $y^2 - 1$: $x = \frac{y^2+1}{y^2-1}$. Men eftersom vi bara har visat att $x = \frac{y^2+1}{y^2-1}$ var ett nödvändigt villkor för att x är en lösning måste vi verifiera implikationen åt andra hållet:

$$x = \frac{y^2+1}{y^2-1} \implies f(x) = f\left(\tfrac{y^2+1}{y^2-1}\right) = \sqrt{\frac{\frac{y^2+1}{y^2-1}+1}{\frac{y^2+1}{y^2-1}-1}}$$

$$= \sqrt{\frac{(y^2+1)+(y^2-1)}{(y^2+1)-(y^2-1)}} = \sqrt{y^2}$$

och vi får $f\left(\frac{y^2+1}{y^2-1}\right) = y$ om och endast om $y \geq 0$ och $y \neq 1$. Därför drar vi slutsatsen att mängden av möjliga y-värden är $M = [0,1) \cup (1,\infty)$.

(c) Redan i (b) har vi visat att $y = f(x) \implies x = \frac{y^2+1}{y^2-1}$ för $y \in M$, så det finns för varje $y \in M$ högst ett x så att $f(x) = y$. Därför är $f : D \to M$ injektiv.

(d) $f^{-1}(y) = \frac{y^2+1}{y^2-1}$.

4.7 (a) Om $n = 1$ är olikheten uppenbar: $(a+b)^{1/n} = a + b \leq a + b = a^{1/n} + b^{1/n}$.

För $n > 1$ ger vi ett motsägelsebevis: Anta att $(a+b)^{1/n} > a^{1/n} + b^{1/n}$. Det medför att

$$\begin{aligned} a + b = \left[(a+b)^{1/n}\right]^n &> (a^{1/n} + b^{1/n})^n \\ &= \sum_{k=0}^{n} \binom{n}{k} a^{(n-k)/n} b^{k/n} \\ &= a + b + \sum_{k=1}^{n-1} \binom{n}{k} a^{(n-k)/n} b^{k/n} \end{aligned}$$

enligt sats 2.13. Det medför att

$$\sum_{k=1}^{n-1} \binom{n}{k} a^{(n-k)/n} b^{k/n} < 0$$

fast alla termer i summan är positiva. Det är en motsägelse och därför måste $(a+b)^{1/n} \leq a^{1/n} + b^{1/n}$.

(b) Till exempel, $a = b = 1$ ger att $(a+b)^{1/n} = (1+1)^{1/n} = 2^{1/n} < 2 = 1 + 1 = 1^{1/n} + 1^{1/n} = a^{1/n} + b^{1/n}$.

5.1 Genom att använda (5.4), (5.5) och (5.10) kan vi beräkna

$$\begin{aligned}\sin(\theta - \varphi) &= \sin(\theta + (-\varphi))\\ &= \sin\theta\cos(-\varphi) + \cos\theta\sin(-\varphi)\\ &= \sin\theta\cos\varphi - \cos\theta\sin\varphi.\end{aligned}$$

5.2 Om vi adderar (5.8) till (5.9) får vi den första likheten och om vi adderar svaret från uppgift 5.1 till (5.10) får vi den andra.

5.3 Först använder vi (5.12) för att skriva om

$$\begin{aligned}\cos(2\theta) - \cos\theta + 1 &= (1 + \cos(2\theta)) - \cos\theta\\ &= 2\cos^2\theta - \cos\theta\\ &= (2\cos\theta - 1)\cos\theta.\end{aligned}$$

Därför är ekvationen ekvivalent med

$$0 = (2\cos\theta - 1)\cos\theta.$$

En lösning kan därför uppstå på två sätt: Antingen är $\cos\theta = 0$, som är ekvivalent med

$$\theta = \frac{\pi}{2} + k\pi \quad \text{för något } k \in \mathbf{Z};$$

eller

$$\cos\theta = \frac{1}{2},$$

som är ekvivalent med

$$\theta = \pm\frac{\pi}{3} + 2k\pi \quad \text{för något } k \in \mathbf{Z}.$$

Alla lösningarna θ till ekvationen är därför

$$\theta = \pm\frac{\pi}{3} + 2k\pi \quad \text{och} \quad \theta = \frac{\pi}{2} + k\pi \quad \text{för } k \in \mathbf{Z}.$$

5.4 (a) Triangelns höjd h uppfyller både $h = b\sin\theta$ och (enligt Pythagoras sats) $h^2 + (a - b\cos\theta)^2 = c^2$ så

$$\begin{aligned}c^2 &= h^2 + (a - b\cos\theta)^2\\ &= b^2\sin^2\theta + a^2 - 2ab\cos\theta + b^2\cos^2\theta\\ &= a^2 + b^2 - 2ab\cos\theta\end{aligned}$$

enligt (5.3).

(b) Cosinussatsen och figuren säger att

$$c^2 = 4 + 4 - 8\cos(\pi/6) = 2(4 - 2\sqrt{3}) = 2(\sqrt{3} - 1)^2$$

och eftersom c är positivt är $c = \sqrt{2}(\sqrt{3} - 1)$. Därför är

$$\sin\left(\frac{\pi}{12}\right) = \frac{(c/2)}{2} = \frac{\sqrt{2}(\sqrt{3} - 1)}{4} = \frac{\sqrt{3} - 1}{2\sqrt{2}}.$$

Från Pythagoras sats uppfyller triangelns höjd h ekvationen $2^2 = c^2/2^2 + h^2$ så

$$h^2 = 4 - \frac{2(4 - 2\sqrt{3})}{4} = \frac{2(4 + 2\sqrt{3})}{4} = \frac{(\sqrt{3} + 1)^2}{2}$$

och därmed är

$$\cos\left(\frac{\pi}{12}\right) = \frac{h}{2} = \frac{(\sqrt{3} + 1)}{2\sqrt{2}}.$$

5.5 Man kan beräkna att:

(a) $\arcsin\left(\frac{1}{2}\right) = \frac{\pi}{6}$

(b) $\arccos\left(-\frac{\sqrt{2}}{2}\right) = \frac{3\pi}{4}$

(c) $\arccos\left(-\frac{\sqrt{3}}{2}\right) = \frac{5\pi}{6}$

(d) $\arccos\left(\cos\left(\frac{17\pi}{5}\right)\right) = \frac{3\pi}{5}$

(e) $\arctan\left(-\sqrt{3}\right) = -\frac{\pi}{3}$

(f) $\sin\left(\arctan\left(\frac{12}{5}\right)\right) = \frac{12}{13}$

5.6 För att genomföra ett motsägelsebevis antar vi negationen av det vi vill bevisa: Vi antar att det finns ett $\theta \in [0, \pi/2]$ sådant att

$$\cos\theta + \sin\theta < 1.$$

För detta θ vet vi att $\cos\theta \geq 0$ och $\sin\theta \geq 0$ så vi måste även ha att

$$1 > (\cos\theta + \sin\theta)^2 = \cos^2\theta + 2\cos\theta\sin\theta + \sin^2\theta \underset{(5.3)}{=} 1 + 2\cos\theta\sin\theta$$

och därmed är $\cos\theta\sin\theta < 0$ – ett motsägelse till faktumet att $\cos\theta \geq 0$ och $\sin\theta \geq 0$. Därför måste antagandet vara falskt och

$$\cos\theta + \sin\theta \geq 1$$

för alla $\theta \in [0, \pi/2]$.

5.7 Sats 5.1 tillsammans med $\sin 0 = 0$ och $\sin(\pi/2) = 1 < \pi/2$ medför att $\sin\theta \le \theta$ för $\theta \in [0, \pi/2]$. Dessutom är $\sin\theta \ge 0$ för $\theta \in [0, \pi/2]$. Därför är $\sin^2\theta \le \theta^2$ och

$$\cos^2\theta = 1 - \sin^2\theta \ge 1 - \theta^2$$

för $\theta \in [0, \pi/2]$.

Nu visar vi att $\cos\theta \ge 1 - \theta$ för $\theta \in [0, \pi/2]$. Vi betraktar två fall: $\theta \in [1, \pi/2]$ och $\theta \in [0, 1)$.

Om $\theta \in [1, \pi/2]$ är $1 - \theta \le 0$ och vi vet redan att $\cos\theta \ge 0$ så

$$\cos\theta \ge 0 \ge 1 - \theta.$$

Om $\theta \in [0, 1)$ kan vi skriva $1-\theta^2 = (1+\theta)(1-\theta) \ge (1-\theta)(1-\theta) = (1-\theta)^2$ så

$$\cos^2\theta \ge 1 - \theta^2 \ge (1-\theta)^2.$$

Eftersom både $\cos\theta$ och $1-\theta$ är positiva för $\theta \in [0,1)$ drar vi slutsatsen att

$$\cos\theta \ge 1 - \theta.$$

5.8 Det finns flera sätt att lösa problemet. Ett sätt är att jämföra uttrycket med

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

där $C > 0$ och $\varphi \in \mathbf{R}$ (jämför med (5.8)). Då räcker det att hitta C och φ så att

$$\begin{aligned} 34\sqrt{3} &= C\cos\varphi \quad \text{och} \\ 34 &= C\sin\varphi \end{aligned}$$

uppfylls. De här likheterna medför att

$$\begin{aligned} C &\underset{(5.3)}{\overset{}{=}} \sqrt{(C\cos\varphi)^2 + (C\sin\varphi)^2} \\ &= \sqrt{(34\sqrt{3})^2 + (34)^2} = 68 \quad \text{och} \\ \tan\varphi &= \frac{C\sin\varphi}{C\cos\varphi} = \frac{34}{34\sqrt{3}} = \frac{1}{\sqrt{3}}, \end{aligned}$$

så

$$\varphi = \frac{\pi}{6} + k\pi \quad (k \in \mathbf{Z}).$$

Ett lämpligt val av φ som tar hänsyn till tecknen av cosinus och sinus är

$$\varphi = \frac{\pi}{6} \quad \text{och vi fick} \quad C = 68$$

så

$$34\sqrt{3}\cos\theta + 34\sin\theta = 68\cos\left(\theta - \frac{\pi}{6}\right).$$

5.9 Betrakta godtyckliga $x, y \in [0, \pi/2)$ sådana att $x < y$. Vi vet från (c) att $0 = \sin 0 \leq \sin x < \sin y$ och från (a) att $\cos x > 0$, så

$$\frac{\sin x}{\cos x} < \frac{\sin y}{\cos x}.$$

Från (a) och (b) vet vi att $\cos x > \cos y > 0$ så

$$\frac{\sin y}{\cos x} < \frac{\sin y}{\cos y}$$

eftersom $\sin y > 0$. Tillsammans får vi att

$$\tan x = \frac{\sin x}{\cos x} < \frac{\sin y}{\cos x} < \frac{\sin y}{\cos y} = \tan y. \tag{A.6}$$

Därmed är tangens strängt växande på $[0, \pi/2)$.

För att visa tangens är strängt växande på $(-\pi/2, \pi/2)$ står det kvar att betrakta två fall. Först om $x \in (-\pi/2, 0)$ och $y \in [0, \pi/2)$ är

$$\tan x = \frac{\sin x}{\cos x} \underset{\substack{\uparrow \\ \text{(a) \& (c)}}}{<} \frac{\sin 0}{\cos x} = 0 \underset{\substack{\uparrow \\ \text{(a) \& (c)}}}{\leq} \frac{\sin y}{\cos y} = \tan y.$$

Om $x, y \in (-\pi/2, 0)$ och $x < y$ får vi att $-x, -y \in (0, \pi/2)$ och $-y < -x$ så

$$\tan x \underset{\substack{\uparrow \\ (5.4)\ \&\ (5.5)}}{=} -\tan(-x) \underset{\substack{\uparrow \\ (\text{A}.6)}}{<} -\tan(-y) \underset{\substack{\uparrow \\ (5.4)\ \&\ (5.5)}}{=} \tan y.$$

Vi har därmed bevisat att tangens är strängt växande på $(-\pi/2, \pi/2)$.

6.1 Vi beräknar:

(a) $\exp(x+y) = \exp(x)\exp(y) = 4\sqrt{3}\sqrt{3} = 12$;

(b) $\exp(2x+y) = \exp(x)^2\exp(y) = (4\sqrt{3})^2\sqrt{3} = 48\sqrt{3}$;

(c) $\exp(x-3y) = \frac{\exp(x)}{\exp(y)^3} = \frac{4\sqrt{3}}{(\sqrt{3})^3} = \frac{4}{3}$.

6.2 Likheten $\exp(x+4) = \exp(x)+\exp(4)$ är ekvivalent med $\exp(4)\exp(x) = \exp(x)+\exp(4)$ som är ekvivalent med $\exp(x) = \exp(4)/(\exp(4)-1)$ eftersom $\exp(4) \neq \exp(0) = 1$. Så det enda möjliga värdet för $\exp(x)$ är $\exp(4)/(\exp(4)-1)$.

6.3 Vi skriver om (6.8) i fallet $x = 1$: Den säger att

$$e = \sup_{n\in\mathbf{Z}_+}\left(1+\frac{1}{n}\right)^n$$

och därför är e en övre begränsning av $\left(1+\frac{1}{n}\right)^n$ för alla heltal $n \geq 1$. Det betyder att

$$\left(1+\frac{1}{n}\right)^n \leq e \tag{A.7}$$

för alla $n \in \mathbf{Z}_+$.

För den övre begränsningen till e tittar vi på sats 6.3. Det står att det finns en funktion B så att $e \leq B(1)$, men om vi läser beviset ser vi att (6.4) är ett uttryck för $B(x)$. Utifrån (6.4) kan vi räkna ut att

$$B(1) = \left(1-\frac{1}{2}\right)^{-2},$$

så

$$e \leq B(1) = \left(1-\frac{1}{2}\right)^{-2}.$$

Om vi lägger ihop olikheten ovan med (A.7) i fallet $n = 2$ får vi att $9/4 \leq e \leq 4$.

6.4 Man kan underskatta mängden E med mängden F_n som är unionen av n icke-överlappande rektangulära mängder $[hj/n, h(j+1)/n) \times [0, e^{hj/n})$ ($j = 0, 1, \ldots, n-1$). Se figur A.2(a). Eftersom $F_n \subseteq E$ är

$|F_n| \le |E|$. Vi kan beräkna[5]

$$|F_n| = \sum_{j=0}^{n-1} \frac{h}{n} e^{hj/n} = \frac{h}{n} \sum_{j=0}^{n-1} e^{hj/n} \underset{\substack{\uparrow \\ \text{sats 1.14}}}{=} \frac{h}{n} \left(\frac{e^h - 1}{e^{h/n} - 1} \right)$$

$$\underset{\substack{\uparrow \\ \text{sats 6.5(d)}}}{\ge} \frac{h}{n} \frac{e^h - 1}{\left(\left(\frac{1}{1-h/n} \right) - 1 \right)} = \left(e^h - 1 \right) \left(1 - \frac{h}{n} \right).$$

Vi kan också överskatta mängden E med mängden G_n som är unionen av n icke-överlappande rektangulära mängder $[h(j-1)/n, hj/n) \times [0, e^{hj/n})$ $(j = 1, 2, \ldots, n)$. Se figur A.2(b). Eftersom $G_n \supseteq P$ är $|G_n| \ge |P|$. Vi kan beräkna

$$|G_n| = \sum_{j=1}^{n} \frac{h}{n} e^{hj/n} = \frac{h}{n} \sum_{j=1}^{n} e^{hj/n}$$

$$\underset{\substack{\uparrow \\ \text{sats 1.14}}}{=} \frac{he^{h/n}}{n} \left(\frac{e^h - 1}{e^{h/n} - 1} \right) = \frac{h}{n} \left(\frac{e^h - 1}{1 - e^{-h/n}} \right)$$

$$\underset{\substack{\uparrow \\ \text{sats 6.5(d)}}}{\le} (e^h - 1) \left(1 + \frac{h}{n} \right).$$

Vi vet att $\sup_n |F_n| \le \sup_{F \subseteq E} |F| \le \inf_{G \supseteq E} |G| \le \inf_n |G_n|$ där $\sup_{F \subseteq E} |F|$ är supremum taget över alla möjliga enkla mängder $F \subseteq E$ och $\inf_{G \supseteq E} |G|$ är infimum taget över alla möjliga enkla mängder $G \supseteq E$. Men beräkningarna ovan medför att $\sup_n |F_n| \ge e^h - 1$ och $\inf_n |G_n| \le e^h - 1$ så $\sup_{F \subseteq E} |F| = \inf_{G \supseteq E} |G| = e^h - 1$ och därmed är $|E| = e^h - 1$. (Jämför med uppgift 3.11.)

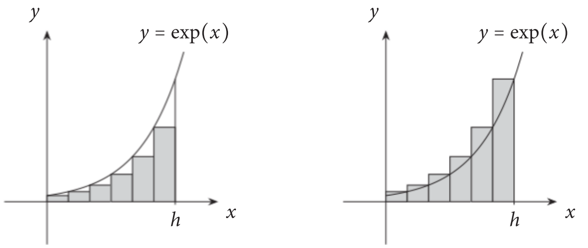
6.5 För att visa att funktionen är bijektiv måste man visa att för varje $y \in [1, \infty)$ finns det en unik $x \in [0, \infty)$ så att

$$y = \frac{e^x + e^{-x}}{2}. \tag{A.8}$$

Likheten kan skrivas om som

$$0 = e^x - 2y + e^{-x}.$$

5 För att använda sats 6.5(d) behöver vi kontrollera att $h/n < 1$. Men i fallet $h/n \ge 1$ får vi samma olikhet eftersom $|F_n| \ge 0 \ge \left(e^h - 1\right)\left(1 - \frac{h}{n}\right)$.



(a) En underskattning av arean av E med hjälp av rektangulära mängder.

(b) En överskattning av arean av E med hjälp av rektangulära mängder.

FIGUR A.2 Två uppskattningar av mängden E med enkla mängder.

Men eftersom $e^x > 0$ kan vi multiplicera likheten med e^x och ser att den är ekvivalent med

$$0 = e^{2x} - 2ye^x + 1,$$

som är ett andragradspolynom i e^x och kan kvadratkompletteras:

$$0 = (e^x - y)^2 + 1 - y^2.$$

Eftersom $y \geq 1$ är möjliga värden av e^x därför

$$y + \sqrt{y^2 - 1} \quad \text{och} \quad y - \sqrt{y^2 - 1}.$$

För att sätta dem lika med e^x där $x \in [0, \infty)$ krävs att

$$y \pm \sqrt{y^2 - 1} \geq 1. \tag{A.9}$$

Båda alternativen är förstås lika om $y = 1$ men om $y > 1$ utesluter kravet (A.9) värdet $y - \sqrt{y^2 - 1}$ eftersom om $y - \sqrt{y^2 - 1} \geq 1$ och $y > 1$ är

$$\begin{aligned} 0 > y - 1 \geq \sqrt{y^2 - 1} &\implies (y-1)^2 \geq y^2 - 1 \\ &\implies y^2 - 2y + 1 \geq y^2 - 1 \\ &\implies 2 \geq 2y \implies y \leq 1 \end{aligned}$$

som är motsägelsefullt med antagandet att $y > 1$. Därför för både $y = 1$ och $y > 1$ är

$$e^x = y + \sqrt{y^2 - 1}$$

och därmed är (A.8) ekvivalent med att

$$x = \ln\left(y + \sqrt{y^2 - 1}\right).$$

För varje $y \in [1, \infty)$ har vi bevisat att (A.8) har en unik lösning $x \in [0, \infty)$ och därför är cosh en bijektiv funktion.

Räkningen visar också att inversen är

$$\cosh^{-1}(x) = \ln\left(x + \sqrt{x^2 - 1}\right).$$

6.6 (a) För $x > 1$ är

$$\begin{aligned}
&\exp(2\ln(\sqrt{x+1}) + \ln(x-1)) \\
&= \exp(\ln(x+1) + \ln(x-1)) \\
&= \exp(\ln((x+1)(x-1))) \\
&= (x+1)(x-1) = x^2 - 1.
\end{aligned}$$

(b) För $x \in \mathbf{R}$ är

$$\begin{aligned}
&3\ln(\exp(1+x)\exp(1-x)) \\
&= 3\ln(\exp(1+x+1-x)) \\
&= 3\ln(\exp(2)) = 6.
\end{aligned}$$

6.7 (a) För att visa att $x \mapsto a^x$ är bijektiv måste man till varje $y \in (0, \infty)$ hitta exakt ett $x \in \mathbf{R}$ så att $y = a^x$. Men $a^x = \exp(x \ln a)$ och exp är ju bijektiv så

$$y = \exp(x\ln a) \iff \ln y = x \ln a \underset{\ln a \neq 0}{\iff} x = \frac{\ln y}{\ln a}.$$

Därför till varje $y \in (0, \infty)$ har ekvationen $y = a^x$ precis en lösning – $x = \ln y / \ln a$ – så $x \mapsto a^x$ är bijektiv.

(b) Vår räkning ovan visar att inversen är

$$\log_a(x) = \frac{\ln x}{\ln a}$$

för alla $x > 0$ så $K = 1/\ln a$.

(c) Vi beräknar

$$\log_a(xy) = \frac{\ln(xy)}{\ln a} \underset{\text{sats 6.20(e)}}{\uparrow}{=} \frac{\ln x + \ln y}{\ln a} = \frac{\ln x}{\ln a} + \frac{\ln y}{\ln a}$$
$$= \log_a(x) + \log_a(y).$$

7.1 Man kan räkna ut att:

(a) $(2+3i)+(6-2i) = 8+i$; (d) $\frac{-4+6i}{1-2i} = -\frac{16}{5} - \frac{2}{5}i$;

(b) $(-2+5i)-(1-7i) = -3+12i$; (e) $6i(1+3i) = -18+6i$; och

(c) $(2-8i)(1+3i) = 26-2i$; (f) $\frac{1-3i}{1+3i} = -\frac{4}{5} - \frac{3}{5}i$.

7.2 Betrakta två komplexa tal $w = u + iv$ och $z = x + iy$ för $u,v,x,y \in \mathbf{R}$. Vi beräknar att

$$\overline{w}\,\overline{z} = (u - iv)(x - iy) = (ux - vy) - i(vx + uy)$$

och

$$\overline{wz} = \overline{(u+iv)(x+iy)} = \overline{(ux - vy) + i(vx + uy)}$$
$$= (ux - vy) - i(vx + uy)$$

så $\overline{wz} = \overline{w}\,\overline{z}$.

Vi kan också beräkna att

$$\overline{w+z} = \overline{(u+iv) + (x+iy)} = \overline{(u+x) + i(v+y)}$$
$$= (u+x) - i(v+y) = (u - iv) + (x - iy) = \overline{w} + \overline{z}.$$

7.3 Vi kan skriva $P(z) = \sum_{j=0}^{n} a_j z^j$ för något $n \in \mathbf{Z}_+$ och $a_j \in \mathbf{R}$ ($j = 0, \dots, n$). Vi vet att

$$0 = P(z) = \sum_{j=1}^{n} a_j z^j$$

för något $z \in \mathbf{C}$. Genom att ta konjugatet av likheten ovan får vi att

$$0 = \overline{0} = \overline{P(z)} = \overline{\sum_{j=0}^{n} a_j z^j} = \sum_{j=0}^{n} \overline{a_j z^j}$$
$$= \sum_{j=0}^{n} \overline{a}_j \overline{z}^j = \sum_{j=0}^{n} a_j \overline{z}^j = P(\overline{z})$$

så $P(\overline{z}) = 0$.

7.4 (a) Vi räknar ut att

$$e^{i\pi} + 1 = (\cos\pi + i\sin\pi) + 1 = -1 + 0 + 1 = 0.$$

(c) Vi kan också beräkna att

$$\begin{aligned} e^{i\theta} + e^{-i\theta} &= (\cos\theta + i\sin\theta) + (\cos(-\theta) + i\sin(-\theta)) \\ &= (\cos\theta + i\sin\theta) + (\cos\theta - i\sin\theta) = 2\cos\theta. \end{aligned}$$

(d) Och även att

$$\begin{aligned} e^{i\theta} - e^{-i\theta} &= (\cos\theta + i\sin\theta) - (\cos(-\theta) + i\sin(-\theta)) \\ &= (\cos\theta + i\sin\theta) - (\cos\theta - i\sin\theta) = 2i\sin\theta. \end{aligned}$$

7.5 Man kan räkna ut att:

(a) $\sqrt{3} + i = 2e^{\pi i/6}$;

(b) $(1+i)^4 = (\sqrt{2}e^{\pi i/4})^4 = 4e^{\pi} = -4$;

(c) $(-12+5i)+(3+4i) = -9+9i = 9\sqrt{2}e^{3\pi i/4}$; och

(d) $3e^{i8\pi/3} \times 7e^{i\pi/3} = 21e^{i\pi} = -21$.

7.6 Enligt sats 7.14 är medelpunkten $(35/8, -7/2)$ och radien är $15/8$.

7.7 (a) För varje $w \in B$ vill vi visa att det finns en entydig lösning $z \in A$ till

$$\exp(z) = w. \tag{A.10}$$

Vi skriver $z = x + iy$ där $x \in \mathbf{R}$ och $y \in [-\pi, \pi)$. Om vi tar absolutbeloppet av (A.10) får vi att

$$|w| = |\exp(z)| = |\exp(x)\exp(iy)| = |\exp(x)| = \exp(x)$$

eftersom $|\exp(iy)| = 1$ för $y \in [-\pi, \pi)$ och $\exp(x) > 0$ för $x \in \mathbf{R}$. Därför medför (A.10) att $x = \ln|w|$. Om vi tittar på argumentet av båda leden i (A.10) ser vi att

$$\arg(w) = \arg(\exp(z)) = \arg(\exp(iy)) = y$$

och eftersom vi letar efter $z \in A$ måste $y = \operatorname{Arg}(w)$.

Vi har visat att om z löser (A.10) måste $z = \ln|w| + i\operatorname{Arg}(w)$ och man kan lätt verifiera att $z = \ln|w| + i\operatorname{Arg}(w)$ löser (A.10):

$$\begin{aligned}\exp(z) &= \exp(\ln|w| + i\operatorname{Arg}(w)) \\ &\underset{\text{sats 7.10}}{=} e^{\ln|w|}(\cos(\operatorname{Arg}(w)) + i\sin(\operatorname{Arg}(w))) \\ &= |w|\cos(\operatorname{Arg}(w)) + i|w|\sin(\operatorname{Arg}(w)) \underset{(7.7)}{=} w\end{aligned}$$

Så vi har visat att $\exp\colon A \to B$ är inverterbar.

(b) Från uträkningarna ovan ges inversen till $\exp\colon A \to B$ av formeln

$$\exp^{-1}(w) = \ln|w| + i\operatorname{Arg}(w)$$

för alla $w \in B$.

7.8 För varje $w \in \mathbf{C} \setminus \{a/c\}$ behöver vi visa att ekvationen $w = f(z)$ har en entydig lösning $z \in \mathbf{C} \setminus \{-d/c\}$. Vi börjar med att för $w \in \mathbf{C} \setminus \{a/c\}$ beräkna

$$\begin{aligned}w = f(z) &\implies w = \frac{az+b}{cz+d} \implies w(cz+d) = az+b \\ &\implies z(-cw+a) = dw-b \underset{w \neq a/c}{\implies} z = \frac{dw-b}{-cw+a}\end{aligned}$$

så det finns för varje $w \in \mathbf{C} \setminus \{a/c\}$ högst en lösning och f är därmed injektiv.

Vi kan vända på alla implikationspilar i beräkningen ovan så länge vi vet att valet $z := (dw-b)/(-cw+a)$ inte kan vara lika med $-d/c$ (och därmed också tillhör definitionsmängden). Vi kontrollerar vad som händer i fallet $z = -d/c$:

$$-\frac{d}{c} = z = \frac{dw-b}{-cw+a} \implies dw - \frac{ad}{c} = dw - b \implies 0 = ad - bc.$$

Det är motsägelsefullt med antagandet $ad - bc \neq 0$ och därför måste

$z \neq -d/c$. Vi kan därmed beräkna att

$$\begin{aligned} z = \frac{dw - b}{-cw + a} &\implies z(-cw + a) = dw - b \\ &\implies w(cz + d) = az + b \\ &\underset{\substack{\uparrow \\ z \neq -d/c}}{\implies} w = \frac{az + b}{cz + d} \implies w = f(z) \end{aligned}$$

och f är också surjektiv.

Så $f\colon \mathbf{C} \smallsetminus \{-d/c\} \to \mathbf{C} \smallsetminus \{a/c\}$ är bijektiv och inversen $f^{-1}\colon \mathbf{C} \smallsetminus \{a/c\} \to \mathbf{C} \smallsetminus \{-d/c\}$ ges av formeln

$$f^{-1}(z) = \frac{dz - b}{-cz + a}.$$

Det grekiska alfabetet

Inom matematik används det grekiska alfabetet ofta. Tabell B.1 är en lista över de grekiska bokstäverna.

TABELL B.1 Det grekiska alfabetet.

Bokstav	Namn	Bokstav	Namn
α, A	alpha	ν, N	ny
β, B	beta	ξ, Ξ	xi
γ, Γ	gamma	o, O	omikron
δ, Δ	delta	π, Π	pi
ϵ eller ε, E	epsilon	ρ, P	rho
ζ, Z	zeta	σ, Σ	sigma
η, H	eta	τ, T	tau
θ, Θ	theta	υ, Υ	ypsilon
ι, I	iota	ϕ eller φ, Φ	fi
κ, K	kappa	χ, X	chi
λ, Λ	lambda	ψ, Ψ	psi
μ, M	my	ω, Ω	omega

Ytterligare material

C.1 Ett axiomsystem för de reella talen

Vår utgångspunkt i boken var att vi kunde utföra aritmetiska operationer och känna till alla enkla räkneregler, men vi skrev inte explicit vad vi menade med det. Det är förstås helt möjligt att göra. Här ger vi ett smakprov av hur det kan gå till.

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

Axiom C.1 (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);*

(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).*

Här är $-a$ och a^{-1} symboler för lösningar till $a+(-a)=0$ och $a\times a^{-1}=1$. Det skiljer sig till exempel från hur vi definierade a^{-1} i (1.16) där division implicit förstås. Här definierar vi division nedan i definition C.3.

Axiom C.2 (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* $ac<bc$.

Axiomen säger inte att reella tal finns, utan de är bara tillräckliga för att härleda alla algebraiska och ordningsegenskaper hos reella tal. Under 1800-talet gav Richard Dedekind en konstruktion av reella tal från de rationella som i sin tur kan konstrueras från naturliga tal. Det finns många talmängder som lyder axiom C.1 – till exempel **Q**, **R** och **C** – och en sådan mängd tillsammans med sina kompositionsregler kallas för en *kropp*. Rationella tal och reella tal lyder både axiom C.1 och C.2 och sådana kroppar kallas för *ordnade kroppar*.

Däremot är **R** den enda ordnade kroppen som lyder fullständighetsaxiomet (axiom 2.28). Det andra axiomet vi betraktade, induktionsaxiomet (axiom 2.6), ses vanligtvis som en egenskap hos naturliga tal (eller likväl positiva heltal) – man kan nå vilket naturligt tal som helst genom att börja i 0 och hoppa från ett naturligt tal till nästa. Axiom 2.6 är ett av de så kallade *Peanos axiom*.

Nedan ger vi några exempel på hur man kan härleda egenskaper hos reella tal från axiom C.1, C.2 och 2.28. Vi börjar med att definiera vad vi menar med subtraktion och division.

Definition C.3 Givet två reella tal a och b definieras operationerna *subtraktion* och *division* enligt:
(a) $a - b := a + (-b)$; respektive
(b) $a/b := a(b^{-1})$ 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}.

Från definition C.3, och axiom C.1 och C.2 kan vi härleda räkneregler för subtraktion och division.

Sats C.4 *För reella tal a, b och c medför påståendet $a = b$ att*
(a) $a + c = b + c$;
(b) $a - c = b - c$;
(c) $ac = bc$; *och*
(d) $a/c = b/c$ *om* $c \neq 0$.

Bevis. Vi bevisar bara (a) och lämnar de andra som övning åt läsaren.

Enligt axiom C.2(a) har vi tre möjligheter och exakt en av dem 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 axiom C.1(e) får vi använda axiom C.2(c) med talet $-c$ och får

$$(a + c) + (-c) < (b + c) + (-c). \tag{C.1}$$

Då är vänsterledet

$$(a+c)+(-c) \underset{\substack{\uparrow \\ \text{ax. C.1(b)}}}{=} a+(c+(-c)) \underset{\substack{\uparrow \\ \text{ax. C.1(e)}}}{=} a+0 \underset{\substack{\uparrow \\ \text{ax. C.1(d)}}}{=} a \tag{C.2}$$

och högerledet är

$$(b+c)+(-c) \underset{\substack{\uparrow \\ \text{ax. C.1(b)}}}{=} b+(c+(-c)) \underset{\substack{\uparrow \\ \text{ax. C.1(e)}}}{=} b+0 \underset{\substack{\uparrow \\ \text{ax. C.1(d)}}}{=} b. \tag{C.3}$$

Då får vi skriva om (C.1) som $a < b$, vilket är en motsägelse till $a = b$ enligt axiom C.2(a).

Sedan antar vi att $a + c > b + c$. Enligt axiom C.1(e) får vi använda axiom C.2(c) med talet $-c$ och får

$$(a+c)+(-c) > (b+c)+(-c). \tag{C.4}$$

Enligt (C.2) och (C.3) får vi skriva om (C.4) som $a > b$, vilket också är en motsägelse till $a = b$ enligt axiom C.2(a).

Därför är den enda möjligheten kvar från axiom C.2(a) att $a+c = b+c$ och (a) är bevisat. □

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

Exempel C.5 Bevisa att $a0 = 0$ för alla reella tal a.

Lösning. Enligt axiom C.1(d) (med $a = 0$) är $0 = 0 + 0$ och så $a0 = a(0+0) = a0 + a0$ enligt axiom C.1(c). Sedan medför axiom C.1(b), C.1(d) och C.1(e) att $0 = a0 + (-(a0)) = (a0 + a0) + (-(a0)) = a0 + (a0 + (-(a0))) = a0$. □

Exempel C.6 Bevisa att $(b+c)a = ba + ca$ för alla reella tal a, b och c.

Lösning. Enligt axiom C.1(a) är $(b+c)a = a(b+c)$ och $a(b+c) = ab+ac$ enligt axiom C.1(c). Sedan medför axiom C.1(a) att $ab + ac = ba + ca$. Därför kan vi dra slutsatsen att $(b+c)a = ba + ca$. □

Exempel C.7 Axiom C.1(e) säger att för ett givet reellt 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

$$a + b = 0. \tag{C.5}$$

Lösning. Anta att $b = b_1$ och $b = b_2$ är två tal som uppfyller (C.5). Vi vill bevisa att $b_1 = b_2$. Från axiom C.2(b) är $a + b_1 = a + b_2$. Enligt axiom C.1(a) vet vi att det medför att $b_1 + a = b_2 + a$ och sedan medför axiom C.1(d), C.1(e) och C.1(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 C.8 Bevisa att $a(-b) = -(ab)$ för alla reella tal a och b.

Lösning. Eftersom vi vet från exempel C.7 att inversen $-(ab)$ är unik, räcker det att visa att $a(-b)$ också är ett inverst element till ab, det vill säga visa att $ab + a(-b) = 0$. Enligt axiom C.1(c), axiom C.1(d) och exempel C.5 vet vi att $ab + a(-b) = a(b + (-b)) = a0 = 0$. □

Exempel C.9 Bevisa att $a < b$ och $c < 0$ medför att $bc < ac$.

Lösning. Observera först att det enligt axiom C.1(e) finns ett $-c$ så att $c + (-c) = 0$. Sedan om $c < 0$ är $0 < -c$ enligt axiom C.2(c) och C.1(d). Därför får vi använda axiom C.2(d) och säga att $a < b$ medför $a(-c) < b(-c)$. Enligt exempel C.8 är $-(ac) < -(bc)$ och sedan $bc = bc + ac - ac < bc + ac - bc = ac + bc - bc = ac$. □

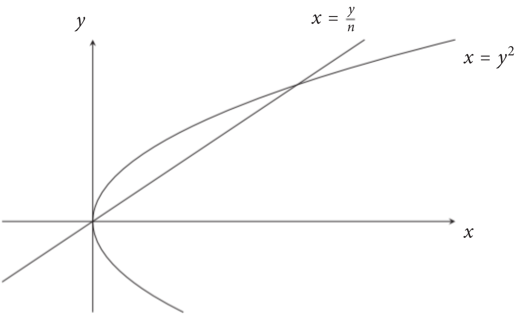
C.2 En mängd utan den arkimediska egenskapen

Den arkimediska egenskapen (sats 2.26) är egentligen en egenskap hos reella och naturliga tal. En tolkning av sats 2.26 (a) är att det inte finns något ”oändligt litet reellt tal”. För att bättre förstå varför vi behöver en sådan sats – som kan kännas så självklar att man knappt behöver bevisa den – är det nyttigt att betrakta ett exempel av en mängd där den inte gäller.

Exempel C.10 Betrakta mängden

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

av alla första- och andragradspolynom i variabeln y som går genom origo tillsammans med det triviala polynomet. Vi kan definiera operationerna



FIGUR C.1 Enligt ordningen $\lesssim$ sitter polynomet $p(y) = y^2$ mellan den triviala polynomet och alla $q_{1/n}(y) = y/n$ ($n \in \mathbf{Z}_+$). Det ses i figuren via lutningarna i origo av y-axeln, och graferna $x = y^2$ och $x = y/n$ ($n \in \mathbf{Z}_+$).

addition och subtraktion på element i P punktvis (det vill säga enligt (3.1)). (Vi bryr oss inte om multiplikation och division här.) Vi definierar också en ordning $\lesssim$ genom följande regel: För polynom $p_1(y) = a_1 y + b_1 y^2$ och $p_2(y) = a_2 y + b_2 y^2$ i P säger vi att

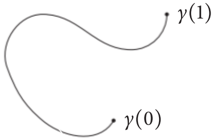
$$p_1(y) \lesssim p_2(y) \text{ om } \begin{cases} \text{antingen } a_1 < a_2, \\ \text{eller } a_1 = a_2 \text{ och } b_1 < b_2. \end{cases}$$

Vi kan betrakta reella tal som en delmängd av P genom motsvarigheten att ett reellt tal α representeras som polynomet $q_\alpha(y) = \alpha y$. Observera att $\alpha_1 < \alpha_2$ om och endast om $q_{\alpha_1}(y) \lesssim q_{\alpha_2}(y)$. Dessutom uppfyller mängden P de vanliga aritmetiska och ordningsregler som gäller addition och subtraktion med $\lesssim$ i stället för $<$ (se axiom C.1 och C.2 i bilaga C.1).

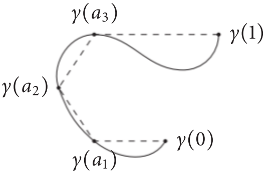
Däremot uppfyller P inte sats 2.26 (a) om a kan vara ett element i P i stället för $\mathbf{R}$ eftersom till exempel polynomet $p(y) = y^2$ är så att $0 \lesssim p(y) \lesssim q_{1/n}(y)$ för alla $n \in \mathbf{Z}_+$ fast $p(y)$ inte är det triviala polynomet. Se figur C.1.

C.3 Kurvor och längd

Cirklar och linjer är exempel på kurvor. Med begreppet ”en kurva” vill vi matematiskt fånga idén att rita med en penna på papper. Vi definierar en



(a) En kurva är värdemängden av en funktion $\gamma\colon[0,1] \to \mathbf{R}^2$.



(b) En uppskattning av en kurvas längd med hjälp av partitionen $\{0,a_1,a_2,a_3,1\}$.

FIGUR C.2 Kurvor i planet.

kurva som värdemängden av en funktion $\gamma\colon[0,1] \to \mathbf{R}^2$. Se figur C.2(a). Funktionen γ beskriver inte bara kurvan på pappret utan också hur man ritar den och den kallas för en *parametrisering* av kurvan. I synnerhet ärver kurvan en ordning på sina element från ordningen som redan finns på intervallet $[0,1]$. Självklart finns det flera parametriseringar som beskriver samma kurva.

När vi ritar en cirkel 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 mäter längden av en kurva: Vi vet bara hur det skulle fungera för en rät linje mellan två punkter där vi tar längden att vara avståndet mellan de två punkterna. Andra kurvor är mer komplicerade. Men vi kan uppskatta längden genom att bryta upp kurvan i små streck och uppskatta längden av varje streck med en rät linje. Välj först 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

$$\sum_{i=1}^{N} \ell(\gamma(a_{i-1}),\gamma(a_i)) \tag{C.6}$$

en uppskattning av kurvans längd, där vi kommer ihåg (3.21) som ger oss en formel för avståndet ℓ mellan två punkter i planet. Eftersom vi underskattar längden av varje litet streck är (C.6) alltid en underskattning av kurvans längd. Men genom att dela upp kurvan i tillräckligt små streck känns det som vi kan uppskatta längden hur exakt vi vill. Därför definierar vi kurvans längd att vara

En rak linje mellan två punkter är den kortaste vägen mellan punkterna. Därför är avståndet mellan två punkter en underskattning av längden av en kurva mellan samma punkter.

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

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

C.4 Motivation bakom komplexa tal

När vi utredde komplexa tal tog vi de vanliga aritmetiska räknereglerna inklusive (7.2) för givet. Här vill vi motivera varifrån (7.2) kommer genom att tänka på aritmetik med komplexa tal som en utvidgning av aritmetik på reella tal.

Därför ställer vi frågan:

> Finns det ett sätt att utöka aritmetiska operationer från den reella linjen till planet?

För att vara mer exakt, betraktar vi den reella linjen som delmängden

$$\{(x,y) \mid x \in \mathbf{R} \text{ och } y = 0\} \tag{C.7}$$

i planet $\mathbf{R}^2$, så punkten $(x,0) \in \mathbf{R}^2$ representerar det reella talet $x \in \mathbf{R}$. Vi frågar om det är möjligt att definiera addition och multiplikation för element i $\mathbf{R}^2$ så att:

(a) Definitionerna stämmer överens med de vanliga aritmetiska operationerna med reella tal för delmängden (C.7) sett som en kopia av $\mathbf{R}$;

(b) Definitionerna medför att godtyckliga $a, b, c \in \mathbf{R}^2$ lyder axiom C.1.

Vi insisterar inte på att man kan jämföra två element i $\mathbf{R}^2$. Man kan faktiskt bevisa att det är omöjligt för axiom C.2 att gälla för komplexa tal.

KARTESISKA KOORDINATER OCH ADDITION I PLANET

Observera att när vi skriver om addition och multiplikation av reella tal sedd som kompositionsregler på mängden (C.7) får vi

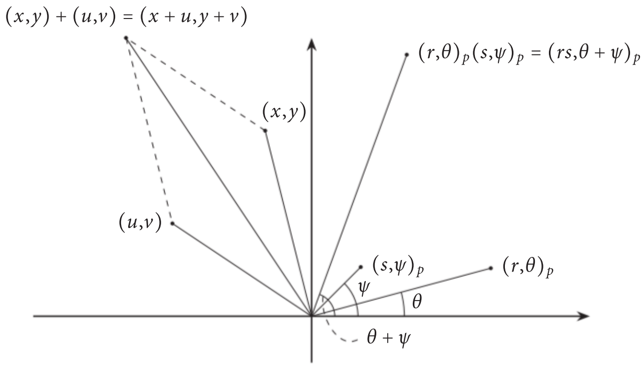
$$(x,0) + (a,0) := (x + a,0) \quad \text{och}$$
$$(x,0)(a,0) := (xa,0)$$

för alla reella $x,a \in \mathbf{R}$. Därför är det naturligt att gissa definitionen

$$(x,y) + (u,v) := (x + u, y + v) \tag{C.8}$$

för addition i $\mathbf{R}^2$. Se figur C.3.

I (C.8) har vi gett en definition av additionen i $\mathbf{R}^2$ som bygger på den bekanta aritmetiska operationen av addition i $\mathbf{R}$. Därför är det möjligt



FIGUR C.3 På vänstersidan finns en illustration av addition i planet. På högersidan finns en illustration av multiplikation i planet.

att verifiera axiomen för den nya additionen i $\mathbf{R}^2$ utifrån axiomen som vi redan tror på för addition i $\mathbf{R}$. Nu kontrollerar att (C.8) lyder allt som gäller addition i axiom C.1:

$$(x,y)+(u,v)=(x+u,y+v)=(u+x,v+y)=(u,v)+(x,y)$$

och

$$\begin{aligned}((x,y)+(u,v))+(r,s) &= (x+u,y+v)+(r,s)\\ &= (u+x+r,v+y+s)\\ &= (x,y)+(u+r,v+s)\\ &= (x,y)+((u,v)+(r,s)),\end{aligned}$$

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 axiom C.1(a), C.1(b) och C.1(d) är bevisat. 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 är det som gäller addition i axiom C.1(e) bevisat.

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 $(x,y)(1,0) = (x,0)$ där axiom C.1(d) kräver att $(x,y)(1,0) = (x,y)$. Så vi måste fundera lite noggrannare på det.

En bättre approach är att betrakta hur saker och ting ser ut i polära koordinater. Varje punkt identifieras med hjälp av två tal, precis som i det kartesiska koordinatsystemet, men här är det första talet avståndet r från punkten till origo och det andra talet θ är vinkeln mellan x-axeln och strecken mellan punkten och origo. Vi skriver

$$(r,\theta)_p$$

med ett index 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 $(r,\theta)_p$. (Jämför med (7.7).)

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 gissar definitionen

$$(r,\theta)_p(s,\varphi)_p := (rs,\theta+\varphi)_p \tag{C.9}$$

för multiplikation i planet. Se figur C.3. Precis som vi gjorde för (C.8) kan vi utifrån axiomen för den bekanta aritmetiska operationen av multiplikation i **R** verifiera att (C.9) lyder allt som gäller multiplikation i axiom C.1. För att göra detta är det praktiskt att först räkna ut hur definitionen (C.9) ser ut i kartesiska koordinater. Betrakta godtyckliga punkter $(x,y) = (r,\theta)_p$ och $(u,v) = (s,\varphi)_p$. Då inhämtar vi att

$$\begin{aligned}
&(x,y)(u,v)\\
&= (r\cos\theta, r\sin\theta)(s\cos\varphi, s\sin\varphi)\\
&\underset{(C.9)}{\uparrow}{=} (rs\cos(\theta+\varphi), rs\sin(\theta+\varphi))\\
&\underset{(5.9)\ \&\ (5.10)}{\uparrow}{=} (rs(\cos\theta\cos\varphi - \sin\theta\sin\varphi), rs(\cos\theta\sin\varphi + \sin\theta\cos\varphi))\\
&= (xu - yv, xv + yu)
\end{aligned}$$

så i kartesiska koordinater är vår gissning

$$(x,y)(u,v) = (xu - yv, xv + yu). \tag{C.10}$$

Utifrån (C.10) kan vi verifiera alla återstående axiom:

$$(x,y)(u,v) = (xu-yv, xv+yu) = (ux-vy, vx+uy) = (u,v)(x,y)$$

och

$$\begin{aligned}
&((x,y)(u,v))(r,s)\\
&= (xu - yv, xv + yu)(r,s)\\
&= ((xu - yv)r - (xv + yu)s, (xu - yv)s + (xv + yu)r)\\
&= (x(ur - vs) - y(us + vr), x(us + vr) + y(ur - vs))\\
&= (x,y)(ur - vs, us + vr) = (x,y)((u,v)(r,s)),
\end{aligned}$$

samt

$$(x,y)(1,0) = (1x - 0y, 0x + 1y) = (x,y)$$

för alla $(x,y), (u,v) \in \mathbf{R}^2$, så vi har verifierat vad gäller multiplikation i axiom C.1(a), C.1(b) och C.1(d). 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 även axiom C.1(c) har verifierats. Det enda axiom som står kvar är existensen av en multiplikativ invers i axiom C.1(e). Vi senarelägger till (C.12) kontrollen av den så att vi kan införa lite notation. Kom ihåg från (C.7) 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

en punkt i planet (x,y) med $x+iy$.

Med denna notation är all aritmetisk räkning enklare att komma ihåg. Nu är aritmetisk räkning precis detsamma som räkning med reella tal eftersom talsystemet uppfyller samma axiom. Om vi bara kommer ihåg att

$$i^2=-1 \quad \text{eftersom enligt (C.10) är} \quad (0,1)(0,1)=(-1,0)$$

är det lätt att se

$$
\begin{aligned}
(x+iy)+(u+iv)&=(u+v)+i(y+v) \quad \text{och}\\
(x+iy)(u+iv)&=ux+xvi+yui+yvi^2\\
&=ux+xvi+yui+yv(-1)\\
&=(ux-yv)+i(xv+yu)
\end{aligned}
\tag{C.11}
$$

precis som vi såg i (C.8) och (C.10) med tyngre notation. Mängden $\mathbf{R}^2$ rustad med notationen $(1,0)=1$ och $(0,1)=i$, och additions- och multiplikationsregler (C.11) är vad vi i kapitel 7 kallade för det *komplexa planet*. Det betecknas

$$\mathbf{C}:=\{x+iy\,|\,x,y\in\mathbf{R}\}.$$

Med notationen från kapitel 7 kan vi använda oss av (7.5) för att se

$$z\frac{\bar{z}}{|z|^2} = 1 \tag{C.12}$$

om $z \neq 0$, så vi har hittat en multiplikativ invers

$$z^{-1} := \frac{\bar{z}}{|z|^2} = \frac{x - iy}{x^2 + y^2}$$

till z för alla $z \neq 0$ och axiom C.1(e) är äntligen bekräftat.

LITTERATURFÖRTECKNING

[1] R.A. Adams & C. Essex, *Calculus*, åttonde upplagan, Pearson, Toronto 2014.

[2] A. Asratian, B.O. Turesson & A. Björn, *Diskret matematik*, Liber, Stockholm 2020.

[3] C.B. Boyer & U.C. Merzbach, *A History of Mathematics*, tredje upplagan, John Wiley & Sons, New Jersey 2011.

[4] G. Chaucer, *The Canterbury Tales and Other Poems*, (red. D. Laing Purves), Lerner Publishing Group, Minneapolis 2018.

[5] E. Cheng, *Cakes, Custard and Category Theory: Easy recipes for understanding complex maths*, Profile Books, London 2015.

[6] I.M. Copi & C. Cohen, *Introduction to Logic*, trettonde upplagan, Pearson Education, New Jersey 2009.

[7] G. Forsling & M. Neymark, *Matematisk analys: En variabel*, andra upplagan, Liber AB, Stockholm 2011.

[8] G.H. Hardy, *A Mathematician's Apology*, Cambridge University Press, Canto utg., Cambridge 1992.

[9] L. Hellström, S. Morander & A. Tengstrang, *Envariabelanalys*, Studentlitteratur, Lund 1991.

[10] C. Hyltén-Cavallius & L. Sandgren, *Matematisk analys II*, Håkan Ohlssons Boktryckeri, Lund 1964.

[11] V.J. Katz, *A History of Mathematics: An introduction*, Harper Collins College Publishers, New York 1993.

[12] E. Nardi, *Tensions in the novice mathematicians induction to mathematical abstraction*, i L. Puig & A. Gutierrez, (red.), Proceeding of the 20th Conference of the International Group for the Psychology of Mathematics Education, 4, (1996) s. 51–58.

[13] H. Petersson, *Undersökande matematik: Differentierade problem*, Studentlitteratur, Lund 2017.

[14] G. Robins & C. Shute, *The Rhind Mathematical Papyrus: An ancient Egyptian text*, British Museum Publications, London 1987.

[15] Trustees of the British Museum, *Facsimile of the Rhind mathematical papyrus in the British Museum*, London 1898.

SAKREGISTER

absolutbeloppet, 70, 71, 181
addition
– av funktioner, 89, 90
– av följder, 75
– av komplexa tal, 181
– av (reella) tal, 18
alla, *se* kvantifikator
antagande, 13, 23
antingen… eller, 16
arcusfunktioner, 144–148
area, 105–110
argument, 183
aritmetik, 19
arkimediska egenskapen, 64–66, 73, 81, 125, 240, 241
avbildning, 197–200
avstånd, *se* längd
avtagande, *se* monoton
axel, 90
axiom, 13, 19, 237, 238

bas, 37, 163, 165, 178
basfallet, 49
begränsad mängd, *se* mängd
Bernoullis olikhet, 152–154
bijektiv, 118, 199
båge, 102

C, *se* komplexa tal
Chaucer, Geoffrey, 7
cirkel, 102–104, 194–196
– area av, 109, 144
– enhets-, 103, 137
– omkrets av, 102
cosinus, 137–148, 191, 192
cosinus hyperbolicus, 178
cosinussatsen, 148
cyclometriska funktioner, *se* arcusfunktioner

de Moivres formel, 191
decimalkomma, 39
decimalutveckling
– icke-entydighet, 80
– oändlig, 66–81, 84
– trunkerad, 67–70
– ändlig, 39, 40
Dedekind, Richard, 237
deduktiv slutledning, 7, 13, 49
definition, 22, 23
delbarhet
– av polynom, 98, 99
– av tal, 22, 37
delmängd, *se* mängd
delsummor, 166
differens, *se även* subtraktion
– av mängder, 46
disjunkta mängder, *se* parvis disjunkta
disjunktion
– stark, 16
– svag, 16
diskriminant, 129

divergent
– följd, 72
– serie, 166
division
– av funktioner, 89, 90
– av följder, 75
– av komplexa tal, 182, 183
– av polynom, 94–99
– av (reella) tal, 18, 22, 23
– lång, *se* liggande stolen

e (Eulers/Napiers tal), 155, 178, 192
ekvivalens, 14
element, 43
eller, 14, 15
. . . (ellipsis), 28, 166
Euler, Leonhard, 155
Eulers identitet, 191
exponentialfunktionen
– den naturliga, 155–173
– med bas *a*, 163, 165
– med komplexa argument, 187–193

faktor, 19
fakultet, 51, 52
fullständighetsaxiomet, 68, 81, 125
funktion, 87
fyrhörning, 105
följd, 47, 64–66
– aritmetisk, 48, 49
– geometrisk, 51
– konstant, 77, 80

Gauss, Carl Friedrich, 31
grad (av polynom), 94
grad (vinkelenhet), 104
graf, 90, 91
grekiska bokstäver, 235
gränsvärde, 70–78

Hardy, Godfrey Harold, 123
hela meningar, 26
heltal, 44
hypotenusa, 105
hörnpunkter, 104, 105

i (imaginära enhet), 179, 192
icke-negativa/positiva, 39, 125
icke-överlappande, *se* parvis disjunkta
imaginära axeln, 180
imaginära tal, 179, 181
imaginärdel, 181
implikation, 14
indelning, *se* partition
index, 29, 30, 47, 48, 67
indirekt bevis, *se* motsägelsebevis
induktion, *se* induktionsbevis
induktionsaxiomet, 49
induktionsbevis, 49–51, 99
induktionsprincipen, *se* induktionsaxiomet
induktionssteget, 49
induktiv slutledning, 7, 49
infimum, *se* största undre begränsning
inget, *se* kvantifikator
injektiv, 118, 132
inre, 105
instängningssatsen, 158, 159
intervall, 45, 58, 63
– ändpunkter, 63
invers funktion, 117–120, 199
– grafen av en, 120
inverterbar, *se* bijektiv
irrationella tal, 38, 122–124

Jordanmått, 105–110
jämna
– funktioner, 138
– tal, 42, 44, 123

kant, *se* sida
kartesisk form, 102
kartesisk produkt, 105
kartesiska koordinater, *se* koordinater, kartesiska
katet, 105
koefficienter, 94, 100
komplexa planet, 180, 247
komplexa tal, 179–184, 243–248
konjugat, 181
kontraposition, 16, 17, 60, 123
konvergent
– följd, 72
– serie, 166
konvergera mot, *se* konvergent
koordinater
– kartesiska, 90
– polära, 183, 191, 245
koordinatplanet, *se* koordinater, kartesiska
kropp, 238
× (kryss), *se* kartesisk produkt; multiplikation
kubik, 27
kubikrot, 132
kurva, 101, 241–243
– sluten, 242
kvadrat, 18, 27, 105
kvadratkomplettering, 59, 129, 186
kvadratrot, 124–130
kvantifikator, 23–26

liggande stolen, 39, 40, 96

likhet
– :=, 27
– mellan funktioner, 88
– mellan komplexa tal, 180
– mellan mängder, 45
– mellan (reella) tal, 18
– trigonometrisk, *se* trigonometriska identiteter
linje, 102, 194–196, *se även* reella linjen
logaritmfunktionen
– den naturliga, 173–175
– i basen *a*, 178
lutning, 102
längd, 102
– av en kurva, 241–243
– av ett intervall, 45
lösningar, 20, 21, 25, 128

maclaurinutveckling, *se* taylorutveckling
matematisk induktion, *se* induktionsbevis
max, *se* största element
medelpunkt, 102
min, *se* minsta element
minsta element, 58, 59, 75, 126
minsta övre begränsning, 59–66, 68
monoton
– funktion, 91–93, 131, 132
– följd, 48, 52, 67, 68, 70, 73, 152
motsägelsebevis, 58, 81, 98, 120
multiplikation
– av funktioner, 89, 90
– av följder, 75
– av komplexa tal, 180, 181
– av (reella) tal, 18
– pi notation, 52

mängd, 43
– begränsad, 59–66, 68, 69, 73, 153
– definitions-, 87
– del-, 45
– enkel, 106
– mål-, 87
– rektangulär, 105
– värde-, 88, 101
– äkta del-, 45
Möbiusavbildning, 198–200

N, *se* naturliga tal
n:e rot, 130–132, *se även* kvadratrot
Napier, John, 155
naturliga tal, 43, 52
nedåt begränsade, *se* mängd, begränsad
negation, 15, 16, 23, 58
noggrannhet, 67
noll, 39, 125, *se även* origo
– delat med noll, 32
– nollte, 52
– upphöjt till noll, 27, 32, 167
nollställe, 98, 128
notation, 26
något, *se* kvantifikator
nämnare, 38
nödvändigt villkor, 20

och, 14
olikhet, 18, 57
– strängt, 18
– trigonometrisk, 143, 143, 144
omkrets, *se* cirkel
operationer
– aritmetiska, *se* aritmetik
– på mängder, 45

origo, 67, 90

parametrisering, 101, 242
(...),[...] och {...}, *se* parenteser
parenteser
– aritmetik, 18
– följder, 47
– koordinater, 90
– mängder, 43, *se även* intervall
partialsummor, *se* delsummor
partition, 242
parvis disjunkta, 106
Pascals triangel, 55
Peanos axiom, 238
periodisk funktion, 138, 191
π (gemen pi), 68, 103, 192, *se även* grekiska bokstäver
Π (versal pi), *se* multiplikation, pi notation; grekiska bokstäver
$\Longrightarrow$ (implikationspil), 14
$\to$ (pil utan fenor), 87
$\mapsto$ (pil utifrån en pinne), 88
plan, 90
$\pm$ (plus/minus), 79, 129, 130
polynom, 94–100, 240
– andragrads-, 94, 128, 129
– det triviala, 94
– förstagrads-, 94, 98
polära koordinater, *se* koordinater, polära
positiva heltal, 44
potens, *se även* talsystem
– av *i*, 180
– heltalspotens, 27, 180
– irrationell, 175–177
– rationell, 130–133

potensserie, 167
primtal, 44
produkt, *se* multiplikation; kartesisk produkt
Pythagoras sats, 110–112
påstående, 13, 23

Q, *se* rationella tal

R, *se* reella tal
radianer, 104
radie, 102
rationella tal, 37, 38, 44
realdel, 181
reella axeln, 180, 181
reella linjen, 57, 90
reella tal, 56, 66–81
regler, 19
rektangel, 105
relationer, 18
rest, 22, 23
restriktion, 145
rot, 128, *se även n*:e rot
rotation, 139
roten ur, *se* kvadratrot
$\mathbf{R}_0^+$, 125
ränta på ränta, 151–154
rätvinkel, 104

sammansättning
– funktioner, 89, 90, 93
– följder, 77
sats, 22, 23
serie, 165–171, 192, 193
sida, 104, 105
siffra, 39
Σ (sigma), *se* summa, sigma notation
sinc, 159
sinus, 137–148, 159, 191, 192

slutsatser, 13, 23
snitt, 45
spegelbild, 35, 120, 138, 139, 181
Stefan-Boltzmanns lag, 117, 132
streck, 101
strängt
– avtagande, *se* monoton, funktion
– olikhet, *se* olikhet, strängt
– växande, *se* monoton, funktion
största element, 57–59, 75, 77, 94, 126
största undre begränsning, 59–66, 69
subtraktion
– av funktioner, 89, 90
– av följder, 75
– av komplexa tal, 181
– av (reella) tal, 18
summa, 28–34, *se även* addition
– av en aritmetisk följd, 31
– av en geometrisk följd, 32
– sigma notation, 29, 165
supremum, *se* minsta övre begränsning
surjektiv, 118
svans
– av en följd, 155
– av en serie, 168
symmetri, 103, 107, 138–142

talsystem, 35
– decimala, 36
– egyptiska, 35
– hindu-arabiska, 35
– positionssystem, 37
– ternära, 37
tangens, 138, 141, 143–145, 147, 148

taylorutveckling, 167
term, 19
tillräckligt villkor, 20
tomma mängden, 47, 106
triangel, 104, 107
– likbent, 109, 143
– liksidig, 142
– rätvinklig, 105, 110, 143
triangelolikheten, 71, 184
trigonometriska identiteter, 138–143
triviala polynomet, *se* polynom, det triviala
täljare, 38

udda
– funktioner, 138
– tal, 42, 44, 123
union, 45
uppåt begränsade, *se* mängd, begränsad
utvidgning
– av absolutbeloppet, 181
– av aritmetiska operationer, 179, 243–248
– av funktion, 128, 187
– av längd, 102, 241–243
– av potens, 132, 176
– av vinkel, 137

variabelbyte, 30, 97
vinkel, 103, 104, 137
växande, *se* monoton

yttre, 105

Z, *se* heltal
$\mathbf{Z}_+$, *se* positiva heltal

∅, *se* tomma mängden

