Matematisk analys

Översikt

Innehållsförteckning

Del 0: Inledning

0.1 Förord till utkastet

Den här texten är tänkt att bli en (omfattande) lärobok i inledande matematik, på högskolenivå. Av de ämnen som behandlas på högskolan har jag för avsikt att täcka följande i princip fullständigt:

- Grundkurs i matematik
- Linjär algebra
- Envariabelanalys
- Flervariabelanalys
- Vektoranalys.

Jag har också för avsikt att ge en något mer begränsad inblick i bland annat följande områden:

- Abstrakt algebra
- Funktionalanalys
- Fourieranalys
- Mått- och integrationsteori
- Partiella differentialekvationer.

Den här texten kommer till viss del att ha många likheter med existerande verk i inledande högskolematematik, men kommer – som synes – att gå längre än de flesta av dessa. Författarens förhoppning är att detta emellertid inte kommer att ske på bekostnad av pedagogik och tydlighet, eller på bekostnad av stringens och fullständighet. Några huvudsakliga idéer som ligger till grund för texten är följande:

- Texten skall präglas av högsta möjliga tydlighet.
- Boken skall vara lämplig som läsning för nybörjare i högskolematematiken. Det skall inte vara svårt att läsa texten, inte ens för studenter som är måttligt förtjusta i matematik.
- Även mer hängivna matematikstudenter skall hitta godsaker i boken.
- Alla skolboksmoment skall exemplifieras tydligt och rikligt. Mångtaliga exempel på konkret problemlösning skall ges; lösta exempel ges i *blå* rutor.
- Vanliga nybörjarfel skall lyftas fram; detta görs i speciella *röda* rutor.
- Texten avstår inte från att när lämpligt göra djupdykningar i mer avancerat material, men sådant material, som inte är nödvändigt för en nybörjare i högskolematematik, markeras tydligt med asterisker (*).
- Och framför allt vill jag betona vikten av god matematisk kommunikation.

Tanken är att jag från den långa texten skall kunna extrahera delmängder – delböcker – lämpliga för olika sorters läsare. Borttagning av alla asteriskmarkerade avsnitt och annat mer fördjupande material skall resultera i en (mindre) bok extra lämplig för nybörjare i matematik, vilka kanske inte är alltför roade av matematikämnet i sig. Å andra sidan skall borttagande av upprepade förklaringar, de röda rutorna och ersättande av vissa avsnitt med alternativa asteriskmarkerade avsnitt resultera i en bok lämpad för de lite mer motiverade matematikstudenterna.

Slutligen har författaren för avsikt att översätta hela boken till engelska när den är färdig. Slutresultatet – om jag kommer dit – är alltså *sex* läroböcker i matematik (vilka?).

0.2 Förkunskaper

Det är nästintill omöjligt att skriva en lärobok i matematik utan att förvänta sig att läsaren redan besitter vissa förkunskaper i ämnet. Den här texten är emellertid avsedd att tjäna som en *inledande* text i högskolematematik, varför den kräver förhållandevis ringa förkunskaper av läsaren. I princip krävs inte mer än högstadiets samtliga och gymnasiets första kurser i matematik. Det underlättar däremot om man läst mer matematik på gymnasiet.

Mer specifikt förväntas läsaren vara mycket väl förtrogen med *tal* och räkning med sådana. Läsaren bör känna till *heltalen* ($\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots$) och *de rationella talen* (såsom $\frac{1}{3}$, 0.43 och -321.1). I praktiken skall läsaren också kunna räkna med *reella* tal. Vidare förutsätts förtrogenhet med de fyra räknesätten (addition, subtraktion, multiplikation och division) när de används på rationella (och reella) tal.

Läsaren förväntas också vara bekant med enkel algebra (med tal). Detta innebär dels att läsaren känner till de lagar som gäller för de fyra räknesätten, dels att läsaren kan formulera uttryck och utsagor med hjälp av symboler och abstrakta ”bokstäver” som platshållare. Enkel ekvationslösning bör man också ha med sig i bagaget.

Tecknen $=, <, \leq, >, \geq, \neq$ bör vara bekanta; dessa bildar utsagor om par av tal.

Endast mycket grundläggande kunskaper om geometri kommer att förutsättas. I praktiken räcker det med att läsaren har en intuitiv förståelse för begreppen längd/avstånd, area och volym samt kännedom om de allra enklaste formlerna inom geometrin, såsom arean av en rektangel och av en cirkel.

Del 1: Grunderna

1.1 Matematiskt språk

1.1.1 Vad är matematik?

Man kan säga att matematik handlar om två saker: *kommunikation* och *argumentation*. Inom ett visst matematisk område (t.ex. analys, algebra eller grafteori) kommer man överens om ett gemensamt språk för att beskriva de objekt man studerar (t.ex. tal, mängder och funktioner). Nu är det förvisso inte så att alla matematikböcker använder exakt samma terminologi och konventioner, men *inom varje verk* utvecklar man i regel en bestämd terminologi, och i de flesta fall (åtminstone när det gäller grundläggande matematik) är olika verk ganska överens om de mest centrala begreppen. Det här handlar alltså om *kommunikation* – man beskriver objekt. Här är några exempel på beskrivningar av objekt som förekommer inom olika delar av matematiken (bara första and andra exemplen torde vara bekanta för en student som kommer direkt från gymnasiet):

- *Talet* 53 *är ett primtal.*
  - → (talteori) Beskriver talet 53.
- *Funktionen* $\sin$ *är strängt växande på* $[0, \pi/2]$.
  - → (analys) Beskriver funktionen $\sin$.
- *Mängden* $\{(x,y) \in \mathbb{R}^2 : x^2 + 2y^4 \leq 9\}$ *är kompakt (=sluten och begränsad).*
  - → (analys) Beskriver mängden $\{(x,y) \in \mathbb{R}^2 : x^2 + 2y^4 \leq 9\}$.
- *Varje icke-konstant polynom med komplexa koefficienter har ett komplext nollställe.*
  - → (algebra) Beskriver kroppen[1] $\mathbb{C}$.
- *Den symmetriska gruppen på fem symboler är inte lösbar.*
  - → (algebra) Beskriver den symmetriska gruppen[2] S_5 på fem symboler.
- *Petersengrafen har kromatiskt nummer* 3.
  - → (grafteori) Beskriver grafen[3] som går under namnet Petersengrafen.

Det är förstås inte bara inom matematiken man beskriver objekt. Det som kännetecknar just den matematiska beskrivningen av objekt är den *exakthet* som råder: matematiska begrepp måste vara (perfekt) precisa – det får inte råda någon som helst tveksamhet angående var ett ord betyder. Också kännetecknande för matematisk beskrivning är att man är väldigt frikostig med att via definitioner införa nya begrepp (jämför med införandet av nya ord inom vanligt språkbruk); det behövs ju också många begrepp för att precist kunna beskriva objekt. Både objekt (t.ex. tal, mängder och funktioner) och deras beskrivande egenskaper kan betraktas som ”begrepp”. Exempel på begrepp (beskrivande egenskaper) som beskriver funktioner är ”kontinuerlig”, ”strängt växande”, ”injektiv”, ”udda” och ”begränsad”, bara för att nämna en mycket liten del av den armada av begrepp som rör funktioner.

---

[1] En *kropp* är en slags algebraisk struktur. $\mathbb{Q}$, $\mathbb{R}$ och $\mathbb{C}$ är kroppar, men inte $\mathbb{Z}$ som får nöja sig med att vara en *kommutativ ring*. (En kropp är en kommutativ ring där varje nollskilt element har en multiplikativ invers.)

[2] En *grupp* är också en slags algebraisk struktur. (Faktiskt så är varje ring är en grupp, liksom varje vektorrum.)

[3] Tyvärr kan ordet ”graf” betyda två helt olika saker inom matematiken. Dels kan man tala om *grafen till en funktion*, t.ex. sinuskurvan $y = \sin x$. Petersengrafen är emellertid inte den sortens graf, utan i stället den sorts graf man studerar inom grafteorin. I princip är en sådan graf en mängd *noder* tillsammans med en mängd *bågar* (eller *kanter*) som förbinder vissa noder med varandra (eller sig själva). Grafteori är ett forskningsfält med tillämpningar inom vitt skilda discipliner. (eller sig själva) a avsnitt. Om man i stället e moment man läser i sin högskolekurs digt bestämt. Alltså kan man fullständigt

När man infört ett språk för att beskriva en sorts objekt kan man också fundera på sambanden mellan olika begrepp. Till exempel kan man (mycket enkelt!) resonera sig fram till att en funktion som är strängt växande, nödvändigtvis, också är injektiv. Däremot är inte varje injektiv funktion strängt växande.

Ett påstående i stil med ”Varje icke-konstant polynom med komplexa koefficienter har ett komplext nollställe” kallas för en matematisk *utsaga*, och är antingen sant eller falskt. Exempel på sanna utsagor är ”$1 + 1 = 2$”, de fem som radas upp i punktlistan ovan och den nyligen diskuterade utsagan ”Varje strängt växande funktion är injektiv”. Exempel på falska utsagor är ”$1 + 1 = 3$”, ”Talet 51 är ett primtal” och ”Varje injektiv funktion är strängt växande”. När man i tal eller skrift argumenterar på ett sådant sätt att det blir *fullkomligt uppenbart* att en viss utsaga är sann, så säger man att man har *bevisat* den, eller *gett ett bevis* för den. Ett bevisat påstående kallas för en *sats*.

Ett ganska sympatiskt sätt att se på matematiken är att matematiken är *den del av svenskämnet där exaktheten i beskrivningar och resonemang är tagen till sin spets.* Detta gör också att man inom just matematiken, men svårligen utanför den, kan dra oerhört sofistikerade slutsatser om mycket invecklade objekt. Innebörden av de flesta satser går (med olika stor möda) att förklara med bara vanliga, allmänna, ord, men i många fall krävs då *väldigt* många ord, och den slutsats som dras i satsen förefaller då så invecklad att en icke-matematiker kan få för sig att man måste vara av övermänsklig intelligens för att kunna dra den slutasen. Så är det egentligen inte; i stället är hemligheten att matematiker delar upp stora problem i många mycket små delar, och undersöker ett litet samband i taget.

1.1.1.1 Matematiska texter

Ett vanligt missförstånd bland personer som är nya i högre matematik är att man *alltid* använder matematiska symboler i stället för vanliga svenska ord och meningar när man bevisar matematiska utsagor. Så är *inte* fallet! Tvärtom, som jag försökt antyda ovan, handlar matematik just om att argumentera med hjälp av mänskligt språk (t.ex. svenska språket). Det finns många exempel på matematiska bevis som kan utföras enbart med hjälp av ord (dock inklusive begrepp som bara används inom det aktuella området), t.ex. inom abstrakt algebra och grafteori. När man författar ett bevis är det mycket viktigt att det är lätt att följa för läsaren, att det framgår mycket klart och tydligt hur författaren resonerar. Detta går nästan inte att åstadkomma utan att använda vanliga ord och uttryck som ”om”, ”eftersom”, ”därför”, ”och”, ”eller” och ”med tanke på att”!

Däremot kan man i många fall infoga matematiska formler (strängar av symboler) antingen som nominalfraser eller som (grammatiska) satser i vanlig text. Till exempel kan man med fördel skriva

> Om $\sin x = \frac{1}{\sqrt{2}}$ och $x \in \left[0, \frac{\pi}{2}\right]$ så är $x = \frac{\pi}{4}$.

i stället för

> Om sinus av x är lika med den multiplikativa inversen av roten ur två och x ligger mellan 0 och halva π (inklusive gränserna) så måste x vara lika med en fjärdedels π.

Första varianten är onekligen mycket lättare att läsa (åtminstone om man känner till det grundläggande matematiska symbolspråket)! När en matematisk formel sprängs in i vanlig brödtext väljer man ibland att skriva den på en egen rad (visuellt: ett eget stycke) för att framhäva den, kanske också för att ge den ett nummer (”ekvationsnumrering”), så att man senare kan hänvisa till den på ett smidigt sätt. Det kan även vara så att formeln tar upp så mycket plats i höjdled att det blir direkt otympligt att låta den stå mitt i en rad i brödtexten. I de flesta av dessa fall väljer man att *centrera* den insprängda formeln:

> Text, text. Basbytesmatrisen är sålunda
>
> $$\begin{pmatrix} a & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}$$
>
> för något $a > 0$. Text, text.

ser onekligen mycket bättre ut än

> Text, text. Basbytesmatrisen är sålunda $\begin{pmatrix} a & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}$ för något $a > 0$. Text, text.

i synnerhet med tanke på raderna ovanför och under. Notera också att ”för” i båda fallen inleds med liten bokstav; det är ju näppeligen så att ordet inleder en ny mening!

Formler brukar alltså *infogas* i vanlig brödtext skriven i det svenska språket. Jag brukar säga att en matematisk text använder det svenska språket som en *röd tråd* genom hela texten; det mänskliga språket förbinder de insprängda matematiska formlerna och anger deras relation till varandra.

1.1.1.1.1 Exempel

(Exemplet nedan handlar om summor och summationssymbolen. Om detta inte är bekant från tidigare matematikstudier behöver läsaren inte misströsta, för vi kommer att ge en fullständig introduktion till summor senare.)

Om $(a_k)_{k=1}^n$ är en följd av tal [d.v.s. vi har de n stycken talen $a_1, a_2, \dots, a_n$] så kan det hända att det finns ett fixt tal d sådant att $a_1 = a_0 + d$, $a_2 = a_1 + d$ o.s.v., d.v.s. givet vilket tal som helst i följden (förutom det sista), så erhålles nästa tal i följden genom addition av det fixa talet d. En talföljd med denna egenskap kallas *aritmetisk* och talet d kallas för följdens *differens* (talet är ju differensen mellan ett element och föregående element i följden). Till exempel är följande talföljd aritmetisk:

$$1, 5, 10, 15, 20, 25, 30, 35.$$

Men den här följden är *inte* aritmetisk:

$$5, 9, 1, 4, -3, 5, -1, 7, 4, 7.$$

Det är lätt att inse att summan $\sum_{k=1}^n a_k = a_1 + a_2 + \cdots + a_n$ av en aritmetisk följd (som kallas för en aritmetisk summa) är lika med produkten av antalet termer och medelvärdet av första och sista termen.

Låt oss nu betrakta en typisk tentamensfråga om just aritmetiska summor.

Exempel

Beräkna summan $\sum_{k=31}^{101}(5k-10)$.

Lösning: Vi ser direkt att summan är aritmetisk (med differensen 5). Första termen är $5\cdot 31-10=145$, sista termen är $5\cdot 101-10=495$ och antalet termer är $101-31+1=71$. Formeln för en aritmetisk summa ger därför att summan har värdet

$$\frac{145+495}{2}\cdot 71=22\,720.$$

Svar: $\sum_{k=31}^{101}(5k-10)=22\,720$.

Följande exempel visar hur man *inte* bör redovisa denna uträkning:

Exempel på hur man *inte* bör göra

Beräkna summan $\sum_{k=31}^{101}(5k-10)$.

Lösning:

$$\sum_{k=31}^{101}(5k-10)$$

$5\cdot 31-10=145,\ 5\cdot 32-10=150,\ 5\cdot 101-10=495$ $\qquad n=101-31+1=71$

$145+495=640$

$$s=\frac{640}{2}\cdot 71=320\cdot 71=22720$$

Svar: $s=22720$

Här har studenten bara kräkts upp en massa formler. Det framgår inte hur studenten har tänkt när han löst uppgiften, och det är svårt att "följa" lösningen. Exempel på fullt legitima "klagomål":

- Vilka är de tre uttrycken till vänster på den andra raden? Var kommer de ifrån? På vilket sätt används de för att beräkna summan? Vilka slutsatser drar man från dessa beräkningar?
- Varför räknas andra termen (150) ut? Man kan naturligtvis inte konstatera att en summa är aritmetisk bara genom att konstatera att andra termen minus första termen *går* att beräkna! Det ser ut som om studenten gör det! Hade även tredje termen beräknats, kan man notera att de två första differenserna i följden är lika. Men det i sig bevisar ju inte att *alla* möjliga differenser av två på varandra följande tal i följden är lika. Nej, att följden är aritmetisk ser man i stället direkt av den explicita formeln $5k-10$ för det k:te elementet.[4]
- Vad är "n" för något?
- Varför beräknas summan $145+495$? Varför inte, säg, $145+150$?
- Vad är "s" för något? Varför är $s=\frac{640}{2}\cdot 71$?

[4] Man kan förstås även vara riktigt formell, sätta $a_k\coloneqq 5k-10$ och notera att $a_{k+1}-a_k=5(k+1)-10-(5k-10)=5k+5-10-5k+10=5$ för alla $k\in\mathbb{Z}$.

Visst, om läsaren av lösningen redan vet hur aritmetiska summor beräknas, så kan han nog lista ut svaren på alla dessa frågor, men i sådana fall hade ju ingen motivering krävts över huvud taget! I sådana fall kunde man ju nöjt sig med att direkt konstatera att $\sum_{k=31}^{101}(5k-10) = 22\,720$. [Och naturligtvis gör man också just det i matematiska sammanhang som inte har med matematiska grundkurser att göra.]

På en matematikskrivning är det klart att läraren som rättar oftast förstår vad studenten egentligen menar, men studenten bör ändå försöka att formulera sina lösningar på ”rätt” sätt. Dels är det ju en övning i matematisk kommunikation (rent av i kommunikation över lag), och dels visar studenten genom en fullgod redovisning att han verkligen förstår hur teorin fungerar, t.ex. att formeln

$$\text{summan} = \frac{\text{första termen} + \text{sista termen}}{2} \cdot \text{antal termer}$$

bara gäller vid beräkning av *aritmetiska* summor. Och detta är ju, av fullt rimliga skäl, något som mycket väl kan vara ett krav även för det lägsta betyget! Se också notisen om hur man konstaterar att följden verkligen är aritmetisk (den andra punkten i punktlistan ovan); även här kan studenten ge tecken på antingen förståelse (välformulerad motivering) eller misstänkt brist på sådan (obefintlig, dålig eller rent av felaktig motivering).

En sista kommentar: Om man i lösningen på ett korrekt sätt inför beteckningen s för summan, t.ex. genom att skriva ”Låt $s \coloneqq \sum_{k=31}^{101}(5k-10)$” eller ”Beteckna summan med s”, så kan man naturligtvis sedan använda den beteckningen för summan, och rent av avsluta lösningsgången med ”$s = 22720$”. Däremot bör *svaret* (”Svar: …”) inte innehålla symbolen ”s”, eftersom svaret skall gå att läsa (och vara begripligt) även utan att man först har läst lösningsgången; man skall kunna förstå svaret bara genom att ha läst själva frågeformuleringen (och i den nämns ju inte symbolen ”s”).

1.1.1.2 Definitioner och satser

En matematisk teori består av definitioner och satser. Först gör man i allmänhet en mängd definitioner, d.v.s. man inför *begrepp*, eller ”nya ord” för att beskriva de objekt man studerar. Sedan försöker man hitta samband mellan dessa begrepp eller undersöka vilka objekt som har vilka egenskaper. Man producerar då *satser*. Ofta, men inte nödvändigtvis, väljer man att rama in definitioner och satser i den löpande texten (bokstavligt talat: man ger alltså styckena kantlinjer), så att det blir lättare att ögna igenom dokumentet. Man ser då direkt *strukturen* i den matematiska teorin; man ser snabbt vilka begrepp som införs, och vilka slutsatser som dras om dem.

Vi ger nu ett typexempel på både definition och sats, med bevis. Notera att beviset avslutas med symbolen ”∎” (U+220E: END OF PROOF). En del författare föredrar termen ”V.S.B.” (=vilket skulle bevisas) eller dess internationella motsvarighet ”Q.E.D.” (lat. *quod erat demonstrandum*).

Definition. En följd $(a_k)_{k=n}^{m}$ av reella tal säges vara *aritmetisk* om $d \coloneqq a_{k+1} - a_k$ inte beror på k, d.v.s. om differensen d mellan två på varandra följande tal alltid är densamma. Talet d kallas i sådana fall för den aritmetiska följdens *differens*.

Sats. Om $(a_k)_{k=n}^{m}$ är en aritmetisk följd med differens d så är

$$\sum_{k=n}^{m} a_k = \frac{a_n + a_m}{2} \cdot (m - n + 1).$$

Bevis. Beteckna summan med s. Vi har då

$$\begin{aligned} s &= a_n + a_{n+1} + \cdots + a_{m-1} + a_m = \\ &= a_m + a_{m-1} + \cdots + a_{n+1} + a_n. \end{aligned}$$

Därför är

$$\begin{aligned} 2s &= (a_n + a_m) + (a_{n+1} + a_{m-1}) + \cdots + (a_{m-1} + a_{n+1}) + (a_m + a_n) = \\ &= (a_n + a_m) + (a_n + a_m) + \cdots + (a_n + a_m) + (a_m + a_n) \end{aligned}$$

eftersom följden är aritmetisk. Antalet termer i högerledet är tydligen $m - n + 1$, så i själva verket är

$$2s = (m - n + 1)(a_n + a_m)$$

och beviset är klart. ■

1.1.1.3 Definitioner är viktiga

Ibland fokuserar studenter *mycket* på satser och *lite* på definitioner, vilket är olyckligt. Definitionerna är nämligen *minst* lika viktiga som satserna. Satserna anger ju egenskaper hos objekt eller samband mellan olika egenskaper, så för att förstå satserna måste man veta vad det är för objekt och egenskaper de uttalar sig om, d.v.s. man måste känna till deras definitioner. För att förstå innebörden av satsen ”varje strängt växande funktion är injektiv” måste man veta exakt vad ”strängt växande” och ”injektiv” betyder. Ett mindre elementärt exempel kommer från den abstrakta algebran, där en av de mer centrala satserna säger att varje PID är en UFD. Naturligtvis har man föga glädje av den satsen som man inte vet vad begreppen PID och UFD betyder.

Dessutom har man nytta av begrepp (och därmed definitioner) även om man inte använder dem i icke-triviala satser. Genom att införa speciella namn på begrepp och egenskaper får vi ett språk med vilket vi på ett exakt sätt kan beskriva saker. Begrepp kan också få oss att tänka till: till exempel kommer vi se att en *stationär punkt* till en reellvärd funktion mycket väl kan undgå att vara både en *lokal extrempunkt* och en *terrasspunkt*.

Man kan säga att definitioner är viktiga, eftersom ”det är med begrepp man begriper”, som doktor Lars Alfred Engström brukade säga.

1.1.2 Logik

Betrakta två matematiska utsagor P och Q. Det kan då hända att Q *följer av* P, d.v.s. att man kan dra slutsatsen att Q är sann om man vet att P är sann. Man säger då att P *implicerar* Q, och i symboler skrives detta $P \Rightarrow Q$ (alternativt: $Q \Leftarrow P$). Med vanliga ord säger man ”om P så Q.” Till exempel, om

P: Katten är i köket
Q: Katten är i huset

så gäller $P \Rightarrow Q$, för om katten är i köket (som är den del av huset), då kan vi med säkerhet också säga att katten är i huset. ”*Om* katten är i köket, *så* är katten i huset.” Notera att både det som står före och det som står efter symbolen ”⇒” måste vara utsagor. Utsagor, som också kan kallas *påståenden*, innehåller alltid ett finit verb (i det här fallet ”är”) när de utläses i ord. $P \Rightarrow Q$ kan i sig ses som en utsaga, och kan antingen sann eller falsk. Till exempel är utsagan $P \Rightarrow Q$ om katten sann (vilket vi nyss konstaterade). Däremot är inte utsagan $Q \Rightarrow P$ sann: bara för att katten är i huset så kan vi inte veta säkert att katten är i köket (den kanske är i vardagsrummet?).

Exemplet med katten är egentligen inte särskilt bra, eftersom begreppen ”katten”, ”är”, ”i”, ”köket” och ”huset” inte är 100 % väldefinierade. Här hoppas jag alltså främst på en pedagogisk poäng!

Ett bättre exempel är, om x är ett reellt tal,

$$2x = 50 \qquad \Longrightarrow \qquad x = 25.$$

Om vi vet att $2x = 50$, *så* måste $x = 25$.

Några fler exempel på implikationer som är sanna:

- $0 < x < 10 \Rightarrow -10 < x < 10$
- $x^2 = 9 \Rightarrow x = -3$ eller $x = 3$
- $x^2 = 9$ och $x > 0 \Rightarrow x = 3$
- $x \geq 5$ och $x \leq 5 \Rightarrow x = 5$
- x är jämnt och x är ett primtal $\Rightarrow x = 2$.

Se noggrant till att du förstår dem. Ibland händer det att två påståenden P och Q medför varandra, d.v.s. att $P \Rightarrow Q$ OCH $Q \Rightarrow P$. Detta kan skrivas $P \Rightarrow Q$ OCH $P \Leftarrow Q$ vilket vanligtvis förkortas $P \Leftrightarrow Q$. Vi säger då att påståendena P och Q är *ekvivalenta*. Till exempel har vi

$$2x = 50 \qquad \Longleftrightarrow \qquad x = 25$$

för om $2x = 50$ så måste $x = 25$, och – omvänt – om $x = 25$ så *är* faktiskt $2x = 50$; här resonerade vi alltså först fram $2x = 50 \Rightarrow x = 25$ och därefter $x = 25 \Rightarrow 2x = 50$.

Av implikationerna i punktlistan ovan är alla utom den första också ekvivalenser (varför?), så vi har

- $x^2 = 9 \Leftrightarrow x = -3$ eller $x = 3$
- $x^2 = 9$ och $x > 0 \Leftrightarrow x = 3$
- $x \geq 5$ och $x \leq 5 \Leftrightarrow x = 5$
- x är jämnt och x är ett primtal $\Leftrightarrow x = 2$.

Det man måste tänka på när man använder symbolerna ⇒, ⇐ och ⇔ är att det som står på respektive sida faktiskt *är* matematiska utsagor. Dessutom måste man, varje gång man använder en av dessa pilar, fundera på om utsagorna verkligen har den relation man påstår att de har. Till exempel, om man skriver ”$x^2 = 9$ och $x > 0 \Leftrightarrow x = 3$” måste man tänka ut att båda implikationerna är sanna, t.ex. genom att tänka ”⇒) om $x^2 = 9$ så är $x = \pm 3$, men om $x > 0$ så måste alltså $x = 3$; ⇐) omvänt, om $x = 3$ så *är* x positivt och *har* kvadraten 9” i huvudet.

Ett mycket vanligt fel

Ett vanligt förkommande fel bland nya studenter i matematik är att de använder ⇔ i stället för likhetstecken ($=$), vilket förstås är absurt:

$$x^2 + 3x + 3 + x + 1 = x^2 + 4x + 4 = (x + 2)^2$$

är ju alldeles riktigt, men

$$x^2 + 3x + 3 + x + 1 \Leftrightarrow x^2 + 4x + 4 \Leftrightarrow (x + 2)^2$$

är obegripligt. Betrakta t.ex. frågan ”Om jag vet att $x^2 + 3x + 3 + x + 1$, kan jag då vara helt säker på att $x^2 + 4x + 4$?” Det går inte att göra den frågan begriplig; jämför med ”Är du helt säker på att bilen?” Bilen vadå? Varken ”$x^2 + 3x + 3 + x + 1$” eller ”bilen” är en utsaga (de är bara nominalfraser). Kom ihåg att en utsaga alltid innehåller ett finit verb när det uttalas i ord, och det är ett påstående som antingen är sant eller falskt. Till exempel är ”$x^2 + 3x + 1 + x + 1 = 5$” och ”bilen är röd” utsagor.

1.1.2.1 Logiska operatorer

Lägg i synnerhet märke till att orden ”och” och ”eller” har väldigt precisa betydelser i samband med matematiska utsagor. Om P och Q är utsagor så är ”P och Q” också en utsaga, som är sann då (och bara då!) *både* P och Q är sanna. På liknande sätt är ”P eller Q” en utsaga som är sann då (och bara då!) *minst en av* P och Q är sann.

1.1.2.2 Nödvändiga och tillräckliga villkor

Om $P \Rightarrow Q$ säger man att Q är ett *nödvändigt* villkor för P och att P är ett *tillräckligt* villkor för Q.

Att Q är ett nödvändigt villkor för P betyder att utsagan Q *måste* vara sann för att P ens skall ha någon möjlighet att vara sann. För om Q är falsk så kan ju inte P vara sann, för om P hade varit sann, så hade ju Q varit sann (ty $P \Rightarrow Q$).

Att P är ett tillräckligt villkor för Q betyder bara att $P \Rightarrow Q$: det räcker att veta att utsagan P är sann, för att man skall kunna dra slutsatsen Q.

Exempel. Ett *nödvändigt* villkor för att jag skall godkänna en uppsats skriven av en svenskelev är att texten saknar uppenbara grammatiska fel, för om texten innehåller uppenbara grammatiska fel, så godkänner jag den inte. Däremot är kravet inte *tillräckligt* – även om texten saknar uppenbara grammatiska fel, så kan jag välja att inte godkänna den, t.ex. om uppsatsen bara består av meningen ”Bananer är goda.” upprepad tusen gånger. Vi har alltså implikationen ”Jag godkänner uppsatsen” ⇒ ”Texten saknar uppenbara grammatiska fel”, men inte omvändningen.

Exempel. Ett *nödvändigt* villkor för att $x \in \mathbb{R}$ löser ekvationen $\sqrt{x-3} + 5x = x^2$ är naturligtvis att $x \geq 3$, för om $x < 3$ så blir ju rotuttrycket odefinierat. Det är däremot inte ett *tillräckligt* villkor, för t.ex. $x = 3$ löser inte ekvationen. Vi har alltså $\sqrt{x-3} + 5x = x^2 \Rightarrow x \geq 3$, men inte omvändningen: det är *inte* så att varje tal ≥ 3 löser ekvationen.

Exempel. Ett *tillräckligt* villkor för att en funktion skall vara injektiv är att den är strängt växande, d.v.s. vi har implikationen ”f är strängt växande ⇒ f är injektiv”. Däremot är kravet inte alls *nödvändigt*: en funktion kan mycket väl vara injektiv utan att för den sakens skull vara strängt växande. Vi har alltså *inte* implikationen ”f är injektiv ⇒ f är strängt växande”.

Exempel. Ett *nödvändigt* villkor för att tecknet semikolon är korrekt använt i sin standardfunktion är att både det som kommer före och det som kommer efter tecknet är fullständiga huvudsatser. Detta krav är emellertid inte *tillräckligt*; huvudsatserna måste nämligen också vara innehållsmässigt närstående.

Om $P \Leftrightarrow Q$, d.v.s. om P och Q är ekvivalenta, så är alltså P ett tillräckligt *och* nödvändigt krav för Q, och tvärtom. Vi kan också säga ”P om Q” (Q tillräckligt villkor) och ”P bara om Q” (Q nödvändigt krav). Med andra ord gäller ”P om och endast om Q”, vilket ofta förkortas ”P omm Q”. På engelska skriver man ”P if and only if Q”, som förkortas ”P iff Q”.

1.1.2.3 Den kontrapositiva utsagan

Det är inte svårt att inse att utsagan

$$\text{Det regnar.} \quad \Longrightarrow \quad \text{Gatan blir blöt.} \tag{1}$$

är ekvivalent med utsagan

$$\text{Gatan blir inte blöt.} \quad \Longrightarrow \quad \text{Det regnar inte.} \tag{2}$$

Mer allmänt gäller att utsagan $P \Rightarrow Q$ är ekvivalent med utsagan ”inte $Q \Rightarrow$ inte P”; den senare utsagan kallas för den *kontrapositiva* formen av den förra utsagan. En implikation är sålunda alltid ekvivalent med sin kontrapositiva form.

Hur inser man då detta? Antag att (1) är sann, d.v.s. ”om det regnar, så blir gatan blöt”. Antag nu att gatan *inte* blir blöt. Då kan det inte regna (d.v.s. (2) gäller), ty *om* det skulle regna, så skulle gatan bli blöt (enligt (1)), men vi antog ju att så inte var fallet.

Tar man den kontrapositiva formen ”två gånger” får man tillbaka den ursprungliga implikationen (varför?), så det är klart att (2) också medför (1), eftersom (1) är den kontrapositiva formen av (2). Alltså är (1) och (2) i själva verket *ekvivalenta*, som utlovat.

Faktiskt är det vanligt att man inom matematiken, när man vill bevisa att $P \Rightarrow Q$, i stället antar att Q *inte* är sann och visar att detta medför att P då *inte* kan vara sann. Detta är en bra idé om det senare beviset är lättare att utföra.

1.1.2.4 Mer om logiska uttryck (*)

De logiska operatorerna ”och”, ”eller” och ”inte” kan förkortas ∧, ∨ respektive ¬. I de allra flesta fall finns det ingen vinst i att använda dessa symboler i stället för de ord de representerar (det gör knappast texten mer lättläst), men när man studerar logiska utsagor *i sig* kan det vara praktiskt. Om P är påståendet ”Kalle är hemma.”, Q är påståendet ”Johan är hemma.” och R är påståendet ”Solen skiner.” så har följande påståenden följande betydelser:

$P \wedge Q \wedge R$	Både Kalle och Johan är hemma, och solen skiner.
$P \wedge Q \wedge (\neg R)$	Både Kalle och Johan är hemma, men solen skiner inte.
$(P \vee Q) \wedge R$	Kalle eller Johan är hemma (d.v.s. minst en av dem), och solen skiner.
$P \wedge (\neg Q)$	Kalle är hemma, men inte Johan.
$\big(P \wedge (\neg Q)\big) \vee \big((\neg P) \wedge Q\big)$	Exakt en av Kalle och Johan är hemma. (*Antingen* Kalle *eller* Johan …)

En syntaktiskt riktig kombination av utsagor och logiska operatorer kallas för ett *logiskt uttryck*, och är alltså i sig en utsaga. Om uttrycket består av n utsagor (t.ex. $n = 3$ i fallet med P, Q och R) så finns det 2^n olika möjligheter att beakta när man analyserar det sammansatta uttryckets sanningsvärde. I vårt fall ($n = 3$) så kan ju dels P vara sann eller falsk (2 kombinationer), dels kan Q vara sann eller falsk (2 till kombinationer) och dels kan R vara sann eller falsk (ytterligare 2 kombinationer), så det finns totalt $2 \cdot 2 \cdot 2 = 2^3 = 8$ olika scenarier. I vart och ett av dessa är sanningsvärdet hos det sammansatta uttrycket förstås entydigt bestämt. Alltså kan man fullständigt beskriva sanningsvärdet hos ett uttryck i n utsagor genom att helt enkelt ange alla 2^n möjligheter var för sig. Om $n = 2$, och de två utsagorna är P och Q, låter sig detta göras i en tabell där kolumnernas rubriker är P och $\neg P$ medan radernas rubriker är Q och $\neg Q$. Tabellcellen i kolumnen P och raden $\neg Q$ är då, som exempel, uttryckets sanningsvärde när utsagan P är sann men Q är falsk.

Som tidigare nämnt är utsagan $P \wedge Q$ sann om och endast om både P och Q är sanna. Å andra sidan är $P \vee Q$ sann omm minst en av utsagorna P och Q är sann. Innebörden av ordet ”inte”, eller operatorn $\neg$, använde vi implicit ovan: Påståendet $\neg P$ är sant omm P är falskt. Vi kan sammanfatta dessa observationer, som också *definierar* operatorerna $\wedge$, $\vee$ och $\neg$, i följande sanningstabeller:

$P \wedge Q$	P	$\neg P$
Q	Sant	Falskt
$\neg Q$	Falskt	Falskt

$P \vee Q$	P	$\neg P$
Q	Sant	Sant
$\neg Q$	Sant	Falskt

P	$\neg P$
Sant	Falskt
Falskt	Sant

Genom att använda dessa tabeller, kan vi successivt konstruera sanningstabeller för mer avancerade logiska uttryck. Som exempel kan vi ta $\big(P \wedge (\neg Q)\big) \vee \big((\neg P) \wedge Q\big)$ från tabellen ovan:

$\big(P \wedge (\neg Q)\big) \vee \big((\neg P) \wedge Q\big)$	P	$\neg P$
Q	Falskt	Sant
$\neg Q$	Sant	Falskt

Utsagan $\big(P \wedge (\neg Q)\big) \vee \big((\neg P) \wedge Q\big)$ är med andra ord sann om exakt en av utsagorna P och Q är sann, d.v.s. om P eller Q är sann *men inte båda*. Språkligt svarar detta mot ”Antingen … eller …”. I vardagsspråk säger man ofta bara ”P eller Q” även i det här fallet, men inom matematiken reserveras ett ensamt ”eller” för varianten som också inkluderar fallet att båda är sanna. De logiska operatorerna vi använder inom matematiken *sammanfattar* därför inte bara det naturliga språkets logik – de *preciserar* det också.

Ibland inför man symbolen $\veebar$ för den logiska operator som har sanningstabellen ovan. $\veebar$ kallas *exklusivt eller* eftersom den *exkluderar* fallet när både P och Q är sanna; ibland kallar man $\vee$ för *inklusivt eller* för att kontrastera.

Notera att det logiska uttrycket $(P \vee Q) \wedge \big(\neg(P \wedge Q)\big)$ har samma sanningstabell som $P \veebar Q$:

$(P \vee Q) \wedge \big(\neg(P \wedge Q)\big)$	P	$\neg P$
Q	Falskt	Sant
$\neg Q$	Sant	Falskt

Att två uttryck i P och Q har samma sanningstabell betyder per definition att de alltid har samma sanningsvärde. Uttrycken är alltså ekvivalenta. Vi har därför

$$P \veebar Q \quad \Longleftrightarrow \quad \big(P \wedge (\neg Q)\big) \vee \big((\neg P) \wedge Q\big) \quad \Longleftrightarrow \quad (P \vee Q) \wedge \big(\neg (P \wedge Q)\big) \tag{3}$$

För att återgå till vårt tidigare exempel kan dessa tre ekvivalenta utsagor utläsas på följande sätt (i tur och ordning, och med varierande grad av förenkling):

- ”exakt en av Kalle och Johan är hemma” eller ”antingen Kalle eller Johan är hemma”
- ”(Kalle är hemma och Johan är ute) ELLER (Kalle är ute och Johan är hemma)”
- ”Kalle eller [=d.v.s. minst en av dem] Johan är hemma, men inte båda”.

Vidare är det enkelt att inse *De Morgans lagar*, som säger att

$$\neg (P \vee Q) \quad \Longleftrightarrow \quad (\neg P) \wedge (\neg Q)$$

och

$$\neg (P \wedge Q) \quad \Longleftrightarrow \quad (\neg P) \vee (\neg Q).$$

Låt oss förklara den första: Om det *inte* är så att P *eller* Q är sann, så måste ju både P och Q vara falska. I vårt exempel har vi följande utsaga: Om varken Kalle eller Johan är hemma, då är Kalle ute och Johan är ute. Låt oss förklara den andra: Om det *inte* är så att både P och Q är sanna, så måste minst en av dem vara falsk. I vårt exempel: Om inte både Kalle och Johan är hemma, så måste minst en av dem vara ute.

Man kan också bevisa De Morgans lagar genom att jämföra sanningstabellerna för uttrycken på båda sidorna av ekvivalenspilen i respektive lag. Det är ett mekaniskt sätt att visa lagarna; man behöver inte tänka, bara göra. Låt oss pröva att applicera den andra av De Morgans lagar på (3) ovan. Vi får då

$$P \veebar Q \quad \Longleftrightarrow \quad (P \vee Q) \wedge \big(\neg (P \wedge Q)\big) \quad \Longleftrightarrow \quad (P \vee Q) \wedge \big((\neg P) \vee (\neg Q)\big).$$

Vi får här ytterligare ett (i själva verket ett *fjärde*) sätt att i ord säga att exakt en av Kalle och Johan är hemma, nämligen ”Kalle eller [=d.v.s. minst en av dem] Johan är hemma och minst en av dem är ute”.

För framtida referens noterar vi lite snabbt följande ekvivalenser och implikationer:

1. $P \vee P \Leftrightarrow P$
2. $P \wedge P \Leftrightarrow P$
3. $\neg(\neg P) \Leftrightarrow P$
4. $P \vee Q \Leftrightarrow Q \vee P$
5. $P \wedge Q \Leftrightarrow Q \wedge P$
6. $P \Rightarrow P \vee Q$
7. $P \wedge Q \Rightarrow P$
8. $P \vee (\neg P)$ är alltid en sann utsaga
9. $P \wedge (\neg P)$ är alltid en falsk utsaga
10. Om $P \Rightarrow Q$ så är $P \wedge Q \Leftrightarrow P$.

Punkterna 4 och 5 säger att $\wedge$ och $\vee$ är *kommutativa*. De är också *associativa*: $(P \wedge Q) \wedge R \Leftrightarrow P \wedge (Q \wedge R)$ och analogt för $\vee$ (visa det, förslagsvis genom att lista alla $2^3 = 8$ möjligheterna i båda fallen). Är det dessutom så att $\wedge$ distribuerar över $\vee$ eller tvärtom? Visa eller motbevisa!

1.1.2.4.1 Sanningstabeller för implikationer

Vi har noterat att $P \Rightarrow Q$ i sig är en utsaga, om P och Q är det; implikationspilen $\Rightarrow$ är alltså också en logisk operator. Vad är dess sanningstabell? För att reda ut det betraktar vi en utsaga som vi menar är sann, t.ex.

$$x > 1 \qquad \Longrightarrow \qquad x > 0.$$

Detta skall vara en *sann* utsaga för varje $x \in \mathbb{R}$. Speciellt ger $x = 5$ att ”sant $\Rightarrow$ sant” är en sann utsaga. Samtidigt ger $x = -3$ att ”falskt $\Rightarrow$ falskt” är en sann utsaga. Slutligen ger $x = 1$ att ”falskt $\Rightarrow$ sant” skall är sant. Låt oss också betrakta en utsaga som vi menar är falsk, t.ex.

$$x > 0 \qquad \Longrightarrow \qquad x < 5.$$

Det som gör att den här utsagan är falsk är förstås att högra påståendet $x < 5$ inte behöver gälla bara för att vänstra påståendet $x > 0$ gäller. Med andra ord vill vi att ”sant $\Rightarrow$ falskt” skall vara en falsk utsaga. Vi kan sammanfatta våra observationer i sanningstabellen för implikationspilen åt höger:

$P \Rightarrow Q$	P	$\neg P$
Q	Sant	Sant
$\neg Q$	Falskt	Sant

På motsvarande sätt inser vi att sanningstabellen för implikation åt vänster har följande utseende:

$P \Leftarrow Q$	P	$\neg P$
Q	Sant	Falskt
$\neg Q$	Sant	Sant

Eftersom $P \Leftrightarrow Q$ per definition betyder $(P \Rightarrow Q) \wedge (Q \Rightarrow P)$ erhåller vi sanningstabellen för ekvivalenspilen genom att kombinera dessa två sanningstabeller med logiskt OCH:

$P \Leftrightarrow Q$	P	$\neg P$
Q	Sant	Falskt
$\neg Q$	Falskt	Sant

Två påståenden P och Q är alltså ekvivalenta om och endast om de har samma sanningsvärde. På så sätt kan man säga att ekvivalenspilen fungerar som ett likhetstecken mellan sanningsvärden.

Notera till sist att sanningstabellen för implikationen $P \Rightarrow Q$ är identisk med den för $\neg(P \wedge \neg Q)$. En implikation är sålunda sann omm det inte är så illa att den är ”motbevisad”, i den meningen att Q är falsk trots att P är sann. Till exempel kan utsagan

$$x > 1 \qquad \Longrightarrow \qquad x > 0$$

formuleras ”det är aldrig så att $x > 1$ och $x \leq 0$”.

1.1.3 Mängder

Mängdbegreppet är mycket centralt inom matematiken. En mängd är en samling objekt (vanligtvis av samma ”typ”, t.ex. reella tal) i vilken ett specifikt objekt kan förekomma högst en gång och där man inte tar hänsyn till ”ordningen” med vilken objekten räknas upp. De objekt som en mängd består av kallas för mängdens *element*. En mängd kan bestå av ändligt eller oändligt många element (t.ex. eleverna i en klass eller de jämna positiva heltalen $2, 4, 6, 8, 10, 12, \ldots$). Om mängden A är ändlig så använder vi beteckningen $|A|$ för antalet element i mängden.

När man i formler anger en ändlig mängd (i text) kan man räkna upp mängdens element (i godtycklig ordning) med kommatecken mellan dem; hela listan skrives inom klammerparenteser. Mängden som består av elementen a, b och c kan sålunda skrivas $\{a, b, c\}$, även om $\{b, c, a\}$ är exakt samma mängd och därmed precis lika rätt. Två mängder säges (förstås) vara lika, omm de innehåller exakt samma element. Så är t.ex. $\{a, b, c\} = \{b, c, a\}$; dessa två mängder är identiska, precis som talen $1 + 1$ och 2.

Om ett objekt a är ett av elementen i en mängd A säger man att a *tillhör* A och skriver $a \in A$. Till exempel har vi $2 \in \{1, 2, 3\}$. Däremot är inte 5 ett element i mängden, vilket skrives $5 \notin \{1,2,3\}$.

Ibland kan man tillåta sig att förkorta en uppräkning av element; man använder då tre punkter (…)[5]. Till exempel är mängden av de 5000 första heltalen,

$$\{1, 2, 3, \ldots, 4999, 5000\},$$

en mängd som består av exakt 5000 element. Mängder som består av oändligt många element kan ibland också anges med ”…”. Till exempel kan vi betrakta mängden av alla jämna heltalskvadrater:

$$\{0^2, 1^2, 2^2, 3^2, 4^2, 5^2, \ldots\} = \{0, 1, 4, 9, 16, 25, \ldots\}.$$

De mest kända namngivna mängderna är förmodligen talmängderna $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$ och $\mathbb{C}$. $\mathbb{N}$ består av alla *naturliga tal*, d.v.s. talen 0, 1, 2, 3, osv. Den här mängden (liksom de övriga av de här talmängderna) består tydligen av oändligt många element. $\mathbb{Z}$ är mängden av alla *heltal*, d.v.s. talen $0, \pm 1, \pm 2, \pm 3, \ldots$. $\mathbb{Q}$ är mängden av alla *rationella tal*, d.v.s. alla tal som kan skrivas som kvoten mellan två heltal. Notera att varje heltal är ett rationellt tal; till exempel är $5 = \frac{5}{1}$. $\mathbb{R}$ är mängden av alla *reella tal*, d.v.s., i princip, alla talen på tallinjen. Dessa inkluderar alla rationella tal liksom övriga, *irrationella* tal som π, $\sqrt{2}$ och e. Slutligen är $\mathbb{C}$ mängden av alla s.k. *komplexa tal* som vi återkommer till senare.

Om A och B är två mängder, och varje element i A också finns i B, så säger vi att A är en *delmängd* av B. I symboler skrives detta $A \subset B$. Vi har alltså $\{1, 2, 3\} \subset \{1, 2, 3, 4, 5, 6\} \subset \mathbb{Z}$. I själva verket gäller $\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R} \subset \mathbb{C}$. Notera att $A \subset A$ för varje mängd A. Om $A \subset B$ men $A \neq B$, d.v.s. det finns (minst) ett element i B som *inte* finns i A, så säger man att A är en *äkta* delmängd av B. Om $A \subset B$ och $B \subset A$ så är $A = B$; i praktiken är det ofta de två förra egenskaperna man visar när man vill visa att två mängder är identiska.

Det finns exakt en mängd som består av noll element. Den mängden kallas *tomma mängden* och skrives vanligtvis $\emptyset$ även om notationen $\{\quad\}$ också förekommer. Den tomma mängden är en delmängd till varje mängd.

[5] I praktiken anger dessa tre punkter att ”det antydda mönstret fortsätter”.

Ibland är det svårt eller omöjligt att specificera en mängd genom att räkna upp elementen i den. Då kan man använda en speciell notation som på engelska kallas för *set-builder notation*. Man specificerar då att mängden består av alla de element som finns i en annan mängd *och som uppfyller ett visst villkor*. Till exempel kan mängden av reella alla tal mellan 0 och 10 (inklusive gränserna) skrivas

$$\{x \in \mathbb{R}: 0 \leq x \leq 10\}$$

som utläses (t.ex.) ”mängden av alla reella tal x sådana att $0 \leq x \leq 10$”.

En variant av den här notationen förekommer när man vill att en mängd skall bestå av alla element *på en viss form*, som i

$$M = \{(x, y): x \in \mathbb{R}, y \in \mathbb{Z}\}$$

som är en mängd som består av alla talpar (x, y) där x är ett reellt tal och y är ett heltal. Till exempel har vi $(2,3) \in M$ och $(\pi, 5) \in M$ men $(\pi, \sqrt{2}) \notin M$ eftersom $\sqrt{2} \notin \mathbb{Z}$.

I många fall betraktar man mängder som alla är delmängder av en given mängd. Till exempel kanske man studerar delmängder av de reella talen $\mathbb{R}$. Ett annat exempel, inom geometrin, är när man studerar delmängder (såsom cirklar, linjer, parabler, kaniner eller andra geometriska figurer) av planet $\mathbb{R}^2$ (som vi alldeles strax kommer att prata mer om!). Den stora mängden, som alla mängder är delmängder av, kallas i sådana fall för *universum* eller *grundmängd*.

1.1.3.1 Mängdoperationer

Givet två mängder kan man bilda en ny mängd på flera sätt:

- **Union.** Unionen $A \cup B$ av två mängder A och B är den mängd som innehåller allt som finns i A och dessutom allt som finns i B. Exempel: $\{1, 2, 3\} \cup \{2, 3, 4\} = \{1,2,3,4\}$.

- **Snitt.** Snittet $A \cap B$ av två mängder A och B är den mängd som innehåller de element som finns i både A och B, d.v.s. de element som A och B har gemensamt. Snittet kan alltså mycket väl vara den tomma mängden även om varken A eller B är den tomma mängden. Exempel: $\{1,2,3\} \cap \{2,3,4\} = \{2,3\}$. Om $A \cap B = \emptyset$ säger man att A och B är *disjunkta* mängder – de har alltså inga gemensamma element.

- **Mängddifferens.** Mängddifferensen $A \setminus B$ erhålles genom att från mängden A ta bort alla element som också finns i B. Exempel: $\{1,2,3\} \setminus \{2,3,4\} = \{1\}$.

Om vår grundmängd är U kan vi skriva

$$\begin{aligned} A \cup B &= \{x \in U: \quad x \in A \quad \text{eller} \quad x \in B\} \\ A \cap B &= \{x \in U: \quad x \in A \quad \text{och} \quad x \in B\} \\ A \setminus B &= \{x \in U: \quad x \in A \quad \text{och} \quad x \notin B\} = \\ &= \{x \in A: \quad x \notin B\}. \end{aligned}$$

Låt oss lista några vanliga (och uppenbara) egenskaper för mängdoperationerna. För alla mängder A och B gäller

- $A \cup A = A$
- $A \cap A = A$

- $A \cup B = B \cup A$
- $A \cap B = B \cap A$
- $A \subset A \cup B$
- $A \cap B \subset A$.

(Har vi dessutom associativa och distributiva lagar?) Givet en grundmängd U definierar man också till varje mängd $A \subset U$ *komplementet* $A^C = U \setminus A$. A^C består alltså av allt som *inte* finns i A. Vi inser direkt att

- $U^C = \emptyset$
- $\emptyset^C = U$
- $A \cup A^C = U$
- $A \cap A^C = \emptyset$
- $A \setminus B = A \cap B^C$.

Givet en grundmängd U kan *utsagor* om elementen och *delmängder* av grundmängden ses som två olika sidor av samma mynt. Att säga att $x \in U$ har en viss egenskap är ju precis samma sak, som att säga att x tillhör mängden av de element som har denna egenskap. Till exempel betyder ”n är ett heltal” exakt samma sak som ”$n \in \mathbb{Z}$”, och ”$x = -3$ eller $x = 3$” betyder exakt samma sak som ”$x \in \{-3, 3\}$”.

Mer formellt, om $P(x)$ är en utsaga om elementet $x \in U$ så kan vi forma mängden $\{x \in U: P(x)\}$. Omvänt, givet en delmängd M så är $x \in M$ en utsaga om elementet x. Börjar man med en mängd och formar motsvarande utsaga och sedan motsvarande mängd, får man samma mängd som man började med; börjar man med en utsaga och formar motsvarande mängd och sedan motsvarande utsaga, får man en utsaga som är ekvivalent med den man började med. (Föregående mening kanske lät invecklad, men slutsatsen är faktiskt mycket självklar.)

I denna analogi mellan logik och mängdlära svarar de logiska operatorerna ”och”, ”eller” och ”inte” mot mängdoperatorerna ”snitt”, ”union” respektive ”komplement”. Ekvivalens svarar mot likhet.

Som exempel kan vi betrakta heltalen $\mathbb{Z}$. Låt $j(n)$ vara utsagan ”n är ett jämnt tal” och låt $p(n)$ vara utsagan ”n är ett positivt tal”. Vi kan då införa mängderna

$$J \coloneqq \{n \in \mathbb{Z}:\ j(n)\}$$
$$P \coloneqq \{n \in \mathbb{Z}:\ p(n)\}$$

av alla jämna respektive positiva tal. Då är

$$J \cup P = \{n \in \mathbb{Z}:\ j(n) \quad \text{eller} \quad p(n)\}$$

mängden av alla heltal som är jämna *eller* positiva medan

$$J \cap P = \{n \in \mathbb{Z}:\ j(n) \quad \text{och} \quad p(n)\}$$

är mängden av alla heltal som är jämna *och* positiva. Vi har också mängden

$$J^C = \{n \in \mathbb{Z}:\ \text{inte}\ \ j(n)\}$$

av alla tal som *inte* är jämna (d.v.s. mängden av alla udda tal).

1.1.3.2 Kartesisk produkt

Vi inför här en ny mängdoperation. Givet två mängder A och B inför vi *den kartesiska produkten*

$$A \times B = \{(a, b) : a \in A, b \in B\}$$

som mängden av alla *ordnade par* (a, b) där $a \in A$ och $b \in B$. ”Ordnade” betyder att $(a, b) \neq (b, a)$ i allmänhet (om inte $a = b$, förstås). Till exempel, om

$$D = \{\text{kanin, råtta, hund}\}$$

är en mängd av djur och

$$K = \{♂, ♀\}$$

är en mängd bestående av symbolerna för könen *hane* och *hona* så är

$$D \times K = \{(\text{kanin}, ♂), (\text{kanin}, ♀), (\text{råtta}, ♂), (\text{råtta}, ♀), (\text{hund}, ♂), (\text{hund}, ♀)\}$$

en mängd som består av de åtta uppräknade elementen på formen (djur, kön), d.v.s. ”alla möjliga kombinationer”. Om A är en mängd och k är ett positivt heltal brukar man också skriva A^k för A ”gånger sig själv” k gånger, där ”multiplikationen” är just den kartesiska produkten, fast bara nästan. Man brukar nämligen ”kollapsa” de ordnade paren. För att förstå vad detta innebär, betrakta exemplet

$$K^2 = K \times K = \{(♂, ♂), (♂, ♀), (♀, ♂), (♀, ♀)\}.$$

Vi får därför

$$(K \times K) \times K = \left\{ \begin{matrix} ((♂, ♂), ♂), ((♂, ♂), ♀), ((♂, ♀), ♂), ((♂, ♀), ♀), \\ ((♀, ♂), ♂), ((♀, ♂), ♀), ((♀, ♀), ♂), ((♀, ♀), ♀) \end{matrix} \right\}$$

medan

$$K \times (K \times K) = \left\{ \begin{matrix} (♂, (♂, ♂)), (♂, (♂, ♀)), (♂, (♀, ♂)), (♂, (♀, ♀)), \\ (♀, (♂, ♂)), (♀, (♂, ♀)), (♀, (♀, ♂)), (♀, (♀, ♀)) \end{matrix} \right\},$$

d.v.s. $(K \times K) \times K \neq K \times (K \times K)$; operationen är alltså inte associativ. I praktiken är man däremot sällan intresserad av den ”gruppering” som uppstår vid upprepad kartesisk produkt, så man definierar helt enkelt

$$K^3 = \left\{ \begin{matrix} (♂, ♂, ♂), (♂, ♂, ♀), (♂, ♀, ♂), (♂, ♀, ♀), \\ (♀, ♂, ♂), (♀, ♂, ♀), (♀, ♀, ♂), (♀, ♀, ♀) \end{matrix} \right\}.$$

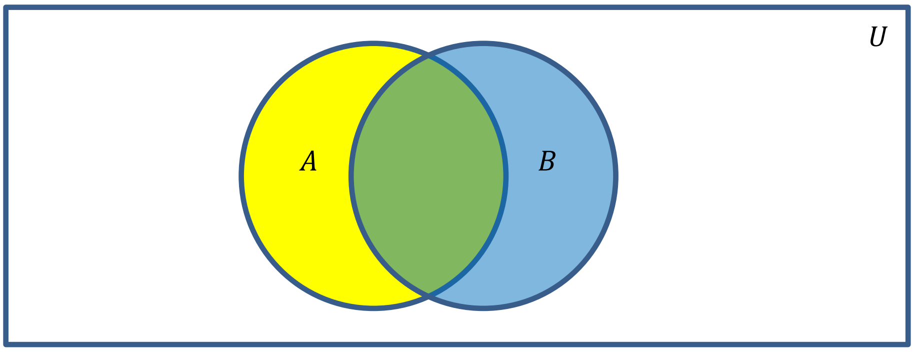
Mer allmänt sätter vi

$$A^k \coloneqq \{(a_1, a_2, \dots, a_k) : \text{varje } a_j \in A\}$$

för varje mängd A. Nu kan vi titta på de kanske mest kända exemplen på kartesiska produkter: $\mathbb{R}^n$ där $n \in \mathbb{Z}^+$. Per definition är $\mathbb{R}^n$ mängden av alla n-tipler $(x_1, x_2, \dots, x_n)$ där varje $x_j \in \mathbb{R}$. I synnerhet kan $\mathbb{R}^2$ betraktas som koordinater i planet medan $\mathbb{R}^3$ kan betraktas som koordinater i vårt tredimensionella rum. Ofta säger man rent av att $\mathbb{R}^2$ ”är” planet och $\mathbb{R}^3$ ”är” rummet.

1.1.3.3 Mängddiagram och mängdalgebra (*)

Betrakta en grundmängd U och två delmängder A och B. Välj ett godtyckligt element $x \in U$. Antingen är $x \in A$ eller så är $x \in A^C$, och antingen så är $x \in B$ eller så är $x \in B^C$. Det finns sålunda bara fyra möjligheter: $x \in A \cap B$, $x \in A^C \cap B$, $x \in A \cap B^c$ eller $x \in A^C \cap B^C$. Dessa fyra mängder är disjunkta och "täcker" U, d.v.s. deras union är U. Man kan illustrera situationen i ett mängddiagram, som det nedan:



I den här bilden representeras mängderna A och B av diskar. Idén är att varje element $x \in U$ kan tänkas svara mot en punkt i diagrammet (som begränsas av en stor rektangel). x tillhör A (respektive B) omm punkten finns i disken som representerar A (respektive B). Notera att alla fyra möjligheter som omnämndes ovan tillåts; diagrammet består ju av fyra delar.

Begå inget tankemisstag här: den här metoden att visualisera relationer mellan mängder fungerar för alla sorters mängder. Det är alltså *inte* nödvändigtvis så att $U = \mathbb{R}^2$ och att delmängderna A och B är fyllda diskar som överlappar varandra precis som i illustrationen! U kanske är mängden av elever i en klass, A delmängden av de elever som spelar något musikinstrument och B delmängden av de elever som utövar någon form av idrott. [Vad är då det gröna området? Vad är det färgade området? Vad är det vita området?]

I diagrammet svarar den mellersta, gröna, delen mot snittet $A \cap B$: en punkt $x \in U$ representeras av en punkt i det här området omm $x \in A \cap B$. På samma sätt svarar $A \cup B$ mot hela det färgade området. $A \cap B^C = A \setminus B$ svarar mot det gula området, och $B \cap A^C = B \setminus A$ mot det blåa. Slutligen noterar vi att $A^C \cap B^C = (A \cup B)^C$ svarar mot det ofärgade området utanför "kikaren".

Den uppmärksamme läsaren noterade just att en av De Morgans lagar tydligen gäller även för mängder. Det är förstås ingen slump, med tanke på att utsagor om element i U och delmängder av U, som tidigare nämnt, är två sidor av samma mynt.

Vilken mängdoperation svarar mot den logiska operatorn exklusivt eller ($\veebar$)? Om vi befinner oss i *antingen* A *eller* B (men *inte* båda!), d.v.s. om $(x \in A) \veebar (x \in B)$, så befinner vi oss tydligen på gult område eller på blått område (men inte på grönt eller vitt område). Den här mängden kallas för den *symmetriska differensen* av A och B, och betecknas $A \triangle B$.

De fyra formlerna för exklusivt eller ger förstås upphov till fyra formler för den symmetriska differensen:

$P \veebar Q$	$A \triangle B$
$\big(P \wedge (\neg Q)\big) \vee \big((\neg P) \wedge Q\big)$	$(A \cap B^C) \cup (A^C \cap B)$ d.v.s. $(A \setminus B) \cup (B \setminus A)$
$(P \vee Q) \wedge \big(\neg(P \wedge Q)\big)$	$(A \cup B) \cap (A \cap B)^C$ d.v.s. $(A \cup B) \setminus (A \cap B)$
$(P \vee Q) \wedge \big((\neg P) \vee (\neg Q)\big)$	$(A \cup B) \cap (A^C \cup B^C)$

Utsagorna i första kolumnen är alla ekvivalenta; mängderna i andra kolumnen är alla identiska (lika).

1.1.4 Kvantifierare

Ofta studerar man en mängd X och ett påstående $P(x)$ om ett godtyckligt element $x \in X$. Det kan då hända att $P(x)$ gäller för *något* (d.v.s. minst ett) $x \in X$. Det kan också hända att $P(x)$ gäller för *samtliga* $x \in X$. Om t.ex. $X = \mathbb{R}$ är de reella talen och påståendena $P_1(x)$ och $P_2(x)$ är ”$x^2 = 7$” respektive ”$x^2 \geq 0$” så gäller $P_1(x)$ för minst ett (i det här fallet precis 2) element i X, medan $P_2(x)$ gäller för alla element i X. Vi skriver då

$$\text{det existerar ett } x \in \mathbb{R} \text{ sådant att } x^2 = 7$$

respektive

$$x^2 \geq 0 \text{ för varje } x \in \mathbb{R}.$$

Uttrycken ”det existerar” och ”för varje/alla” kallas för *kvantifierare*, och inom matematiken finns speciella symboler för dessa, nämligen ∃ (”det existerar”) och ∀ (”för varje/alla”). Vi kan därför skriva

$$\exists x \in \mathbb{R}: x^2 = 7$$

(där kolonet skall utläsas ”sådant att”) samt

$$x^2 \geq 0, \qquad \forall x \in \mathbb{R}.$$

Ibland används också kvantifieraren ∃! (”det existerar ett unikt”) om mängden innehåller exakt ett (1) element med egenskapen. T.ex. har vi (där $\mathbb{R}^+$ är mängden av alla *positiva* reella tal)

$$\exists!\, x \in \mathbb{R}^+: x^2 = 7.$$

I praktiken är man ofta ”slarvig” och utelämnar kvantifierare trots att de är nödvändiga för att texten skall vara lättläst och korrekt. En artikel ämnad åt gymnasieelever kan mycket väl inledas med texten ”Vi vet ju alla att $2a + 5a = 7a$…”, vilket strängt taget är obegripligt, eftersom man inte sagt vad a är för något. Att det är ett *tal* är nog underförstått, men *vilket*? Det man egentligen menar är ”Vi vet ju alla att $2a + 5a = 7a$ för varje reellt tal a…”. Det första uttalandet handlar om det specifika talet a, medan det andra handlar om mängden $\mathbb{R}$.

I *praktiken* är detta nästan aldrig ett problem, eftersom mänskliga läsare av texter ofta förstår av sammanhanget vad som egentligen avses. Som tumregel kan sägas att en *för alla*-kvantifierare är underförstådd om en utsaga handlar om en sedan tidigare okänd storhet (som a i vårt exempel) och man *inte* explicit säger för vilka a utsagan är sann, eller säger att det rör sig om ekvationslösning (d.v.s. bestämning av de värden som gör utsagan sann).

1.1.4.1 Mer om kvantifierare (*)

Om X är en ändlig mängd kan kvantifierarna uttryckas med de logiska operatorerna ”ELLER” respektive ”OCH”. Om t.ex. $X = \{a, b, c\}$ så kan utsagan

$$\exists x \in X : x > 0$$

också skrivas

$$a > 0 \text{ eller } b > 0 \text{ eller } c > 0.$$

Utsagan

$$x > 0, \qquad \forall x \in X$$

kan å sin sida skrivas

$$a > 0 \text{ och } b > 0 \text{ och } c > 0.$$

OCH:et av noll stycken påståenden (den tomma konjunktionen) definieras som sann, eftersom det logiska värdet ”sant” fungerar som ett neutralt element med avseende på OCH. ELLER:et av noll stycken påståenden (den tomma disjunktionen) definieras av motsvarande anledning som falsk. Det betyder speciellt att om $X = \emptyset$ och $P(x)$ är en utsaga om elementen i X så är utsagan

$$\exists x \in X : P(x)$$

falsk medan utsagan

$$P(x), \qquad \forall x \in X$$

är sann.

Exempel

I skrivande stund finns det inga människor med vingar, så mängden av människor med vingar är tom. Påståendet ”det finns en människa med vingar, som kan flyga snabbare än ljudet” är därför falskt. Å andra sidan är påståendet ”alla människor med vingar kan flyga snabbare än ljudet” sant.

Ett alternativt sätt att motivera varför $P(x)\ \forall x \in \emptyset$ alltid är en sann utsaga är genom att utgå från implikationens sanningstabell. Utsagan

$$P(x), \qquad \forall x \in X$$

kan ju också skrivas

$$x \in X \qquad \Longrightarrow \qquad P(x)$$

så med $X = \emptyset$ blir $x \in X$ ett falskt påstående (för varje x i ett universum som innehåller X), varför implikationen är sann (för varje x).

Övning

1. Motivera varför $\emptyset \subset B$ gäller för varje mängd B genom att notera att en inklusion $A \subset B$ också kan skrivas $x \in A \Rightarrow x \in B$ och se vad som händer då $A = \emptyset$.

1.2 Tal, algebra och geometri

Matematik är *inte* vetenskapen om tal, utan snarare vetenskapen om exakta logiska resonemang i största allmänhet. Icke desto mindre spelar tal en betydande roll för allt mänskligt tänkande, både naturliga tal (0, 1, 2, …) som anger antal och reella tal som ofta är mätvärden för fysikaliska storheter, såsom avstånd, area och volym. Läsaren förväntas vara bekant med räkning med heltal, rationella tal och reella tal åtminstone på högstadienivå. Följande genomgång är alltså väldigt ytlig eftersom läsaren förväntas känna till så gott som alla utelämnade detaljer.

Det finns flera olika sorters tal. De enklaste talen är *de naturliga talen* i

$$\mathbb{N} = \{0, 1, 2, 3, \dots\}$$

som t.ex. kan ange *antal*. Givet två naturliga tal m och $n \in \mathbb{N}$ kan man också alltid beräkna *summan* $m + n$ och produkten $m \cdot n$; produkten skrivs oftast utan multiplikationspunkten. Om $m \geq n$, d.v.s. om m är ett minst lika stort tal som n, kan vi också beräkna *differensen* $m - n$. Om $n \neq 0$ kan man i vissa fall också finna kvoten $\frac{m}{n}$ som definieras som det unika (om det ens existerar) naturliga tal som multiplicerat med n blir m.

Vi har också *heltalen* i

$$\mathbb{Z} = \{\dots, -3, -2, -1, 0, 1, 2, 3, \dots\}.$$

Här kan vi *alltid* beräkna differensen mellan två tal; t.ex. är $5 - 12 = -7$. De negativa talen $-1, -2, -3, \dots$ kan betraktas som en tankeabstraktion. Även om man inte kan ha -20 kronor i handen, så kan man ha det saldot på sitt bankkonto.

Vidare har vi *de rationella talen*

$$\mathbb{Q} = \left\{\frac{p}{q}: \quad p, q \in \mathbb{Z}, \quad q \neq 0\right\}$$

som utöver alla heltal också innehåller alla bråktal, som t.ex. $\frac{1}{2} = 0.5$, $\frac{3}{4} = 0.75$, $\frac{1}{3} = 0.333\dots$ och $\frac{50}{4} = 12.5$. Med dessa tal kan man alltid beräkna kvoter $\frac{r}{q}$ om bara $q \neq 0$. När ett rationellt tal skrives med hjälp av decimalpunkt och decimaler är antalet siffror som krävs alltid ändligt eller så upprepas, efter en ändlig initial följd av siffror, en ändlig följd av siffror om och om igen. I det senare fallet kan detta markeras genom att följden som upprepas i oändlighet skrives med ett streck ovanför:

$$\begin{aligned}
&\frac{1}{2} = 0.5, \qquad \frac{3}{4} = 0.75, \qquad \frac{50}{4} = 12.5, \qquad \frac{3}{8} = 0.375 \\
&\frac{1}{3} = 0.333\dots = 0.\overline{3} \\
&\frac{1}{7} = 0.142857142857142857\dots = 0.\overline{142857} \\
&\frac{23}{17} = 1.352941176470588235294117647058823529411764705882\dots = \\
&\qquad\qquad = 1.\overline{3529411764705882}.
\end{aligned}$$

I de flesta praktiska sammanhang räcker de rationella talen. Med de rationella talen kan man ju med *godtycklig* noggrannhet ange alla tänkbara fysikaliska storheter; med ”godtycklig noggrannhet” menas att felet kan fås hur litet (positivt) man vill: $< 0.1, < 0.01, < 0.001, < 10^{-100}$. Ur en teoretisk synvinkel är de rationella talen emellertid otillräckliga. Bland annat kan man resonera sig fram till att en kvadrat med sidlängden 1 har en diagonal vars längd *inte* exakt kan uttryckas som ett rationellt tal, d.v.s. *varje* rationellt tal kan visas vara antingen större eller mindre än den exakta längden (även om felet i praktiken kan fås godtyckligt litet).

Intuitivt kan man säga att det som ”saknas” är tal som består av oändliga icke-periodiska följder av siffror efter decimalpunkten. Lägger man till alla sådana tal till de rationella talen får man *de reella talen*. De reella talen kan sägas vara alla tänkbara punkter på den reella tallinjen, och inkluderar speciellt de rationella talen och (därmed) heltalen.

1.2.1 Algebra

Läsaren är sedan grundskolan bekant med de flesta elementära algebraiska räknelagar som gäller för (rationella och reella) tal. Till exempel har vi de *kommutativa lagarna*

$$a + b = b + a, \qquad ab = ba,$$

de *associativa lagarna*

$$a + (b + c) = (a + b) + c, \qquad a(bc) = (ab)c$$

och de *distributiva lagarna*

$$a(b + c) = ab + ac, \qquad (a + b)c = ac + bc.$$

Från dessa följer bland annat $(a + b)(c + d) = ac + ad + bc + bd$ ty

$$(a + b)(c + d) = a(c + d) + b(c + d) = ac + ad + bc + bd.$$

Speciellt har vi de s.k. *kvadreringsreglerna*

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

och *konjugatregeln*

$$(a + b)(a - b) = a^2 - b^2.$$

En praktiskt viktig observation är att $ab = 0 \Rightarrow a = 0$ eller $b = 0$. Mer allmänt har vi

$$ab = ac \qquad \Longrightarrow \qquad \begin{cases} a = 0 \\ \text{eller} \\ b = c. \end{cases}$$

En möjlighet är ju onekligen att $a = 0$ (i vilket fall b och c inte alls måste vara lika). Om så *inte* är fallet kan vi dividera båda talen i likheten med a varvid vi erhåller likheten $b = c$. Ett alternativt bevis är som följer:

$$ab = ac \qquad \Longrightarrow \qquad ab - ac = 0 \qquad \Longrightarrow \qquad a(b - c) = 0$$

så att $a = 0$ eller $b - c = 0$, d.v.s. $a = 0$ eller $b = c$. En annan observation är att

$$x = -x \qquad \Longrightarrow \qquad x = 0.$$

Det finns flera sätt att inse detta. Antingen noterar man helt enkelt att det enda tal x som är lika med minus sig självt är noll (tänk på vad som händer på tallinjen om man byter ut x mot $-x$). Alternativt så gäller ju att om $x \neq 0$ så kan vi dividera med x i båda leden, varvid vi erhåller $1 = -1$ som är absurt. Tredje alternativet är att skriva $x = -x \Rightarrow x + x = 0 \Rightarrow 2x = 0 \Rightarrow x = 0$.

1.2.1.1 Neutrala element och inverser

Av de fyra räknesätten addition, subtraktion, multiplikation och division kan man säga att endast additionen och multiplikationen är ”fundamentala”, i den meningen att de två andra kan beskrivas med hjälp av dessa: addition ger upphov till subtraktion medan multiplikation ger upphov till division.

Låt oss först betrakta additionen. Man säger att talet $0 \in \mathbb{R}$ är ett *neutralt element* med avseende på addition eftersom $0 + a = a + 0 = a$ för varje $a \in \mathbb{R}$; addition av talet 0 har så att säga ”ingen effekt”. Det är vidare lätt att inse att det till varje tal $a \in \mathbb{R}$ hör exakt ett tal $a' \in \mathbb{R}$ sådant att $a + a' = a' + a = 0$, d.v.s. varje tal $a \in \mathbb{R}$ har ett ”motsatt” tal sådant att addition av dessa två tal ger precis det neutrala elementet 0. Talet a' betecknas normalt sett $-a$ och kallas för a:s *additiva invers* (eller det *motsatta* talet). Till exempel är den additiva inversen till talet 6 lika med -6, eftersom $6 + (-6) = (-6) + 6 = 0$.

För varje par $a, b \in \mathbb{R}$ definieras sedan deras *differens* $a - b$ som $a + (-b)$, d.v.s. verkan av subtraktionsoperationen $-$ på två tal a och b definieras som *summan* av det första talet och den additiva inversen av det andra.

På ett i stort sett analogt sätt definieras division med hjälp av multiplikation. Man säger att talet $1 \in \mathbb{R}$ är ett *neutralt element* med avseende på multiplikation eftersom $1 \cdot a = a \cdot 1 = a$ för varje $a \in \mathbb{R}$; multiplikation med talet 1 har ”ingen effekt”. Det är lätt att inse att till varje *nollskilt* tal $a \in \mathbb{R} \setminus \{0\}$ hör precis ett tal $a' \in \mathbb{R}$ sådant att $a \cdot a' = a' \cdot a = 1$, d.v.s. varje tal $a \in \mathbb{R} \setminus \{0\}$ har ett ”inverst” tal sådant att multiplikation av dessa tal ger precis det neutrala elementet 1. Talet a' betecknas normalt sett a^{-1} eller $\frac{1}{a}$ och kallas för a:s *multiplikativa invers* (eller det *inversa* talet). Till exempel är den multiplikativa inversen till talet 2 lika med $\frac{1}{2}$ eftersom $2 \cdot \frac{1}{2} = \frac{1}{2} \cdot 2 = 1$.

För varje par $a, b \in \mathbb{R}$, med $b \neq 0$, definieras då *kvoten* $\frac{a}{b}$ som $a \cdot b^{-1}$, d.v.s. verkan av divisionsoperationen på två tal a och $b \neq 0$ definieras som *produkten* av det första talet och den multiplikativa inversen av det andra.

Notera speciellt att talet 0 saknar multiplikativ invers. Det finns ju inget tal $a' \in \mathbb{R}$ sådant att $0 \cdot a' = 1$. Därför kan man heller inte dividera *med* noll. Däremot kan noll divideras med något annat (med noll som resultat): t.ex. är $\frac{0}{5} = 0 \cdot \frac{1}{5} = 0$.

1.2.1.2 Räkning med bråk

Läsaren förväntas – som nämnt i inledningen – redan vara väl förtrogen med räkning med bråk. Till exempel känner läsaren redan till att

$$\frac{a}{b} = \frac{ka}{kb}, \qquad k \cdot \frac{a}{b} = \frac{ka}{b}, \qquad \frac{a}{b} + \frac{c}{b} = \frac{a + c}{b}, \qquad \frac{a}{b} + \frac{c}{d} = \frac{ad}{bd} + \frac{cb}{db} = \frac{ad + cb}{bd}, \qquad \frac{1}{\frac{a}{b}} = \frac{b}{a},$$

$$\frac{a}{b} \cdot \frac{c}{d} = \frac{ac}{bd}, \qquad \frac{a}{b} \Big/ \frac{c}{d} = \frac{ad}{bc}, \qquad \frac{\left(\frac{a}{b}\right)}{c} = \frac{a}{bc}$$

där alla nämnare förutsätts vara nollskilda.

1.2.1.2.1 Bevis av några av resultaten (*)

Som en övning i algebra ger vi här några exempel på hur resultaten ovan kan visas utifrån de mest grundläggande egenskaperna hos de reella talen, nämligen de kommutativa, associativa och distributiva lagarna, vilka kan betraktas som axiom för de reella talen. Det enda vi behöver komma ihåg är då vad division betyder: $\frac{a}{b}$ är bara en symbol för produkten $a \cdot b^{-1}$ mellan a och b:s multiplikativa invers.

Sats. För alla $a, b, c, k \in \mathbb{R}$ gäller

$$k \cdot \frac{a}{b} = \frac{ka}{b}, \qquad \frac{a}{b} = \frac{ka}{kb}, \qquad \frac{a}{b} + \frac{c}{b} = \frac{a+c}{b}, \qquad \frac{a}{b} + \frac{c}{d} = \frac{ad}{bd} + \frac{cb}{db} = \frac{ad+cb}{bd}$$

där samtliga nämnare förutsätts vara nollskilda.

Bevis. Den associativa lagen ger

$$k \cdot \frac{a}{b} \overset{\text{def}}{=} k \cdot (a \cdot b^{-1}) = (k \cdot a) \cdot b^{-1} \overset{\text{def}}{=} \frac{ka}{b}.$$

Den multiplikativa inversen till ka är $k^{-1} \cdot a^{-1}$ eftersom $ka \cdot (k^{-1} \cdot a^{-1}) = kk^{-1} \cdot aa^{-1} = 1 \cdot 1 = 1$. Därför är

$$\frac{ka}{kb} \overset{\text{def}}{=} (ka) \cdot (kb)^{-1} = (ka) \cdot k^{-1} \cdot b^{-1} = kk^{-1} \cdot ab^{-1} = ab^{-1} \overset{\text{def}}{=} \frac{a}{b}.$$

Vidare har vi

$$\frac{a}{b} + \frac{c}{b} \overset{\text{def}}{=} a \cdot b^{-1} + c \cdot b^{-1} = (a+c)b^{-1} \overset{\text{def}}{=} \frac{a+c}{b}.$$

Det fjärde resultatet följer direkt av de två föregående. ■

Övning

1. Visa de fyra återstående resultaten i avsnitt 1.2.1.2.

1.2.1.3 Potenser

Om $a \in \mathbb{R}$ och $n \in \mathbb{Z}^+$ (ett *positivt* heltal) inför vi beteckningen

$$a^n := \underbrace{a \cdot a \cdot \cdots \cdot a}_{n \text{ faktorer}}.$$

Det följer omedelbart att $a^n \cdot a^m = a^{n+m}$ och $(a^n)^m = a^{nm}$. Till exempel är ju

$$a^n \cdot a^m = \underbrace{a \cdot a \cdot \cdots \cdot a}_{n \text{ faktorer}} \cdot \underbrace{a \cdot a \cdot \cdots \cdot a}_{m \text{ faktorer}} = \underbrace{a \cdot a \cdot \cdots \cdot a}_{n+m \text{ faktorer}} = a^{n+m}.$$

Dessutom, om $n > m$ så är $\frac{a^n}{a^m} = a^{n-m}$. Om vi vidare definierar[6]

$$a^0 := 1$$

för varje $a \neq 0$ och

$$a^{-n} := \frac{1}{a^n}$$

för varje $a \neq 0$ och $n \in \mathbb{Z}^+$ så följer direkt

Sats. För varje $m, n \in \mathbb{Z}$ och $a \neq 0$ gäller

$$a^n a^m = a^{n+m}, \qquad (a^n)^m = a^{nm}, \qquad \frac{a^n}{a^m} = a^{n-m}.$$

Det är också mycket lätt att inse (visa det!) att om $a, b \in \mathbb{R}$ [i andra fallet med $b \neq 0$] så gäller för varje $n \in \mathbb{Z}$ att

$$(ab)^n = a^n b^n, \qquad \left(\frac{a}{b}\right)^n = \frac{a^n}{b^n}.$$

Man kan också enkelt definiera potenser med exponenter som *inte* är heltal, utan allmänna *rationella* tal, på ett sådant sätt att räknereglerna ovan fortsätter att gälla och på ett sådant sätt att a^r bevarar sin natur om r är ett heltal. Däremot kommer vi i dessa fall att kräva att basen är icke-negativ. Vi vill t.ex. att räknelagen

$$\left(a^{\frac{1}{q}}\right)^q = a^{\left(q \cdot \frac{1}{q}\right)} = a^1 = a, \qquad q \in \mathbb{Z}^+$$

skall gälla. Vi måste därför definiera $a^{1/q}$, där $a \geq 0$ och $q \in \mathbb{Z}^+$, som det (existerande och unika) icke-negativa reella tal som upphöjt till q blir a; vi säger att $a^{1/q}$ är den q:te roten ur a. Vidare inser vi att om $p, q \in \mathbb{Z}^+$ så ger de vanliga räknereglerna (som vi kräver skall gälla) att $\left(a^{1/q}\right)^p = a^{p/q}$ varför vi måste *definiera* a^r, där $r = p/q$ är ett rationellt tal och $p, q \in \mathbb{Z}^+$, som $\left(a^{1/q}\right)^p$. Om vi sedan låter $a^{-r} = 1/a^r$ för varje positivt rationellt tal r och $a > 0$, så har vi definierat a^r för (bland annat) varje $a > 0$ och $r \in \mathbb{Q}$ på ett sådant sätt att räknelagarna ovan fortfarande gäller.

Om $a \geq 0$ och $q \in \mathbb{Z}^+$ inför vi symbolen $\sqrt[q]{a}$ för den q:te roten ur a. Speciellt låter vi $\sqrt{a} := \sqrt[2]{a}$ vara *kvadratroten* ur a. Vi har $\sqrt{0} = 0$, $\sqrt{4} = 2$ och $\sqrt{100} = 10$. Man kan bevisa att $\sqrt{2}$ är ett irrationellt tal som avrundat till 10 decimaler är 1.4142135624.

Övning

1. Låt $a, b \in \mathbb{R}^+$ och $q \in \mathbb{Z}^+$. Visa att $(ab)^{1/q} = a^{1/q} b^{1/q}$. Speciellt är $\sqrt{ab} = \sqrt{a}\sqrt{b}$.

[6] Notera i synnerhet att vi *inte* definierar 0^0. Anledningen är att det inte finns något naturligt sätt att göra det på.

1.2.1.4 Kvadratkomplettering

Vi avslutar algebraavsnittet med en genomgångav av den mycket viktiga tekniken *kvadratkomplettering*. Vi kan som bekant *utveckla* en kvadrat:

$$(x+2)^2+5=x^2+4x+4+5=x^2+4x+9.$$

Notera att högerledet innehåller både x^2-term och x-term, medan variabeln x bara bor innanför en kvadrat i vänsterledet. Detta gör att vänsterledet är lättare att analysera än högerledet; till exempel ser vi direkt att uttrycket $(x+2)^2+5$ alltid är större än eller lika med 5, och att detta minsta värde erhålles precis då $x=-2$; anledningen är ju att $(x+2)^2 \geq 0$ med likhet bara då $x=-2$. Däremot kan vi inte dra denna slutsats bara genom att titta på uttrycket x^2+4x+9.

Därför är det i praktiken ofta önskvärt att ”gå åt andra hållet”, d.v.s. att skriva om ett uttryck som högerledet på formen i vänsterledet. Denna omskrivning, som alltså går ut på att skriva ett uttryck ax^2+bx+c på formen $A(x+B)^2+C$, går under namnet *kvadratkomplettering*.

Hur utför man då kvadratkomplettering? Säg att vi vill kvadratkomplettera x^2+4x+9. Vi vill med andra ord skriva det på formen $(x+c)^2+d$ för några tal c och d. Det vi gör är att vi först resonerar oss fram till att $c=2$ (d.v.s. *halva* 4:an i x^2+4x+9), ty då blir ju utvecklingen av $(x+c)^2=(x+2)^2=x^2+4x+4$. De två första termerna vill vi ha, men inte den sista fyran, så vi tar bort den: $(x+2)^2-4=x^2+4x$. Det följer då att $(x+2)^2-4+9=x^2+4x+9$, och vi är alltså klara: $x^2+4x+9=(x+2)^2+5$.

Exempel

Kvadratkomplettera x^2+3x+1.

Lösning: $x^2+3x+1=\left(x+\frac{3}{2}\right)^2-\left(\frac{3}{2}\right)^2+1=\left(x+\frac{3}{2}\right)^2-\frac{9}{4}+\frac{4}{4}=\left(x+\frac{3}{2}\right)^2-\frac{5}{4}.$

Exempel

Kvadratkomplettera x^2-x-2.

Lösning: $x^2-x-2=\left(x-\frac{1}{2}\right)^2-\left(\frac{1}{2}\right)^2-2=\left(x-\frac{1}{2}\right)^2-\frac{1}{4}-\frac{8}{4}=\left(x-\frac{1}{2}\right)^2-\frac{9}{4}.$

Exempel

Kvadratkomplettera $5x^2+6x-2$.

Lösning:

$$5x^2+6x-2=5\left(x^2+\frac{6}{5}-\frac{2}{5}\right)=5\left(\left(x+\frac{3}{5}\right)^2-\left(\frac{3}{5}\right)^2-\frac{2}{5}\right)=5\left(\left(x+\frac{3}{5}\right)^2-\frac{9}{25}-\frac{10}{25}\right)=$$
$$=5\left(\left(x+\frac{3}{5}\right)^2-\frac{9}{25}-\frac{10}{25}\right)=5\left(\left(x+\frac{3}{5}\right)^2-\frac{19}{25}\right)=5\left(x+\frac{3}{5}\right)^2-\frac{19}{5}.$$

1.2.2 Intervall

Inom matematiken förekommer väldigt ofta mängder av alla reella tal som ligger *mellan* två givna reella tal på tallinjen, eller som ligger till vänster eller höger om något givet reellt tal. Ibland inkluderas och ibland exkluderas själva gränspunkterna. Man har därför infört speciella symboler för sådana mängder, som kallas *intervall*:

- $[a, b] \coloneqq \{x \in \mathbb{R}: a \leq x \leq b\}$
- $]a, b] \coloneqq \{x \in \mathbb{R}: a < x \leq b\}$
- $[a, b[\coloneqq \{x \in \mathbb{R}: a \leq x < b\}$
- $]a, b[\coloneqq \{x \in \mathbb{R}: a < x < b\}$.

Man skriver också

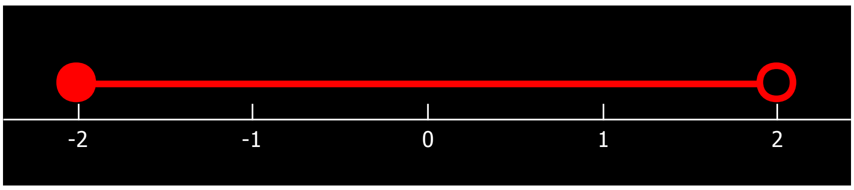
- $]-\infty, b] = \{x \in \mathbb{R}: x \leq b\}$
- $]-\infty, b[= \{x \in \mathbb{R}: x < b\}$
- $[a, \infty[= \{x \in \mathbb{R}: x \geq a\}$
- $]a, \infty[= \{x \in \mathbb{R}: x > a\}$

samt, ibland, $]-\infty, \infty[= \mathbb{R}$ som också räknas som ett ”intervall”. I de fall både a och b ingår i notationen så förutsätts att $b > a$. De fyra första typerna av intervall säges vara *begränsade* eftersom de har ändlig längd $b - a$. De fyra sista (liksom $\mathbb{R}$) kallas *obegränsade*.

Talen a och b kallas för intervallens *ändpunkter*; ett intervall kan sålunda ha noll, en eller två ändpunkter. I samtliga fall anger en inåtpekande parentes att den bredvidliggande ändpunkten (a eller b) *tillhör* intervallet, medan en utåtpekande parentes anger motsatsen. Om ett intervalls *samtliga* ändpunkter tillhör intervallet, säges intervallet vara *slutet*. Om ett intervalls samtliga ändpunkter ligger utanför intervallet, säges det vara *öppet*. Vi har därför

- Slutna intervall: $[a, b]$, $]-\infty, b]$, $[a, \infty[$, $\mathbb{R}$
- Öppna intervall: $]a, b[$, $]-\infty, b[$, $]a, \infty[$, $\mathbb{R}$
- Varken slutna eller öppna: $]a, b]$, $[a, b[$
- Slutet och öppet: $\mathbb{R}$.

Intervall kan illustreras grafiskt genom att man ritar en tallinje och markerar de punkter som tillhör intervallet. För att markera om en ändpunkt tillhör intervallet eller inte brukar man avsluta linjestycket som markerar intervallet med en fylld eller ofylld disk, där den är fylld omm ändpunkten tillhör intervallet.



Figur 1. Intervallet $[-2, 2[$.

Följande definition är viktig.

Definition. Om $x \in \mathbb{R}$ så säger man att en mängd $N \subset \mathbb{R}$ är en *omgivning* till x omm N innehåller något öppet intervall centrerat kring x, d.v.s. omm det finns ett tal $\epsilon > 0$ sådant att intervallet $]x-\epsilon, x+\epsilon[\subset N$.

Om N är en omgivning till x gäller speciellt att $x \in N$, men N innehåller också en hel sträcka av punkter "nära" x både till vänster och till höger. Till exempel är $[-0.1, 0.1]$ en omgivning av origo (talet 0), t.ex. eftersom det innehåller intervallet $]0-0.01, 0+0.01[$.

För framtida bruk noterar vi att symbolerna $\mathbb{Z}^+ \coloneqq \mathbb{Z} \cap]0,\infty[$ samt $\mathbb{R}^+ \coloneqq \mathbb{R} \cap]0,\infty[$ brukar användas för mängderna av de *positiva* heltalen respektive de *positiva* reella talen. Notera att $\mathbb{N} = \mathbb{Z}^+ \cup \{0\}$, $\mathbb{Z}^+ = \mathbb{N} \setminus \{0\}$.

Exempel

Vi ser att

$$\begin{aligned}
&[-10,10] \cup]0,20[= [-10,20[\\
&[-1,1] \setminus \{0\} = [-1,0[\cup]0,1] \\
&[-2,2] \setminus [-1,1[= [-2,-1[\cup [1,2] \\
&[-1,1]^C =]-\infty,-1[\cup]1,\infty[\\
&]-1,0] \cap [0,1[= \{0\} \\
&]-1,0[\cap [0,1[= \emptyset \\
&[0,5] \cap \mathbb{Z} = \{0,1,2,3,4,5\}.
\end{aligned}$$

Av dessa mängder är det bara den första som är ett intervall.

Definition. Ett slutet och begränsat intervall säges vara *kompakt*.

Exempel

Intervallet $[0,1]$ är kompakt, men inte $]0,1]$ (som inte är slutet) eller $[0,\infty[$ (som inte är begränsat).

1.2.3 Delmängder av $\mathbb{R}$ (*)

Även om de flesta delmängderna av $\mathbb{R}$ förstås inte är intervall, så tenderar de delmängder man stöter på i praktiken att vara just intervall eller unioner av intervall. Nedan är några exempel på delmängder av $\mathbb{R}$ vilka *inte* är intervall:

$$\{0\}, \qquad \{-\sqrt{2}, \sqrt{2}\}, \qquad \mathbb{R} \setminus \{0\}, \qquad [-1,1] \setminus \{0\}, \qquad [0,1] \cup [2,3], \qquad \mathbb{Z}, \qquad \mathbb{Q}.$$

Det finns även många mycket mer fantasifulla varianter, som t.ex. den så kallade *Cantormängden*, som vi inte kommer att gå in på här. Vi inför följande allmänna terminologi för godtyckliga delmängder av $\mathbb{R}$.

Definition. Låt $X \subset \mathbb{R}$ och $x \in \mathbb{R}$. Om det finns en omgivning till x vilken helt ligger i X så säges x vara en *inre punkt* till X; om det finns en omgivning till x som helt ligger utanför X säges x vara en *yttre punkt* till X. Om i stället varje omgivning till x innehåller punkter både från X och från X^C säges x vara en *randpunkt* till X.

Notera speciellt att om x är en inre punkt till X så har vi $x \in X$, och om x är en yttre punkt till X så gäller $x \notin X$. Notera också att varje $x \in \mathbb{R}$ antingen är en inre punkt, en yttre punkt eller en rand-punkt till X (d.v.s. *exakt ett* av alternativen gäller).

För de *intervall* vi tittat på gäller det uppenbarligen så att ändpunkterna a och b är randpunkter, och a och b är också de *enda* randpunkterna till ett intervall. Alla punkter *i* ett intervall, förutom eventu-ella ändpunkter, är inre punkter till intervallet. Alla punkter *utanför* ett intervall, förutom eventuella ändpunkter, är yttre punkter till intervallet.

Exempel

Betrakta intervallet $I ≔]0,1]$.

Punkten $x = 0$ är en randpunkt, för varje omgivning av 0 innehåller punkter både i och utanför I. Av samma anledning är $x = 1$ en randpunkt. Punkten $x = 0.7$ är däremot en inre punkt, för t.ex. omgiv-ningen $]0.6, 0.8[$ ligger helt i I. Punkten $x = 1.01$ är en yttre punkt; t.ex. ligger omgivningen $]1.009, 1.011[$ helt i I^C.

Exempel

Låt $X ≔ [-1,1] \cup]3,\infty[$. Då mängden av randpunkter till X lika med $\{-1,1,3\}$, mängden av inre punkter till X är $]-1,1[\cup]3,\infty[$ och mängden av yttre punkter till X är $]-\infty,-1[\cup]1,3[$.

Definition. En delmängd $X \subset \mathbb{R}$ säges vara *öppen* omm varje $x \in X$ i delmängden är en inre punkt. En delmängd X säges vara *sluten* omm dess komplement $X^C (= \mathbb{R} \setminus X)$ är öppet.

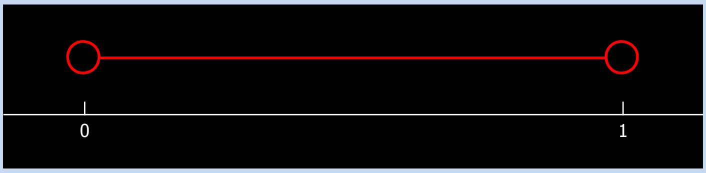
En delmängd är med andra ord öppen om och endast om varje punkt i delmängden har en omgivning som helt ligger i delmängden. Man inser direkt att ett öppet intervall är en öppen mängd och att ett slutet intervall är en sluten mängd, så terminologin är konsistent. $\emptyset$ och $\mathbb{R}$ är båda både öppna och slutna (varför?).

Vi inser direkt följande sats:

Sats. En delmängd $X \subset \mathbb{R}$ är öppen omm varje randpunkt tillhör X^C; X är sluten omm varje rand-punkt tillhör X.

Exempel

Intervallet $I_1 \coloneqq]0,1]$ från exemplet ovan är *inte* en öppen mängd eftersom $1 \in I_1$ inte är en inre punkt. Däremot är $I_2 \coloneqq]0,1[$ en öppen mängd.



Vilket $x \in I_2$ du än väljer, så finns det ju en omgivning till punkten helt innesluten i I_2. Väljer du t.ex. $x = 0.0001$ så fungerar omgivningen $]0.00009, 0.00011[\subset I_2$.

Exempel

Mängderna

$$\mathbb{R}, \quad \mathbb{R}^+ (=]0,\infty[), \quad \mathbb{R} \setminus \{0\}, \quad]-1,1[, \quad]-\infty,0[\cup]1,2[$$

är samtliga öppna. Mängderna

$$\mathbb{R}, \quad [0,\infty[, \quad \{1,2\}, \quad [-1,1], \quad]-\infty,0] \cup [1,2]$$

är samtliga slutna, eftersom deras komplement

$$\emptyset, \quad]-\infty,0[, \quad \mathbb{R} \setminus \{1,2\}, \quad]-\infty,-1[\cup]1,\infty[, \quad]0,1[\cup]2,\infty[$$

är öppna (alt. eftersom randpunkterna tillhör mängderna). Mängderna

$$[0,1[, \quad]-\infty,0[\cup \{5\}$$

är slutligen exempel på mängder som varken är öppna eller slutna.

Definition. En delmängd $X \subset \mathbb{R}$ säges vara *begränsad* omm den är helt innesluten i någon omgivning av origo, d.v.s. omm $X \subset [-R,R]$ för något $R > 0$.

En mängd är alltså begränsad om den får plats i något intervall av ändlig längd. Till exempel är begränsade intervall begränsade mängder, så terminologin är konsistent.

Definition. En sluten och begränsad delmängd av $\mathbb{R}$ säges vara *kompakt*.

Speciellt är kompakta intervall kompakta mängder, så terminologin är konsistent även här.

Övningar

1. Visa att unionen av en godtycklig samling öppna mängder är öppen.
2. Visa att snittet av en godtycklig samling slutna mängder är slutet.
   [Ledning: Använd föregående uppgift och en av De Morgans lagar.]

3. Visa att snittet av en *ändlig* samling öppna mängder är öppet.
4. Visa att unionen av en *ändlig* samling slutna mängder är sluten.
5. Visa att snittet av en godtycklig samling öppna mängder *inte* måste vara öppen. [Ledning: Betrakta $A_n := \left]-\frac{1}{n}, \frac{1}{n}\right[$.]
6. Visa att unionen av en godtycklig samling slutna mängder *inte* måste vara sluten. [Ledning: Betrakta $A_n := \left[\frac{1}{n}, 1\right]$.]

1.2.4 Följder

En *följd* är en uppräkning, eller numrerad lista, av något slags objekt, där ordningen spelar roll och där ett och samma objekt kan förekomma flera gånger i uppräkningen. En följd kan bestå av ändligt eller oändligt många element (men vi tillåter inte ”följer” som består av noll objekt). Vi är här mest intresserade av följder av *tal*. Ett exempel på en talföljd som består av sju element är

$$7, \quad 3, \quad -2, \quad 6, \quad 8, \quad -2, \quad 5.$$

Elementen i en följd brukar *indexeras* med hjälp av heltal. Det mest naturliga är att använda 0 eller 1 som etikett för det första elementet i följden, och sedan ge de kommande elementen successivt högre etiketter. I vårt exempel kan vi sätta $a_1 := 7$, $a_2 := 3$, …, $a_7 := 5$. Vi kan då använda $(a_k)_{k=1}^{7}$ som en rent symbolisk representation av följden. Rent syntaktiskt kan symbolen också tänkas stå för *mängden* av talen i följden. Mer allmänt är $(a_k)_{k=m}^{n}$, där m och n är heltal med $n \geq m$, en symbol för den ändliga följden

$$a_m, \quad a_{m+1}, \quad a_{m+2}, \quad \dots, \quad a_n$$

bestående av $n - m + 1$ element. $(a_k)_{k=m}^{\infty}$ är en symbol för den oändliga följden

$$a_m, \quad a_{m+1}, \quad a_{m+2}, \quad \dots$$

och $(a_k)_{k=-\infty}^{\infty}$ är en symbol för den (”dubbelt”) oändliga följden

$$\dots, \quad a_{-2}, \quad a_{-1}, \quad a_0, \quad a_1, \quad a_2, \quad \dots .$$

Ett konkret exempel på en oändlig följd är

$$0, \quad 1, \quad 4, \quad 9, \quad 16, \quad 25, \quad 36, \quad \dots$$

som är följden av alla heltalskvadrater. Denna följd kan skrivas $(a_k)_{k=0}^{\infty}$ där $a_k = k^2$, d.v.s. vi har en explicit *formel* för det k:te elementet i följden. Notera att det här är väldigt praktiskt att börja indexeringen med just 0 (och inte t.ex. 1).

Som ytterligare exempel ger vi följderna $(a_k)_{k=0}^{\infty}$ och $(b_k)_{k=0}^{\infty}$ där $a_k = 1 + 2k$ och $b_k = 2^k$. Dessa inleds

$$1, \quad 3, \quad 5, \quad 7, \quad 9, \quad 11, \quad \dots$$

respektive

$$1, \quad 2, \quad 4, \quad 8, \quad 16, \quad 32, \quad \dots .$$

$(a_k)_{k=0}^{\infty}$ har tydligen den speciella egenskapen att *differensen* mellan ett tal och föregående alltid är densamma (nämligen 2), medan $(b_k)_{k=0}^{\infty}$ har den speciella egenskapen att *kvoten* mellan ett tal och föregående alltid är densamma (nämligen 2). Det finns speciella namn för dessa egenskaper.

Definition. En följd $(a_k)_{k=m}^{n}$ av reella tal säges vara *aritmetisk* omm $d \coloneqq a_{k+1} - a_k$ inte beror på k, d.v.s. om differensen d mellan två på varandra följande tal alltid är densamma. Talet d kallas i sådana fall för den aritmetiska följdens *differens*.

En följd $(a_k)_{k=m}^{n}$ av reella tal säges vara *geometrisk* omm $q \coloneqq a_{k+1}/a_k$ inte beror på k, d.v.s. om kvoten q mellan två på varandra följande tal alltid är densamma. Talet q kallas i sådana fall för den geometriska följdens *kvot*.

I definitionen ovan tillåts att m ersätts med $-\infty$ och att n ersätts med ∞ (i stället för heltal), så även oändliga följder kan vara aritmetiska och geometriska.

1.2.5 Summor

Om $(a_k)_{k=m}^{n}$ är en ändlig följd [vilket medför att m och n är heltal med $n \geq m$] av så inför vi symbolen

$$\sum_{k=m}^{n} a_k \coloneqq a_m + a_{m+1} + \cdots + a_n$$

för summan av alla tal i följden; detta är alltså en summa med $n - m + 1$ stycken termer. Mer allmänt, om a_k är element i *någon* följd av tal och om m och n är heltal inför vi symbolen

$$\sum_{k=m}^{n} a_k$$

för summan av alla tal a_k där $k \in [m, n] \cap \mathbb{Z}$. Om t.ex. $(a_k)_{k=0}^{\infty}$ är en följd av oändligt många tal så är

$$\sum_{k=5}^{9} a_k = a_5 + a_6 + a_7 + a_8 + a_9$$

en summa av fem av talen i följden. Om $n \geq m$ är $\sum_{k=m}^{n} a_k$ alltså en summa med $|[m, n] \cap \mathbb{Z}| = n - m + 1$ stycken termer. Om $n < m$ är $[m, n] \cap \mathbb{Z} = \emptyset$ och summan består därför av $|\emptyset| = 0$ termer; denna s.k. *tomma summa* är definitionsmässigt lika med 0 (det neutralal elementet med avseende på addition).

Exempel

Om $a_k = 2k + 1$ för varje $k \in \mathbb{Z}$ så är

$$\sum_{k=0}^{5} a_k = a_0 + a_1 + a_2 + a_3 + a_4 + a_5 = 1 + 3 + 5 + 7 + 9 + 11 = 36.$$

Följden $(a_k)_{k=0}^{5}$ i exemplet är aritmetisk. Summor av ändliga aritmetiska följder, så kallade *aritmetiska summor*, kan beräknas med en speciell formel, som vi bevisade i avsnitt 1.1.1.2.

Sats. Om $(a_k)_{k=m}^{n}$ är en aritmetisk följd med differens d så är

$$\sum_{k=m}^{n} a_k = \frac{a_m + a_n}{2} \cdot (n - m + 1).$$

Vi kan alltså beräkna vår senaste exempelsumma på ett alternativt sätt:

$$\sum_{k=0}^{5} a_k = \frac{a_0 + a_5}{2} \cdot (5 - 0 + 1) = \frac{1 + 11}{2} \cdot 6 = 36.$$

Satsen är lätt att komma ihåg, för $(a_m + a_n)/2$ är ju *medelvärdet* av första och sista termen medan $n - m + 1$ helt enkelt är *antalet* termer. Summan av en aritmetisk följd är därför lika med medelvärdet av första och sista termen, multiplicerat med antalet termer.

Exempel

Beräkna summan $\sum_{k=0}^{1000} 5k$.

Lösning: Summan är aritmetisk med differensen 5, första termen 0, sista termen 5000 och antalet termer 1001. Sålunda är summans värde

$$\sum_{k=0}^{1000} 5k = \frac{0 + 5000}{2} \cdot 1001 = 2500 \cdot 1001 = 2502500.$$

Även ändliga *geometriska* följder kan summeras med en speciell formel.

Sats. Om $(a_k)_{k=m}^{n}$ är en geometrisk följd med kvot $k \neq 1$ så är

$$\sum_{k=m}^{n} a_k = a_m \frac{1 - k^{n-m+1}}{1 - k}.$$

Bevis. Beteckna summans värde med s. Vi har då

$$s = a_m + a_m k + a_m k^2 + \cdots + a_m k^{n-m}$$

varför

$$sk = a_m k + a_m k^2 + a_m k^3 + \cdots + a_m k^{n-m+1}$$

och

$$s(1 - k) = s - sk = a_m - a_m k^{n-m+1} = a_m (1 - k^{n-m+1})$$

varifrån den utlovade formeln direkt följer. ■

Exempel

Beräkna summan $\sum_{k=0}^{99} 2^k$.

Lösning: Summan är geometrisk med kvoten 2, första termen 1 och antalet termer 100. Därför är

$$\sum_{k=0}^{99} 2^k = \frac{1-2^{100}}{1-2} = \frac{1-2^{100}}{-1} = 2^{100} - 1 (\approx 1.27 \cdot 10^{30}).$$

De allra flesta följder av tal är förstås varken geometriska eller aritmetiska. Ett exempel på en följd som är varken eller är $(a_k)_{k=1}^{\infty}$ där a_k är det k:te primtalet. I allmänhet finns ingen formel som ger summan av en följd på sluten form. Nedanstående exempel handlar om summor som trots att de varken är aritmetiska eller geometriska kan beräknas medelst enkla omskrivningar eller observationer. Men först några grundläggande observationer om summor.

1.2.5.1 Räkneregler för summor

Först har vi uppdelning:

$$\sum_{k=m}^{n} a_k = \sum_{k=m}^{\ell} a_k + \sum_{k=\ell+1}^{n} a_k$$

om ℓ är ett tal mellan m och n. Till exempel är ju

$$\sum_{k=1}^{10} a_k = \sum_{k=1}^{3} a_k + \sum_{k=4}^{10} a_k$$

eftersom

$$a_1 + a_2 + a_3 + a_4 + a_5 + a_6 + a_7 + a_8 + a_9 + a_{10} = \\ = a_1 + a_2 + a_3 + a_4 + a_5 + a_6 + a_7 + a_8 + a_9 + a_{10}.$$

Sedan har vi linearitet:

$$\sum_{k=m}^{n} (a_k + b_k) = \sum_{k=m}^{n} a_k + \sum_{k=m}^{n} b_k$$
$$\sum_{k=m}^{n} c a_k = c \sum_{k=m}^{n} a_k.$$

Som exempel ger vi

$$\sum_{k=1}^{3} (a_k + b_k) = \sum_{k=1}^{3} a_k + \sum_{k=1}^{3} b_k$$

eftersom

$$a_1 + b_1 + a_2 + b_2 + a_3 + b_3 = a_1 + a_2 + a_3 + b_1 + b_2 + b_3$$

samt

$$\sum_{k=1}^{3} 7a_k = 7 \sum_{k=1}^{3} a_k$$

eftersom

$$7a_1 + 7a_2 + 7a_3 = 7(a_1 + a_2 + a_3).$$

1.2.5.2 Exempel

Exempel

Beräkna summan $\sum_{k=3}^{29} \frac{2^k(1+3k)+1}{2^k}$.

Lösning: Eftersom

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

har vi

$$\sum_{k=3}^{29} \frac{2^k(1+3k)+1}{2^k} = \sum_{k=3}^{29} 1 + \sum_{k=3}^{29} 3k + \sum_{k=3}^{29} \left(\frac{1}{2}\right)^k$$

där första summan trivialt ger

$$\sum_{k=3}^{29} 1 = \underbrace{1 + 1 + \cdots + 1}_{\text{27 termer}} = 27 \cdot 1 = 27,$$

andra summan är aritmetisk med

$$\sum_{k=3}^{29} 3k = \frac{9 + 87}{2} \cdot 27 = 1296$$

och tredje summan är geometrisk med

$$\sum_{k=3}^{29} \left(\frac{1}{2}\right)^k = \frac{1}{8} \cdot \frac{1 - \left(\frac{1}{2}\right)^{27}}{1 - \frac{1}{2}} = \frac{1}{4}\left(1 - \frac{1}{2^{27}}\right) = \frac{1}{4} - \frac{1}{2^{29}}.$$

Därför är

$$\sum_{k=3}^{29} \frac{2^k(1+3k)+1}{2^k} = 27 + 1296 + \frac{1}{4} - \frac{1}{2^{29}} = \frac{5293}{4} + \frac{1}{2^{29}}.$$

Ibland kan det hända att nästan alla termer i en summa tar ut varandra. Betrakta t.ex. följden $(a_k)_{k=1}^{5}$ definierad av

$$a_k = \frac{1}{k} - \frac{1}{k+1}.$$

Summan är

$$\sum_{k=1}^{5} a_k = \left(\frac{1}{1} - \frac{1}{2}\right) + \left(\frac{1}{2} - \frac{1}{3}\right) + \left(\frac{1}{3} - \frac{1}{4}\right) + \left(\frac{1}{4} - \frac{1}{5}\right) + \left(\frac{1}{5} - \frac{1}{6}\right) = \frac{1}{1} - \frac{1}{6} = \frac{5}{6}.$$

En summa som uppvisar ett sådant fenomen kallas för *teleskoperande*.

Exempel

Beräkna summan $\sum_{k=1}^{100} \frac{1}{k^2+k}$.

Eftersom

$$\frac{1}{k^2+k} = \frac{1}{k} - \frac{1}{k+1}$$

är summan teleskoperande och

$$\sum_{k=1}^{100} \frac{1}{k^2+k} = \sum_{k=1}^{100} \left(\frac{1}{k} - \frac{1}{k+1}\right) = \frac{1}{1} - \frac{1}{101} = \frac{100}{101}$$

varmed uppgiften är löst.

1.2.5.3 Fler exempel

Vårt första exempel är ganska fascinerande. Notera att

$$\begin{aligned}
&1 = 1 = 1^2 \\
&1 + 3 = 4 = 2^2 \\
&1 + 3 + 5 = 9 = 3^2 \\
&1 + 3 + 5 + 7 = 16 = 4^2 \\
&1 + 3 + 5 + 7 + 9 = 25 = 5^2 \\
&1 + 3 + 5 + 7 + 9 + 11 = 36 = 6^2 \\
&\vdots
\end{aligned}$$

Är det här verkligen en slump?

Exempel

Visa att summan av de k första udda positiva heltalen alltid är en jämn kvadrat, nämligen k^2.

Lösning: Om vi betecknar nämnd summa med s_k har vi

$$s_k = \sum_{i=1}^{k} (2i - 1) = \frac{1 + 2k - 1}{2} \cdot k = k^2$$

med formeln för en aritmetisk summa.

Exempel

Beräkna

$$\sum_{n=0}^{100} \sum_{k=0}^{n} 2^{-k}.$$

Lösning: Först har vi

$$\sum_{k=0}^{n} 2^{-k} = \sum_{k=0}^{n} \frac{1}{2^k} = \sum_{k=0}^{n} \left(\frac{1}{2}\right)^k = \frac{1-\left(\frac{1}{2}\right)^{n+1}}{1-\frac{1}{2}} = 2\left(1-\frac{1}{2^{n+1}}\right) = 2-\frac{1}{2^n}$$

enligt formeln för en geometrisk summa. Därför är

$$\sum_{n=0}^{100}\sum_{k=0}^{n} 2^{-k} = \sum_{n=0}^{100}\left(2-\frac{1}{2^n}\right) = \sum_{n=0}^{100} 2 - \sum_{n=0}^{100}\left(\frac{1}{2}\right)^n$$

där $\sum_{n=0}^{100} 2 = 202$ och

$$\sum_{n=0}^{100}\left(\frac{1}{2}\right)^n = \frac{1-\left(\frac{1}{2}\right)^{101}}{1-\frac{1}{2}} = 2\left(1-\frac{1}{2^{101}}\right) = 2-\frac{1}{2^{100}}.$$

Alltså är

$$\sum_{n=0}^{100}\sum_{k=0}^{n} 2^{-k} = 200 + \frac{1}{2^{100}}.$$

Exempel (*)

Notera att den summa vi just beräknat är summan av alla raderna i listningen

$$\begin{array}{l} 2^{-0} \\ 2^{-0}+2^{-1} \\ 2^{-0}+2^{-1}+2^{-2} \\ 2^{-0}+2^{-1}+2^{-2}+2^{-3} \\ \vdots \\ 2^{-0}+2^{-1}+2^{-2}+2^{-3}+\cdots+2^{-100} \end{array}$$

som består av precis 101 stycken rader. Vi kan förstås också summera dessa tal kolonnvis, varvid vi erhåller samma summa:

$$101\cdot 2^{-0} + 100\cdot 2^{-1} + 99\cdot 2^{-2} + \cdots + 1\cdot 2^{-100} = 101 + 50 + 24.75 + \cdots + 2^{-100}$$

som med summatecken kan skrivas

$$\sum_{k=0}^{100}(101-k)2^{-k}.$$

Det följer att

$$01\, följer\, att$$

Vi har alltså lyckats beräkna summan $\sum_{k=0}^{100} k2^{-k}$ som varken är aritmetisk, geometrisk eller teleskoperande!

1.2.6 Produkter (fakultet, binomialkoefficienter och binomialsatsen)

På samma sätt som summatecknet $\sum$ kan användas för att beteckna summan av en följd av tal, så betecknar produkttecknet $\prod$ produkten av en följd av tal. Till exempel, om $a_k = 2k$ så är

$$\prod_{k=1}^{3} a_k = a_1 a_2 a_3 = 2 \cdot 4 \cdot 6 = 48.$$

$\prod_{k=m}^{n} a_k$ definieras i allmänhet som produkten av alla tal a_k där $k \in [m, n] \cap \mathbb{Z}$. Det rör sig tydligen om en produkt av $n - m + 1$ faktorer, om bara $n \geq m$. Om $n < m$ är $[m, n] \cap \mathbb{Z} = \emptyset$ och då definierar vi $\prod_{k=m}^{n} a_k = 1$; den s.k. *tomma produkten* har värdet 1, som ju är det neutrala elementet med avseende på multiplikation. Vi har då likheten

$$a^n = \prod_{k=1}^{n} a$$

som gäller för varje $a \in \mathbb{R}$ och $n \in \mathbb{Z}^+$; formen gäller även i fallet $n = 0$, om $a \neq 0$. Vi inför nu symbolen *fakultet* (!) som definieras av

$$n! \coloneqq \prod_{k=1}^{n} k\,, \qquad \forall n \in \mathbb{N}$$

$n!$ är sålunda produkten av talen $1, 2, 3, \dots, n-1, n$. Till exempel har vi $5! = 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 120$. Notera att $0! = 1$. Det finns en enkel kombinatorisk tolkning av fakultetsoperationen. Om vi har k olika symboler (av något slag) finns det nämligen $k!$ olika sätt att räkna upp dem.

Exempel

Sex personer står utanför en buss. På hur många olika sätt kan dessa sex personer bilda en kö?

Lösning: Det finns 6 olika möjligheter att välja första personen i kön. Sedan finns 5 olika kandidater till andraplatsen. Därför finns det $6 \cdot 5$ olika sätt att välja de två första platserna. (”Det finns sex sätt att välja förstaplatsen, *och för vart och ett* av dessa finns fem sätt att välja andraplatsen.”) Efter det återstår 4 kandidater till tredjeplatsen, och så vidare. Totalt finns det alltså $6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 6!$ olika sätt att bilda en kö med sex personer.

Svar: Det finns $6! = 720$ olika sätt att bilda en kö med sex personer.

Om $n \geq m$ har vi

$$\frac{n!}{m!} = n(n-1)(n-2) \cdot \dots \cdot (m+1);$$

till exempel är

$$\frac{27!}{24!} = \frac{27 \cdot 26 \cdot 25 \cdot 24 \cdot 23 \cdot \dots \cdot 2 \cdot 1}{24 \cdot 23 \cdot \dots \cdot 2 \cdot 1} = 27 \cdot 26 \cdot 25.$$

Definition. Om $n, k \in \mathbb{N}$ och $n \geq k$ så är *binomialkoefficienten*

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

Ett förekommande missförstånd

Lägg märke till att parenteserna i $\binom{n}{k}$ är en del av notationen. Om du skriver $\begin{matrix} n \\ k \end{matrix}$ kommer ingen att förstå vad du menar! Binomialkoefficienten $\binom{n}{k}$ skall förstås inte heller blandas ihop med kvoten $\frac{n}{k}$.

Även binomialkoefficienten har en kombinatorisk tolkning: $\binom{n}{k}$ är antalet delmängder med k element som existerar till en grundmängd på n element.

Exempel

En ingenjörsklass består av 30 elever. På hur många olika sätt kan man erhålla ett fotbollslag bestående av 11 spelare ur klassen?

Lösning: Om man tar hänsyn till spelarnas *ordning* (t.ex. tröjnumren 1, 2, …, 11) finns det förstås 30 sätt att välja spelare nummer 1, 29 sätt att välja spelare nummer 2, osv., så att det finns $30 \cdot 29 \cdot \cdots \cdot 21 \cdot 20 = \frac{30!}{19!} = \frac{30!}{(30-11)!} = 2\,180\,547\,008\,640\,000$ olika sätt att bilda ett numrerat lag.

Många av dessa lag består alltså av samma mängd spelare, fast med olika tröjnummer. Mer precist finns det förstås $11!$ sätt att tilldela en mängd bestående av 11 spelare olika tröjor, så varje möjlig mängd av spelare förekommer exakt $11!$ gånger bland alla $\frac{30!}{(30-11)!}$ kombinationer av numrerade lag.

Om vi med N betecknar antalet möjliga lag utan numrering, d.v.s. alla möjliga mängder av 11 spelare, har vi därför

$$11!\,N = \frac{30!}{(30-11)!}$$

varför

$$N = \frac{30!}{11!\,(30-11)!} = \binom{30}{11} = 54627300.$$

Svar: Det finns $\frac{30!}{(30-11)!} = 2\,180\,547\,008\,640\,000$ olika sätt att bilda ett numrerat lag och $\binom{30}{11} = 54\,627\,300$ olika sätt att bilda ett onumrerat lag.

Man kan speciellt lägga märke till att binomialkoefficienten $\binom{n}{k}$ alltid är ett positivt heltal. Följande sats är viktig därför att den ligger till grund för en otroligt smidig algoritm för att med papper och penna blixtsnabbt beräkna binomialkoefficienter för måttligt stora n.

Sats. Om $n, k \in \mathbb{Z}^+$ med $n \geq k + 1$ så gäller

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

Bevis. Vi har

$$\binom{n-1}{k-1} + \binom{n-1}{k} = \frac{(n-1)!}{(k-1)!\,(n-k)!} + \frac{(n-1)!}{k!\,(n-k-1)!} = \frac{n!\,k}{nk!\,(n-k)!} + \frac{n!\,(n-k)}{nk!\,(n-k)!} = \\ = \frac{n!\,k + n!\,(n-k)}{nk!\,(n-k)!} = \frac{n!\,k + n \cdot n! - n!\,k}{nk!\,(n-k)!} = \frac{n!}{k!\,(n-k)!} = \binom{n}{k}$$

som utlovat. ∎

Den algoritm som nämndes före satsen handlar om att konstruera en horisontellt centrerad triangulär tabell med binomialkoefficienter, den så kallade *Pascals triangel*. Om vi kallar första raden för ”rad 0” så gäller att på rad $n \in \mathbb{N}$ i tabellen finns de $n + 1$ stycken binomialkoefficienterna $\binom{n}{k}$, $k = 0, 1, \ldots, n$, uppradade:

$$\begin{matrix} & & & & \binom{0}{0} & & & & \\ & & & \binom{1}{0} & & \binom{1}{1} & & & \\ & & \binom{2}{0} & & \binom{2}{1} & & \binom{2}{2} & & \\ & \binom{3}{0} & & \binom{3}{1} & & \binom{3}{2} & & \binom{3}{3} & \\ \binom{4}{0} & & \binom{4}{1} & & \binom{4}{2} & & \binom{4}{3} & & \binom{4}{4} \\ & & & & \vdots & & & & \end{matrix}$$

Föregående sats säger att en binomialkoefficient i triangeln (förutom de längs kanterna) är lika med summan av de två ovanstående binomialkoefficienterna. Eftersom de två yttersta binomialkoefficienterna på varje rad trivialt har värdet 1 kan vi därmed mycket snabbt fylla i alla binomialkoefficienternas värden:

$$\begin{matrix} & & & & & 1 & & & & & \\ & & & & 1 & & 1 & & & & \\ & & & 1 & & 2 & & 1 & & & \\ & & 1 & & 3 & & 3 & & 1 & & \\ & 1 & & 4 & & 6 & & 4 & & 1 & \\ 1 & & 5 & & 10 & & 10 & & 5 & & 1 \\ & & & & & \vdots & & & & & \end{matrix}$$

Sitt namn har binomialkoefficienterna fått från en speciell tillämpning: *binomialutveckling*. Ett *binom* är per definition ett polynom som består av precis två termer, men binomialutveckling handlar mer generellt om att utveckla en heltalspotens av en summa bestående av två termer, d.v.s. utveckling av uttrycket

$$(a + b)^n$$

där $n \in \mathbb{N}$. Läsaren är redan bekant med specialfallet $n = 2$:

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

kanske är även fallet $n = 3$ bekant:

$$(a+b)^3 = a^3 + 3a^2b + 3ab^2 + b^3.$$

Lägg märke till att exponenten på a:et börjar på 3 och minskar med 1 för varje ny term, medan exponenten för b:et är 3 i sista termen, och minskar med ett för varje ny term åt vänster. Den observante läsaren har dessutom noterat att koefficienterna i dessa två utvecklingar ($n = 2$ respektive $n = 3$) är precis talen på rad 2 respektive 3 i Pascals triangel, och det är inte en slump.

Sats ***(binomialsatsen).*** För varje $a, b \in \mathbb{R}$ och varje $n \in \mathbb{Z}^+$ gäller

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^{n-k} b^k.$$

Bevis. Potensen är

$$(a+b)^n = \underbrace{(a+b)(a+b)\cdots(a+b)}_{n \text{ faktorer}}$$

och vid utveckling av denna erhålles en summa bestående av 2^n termer (före förenkling). Varje term är en produkt av n faktorer, nämligen termerna i potensen, en från varje parentes. De 2^n termerna i utvecklingen är precis alla kombinationer man kan erhålla. Det är sålunda klart att

$$(a+b)^n = \sum_{k=0}^{n} C_{nk} a^{n-k} b^k$$

för någon uppsättning heltal C_{nk}. Talet C_{nk} är antalet gånger som produkten $a^{n-k}b^k$ erhålles i utvecklingen, d.v.s. antalet gånger som vi väljer precis k stycken b:n. Men detta antal är ju så många sätt på vilka vi kan välja ut k stycken parenteser i potensen [de parenteser från vilka vi väljer b:et], utan hänsyn till ordningen på dem. Och detta antal är ju precis binomialkoefficienten, så

$$C_{nk} = \binom{n}{k}.$$

■

Exempel

Utveckla $(a+b)^5$.

Lösning: Med binomialsatsen och Pascals triangel får vi

$$(a+b)^5 = a^5 + 5a^4b + 10a^3b^2 + 10a^2b^3 + 5ab^4 + b^5.$$

Exempel

Utveckla $(x^2 - y)^4$.

Lösning: Med binomialsatsen och Pascals triangel får vi

$$(x^2 - y)^4 = \big(x^2 + (-y)\big)^4 = (x^2)^4 + 4(x^2)^3(-y) + 6(x^2)^2(-y)^2 + 4x^2(-y)^3 + (-y)^4 =$$
$$= x^8 - 4x^6y + 6x^4y^2 - 4x^2y^3 + y^4.$$

1.2.7 Medelvärden

Betrakta följande följd av värden:

$$17, \qquad 22, \qquad 19, \qquad 21, \qquad 40, \qquad 25, \qquad 20.$$

Det skulle kunna röra sig om temperaturer (i °C) eller månadslönser (i kkr). Vad är *medelvärdet* av dessa tal? Intuitivt är medelvärdet av en samling tal ett enskilt tal, som ligger mellan minsta och största talet, och som på något sätt är ”representativt” för hela samlingen. I vårt exempel bör medelvärdet t.ex. vara närmare det minsta värdet (17) än det största värdet (40), eftersom de flesta tal ligger i intervallets ”nedre del”. Inom matematiken finns det två synnerligen enkla precisa medelvärdesbegrepp, vilka vi nu introducerar.

Definition. Låt $(a_k)_{k=1}^n$ vara en följd av tal. Det *aritmetiska medelvärdet* av talen är talet

$$\frac{1}{n}\sum_{k=1}^{n} a_k$$

och, om alla tal är positiva, är det *geometriska medelvärdet* talet

$$\sqrt[n]{\prod_{k=1}^{n} a_k}.$$

Speciellt har vår exempelföljd det aritmetiska medelvärdet 23.4 och det geometriska medelvärdet 22.6 (avrundat till en decimal). Det är inte en slump att det aritmetiska medelvärdet är större än det geometriska; detta gäller alltid. Vi visar detta resultat nedan i specialfallet där följden består av precis *två* tal.

Sats. Låt $a, b \in \mathbb{R}^+$ och låt $A \coloneqq \frac{1}{2}(a + b)$ vara det aritmetiska och $G \coloneqq \sqrt{ab}$ det geometriska medelvärdet. Då är $A \geq G$ med likhet omm $a = b$.

Bevis. Vi har

$$4(A^2 - G^2) = (a + b)^2 - 4ab = (a^2 + 2ab + b^2) - 4ab = a^2 - 2ab + b^2 = (a - b)^2 \geq 0$$

vilket visar att $A^2 - G^2 \geq 0$, så $A \geq G$. Vidare gäller likhet precis då $a = b$. ■

1.2.8 Geometriska mängder

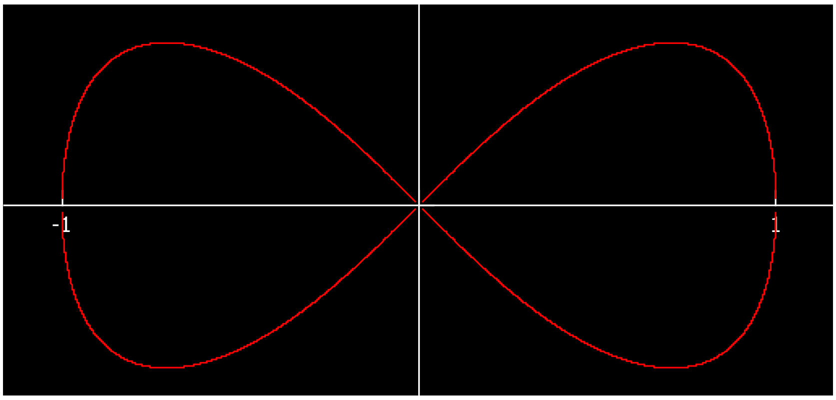
Inte sällan pratar man om t.ex. ekvationer som om de vore geometriska figurer. Man kan t.ex. höra talas om ”kurvan”

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

Vad menas med det? ”$x^4 + y^4 = x^2 - y^2$” är ju en ekvation, ett påstående vars sanningsvärde beror på vilka tal x och y man stoppar in i den. Det man egentligen menar är att kurvan är *mängden* av alla punkter (x, y) i planet $\mathbb{R}^2$ vilka uppfyller ekvationen, d.v.s. kurvan är

$$\{(x, y) \in \mathbb{R}^2: \quad x^4 + y^4 = x^2 - y^2\}.$$

Och en delmängd av planet, ja, det är ju precis vad en geometrisk figur är för något! Om vi kallar kurvan i exemplet för Γ så inser vi t.ex. att $(0, 0) \in \Gamma$, $(1, 0) \in \Gamma$ och $(-1, 0) \in \Gamma$. Däremot är t.ex. $(1, 1) \notin \Gamma$. Med ett datorprogram kan vi (approximativt) plotta kurvan; t.ex. kan datorprogrammet dela in området $[-1, 1]^2 \subset \mathbb{R}^2$ i $10\,000 \times 10\,000$ småbitar och låta pixeln i rutan (x, y) vara färgad omm ekvationen är uppfylld där (så när som på en liten feltolerans). Vi erhåller då följande bild:



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

1.2.8.1 Enkla kurvor

Man måste framför allt vara väl förtrogen med ekvationer för linjer och cirklar (och ellipser).

1.2.8.1.1 Räta linjer

Om $c \in \mathbb{R}$ är något fixt tal är det klart att $x = c$ är en vertikal linje medan $y = c$ är en horisontell linje. Från tidigare matematikstudier är läsaren också väl bekant med

$$y = kx + m$$

som är en (i allmänhet sned) rät linje med *lutningen* (*riktningskoefficienten*) k, vilket innebär att kurvan går k steg i y-led för varje steg i x-led; m är y-värdet där linjen skär y-axeln (där ju $x = 0$). Alla räta linjer, förutom de vertikala, kan skrivas på formen $y = kx + m$.

Den mest allmänna ekvationen för en rät linje är

$$ax + by = d$$

där a, b och d är konstanter. Varje sådan kurva är en rät linje, och varje rät linje kan skrivas på den formen för lämpliga val av konstanterna. Ekvationen kallas för linjens *normalekvation*.

Slutligen kan det vara praktiskt att känna till den s.k. *enpunktsformen* för en rät linje, d.v.s. formen

$$y - y_0 = k(x - x_0)$$

där k, x_0 och y_0 är konstanter. Ekvationen beskriver återigen en linje, och varje icke-vertikal linje kan skrivas på den här formen. Man ser direkt att (x_0, y_0) uppfyller ekvationen, så punkten (x_0, y_0) tillhör linjen. Vi känner också igen k som linjens riktningskoefficient:

$$y - y_0 = k(x - x_0) \qquad \Longleftrightarrow \qquad y = kx + \underbrace{y_0 - kx_0}_{\text{konst.}}.$$

Enpunktsformen tillåter alltså att vi direkt skriver ner ekvationen för en (icke-vertikal) linje bara vi känner till *en* punkt på linjen, samt linjens lutning.

1.2.8.1.2 Cirklar (och ellipser)

Avståndet mellan origo och punkten $(x, y) \in \mathbb{R}^2$ är $\sqrt{x^2 + y^2}$ enligt Pythagoras sats. Därför är ekvationen $\sqrt{x^2 + y^2} = 1$ uppfylld precis då (x, y) har avståndet 1 till origo. Men den här ekvationen är ekvivalent med $x^2 + y^2 = 1$, så därför är

$$S^1 \coloneqq \{(x, y) \in \mathbb{R}^2: \quad x^2 + y^2 = 1\}$$

cirkeln med radie 1 kring origo – den består ju av alla punkter som har precis avståndet 1 till origo. Den här cirkeln kallas för *enhetscirkeln*, just eftersom den har radie 1. Beteckningen S^1 är standard för enhetscirkeln.

Mer allmänt inser vi att för varje $r > 0$ är

$$x^2 + y^2 = r^2$$

en cirkel kring origo med radie r (ekvationen är ekvivalent med $\sqrt{x^2 + y^2} = r$). Avståndet mellan punkten (x_0, y_0) och (x, y) är $\sqrt{(x - x_0)^2 + (y - y_0)^2}$, så ännu mer allmänt ser vi att

$$(x - x_0)^2 + (y - y_0)^2 = r^2$$

är en cirkel kring punkten (x_0, y_0) med radie r.

Ekvationen (där $a, b > 0$)

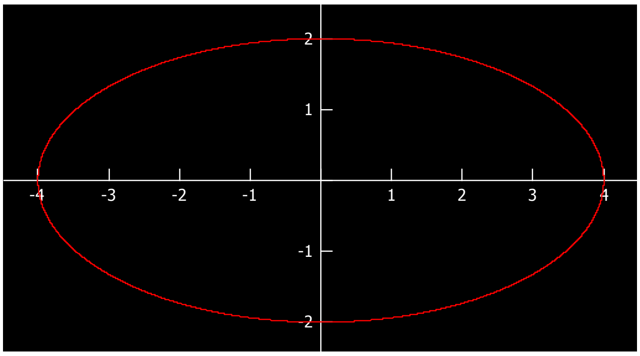
$$\left(\frac{x}{a}\right)^2 + \left(\frac{y}{b}\right)^2 = 1$$

beskriver en kurva som *inte* är en cirkel, om inte $a = b$ är lika, i vilket fall vi återfår en cirkel med radien a. I stället har vi en "tillplattad" cirkel, en *ellips* med *halvaxlarna* a och b. Om vi sätter $y = 0$ ser vi att ekvationen ger $x^2 = a^2$, d.v.s. $x = \pm a$; om vi sätter $x = 0$ får vi $y = \pm b$. Alltså ligger punkterna $(\pm a, 0)$ och $(0, \pm b)$ på ellipsen, och det är vidare lätt att se att ellipsen är en delmängd av $[-a, a] \times [-b, b]$ (hur?).

Nedan visas ellipsen

$$\left(\frac{x}{4}\right)^2 + \left(\frac{y}{2}\right)^2 = 1$$

som sträcker sig mellan -4 och $+4$ i x-led och mellan -2 till $+2$ i y-led.



Figur 3. En ellips.

1.2.8.1.3 Diskar och fyllda rektanglar

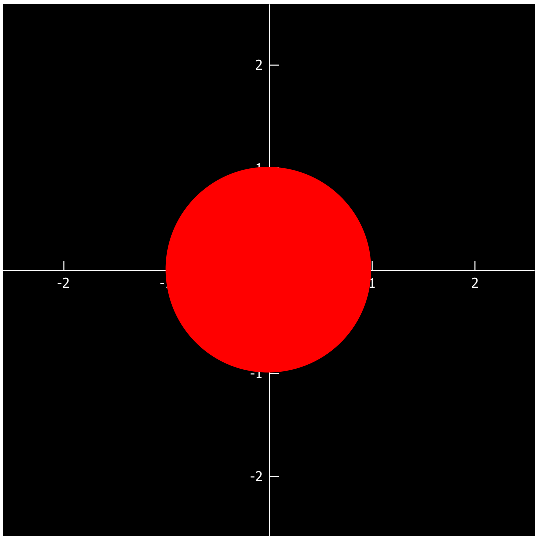
En geometrisk mängd behöver inte vara en *kurva*, d.v.s. en *endimensionell* ”sak”. Som exempel är

$$M_1 \coloneqq \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\}$$

den slutna *enhetsdisken*, medan

$$M_1 \coloneqq \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 < 1\}$$

är den öppna enhetsdisken; skillnaden mellan dessa är att diskens rand, enhetscirkeln, tillhör M_1 men tillhör komplementet till M_2.[7] I en datorplot märks ingen skillnad:



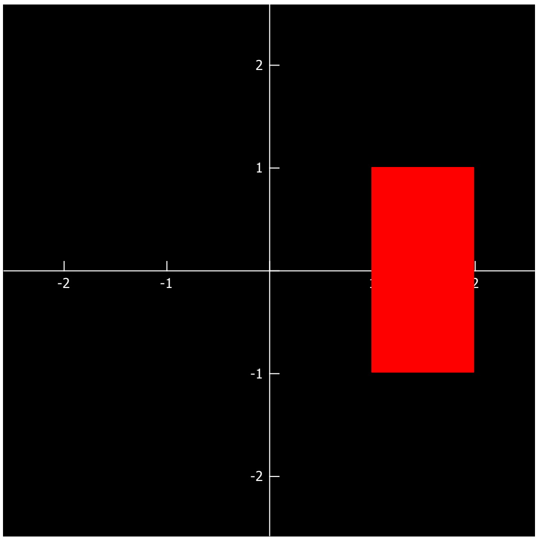
Figur 4. Enhetsdisken.

Ett väldigt smidigt sätt att ange en fylld rektangel i planet är att använda den kartesiska produkten mellan två intervall. Så är t.ex.

$$R \coloneqq [1,2] \times [-1, 1]$$

mängden av alla punkter (x, y) där $x \in [1, 2]$ och $y \in [-1, 1]$, d.v.s. rektangeln med hörn i $(1, 1)$, $(2, 1)$, $(2, -1)$ och $(1, -1)$:

[7] Notera likheten med terminologin för delmängder av $\mathbb{R}$. Vi kommer i delen om flervariabelanalys att noggrant behandla delmängder av planet (och rummet).



Figur 5. En rektangel.

Den här fyllda rektangeln kallas *sluten* eftersom kanten tillhör mängden. Motsvarande öppna fyllda rektangel är $]1,2[\times]-1,1[$. Speciellt kallas $[0,1]^2$ (resp. $]0,1[^2$) för den slutna (resp. öppna) fyllda *enhetskvadraten*.

1.2.8.1.4 Samspelet mellan algebra och geometri

Exempel

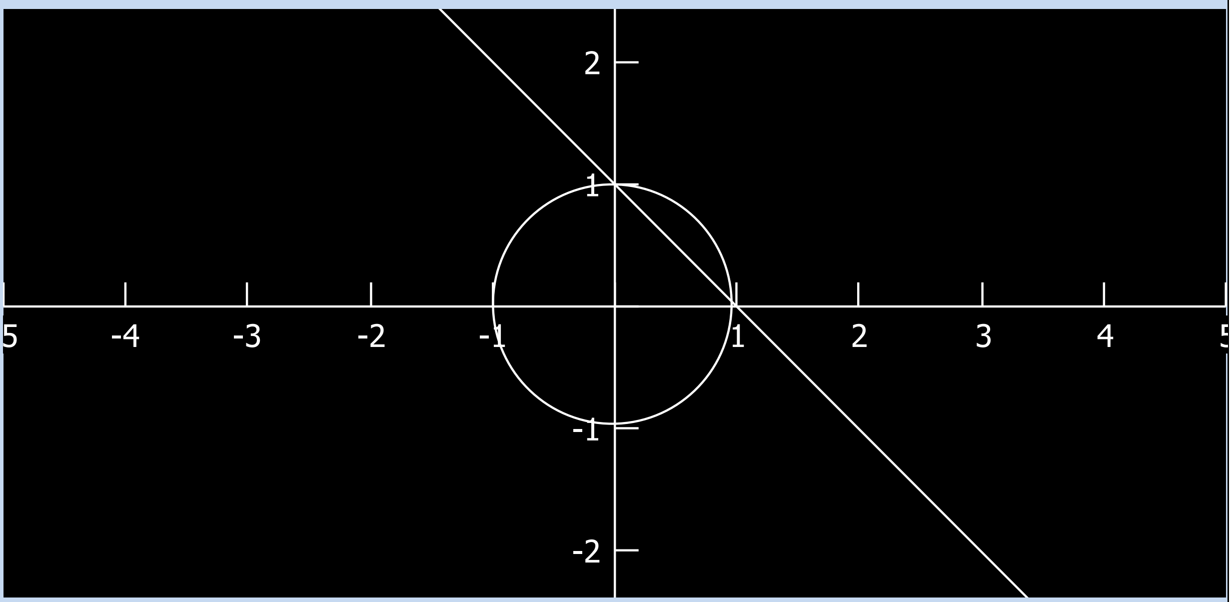
Som ett exempel på samspelet mellan algebra och geometri, betrakta de två mängderna

$$S^1 = \{(x,y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}$$
$$L = \{(x,y) \in \mathbb{R}^2 : x + y = 1\}.$$

S^1 är enhetscirkeln och L är en rät linje som också kan skrivas $y = 1 - x$. Vi är intresserade av snittet

$$S^1 \cap L = \{(x,y) \in \mathbb{R}^2 : x^2 + y^2 = 1 \quad \text{och} \quad x + y = 1\}.$$

I bilden nedan visas S^1 och L.



Det är uppenbart att $S^1 \cap L = \{(0,1),(1,0)\}$, så snittet av två mängder som båda består av oändligt många punkter kan tydligen vara en mängd som består av precis två punkter. Det vi har gjort är i praktiken att vi löst *ekvationssystemet*

$$\begin{cases} x^2 + y^2 = 1 \\ x + y = 1, \end{cases}$$

d.v.s. vi har funnit de punkter $(x,y) \in \mathbb{R}^2$ som uppfyller båda ekvationerna:

$$x^2 + y^2 = 1 \text{ och } x + y = 1 \iff (x,y) = (1,0) \text{ eller } (x,y) = (0,1).$$

1.2.8.1.5 Exempel

Exempel

Beskriv mängden $\{(x,y) \in \mathbb{R}^2: \quad x^2 - 2x + y^2 - 2y = -1\}$.

Lösning: Kvadratkomplettering ger

$$x^2 - 2x + y^2 - 2y = (x-1)^2 - 1 + (y-1)^2 - 1$$

så att

$$x^2 - 2x + y^2 - 2y = -1 \iff$$
$$\iff (x-1)^2 + (y-1)^2 = 1.$$

Mängden är alltså cirkeln med radie 1 kring $(1,1)$.

Exempel

Bestäm snittet $A \cap B$ mellan mängderna $A \coloneqq \{(x,y) \in \mathbb{R}^2: \quad x^2 + y^2 - 2y = 3\}$ och $B \coloneqq \{(x,y) \in \mathbb{R}^2: \quad x^2 - 10x + y^2 = -24\}$.

Lösning: Notera att

$$x^2 + y^2 - 2y = 3 \iff x^2 + (y-1)^2 = 4$$

medan

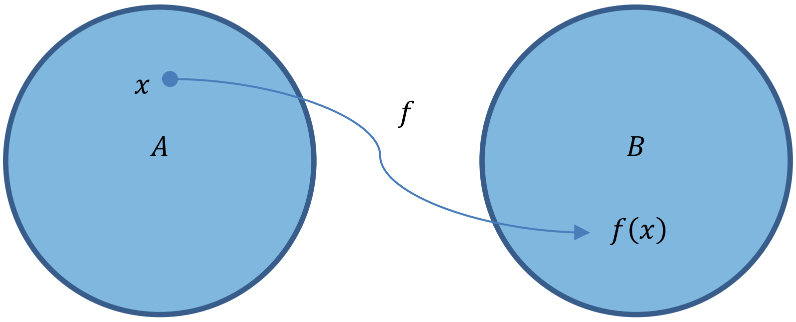
$$x^2 - 10x + y^2 = -24 \iff (x-5)^2 + y^2 = 1.$$

A är alltså cirkeln med radie 2 kring $(0,1)$ medan B är cirkeln med radie 1 kring $(5,0)$. Avståndet mellan cirklarnas medelpunkter är $\sqrt{(5-0)^2 + (0-1)^2} = \sqrt{26}$. Samtidigt är summan av cirklarnas radier $2 + 1 = 3 = \sqrt{9} < \sqrt{26}$. Eftersom summan av radierna är mindre än avståndet mellan medelpunkterna måste cirklarna vara disjunkta.

Svar: $A \cap B = \emptyset$.

1.3 Funktioner

Betrakta två mängder X och Y. En *funktion* från X till Y är en *regel* som till *varje* $x \in X$ entydigt ordnar ett $y \in Y$. En funktion kan alltså ses som en maskin som man kan stoppa in ett element från X i, och varje gång man gör det, får man ut ett bestämt element i Y. Varje gång man stoppar in samma x, får man ut samma y, så utvärdet är entydigt bestämt av valet av det element man stoppar in.



Om f är en funktion från X till Y så skriver man $f: X \to Y$. Mängden X kallas för f:s *definitionsmängd* och brukar betecknas D_f. Mängden Y kallas för funktionens *målmängd*.

Observera!

Om vi t.ex. har en funktion vid namn ”sin” så betecknas alltså dess definitionsmängd $D_{\sin}$.

Mängden av element som f ger ifrån sig (när man, i tur och ordning, stoppar in alla element i X) kallas för f:s *värdemängd* och brukar betecknas V_f. Notera noggrant att $D_f = X$ medan $V_f = \{f(x): x \in D_f\} \subset Y$ mycket väl kan vara en *äkta* delmängd av Y, d.v.s. $V_f \neq Y$.

Det element i Y som f ger ifrån sig då man stoppar in $x \in X$ i f betecknas $f(x)$ och kallas ibland för *bilden av* x (under f).

För att ange en funktion, måste man alltså ange mängderna X och Y likväl som den regel som associerar ett element i Y till varje element i X. Låt oss betrakta ett konkret exempel, genom att sätta $X = Y = \mathbb{R}$, d.v.s. vi betraktar en funktion $f: \mathbb{R} \to \mathbb{R}$ som för *varje* reellt tal $x \in \mathbb{R}$ ger ifrån sig ett bestämt reellt tal $f(x) \in \mathbb{R}$. Som funktionsregel använder vi

$$f(x) \coloneqq x^2 + 1,$$

d.v.s. f tar in ett reellt tal och ger ifrån sig det tal som erhålles genom att det givna talet först kvadreras varvid resultatet adderas med 1. Till exempel får vi då $f(0) = 1$, $f(1) = 2$, $f(2) = 5$, $f(3) = 10$ och $f(-10) = 101$. I det här fallet är definitionsmängden $D_f = \mathbb{R}$ och värdemängden $V_f = [1, \infty[$. I det här fallet är värdemängden en *äkta* delmängd till målmängden $\mathbb{R}$; det finns t.ex. *inget* tal $x \in \mathbb{R}$ sådant att $f(x) = 0$.

För att ange regeln för funktionen f kan man också skriva $f: x \mapsto x^2 + 1$, där tecknet ”$\mapsto$” utläses ”avbildas på”. Man säger alltså att $x \in X$ ”hamnar på” $x^2 + 1 \in Y$.

Observera!

Notera i synnerhet att $f: X \to Y$ anger att f är en funktion från mängden X till mängden Y, medan $f: x \mapsto$ (uttryck i x) anger själva regeln för funktionen; här är $x \in X$ medan uttrycket måste resultera i ett element i Y. Notera också att notationen $f: X \to Y$ per automatik medför att $X = D_f$ är mängden av alla tal som kan stoppas in i f, så $f(x)$ måste vara definierat för varje $x \in X$. Däremot behöver inte $Y = V_f$, d.v.s. varje element i Y behöver inte erhållas under f för något $x \in X$.

I många fall anger man inte explicit definitionsmängden för en funktion f. I sådana fall är det underförstått att definitionsmängden är den största delmängd av det universum man för tillfället betraktar, i vilken regeln för funktionen är definierad. Till exempel, om vi arbetar med reella tal och $f: x \mapsto x^2$ och $g: x \mapsto 1/x$ så är det underförstått att $D_f = \mathbb{R}$ och $D_g = \mathbb{R} \setminus \{0\}$.

Två funktioner f och g säges vara lika (vilket skrivs $f = g$) omm de har samma definitionsmängd $D_f = D_g$, samma målmängd och $f(x) = g(x)$ för varje $x \in D_f (= D_g)$.

Exempel

Betrakta funktionerna

$$\begin{aligned}&f: \mathbb{R} \to \mathbb{R}, x \mapsto x^2 + 1\\&g: [0, \infty[, x \mapsto x^2 + 1\\&h: \mathbb{R} \to \mathbb{R}, x \mapsto (x+1)^2 - 2x\\&k: \mathbb{R} \to [0, \infty[, x \mapsto x^2 + 1\end{aligned}$$

Då gäller att $f = h$ men $f \neq g$ och $g \neq h$, eftersom definitionsmängderna är olika.

Vi har också t.ex. $f \neq k$ eftersom målmängderna är olika.

I många situationer är det av ringa vikt vad målmängden till en funktion är. Det kan i sådana fall vara motiverat att även betrakta två funktioner som lika om det enda som skiljer dem åt är att de har olika målmängd, som i exemplet ovan med f och k (men båda målmängderna är förstås supermängder till den gemensamma värdemängden). Vissa författare definierar därför likhet mellan funktioner genom att säga att två funktioner f och g är lika omm de har samma definitionsmängd $D_f = D_g$ och $f(x) = g(x)$ för varje $x \in D_f$. De två definitionerna har sina respektive fördelar och är olika lämpliga inom olika matematiska discipliner. (Generellt kan sägas att den ”stränga” definitionen av likhet är lämplig inom algebran, medan den ”snälla” definitionen av likhet är lämplig inom analys.)

Exempel

Betrakta funktionerna

$$\begin{aligned}&f: \mathbb{R} \to \mathbb{R}, x \mapsto x^2\\&g: \mathbb{R} \to [0, \infty[, x \mapsto x^2.\end{aligned}$$

f och g har samma definitionsmängd ($D_f = D_g = \mathbb{R}$) och värdemängd ($V_f = V_g = [0, \infty[$). Dessutom gäller $f(x) = g(x)$ för varje $x \in D_f = D_g$. I många situationer kan det då vara naturligt att säga att f och g är *samma* funktion. I andra situationer, emellertid, är det mer naturligt att säga att $f \neq g$. T.ex. har funktionen g en egenskap som f saknar, nämligen att V_g är lika med målmängden, d.v.s. *varje* element i målmängden $[0, \infty[$ erhålles som $g(x)$ för något $x \in D_f$.

1.3.1 Sammansättning av funktioner

Om $f: X \to Y$ och $g: Y \to Z$ kan man förstås bilda en ny funktion $h: X \to Z$ genom att sätta $h(x) := g\big(f(x)\big)$ för varje $x \in X$. Funktionen h kallas för *sammansättningen* av g och f och betecknas $g \circ f$, så $h = g \circ f$. Notera speciellt ordningen mellan funktionerna: $g \circ f$ är den funktion som *först* applicerar f på argumentet, och *sedan* applicerar g på resultatet.

Exempel

Låt $f: \mathbb{R} \to \mathbb{R}^3, t \mapsto (x, y, z)$ ge en flugas position (x, y, z) vid tiden t och låt $g: \mathbb{R}^3 \to \mathbb{R}, (x, y, z) \mapsto T$ ge temperaturen T i punkten (x, y, z). Då är $g \circ f: \mathbb{R} \to \mathbb{R}, t \mapsto T$ den funktion som ger temperaturen T där flugan är vid tiden t.

Låt $h: \mathbb{R}^3 \to \mathbb{R}$ vara definierad genom $h(x, y, z) = z$, d.v.s. h tar in en punkt i rummet och ger ifrån sig punktens z-koordinat (höjden över golvet, säg). Då är $h \circ f: \mathbb{R} \to \mathbb{R}, t \mapsto z$ den funktion som ger flugans z-koordinat (höjd över golvet) vid tiden t.

Exempel

Låt $f, g: \mathbb{R} \to \mathbb{R}$ vara definierade av

$$\begin{aligned} f(x) &= x^2 + 1, \qquad \forall x \in \mathbb{R} \\ g(x) &= 5x, \qquad \forall x \in \mathbb{R}. \end{aligned}$$

Då ges $g \circ f$ och $f \circ g: \mathbb{R} \to \mathbb{R}$ av

$$(g \circ f)(x) = g\big(f(x)\big) = g(x^2 + 1) = 5x^2 + 5$$

respektive

$$(f \circ g)(x) = f\big(g(x)\big) = f(5x) = 25x^2 + 1$$

varför $g \circ f \neq f \circ g$ (pröva t.ex. $x = 1$). Funktionssammansättning är sålunda *inte* en kommutativ operation.

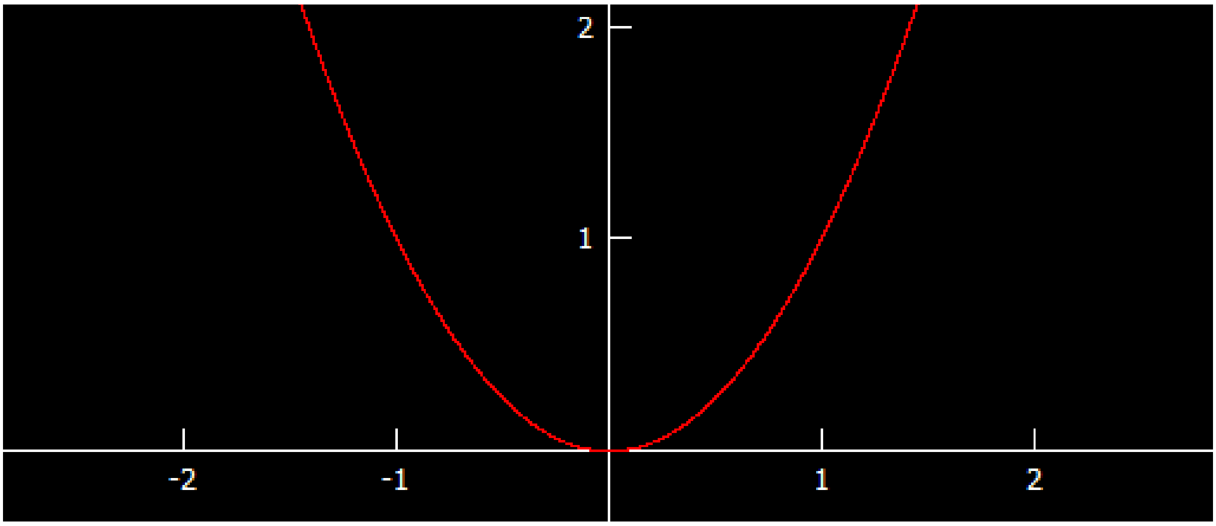
1.3.2 Grafer

Inom envariabelanalysen studerar man företrädelsevis funktioner $f: D_f \to \mathbb{R}$ där $D_f \subset \mathbb{R}$, d.v.s. funktioner som tar in ett reellt tal och ger ifrån sig ett annat. Sådana funktioner kan oftast illustreras ganska väl med hjälp av sina *grafer*, som vi här introducerar.

Grafen till en funktion $f: D_f \to \mathbb{R}$, där $D_f \subset \mathbb{R}$, är mängden

$$\{(x, y) \in \mathbb{R}^2: \quad x \in D_f, \quad y = f(x)\}.$$

Grafen är alltså den geometriska figuren som har ekvationen $y = f(x)$, för att återknyta till sådant vi pratat om tidigare. Som första exempel ger vi grafen till funktionen $x \mapsto x^2$, d.v.s. kurvan $y = x^2$, som kallas för en *parabel*:



Kurvan $y = x^2$ tydliggör för övrigt det faktum att det för varje icke-negativt tal y finns ett entydigt icke-negativt tal x vars kvadrat blir y; till exempel finns det precis ett icke-negativt tal vars kvadrat är 9, nämligen 3 (även -3 kvadreras till 9, men -3 är ju ett negativt tal).

1.3.2.1 Grafer som plana kurvor

Grafen till en funktion $f\colon D_f \to \mathbb{R}$ ($D_f \subset \mathbb{R}$) är alltså kurvan $y = f(x)$. Man kan ställa sig frågan om *varje* plan kurva är grafen till någon funktion. Svaret är ”nej”. Till exempel finns det ingen funktion som har enhetscirkeln som graf. Det är i det närmaste uppenbart: En graf är ju en kurva med en ekvation på formen $y = f(x)$, d.v.s. för varje fixt $x \in D_f$, så måste y vara just $f(x)$ för att (x, y) skall tillhöra grafen. Det innebär att varje rät linje i planet, parallell med y-axeln (d.v.s. med x-koordinaten fix), skär grafen max en gång (nämligen då $y = f(x)$ under förutsättning att $x \in D_f$).

Däremot är enhetscirkeln *unionen av två grafer*. Vi har ju

$$x^2 + y^2 = 1 \qquad \Longleftrightarrow \qquad y^2 = 1 - x^2 \qquad \Longleftrightarrow \qquad \begin{cases} y = -\sqrt{1 - x^2} \\ \text{eller} \\ y = +\sqrt{1 - x^2} \end{cases}$$

så enhetscirkeln är tydligen *unionen* av kurvorna $y = -\sqrt{1 - x^2}$ och $y = +\sqrt{1 - x^2}$, d.v.s. unionen av graferna till de två funktionerna $x \mapsto -\sqrt{1 - x^2}$ och $x \mapsto +\sqrt{1 - x^2}$, båda med definitionsmängd $[-1, 1]$ och värdemängderna $[-1, 0]$ respektive $[0, 1]$.

1.3.3 Några enkla funktioner

Vi ger här exempel på några enkla (och flera av dem ofta förekommande) funktioner.

Signumfunktionen $\operatorname{sgn}\colon \mathbb{R} \to \mathbb{R}$ ger ifrån sig *tecknet* av sitt argument (=det som man stoppar in i funktionen), d.v.s. sgn definieras av

$$\operatorname{sgn} x \coloneqq \begin{cases} -1 & \text{om } x < 0 \\ 0 & \text{om } x = 0 \\ +1 & \text{om } x > 0. \end{cases}$$

Här har vi också smugit in en ny notation, nämligen den för en *falldefinierad* funktion, där man delar in definitionsmängden D_f i olika delar och anger en separat funktionsregel för respektive del. I det här fallet skall notationen tolkas enligt följande: $\operatorname{sgn} x = -1$ för varje $x \in \,]-\infty, 0[$, $\operatorname{sgn} 0 = 0$ och $\operatorname{sgn} x = +1$ för varje $x \in \,]0, \infty[$.

Lägg märke till att vi skriver funktionsvärdet ”$\operatorname{sgn} x$” i stället för ”$\operatorname{sgn}(x)$”. Man brukar utelämna parenteserna vid välkända standardfunktioner om de inte behövs för tydligheten skull.

Absolutbeloppsfunktionen $|\cdot|\colon \mathbb{R} \to \mathbb{R}$ är så vanligt förekommande att den fått en egen symbol. Funktionsvärdet skrivs $|x|$, där x är argumentet [jfr $f(x)$]. Per definition är

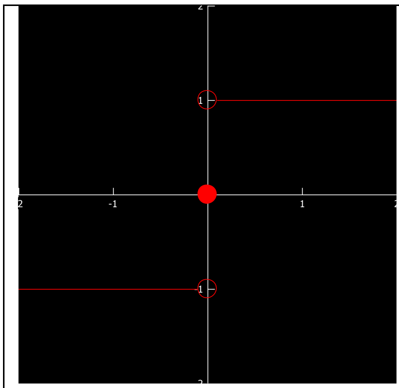
$$|x| := \begin{cases} x & \text{om } x \geq 0 \\ -x & \text{om } x < 0, \end{cases}$$

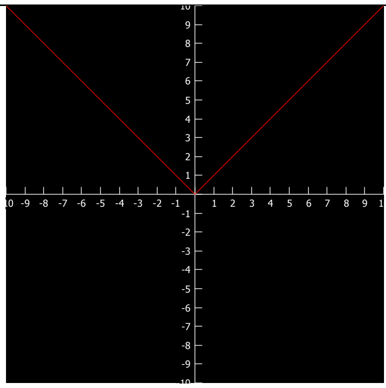
d.v.s. $|x|$ är lika med x fast med det eventuella minustecknet borttaget. Så är t.ex. $|5| = 5$, $|11| = 11$, $|2.1| = 2.1$, $|0| = 0$ men $|-3.5| = -(-3.5) = 3.5$, $|-10| = 10$ o.s.v. Vi har alltså, i själva verket, $|\cdot|\colon \mathbb{R} \to [0, \infty[$, d.v.s. $|x| \geq 0$ för varje $x \in \mathbb{R}$.

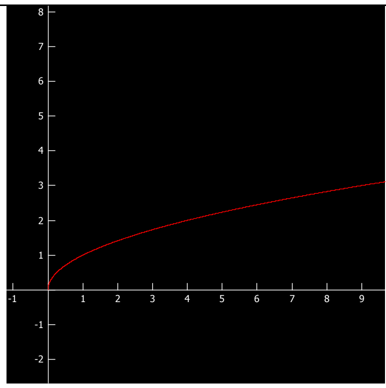
En användbar geometrisk karaktärisering av absolutbeloppsfunktionen är att $|x - y|$ är lika med *avståndet* mellan talen x och $y \in \mathbb{R}$ på den reella tallinjen.

Kvadratrotsfunktionen $\sqrt{}\colon [0, \infty[\,\to [0, \infty[$ har också fått en egen symbol: bilden av x under kvadratrotsfunktionen skrives $\sqrt{x}$. För varje $x \geq 0$ är $\sqrt{x}$ det (existerande och unika) icke-negativa tal som upphöjt till 2 blir x. Till exempel är $\sqrt{9} = 3$ eftersom 3 är det entydigt bestämda icke-negativa tal som upphöjt till 2 blir 9.

Nedan visas graferna till sgn, $|\cdot|$ och $\sqrt{}$.


$y = \operatorname{sgn} x$


$y = |x|$


$y = \sqrt{x}$

Ett vanligt fel

Notera två saker:

$D_{\sqrt{}} = [0, \infty[$, d.v.s. man kan bara stoppa in icke-negativa tal i kvadratrotsfunktionen.

$V_{\sqrt{}} = [0, \infty[$, d.v.s. kvadratrotsfunktionen ger bara ifrån sig icke-negativa tal.

Detta är två helt olika saker, som studenter emellanåt blandar ihop.

Exempel: När är uttrycket $\sqrt{x-3}+5$ definierat? Jo, det enda som kan gå fel är att argumentet till rotfunktionen blir negativt, så för att uttrycket skall vara definierat krävs och räcker att $x-3 \geq 0$, d.v.s. att $x \geq 3$. Här använder vi alltså att $D_{\sqrt{}} = [0, \infty[$.

Å andra sidan kan man lätt se att $x^2 + \sqrt{2x+6} \geq 0$ för alla x sådana att uttrycket över huvud taget är definierat, eftersom $x^2 \geq 0$ för alla $x \in \mathbb{R}$ och $V_{\sqrt{}} = [0, \infty[$.

Ett till vanligt fel

Notera ännu en gång att $D_{\sqrt{}} = [0, \infty[$. Det man stoppar in i kvadratrotsfunktionen måste alltså vara ett reellt tal, och inte vilket som helst, utan ett icke-negativt sådant! Faktum är att det är absolut nödvändigt, för vi har inte sagt vad som menas med t.ex. $\sqrt{-9}$ [roten ur ett negativt tal] eller $\sqrt{2+i}$ [roten ur ett icke-reellt komplex tal]. Och det är inte alls självklart vad det skulle betyda!

Ytterligare ett vanligt fel

Lägg märke till att det är *fel* att säga att x^2 alltid är positivt om $x \in \mathbb{R}$; det är inte sant att $x^2 > 0$ för alla $x \in \mathbb{R}$, eftersom $0^2 = 0$. Däremot är det sant att x^2 är icke-negativt, d.v.s. $x^2 \geq 0$ för varje $x \in \mathbb{R}$.

Talet 0 är varken negativt eller positivt. Mängden $]0, \infty[$ består av de *positiva* reella talen, mängden $[0, \infty[$ av de *icke-negativa*.

Följande observation är viktig:

Observation. $\sqrt{x^2} = |x|$ för varje $x \in \mathbb{R}$

Bevis. Antag att $x > 0$. Per definition är $\sqrt{x^2}$ det tal ≥ 0 vars kvadrat är x^2. Det finns två tal som har x^2 som kvadrat, nämligen $-x$ och $+x$. Eftersom $x > 0$ är det $+x$ som är det icke-negativa av dem. Alltså är $\sqrt{x^2} = x$. Men eftersom $x > 0$ är också $|x| = x$, så $\sqrt{x^2} = |x|$.

Om i stället $x = 0$ så är $x^2 = 0$ och det enda icke-negativa tal som kvadreras till 0 är 0. Samtidigt är $|x| = |0| = 0$, så återigen har vi $\sqrt{x^2} = |x|$.

Sista möjligheten är $x < 0$. Av de två tal som kvadreras till x^2, d.v.s. av $-x$ och $+x$, är det då bara $-x$ som är icke-negativt, så $\sqrt{x^2} = -x$. Men då $x < 0$ är ju också $|x| = -x$ så $\sqrt{x^2} = |x|$ även i det fallet.■

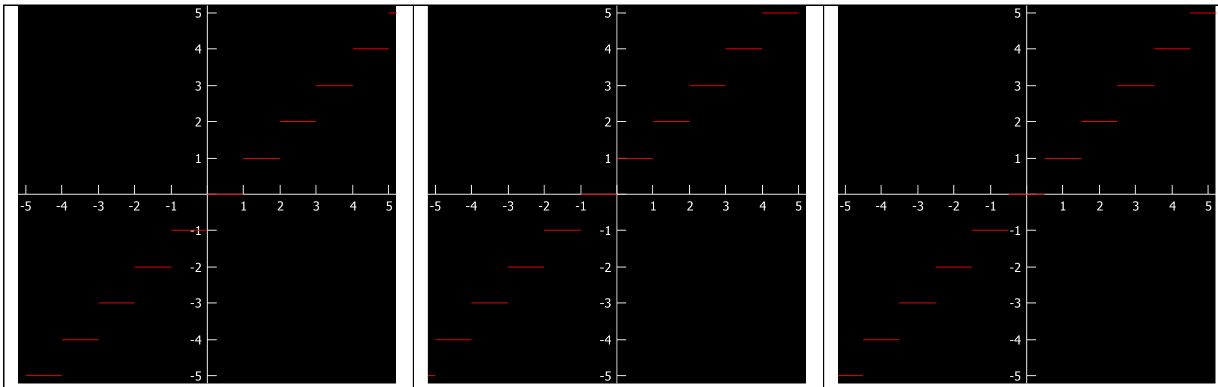
Till exempel är $|5| = 5$ och $\sqrt{5^2} = \sqrt{25} = 5$; $|-7| = 7$ och $\sqrt{(-7)^2} = \sqrt{49} = 7$.

Ett vanligt fel

Ett ganska vanligt förkommande fel hos nya matematikstudenter är antagandet att $\sqrt{x^2} = x$ för alla $x \in \mathbb{R}$. Detta gäller ju *inte* för negativa x. Nyss såg vi ju att $\sqrt{(-7)^2} = 7$. Det riktiga sambandet är $\sqrt{x^2} = |x|$.

Golv och tak. Här har vi också ett par funktioner med egna symboler. *Golvfunktionen* $\lfloor\cdot\rfloor: \mathbb{R} \to \mathbb{Z}$ avrundar sitt argument *nedåt* (till vänster på tallinjen) till närmsta heltal, och *takfunktionen* $\lceil\cdot\rceil: \mathbb{R} \to \mathbb{Z}$ avrundar sitt argument *uppåt* (till höger på tallinjen) till närmsta heltal. Så är t.ex. $\lfloor 5.3 \rfloor = 5$ medan $\lceil 5.3 \rceil = 6$.

Närmaste heltal. Funktionen $\mathrm{round}: \mathbb{R} \to \mathbb{Z}$ är nära besläktad med golvet och taket. $\mathrm{round}\, x$ är det heltal som ligger närmast x, d.v.s. round avrundar sitt argument till närmaste heltal, så $\mathrm{round}\, 5.3 = 5$ men $\mathrm{round}\, 5.7 = 6$. Om argumentet ligger precis mellan två heltal avrundas det uppåt (till höger på tallinjen), så $\mathrm{round}\, 5.5 = 6$.



$y = \lfloor x \rfloor$ $y = \lceil x \rceil$ $y = \mathrm{round}\, x$

1.3.4 Elementära funktioner

Inom envariabelanalysen studerar man funktioner som tar in och ger ifrån sig reella tal, d.v.s. funktioner $f: D_f \to \mathbb{R}$ där $D_f \subset \mathbb{R}$. Det finns en rad mycket vanligt förekommande, ganska enkla, funktioner som används överallt inom analysen och dess tillämpningar. Till dessa s.k. *elementära funktioner* räknas *polynom*, *rationella funktioner*, *potensfunktioner*, *exponentialfunktioner*, *logaritmfunktioner*, *trigonometriska funktioner* samt *arcusfunktioner*. Vi kommer nu att ge en inledande studie av dessa elementära funktioner.

1.3.4.1 Polynom och rationella funktioner

Om $(a_k)_{k=0}^n$ är en följd av givna tal så är

$$f(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_2 x^2 + a_1 x + a_0$$

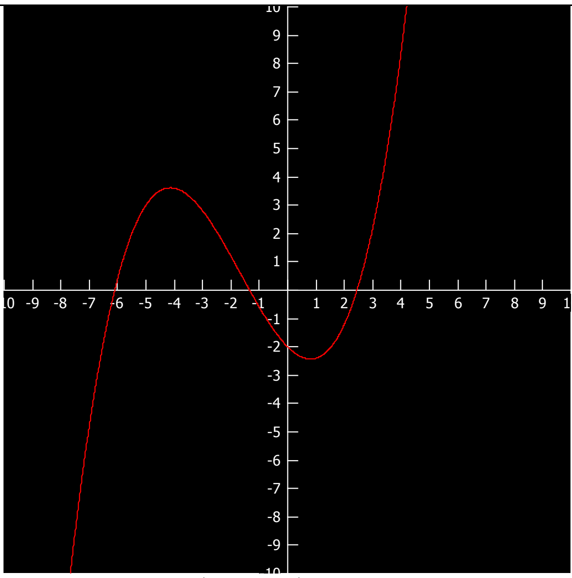
ett *polynom*. Ett exempel är

$$f(x) = \frac{1}{10} x^3 + \frac{1}{2} x^2 - x - 2$$

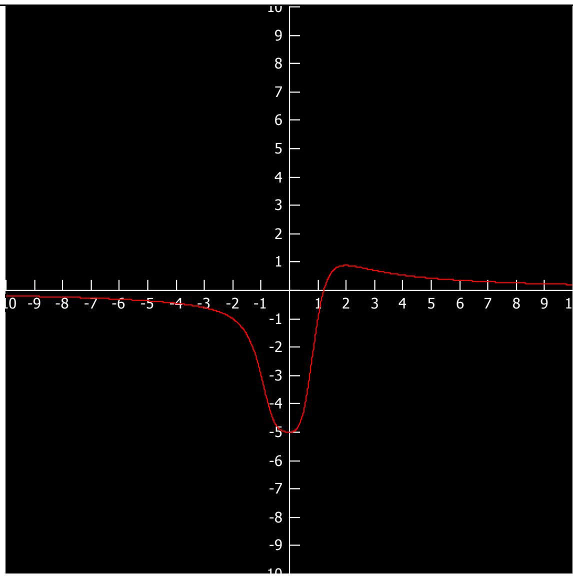
En *rationella funktion* är en kvot mellan två polynom, som t.ex.

$$\frac{2x^3 + x^2 - 5}{x^4 + 1}.$$

Nedan visas graferna till ett polynom och en rationell funktion:



$$y = \frac{1}{10}x^3 + \frac{1}{2}x^2 - x - 2$$



$$y = \frac{2x^3 + x^2 - 5}{x^4 + 1}$$

Det är enkelt att förklara dessa funktioners beteenden långt från origo. I polynomfallet gäller förstås att

$$\frac{1}{10}x^3 + \frac{1}{2}x^2 - x - 2 \approx \frac{1}{10}x^3$$

om x är ett till beloppet stort tal. Om x är stort positivt kommer alltså $f(x)$ att vara stort positivt, och om x är stort negativt kommer $f(x)$ att vara stort negativt (eftersom exponenten 3 är udda). I fallet med den rationella funktionen har vi i stället, långt från origo, uppskattningen

$$\frac{2x^3 + x^2 - 5}{x^4 + 1} \approx \frac{2x^3}{x^4} = \frac{2}{x}$$

så funktionsvärdet blir väldigt litet både när x blir stort positivt och när det blir stort negativt. Den rationella funktionen

$$x \mapsto \frac{2x^3 + x^2 - 5}{x^4 + 1}$$

är speciell på det sättet att nämnaren aldrig är noll. En följd blir då att funktionen är definierad på hela $\mathbb{R}$. Vi ställer oss nu frågan vad som händer nära punkter där nämnaren *blir* noll i en allmän rationell funktion. Vi betraktar därför tre konkreta exempel på sådana rationella funktioner:

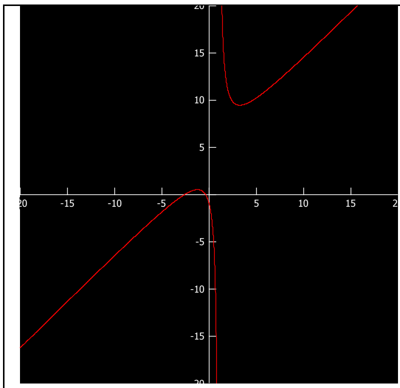
$$f(x) = \frac{x^2 + 3x + 1}{x - 1}, \qquad D_f = \mathbb{R} \setminus \{1\}$$
$$g(x) = \frac{x^2 + x - 2}{x - 1}, \qquad D_g = \mathbb{R} \setminus \{1\}$$
$$h(x) = \frac{x^2 + 3x - 1}{x^2}, \qquad D_h = \mathbb{R} \setminus \{0\}.$$

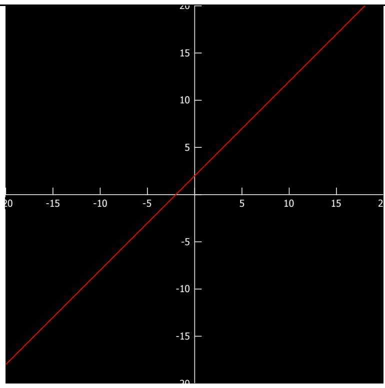
I första fallet (f) kommer täljaren att närma sig 4 när x närmar sig 1; samtidigt kommer nämnaren att närma sig 0. Om x närmar sig 1 *från höger* på tallinjen, d.v.s. om x är 1.1, 1.01, 1.001, …, så kommer nämnaren $x-1$ att närma sig 0 från höger $(0.1, 0.01, 0.001, \ldots)$. Kvoten $f(x)$ kommer då alltid att vara positiv, men den kommer att bli större och större. Grafen kommer därför att "sticka iväg" uppåt, mot "oändligheten" när vi närmar oss punkten $x=1$ från höger. En liknande analys ger att grafen "sticker iväg" nedåt, mot "minus oändligheten", när vi närmar oss punkten $x=1$ från vänster.

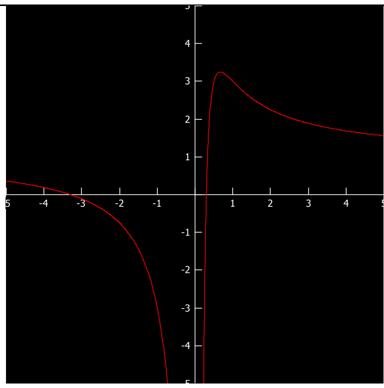
Man skulle kunna tro att detta alltid händer när nämnaren i en rationell funktion har ett nollställe, men så är inte fallet. Om vi tittar på det andra exemplet, har vi

$$g(x) = \frac{x^2+x-2}{x-1} = \frac{(x+2)(x-1)}{x-1} = x+2, \qquad \forall x \neq 1.$$

För alla $x \in \mathbb{R} \setminus \{1\}$ gäller därför att $g(x) = x+2$, en rät linje. Om vi plottar grafen till g kommer vi troligtvis inte ens att ana att något speciellt inträffar i punkten $x=1$ (eftersom en punkt har "storlek noll" som sträcka betraktad). Slutligen inser vi beträffande h, med en analys liknande den för f, att vi kommer att närma oss "minus oändligheten" när vi närmar oss 0 både från vänster och från höger. Graferna till f, g och h visas nedan.


$y = f(x)$

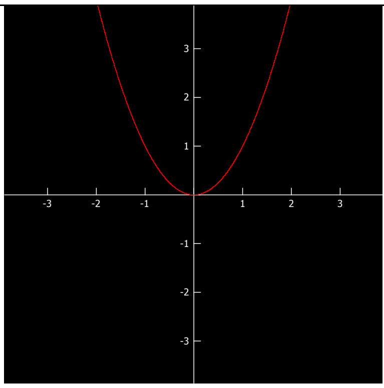

$y = g(x)$

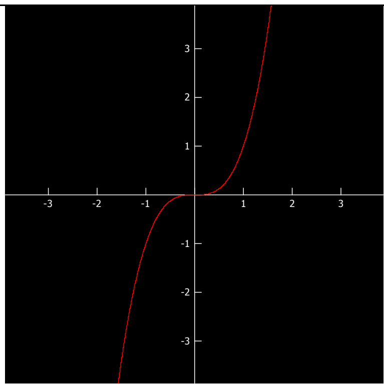

$y = h(x)$

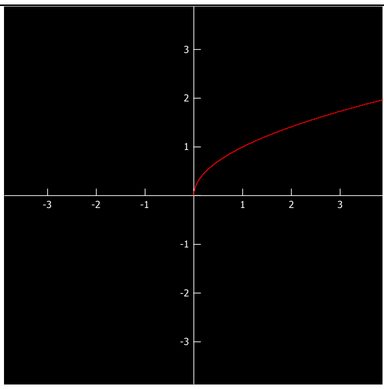
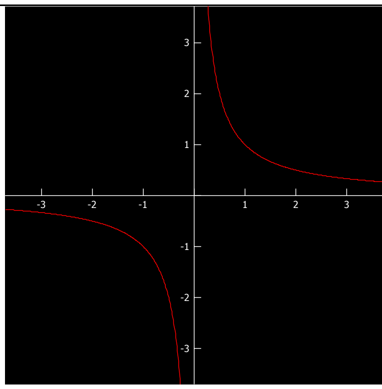
Man kan också notera att $f(x) \approx g(x) \approx x$ för stora $|x|$ medan $h(x) \approx 1$ för stora $|x|$.

1.3.4.2 Potensfunktioner

Om $\alpha \in \mathbb{R}$ är ett fixt tal så är $x \mapsto x^\alpha$ en *potensfunktion*. Hit räknas alltså bland annat identitetsfunktionen $x \mapsto x$, kvadreringsfunktionen $x \mapsto x^2$, kubfunktionen $x \mapsto x^3$, funktionen $x \mapsto \frac{1}{x} (= x^{-1})$ och kvadratrotsfunktionen $x \mapsto x^{1/2} (= \sqrt{x})$. Nedan visas några exempel på grafer.

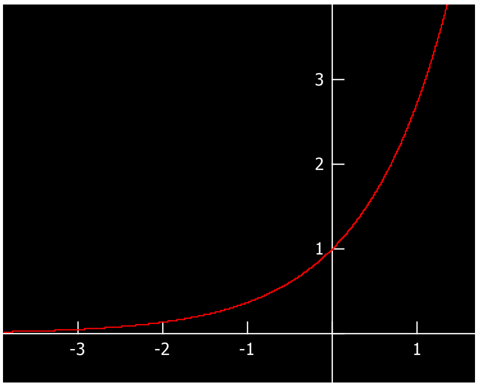

$y = x^2$


$y = x^3$

 $y = \sqrt{x}$	 $y = 1/x$

1.3.4.3 Exponentialfunktioner

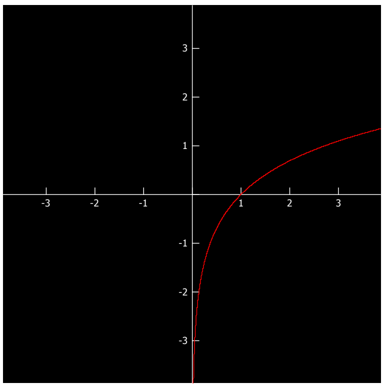
Om $\alpha \in \mathbb{R}^+$ är ett fixt tal så är $x \mapsto \alpha^x$ en *exponentialfunktion*. Den i särklass mest använda exponentialfunktionen är den som har det speciella irrationella talet $e = 2.718281\ldots$ som bas; ibland avses rent av just denna när man bara pratar om ”exponentialfunktionen”. Funktionen $x \mapsto e^x$ kallas ofta ”exp” och vi kan därför skriva $e^x = \exp x$. Nedan visas grafen till exp; notera att $\exp: \mathbb{R} \to \mathbb{R}^+$.



Figur 6. $y = e^x$.

1.3.4.4 Logaritmfunktioner

Om $y \in \mathbb{R}^+$ är ett givet positivt tal så finns det exakt ett $x \in \mathbb{R}$ sådant att $y = e^x$. Detta tal betecknas $\ln y$, så vi har alltså $x = \ln y$. På detta sätt har vi introducerat en funktion $\ln: \mathbb{R}^+ \to \mathbb{R}$ som kallas för *den naturliga logaritmfunktionen*. Nedan visas grafen till ln.

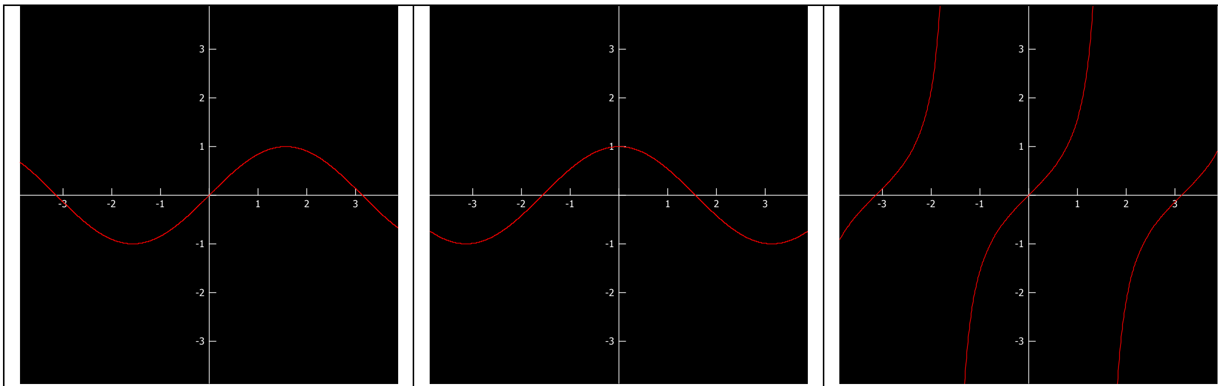


Figur 7. $y = \ln x$.

Om α är ett givet positivt tal kan vi betrakta en allmän exponentialfunktion $x \mapsto \alpha^x$. Även här finns det för varje $y \in \mathbb{R}^+$ precis ett tal x sådant att $y = \alpha^x$, och detta tal betecknas $\log_\alpha y$, så vi har $x = \log_\alpha y$. Vi har därmed definierat en funktion $\log_\alpha: \mathbb{R}^+ \to \mathbb{R}$ för varje $\alpha \in \mathbb{R}^+$, nämligen *logaritmfunktionen med α som bas*. Notera att $\ln y = \log_e y$. Ett annat specialfall är $\log_{10}$ som ibland skrives $\log$, kort och gott. Tyvärr använder en del författare och datorprogram beteckningen ”log” för det vi skulle kalla ”ln”, så viss försiktighet rekommenderas.

1.3.4.5 Trigonometriska funktioner

De trigonometriska funktionerna bör vara bekanta från tidigare studier. Vi kommer ändå att ägna ett helt kapitel åt ämnet, och här nöjer vi oss med att repetera graferna till de tre ”vanligaste” trigonometriska funktionerna, sinus (sin), cosinus (cos) och tangens (tan). Anledningen är att vi då kan använda också dessa som exempel i nedanstående avsnitt om funktioners egenskaper.



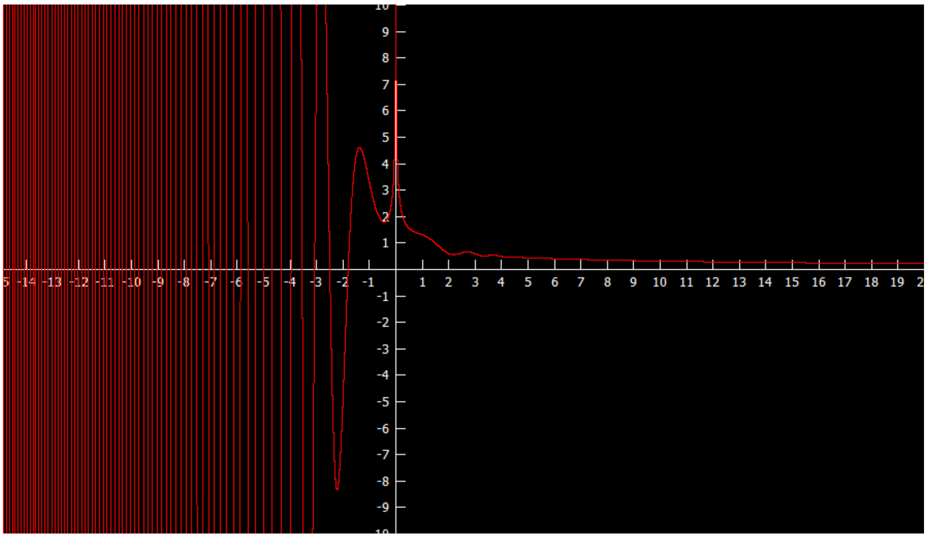
$y = \sin x$ | $y = \cos x$ | $y = \tan x$

1.3.4.6 Uttryck i elementära funktioner

Ett *uttryck i elementära funktioner* är en formel (som ger upphov till en funktion) vilken är uppbyggd av aritmetiska operationer (addition, subtraktion, multiplikation, division, upphöjt till) av sammansättningar av elementära funktioner, såsom

$$\frac{\sin x^2}{e^x} + \frac{1}{\sqrt{|x|}}$$

med den fascinerande grafen



Vissa sådana uttryck (funktioner) förekommer så ofta att de får egna namn, t.ex. *sinus hyperbolicus* och *cosinus hyperbolicus*, som är funktionerna

$$x \mapsto \frac{e^x - e^{-x}}{2}, \quad x \mapsto \frac{e^x + e^{-x}}{2}.$$

Grafen till cosinus hyperbolicus stöter du kanske på varje dag – det är den form som en ideal tvättlina (utan tvätt) har när den bara påverkas av gravitationskraften. De hyperboliska funktionerna kommer också att studeras ingående i ett eget kapitel.

1.3.5 Egenskaper hos funktioner

I det här avsnittet introducerar vi begrepp med vilka vi kan beskriva och klassificera funktioner. Vi börjar först med ”allmänna” funktionsbegrepp, som kan appliceras på alla sorters funktioner. Sedan tittar vi speciellt på funktioner som tar in och ger ifrån sig just reella tal, vilket är vad man främst studerar inom envariabelanalysen.

1.3.5.1 Allmänna egenskaper hos funktioner

1.3.5.1.1 Injektivitet

Definition. Låt $f: X \to Y$. Omm utsagan

$$x \neq y \qquad \Longrightarrow \qquad f(x) \neq f(y)$$

är sann för varje $x, y \in X$ så säges f vara *injektiv*.

En ekvivalent utsaga (den *kontrapositiva* utsagan) är

$$f(x) = f(y) \qquad \Longrightarrow \qquad x = y.$$

En injektiv funktion är med andra ord en funktion sådan att det inte finns två *olika* element i X som avbildas på *samma* element i Y.

Exempel

Funktionen $f: \mathbb{R} \to \mathbb{R}, x \mapsto x^2$ är *inte* injektiv, t.ex. eftersom $f(-5) = (-5)^2 = 25$ och $f(5) = 5^2 = 25$, d.v.s. både -5 och $+5 \in \mathbb{R}$ avbildas på $25 \in \mathbb{R}$. Däremot är funktionen $g: \mathbb{R} \to \mathbb{R}, x \mapsto x + 1$ injektiv ty om $x \neq y$ så är det självklart så att $x + 1 \neq y + 1$, d.v.s. $g(x) \neq g(y)$.

Om f är injektiv och om $y \in V_f$ är något som funktionen ”spottat ur sig”, kan man alltså, givet fullständig kunskap om f, med säkerhet säga exakt vad som stoppades in i f. Detta följer direkt av definitionen för injektivitet: om $y \in V_f$ så är $y = f(x)$ för något $x \in X$. Men eftersom f är injektiv så kan det inte finnas två (eller flera) *olika* sådana x, som båda avbildas på y.

Om $f: X \to Y$ är injektiv, så finns det med andra ord för varje $y \in V_f$ ett *unikt* $x \in D_f$ sådant att $y = f(x)$. Detta element betecknas ofta $x = f^{-1}(y)$. Regeln som till varje $y \in V_f$ ordnar $f^{-1}(y) \in X$ kallas för f:s invers, betecknas (som antytt) f^{-1} och är tydligen en funktion $V_f \to D_f$.

Exempel

Betrakta f och g från föregående exempel. Funktionen f är inte injektiv. Om vi t.ex. vet att $f(x) = 9$ så kan vi inte med säkerhet säga exakt vilket element x är – det finns ju flera element som avbildas på 9, nämligen -3 och $+3$. Funktionen g är däremot injektiv med invers $g^{-1}: \mathbb{R} \to \mathbb{R}, x \mapsto x - 1$. Om vi vet att $g(x) = 27$ kan vi sålunda med säkerhet säga att $x = g^{-1}(27) = 26$.

Om $f: X \to Y$ är en injektiv funktion gäller uppenbarligen att $f^{-1}\big(f(x)\big) = x$ för varje $x \in D_f$ och $f\big(f^{-1}(x)\big) = x$ för varje $x \in V_f = D_{f^{-1}}$. Alternativt uttryckt har vi (förklara!)

$$y = f(x) \qquad \Longleftrightarrow \qquad x = f^{-1}(y).$$

Exempel

Funktionen $f: \mathbb{R} \to \mathbb{R}$ definierad av $f(x) = 3x + 7$ är injektiv och har inversen $f^{-1}(x) = \frac{x-7}{3}$. Vi ser enkelt att $f\big(f^{-1}(x)\big) = f^{-1}\big(f(x)\big) = x$.

Kvadratrotsfunktionen $\sqrt{\cdot}: [0, \infty[$ är inversen till kvadreringsfunktionen $[0, \infty[\to [0, \infty[, x \mapsto x^2$. (**OBSERVERA** den begränsade definitionsmängden – den vanliga kvadreringsfunktionen $\mathbb{R} \to [0, \infty[, x \mapsto x^2$ är ju **inte** injektiv!)

Den naturliga logaritmen $\ln: [0, \infty[\to \mathbb{R}$ är inversen till exponentialfunktionen $\exp: \mathbb{R} \to [0, \infty[$.

I specialfallet $f: D_f \to \mathbb{R}$, där $D_f \subset \mathbb{R}$, är f injektiv om och endast om varje rät linje $y = c$ skär grafen till f i högst en punkt; detta följer direkt av definitionen av injektivitet.

1.3.5.1.2 Surjektivitet

Om $f: X \to Y$ så gäller att $D_f = X$ och $V_f \subset Y$. Till exempel är kvadreringsfunktionen $f: \mathbb{R} \to \mathbb{R}, x \mapsto x^2$ sådan att $D_f = \mathbb{R}$ [du kan stoppa in varje reellt tal i den] medan $V_f = [0, \infty[\subset \mathbb{R}$ är en *äkta* delmängd av $\mathbb{R}$. Exempelvis finns det inget tal $x \in \mathbb{R}$ sådant att $f(x) = x^2 = -1 \in \mathbb{R}$.

Om $V_f = Y$ säges f vara *surjektiv*. Det betyder att det till *varje* $y \in Y$ finns *minst* ett $x \in X$ sådant att $y = f(x)$; man kan alltså erhålla *varje* element i Y. Funktionen $f: \mathbb{R} \to \mathbb{R}: x \mapsto x^2$ är alltså inte surjektiv, men både $g: \mathbb{R} \to \mathbb{R}, x \mapsto x + 1$ och $h: \mathbb{R} \to [0, \infty[, x \mapsto x^2$ är surjektiva.

Exempelparet f och h kan få begreppet surjektivitet att förefalla fånigt. För en godtycklig funktion $f: X \to Y$ är ju funktionen $\tilde{f}: X \to V_f$ definierad av samma regel som f en surjektiv funktion sådan att $f(x) = \tilde{f}(x)$ för varje $x \in X$ – det rör sig "i all väsentlighet" om samma funktion. Så, "i all väsentlighet" förefaller alla funktioner vara surjektiva.

Begreppet är emellertid ofta användbart. Ofta är det nämligen naturligt att betrakta en funktion som en regel från X till Y även om $V_f \neq Y$ är en tänkbar möjlighet eller ett uppenbart faktum. Ofta är X och Y nämligen väldigt naturliga universum, och det kan också vara svårt att bestämma V_f samtidigt som den precisa värdemängden kanske inte ens är relevant. Som ett belysande exempel kan vi betrakta de två polynomfunktionerna $f: \mathbb{R} \to \mathbb{R}$ och $g: \mathbb{R} \to \mathbb{R}$ definierade av

$$f(x) \coloneqq x^3 + 3x^2 - 2x + 5, \qquad g(x) \coloneqq x^4 - 6x^3 + 10x^2 - 6x + 2.$$

Dessa funktioner är definierade för alla reella tal x, och det är fullkomligt uppenbart att V_f och V_g båda är delmängder av $\mathbb{R}$. Vi säger att både f och g är funktioner "som tar in reella tal och ger ifrån sig reella tal". Däremot krävs räknearbete för att inse att[8]

$$V_f = \mathbb{R}, \qquad V_g = [-7, \infty[.$$

Det är således i det här fallet naturligt att säga att $f, g: \mathbb{R} \to \mathbb{R}$ samtidigt som det kan vara av stor vikt att f är surjektiv medan g inte är det.

1.3.5.1.3 Bijektivitet

En funktion $f: X \to Y$ som både är injektiv och surjektiv säges vara *bijektiv*. I sådant fall avbildas varje $x \in X$ på något $y \in Y$, och för varje $y \in Y$ finns det minst ett (surjektivitet) och högst ett (injektivitet), d.v.s. *exakt* ett, $x \in X$ sådant att $f(x) = y$.

Det råder sålunda ett 1–1-förhållande mellan elementen i X och Y; de hör ihop parvis. Om endera X eller Y är en ändlig mängd, måste de alltså båda vara ändliga, och $|X| = |Y|$, d.v.s. de består av precis lika många element.

Angående terminologin så kallas ibland en injektiv, surjektiv eller bijektiv funktion för en *injektion*, *surjektion* respektive *bijektion*.

Notera att om $f: X \to Y$ är en bijektion från X till Y så är $f^{-1}: Y \to X$ en bijektion från Y till X.

Exempel

Låt $X = [0,10] \cap \mathbb{Z}$, $Y = [0, 100] \cap \mathbb{Z}$ och $Z = [50, 60] \cap \mathbb{Z}$.

Det finns då *ingen* surjektion $X \to Y$ (varför?). Däremot är såväl $f: X \to Y, x \mapsto x$ som $g: X \to Y, x \mapsto 10x$ injektiva.

Det finns *ingen* injektion $Y \to X$ (varför?). Däremot är funktionen $h: Y \to X, y \mapsto \lfloor y/10 \rfloor$ en surjektion.

Funktionen $b_1: X \to Z, x \mapsto x + 50$ är en bijektion, liksom $b_2: X \to Z, x \mapsto 60 - x$. Och mycket riktigt har vi $|X| = |Z| = 11$.

1.3.5.2 *Egenskaper hos reella funktioner*

Envariabelanalys handlar om funktioner $f: D_f \to \mathbb{R}$ där $D_f \subset \mathbb{R}$, d.v.s. funktioner som "tar in och ger ifrån sig" reella tal. Å andra sidan handlar flervariabelanalys om funktioner $F: D_F \to \mathbb{R}^m$ där $D_F \subset \mathbb{R}^n$, d.v.s. funktioner som tar in punkter i $\mathbb{R}^n$ och ger ifrån sig punkter i $\mathbb{R}^m$.

I det här avsnittet studerar vi envariabelanalysens funktioner som tar in och ger ifrån sig reella tal. Vi betraktar därför genomgående en funktion $f: D_f \to \mathbb{R}$ med $D_f \subset \mathbb{R}$.

[8] Att $V_f = \mathbb{R}$ är dock trivialt eftersom polynomet är av udda grad.

1.3.5.2.1 Begränsade funktioner

Definition. Om det finns ett tal $M \in \mathbb{R}$ sådant att $f(x) \leq M$ för alla $x \in D_f$ säger vi att f är *uppåt begränsad*; om det finns ett tal $M \in \mathbb{R}$ sådant att $f(x) \geq M$ för alla $x \in D_f$ säger vi att f är *nedåt begränsad*. En funktion som är både uppåt och nedåt begränsad kallas *begränsad*.

Notera att följande påståenden är ekvivalenta:

- f är begränsad
- det finns ett $M > 0$ sådant att $|f(x)| \leq M$ för alla $x \in D_f$
- V_f är en begränsad mängd
- grafen till f är en delmängd av något band $\mathbb{R} \times [-M, M] \subset \mathbb{R}^2$ av ändlig höljd $2M$.

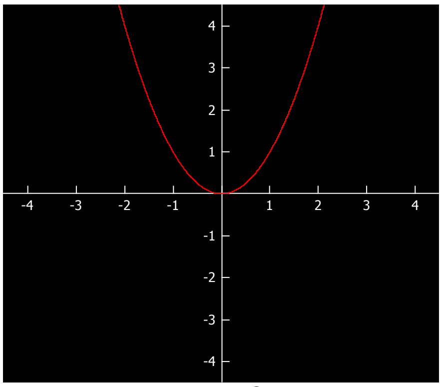
Icke-konstanta polynom är aldrig begränsade, t.ex. kan ju x^2 bli hur stort som helst. Däremot är de trigonometriska funktionerna sin och cos begränsade (men inte tan): $|\sin x| \leq 1$ för alla $x \in \mathbb{R}$, d.v.s. $V_{\sin} \subset [-1,1]$ (i själva verket är $V_{\sin} = [-1,1]$), och liknande för cos.

1.3.5.2.2 Jämna och udda funktioner

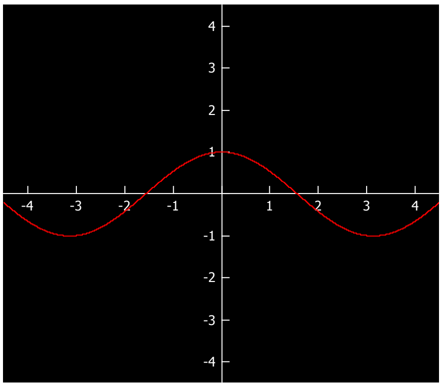
Om identiteten $f(-x) = f(x)$ gäller där definierat säges f vara *jämn*. Funktionerna $x \mapsto x^n$ (n jämnt heltal) och cos är exempel på jämna funktioner. Om identiteten $f(-x) = -f(x)$ gäller där definierat säges f vara *udda*. Funktionerna $x \mapsto x^n$ (n udda heltal) och sin är exempel på udda funktioner.

Nedan visas graferna till några jämna respektive udda funktioner. Se till att du förstår de två sorters symmetri kring y-axeln som jämna respektive udda funktioner uppvisar i sina grafer.

Jämna funktioner

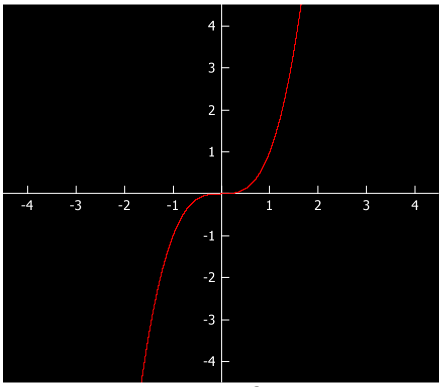


$y = x^2$

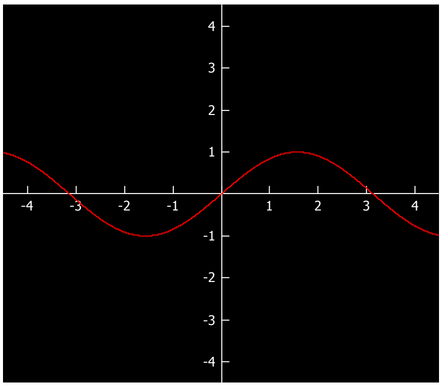


$y = \cos x$

Udda funktioner



$y = x^3$



$y = \sin x$

Det är lätt att inse att t.ex. produkten av en jämn och udda funktion är udda. Låt oss bevisa det. Antag att j är jämn och u är udda och betrakta produkten[9] $p \coloneqq ju$. Då är

$$p(-x) = j(-x)u(-x) = j(x) \cdot \big(-u(x)\big) = -j(x)u(x) = -p(x)$$

så att p är udda. På liknande sätt inser vi att ”jämn · jämn = jämn” och ”udda · udda = jämn”, samt att motsvarande regler gäller om ”produkt” ersätts med ”kvot”. Till exempel är tan udda, eftersom den är kvoten mellan en udda funktion (sin) och en jämn funktion (cos).

1.3.5.2.3 (Strängt) växande/avtagande/monoton

Definition. Låt f vara en funktion $D_f \to \mathbb{R}$ med $D_f \subset \mathbb{R}$. Om

$$x_2 > x_1 \qquad \Longrightarrow \qquad f(x_2) \geq f(x_1)$$

för varje $x_1, x_2 \in D_f$ säges f vara *växande*. Om till och med

$$x_2 > x_1 \qquad \Longrightarrow \qquad f(x_2) > f(x_1)$$

för varje $x_1, x_2 \in D_f$ säges f vara *strängt växande*.

Om

$$x_2 > x_1 \qquad \Longrightarrow \qquad f(x_2) \leq f(x_1)$$

för varje $x_1, x_2 \in D_f$ säges f vara *avtagande*. Om till och med

$$x_2 > x_1 \qquad \Longrightarrow \qquad f(x_2) < f(x_1)$$

för varje $x_1, x_2 \in D_f$ säges f vara *strängt avtagande*.

Om f är växande *eller* avtagande, säges f vara monoton; om f är strängt växande *eller* strängt avtagande, säges f vara strängt monoton.

Tolkningarna av dessa begrepp torde vara uppenbara. En växande funktion, t.ex., är en funktion sådan att funktionsvärdet $f(x)$ antingen är oförändrat eller blir större när x ökar. Om det *alltid* blir större när x ökar, är funktionen till och med *strängt* växande.

Lägg märke till att en konstant funktion $f(x) \coloneqq c$ för varje $x \in D_f$ ($c \in \mathbb{R}$ fix konstant) både är växande och avtagande (och monoton), men naturligtvis inte *strängt* monoton (om D_f består av två eller fler tal).

Sats. Om f är strängt växande så är f injektiv.

Bevis. Välj två element $x, y \in D_f$. Om $x \neq y$ så är antingen $x < y$ eller $x > y$. I första fallet är $f(x) < f(y)$, och i andra fallet är $f(x) > f(y)$, på grund av att f är strängt växande. I båda fallen är alltså $f(x) \neq f(y)$, d.v.s. $x \neq y \Rightarrow f(x) \neq f(y)$ så f är injektiv. ■

Ett analogt resonemang fungerar om f i stället är strängt *avtagande*, så i själva verket har vi

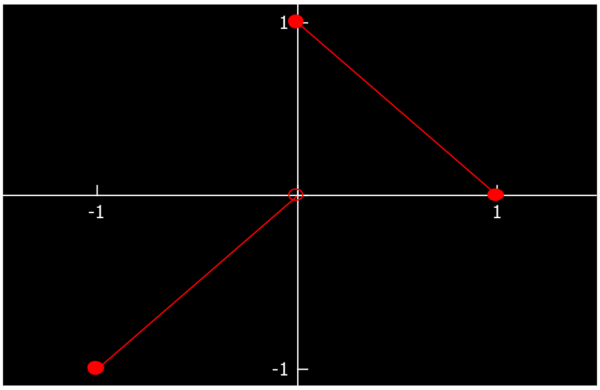
[9] Produkten $p: D_j \cap D_u \to \mathbb{R}$ definieras av $p(x) = j(x)u(x)$ för varje $x \in D_j \cap D_u$.

Sats. Om f är strängt monoton så är f injektiv.

Däremot gäller inte omvändingen; ett motexempel ges av funktionen $f: [-1, 1] \to \mathbb{R}$ definierad av

$$f(x) := \begin{cases} x & \text{om } x \in [-1,0[\\ 1-x & \text{om } x \in [0,1] \end{cases} \tag{1}$$

som är injektiv men varken växande eller avtagande.



Det är lätt att inse att summan av två (strängt) växande (resp. avtagande) funktioner också är (strängt) växande (resp. avtagande). Till exempel, om både f och g är strängt växande och summan[10] $h := f + g$ så gäller

$$x_2 > x_1 \qquad \Longrightarrow \qquad h(x_2) = f(x_2) + g(x_2) > f(x_1) + g(x_1) = h(x_1).$$

Däremot behöver inte t.ex. *produkten* av två strängt växande funktioner vara strängt växande; ett motexempel ges av $f(x) = x$ och $g(x) = x$, som båda är strängt växande, men där produkten $h(x) = f(x)f(x) = x^2$ inte är det.

(Strängt) växande/avtagande/monoton, som vi definierat det ovan, är alltså en egenskap hos en funktion. Man kan också tala om motsvarande egenskaper hos en funktion och en godtycklig delmängd av funktionens definitionsmängd. Om t.ex. $S \subset D_f$ är en delmängd av f:s definitionsmängd och

$$x_2 > x_1 \qquad \Longrightarrow \qquad f(x_2) > f(x_1)$$

för varje $x_1, x_2 \in S$ säges f vara *strängt växande på* S. Liknande definitioner görs förstås för övriga begrepp. Som exempel kan sägas att vår funktion (1) ovan är strängt växande på $[-1, 0]$ och strängt avtagande på $[0,1]$. Kvadreringsfunktionen $x \mapsto x^2$ är å sin sida strängt avtagande på $]-\infty, 0]$ och strängt växande på $[0, \infty[$. Se till att du förstår varför de stängda ändpunkterna i dessa fyra intervall faktiskt kan vara stängda.

Övning

1. Visa att om f är udda och $0 \in D_f$ så är $f(0) = 0$.

[10] Summan $h: D_f \cap D_g \to \mathbb{R}$ är (naturligtvis) definierad av $h(x) = f(x) + g(x)$ för alla $x \in D_f \cap D_g$.

1.3.6 Bilder och urbilder

Vi avslutar funktionskapitlet med att introducera ett par ganska viktiga begrepp. Möjligtvis kan dessa ses lite som ”överkurs”, men begreppen är så pass vanligt förekommande att det ändå är en god idé att ta dem till hjärtat.

1.3.6.1 Bilder

Betrakta en funktion $f: X \to Y$. Den ger då ifrån sig ett element $f(x) \in Y$ för varje $x \in X$. Om $A \subset X$ är en *mängd* av element i X, så kan vi förstås *a priori* INTE skriva ”$f(A)$”, ty A är inte ett element i X.

Exempel. Om $f: \mathbb{R} \to \mathbb{R}$ är definierad av $f(x) = x + 2$ för varje $x \in \mathbb{R}$, så kan vi räkna ut $f(2) = 4$ och $f(5) = 7$, men inte $f(\{1,2,3\})$. Anledningen är att f förväntar sig ett *tal*, men $\{1,2,3\}$ är inte ett tal, utan en mängd av tal; det rör sig alltså om fel sorts objekt. f tar in tal, inte mängder av tal. (Å andra sidan kan medelvärdesfunktionen, som tar in en mängd tal och ger ifrån sig deras aritmetiska medelvärde, appliceras på $\{1,2,3\}$ (med värdet 2), men inte på 2 eller 5.)

Däremot är det en konvention att göra följande definition:

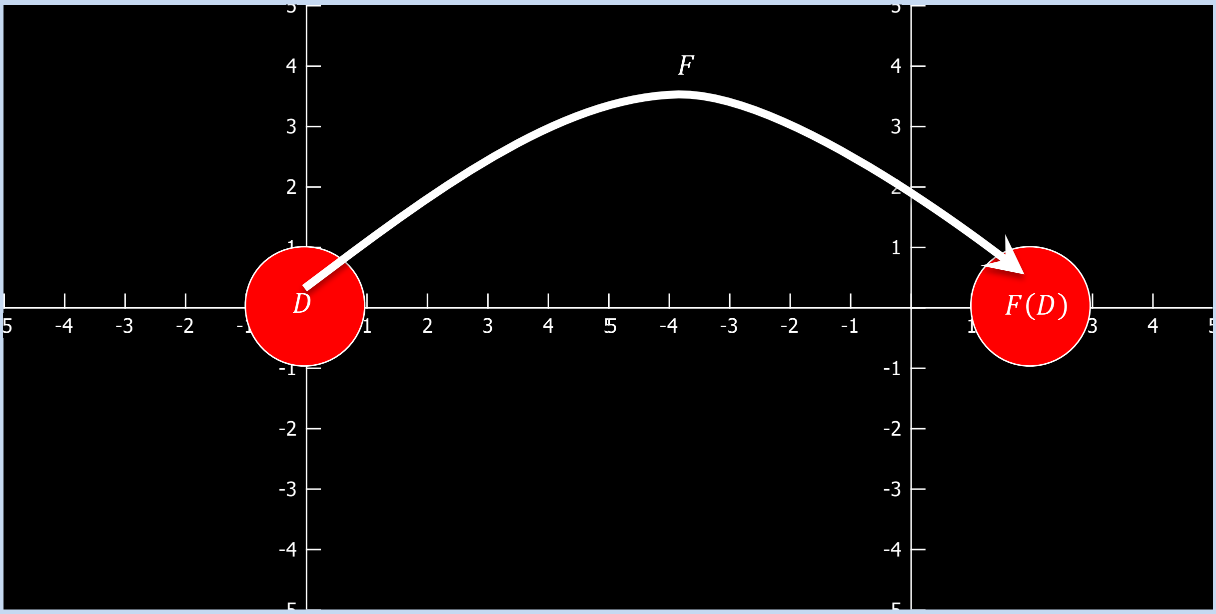
Definition. Om $f: X \to Y$ och $A \subset X$ så är $f(A) = \{f(a): a \in A\}$, d.v.s. $f(A)$ är *mängden* av element i Y som erhålles när man stoppar in alla element i A i f. $f(A)$ kallas *bilden* av A under f.

Exempel. Om $f(x) = x + 2$ så är $f(\{1,2,3\}) = \{3,4,5\}$, $f([0,1[) = [2,3[$, $f(\mathbb{R}^+) =]2,\infty[$ och $f(\mathbb{R}) = \mathbb{R}$.

Exempel. Om $f(x) = x^2$ är $f(\mathbb{R}) = [0,\infty[$ och $f(\{-2,2\}) = \{4\}$.

Exempel

Låt $F: \mathbb{R}^2 \to \mathbb{R}^2$ vara definierad av $F(x,y) = (x+2, y)$, så att t.ex. $F(7,5) = (9,5)$. F tar alltså in en punkt i planet och flyttar den två längdenheter åt höger. Om $D = \{(x,y) \in \mathbb{R}^2: \quad x^2 + y^2 \leq 1\}$ är den fyllda enhetsdisken så är då $F(D)$ den fyllda disken av radie 1 kring punkten $(2,0)$:



1.3.6.2 Urbilder

Vi gör också följande definition:

Definition. Om $f: X \to Y$ och $B \subset Y$ så är $f^{-1}(B) = \{x \in X: \quad f(x) \in B\}$, d.v.s. $f^{-1}(B)$ består precis av de element i X vilka avbildas in i B. $f^{-1}(B)$ kallas för *urbilden* till B under f.

Exempel. Om $f(x) = x + 2$ så är $f^{-1}(\{5\}) = \{3\}$, $\quad f^{-1}(\{1,2,3\}) = \{-1, 0, 1\}$, $\quad f^{-1}([0,1[) = [-2, -1[$ och $f^{-1}(\mathbb{R}) = \mathbb{R}$.

Exempel. Om $f(x) = x^2$ så är $f^{-1}(\{4\}) = \{-2, 2\}$, $\quad f^{-1}([4,9[) =]-3, -2] \cup [2, 3[$ och $f^{-1}([0, \infty[) = \mathbb{R}$.

En funktion $f: X \to Y$ behöver som bekant inte vara injektiv, d.v.s. den behöver inte ha en *invers*. Notera att en funktion är injektiv om och endast om varje urbild bestående av ett element i Y består av max ett element i X. Till exempel är $f(x) = x^2$ inte injektiv, ty t.ex. urbilden $f^{-1}(\{4\}) = \{-2, 2\}$ består av två element. Om funktionen är injektiv, däremot, så är urbilden av ett enskilt element $y \in Y$ precis mängden av invsen $f^{-1}(y)$, d.v.s. $f^{-1}(\{y\}) = \{f^{-1}(y)\}$; till exempel, om $f(x) = x + 2$ så är $f^{-1}(\{5\}) = \{3\} = \{f^{-1}(5)\}$ där första ”f^{-1}” är urbildsfunktionen medan andra är inversen.

Är det inte farligt att använda samma symbol för invers och urbildsfunktion? Nej, för det finns en viktig syntaktisk skillnad: inversen (om den existerar) tar in element i Y, medan urbildsfunktionen (som alltid existerar) tar in delmängder av Y, så de kan inte blandas ihop. Och, som vi nyss såg, urbilden av en singelton (=mängd bestående av bara ett element) i Y är ”väsentligen” samma sak som inversen av elementet, ifall den senare existerar.

Så urbildsfunktionen generaliserar inversen, dels för att *varje* funktion har en urbildsfunktion (även funktioner som inte är injektiva), dels för att man kan beräkna urbilden av en hel *mängd* element i Y (inte bara ett ensamt element).

1.4 Ekvationslösning

1.4.1 Formell inledning

I det här avsnittet ger vi en ganska strikt, men ändå belysande, beskrivning av begreppen ”ekvation” och ”ekvationslösning”.

Likhetstecknet ”$=$” används mellan två matematiska objekt (såsom tal, vektorer, matriser, mängder eller funktioner), och den erhållna formeln utgör en utsaga, som (liksom alla utsagor) antingen är sann eller falsk. Utsagan är sann om och endast om de två objekten är identiska; detta kan sägas vara likhetstecknets definition. Till exempel är $1+1=2$ en sann utsaga, eftersom talen $1+1$ och 2 är exakt lika; de båda uttrycken representerar exakt samma element i $\mathbb{R}$. Utsagan $1+1=3$ är däremot falsk, eftersom elementen 2 och $3 \in \mathbb{R}$ *inte* är identiska.

Den utsaga som erhålles kallas för en *ekvation*. Ibland kan den innehålla en symbol som står för en ett tills vidare okänt objekt (en *obekant variabel*), som i exemplet

$$5x = 3x + 7 \tag{1}$$

som är en ekvation innehållande reella tal. Sanningsvärdet för denna utsaga beror på vilket reellt tal symbolen x står för. Om t.ex. $x = 2$ lyder ekvationen $10 = 13$ som är en falsk utsaga, men om $x = \frac{7}{2}$ lyder ekvationen $\frac{35}{2} = \frac{35}{2}$ som är en sann utsaga. Vi säger att $x = \frac{35}{2}$ är en *lösning* eller en *rot* till ekvationen (själva lösningen eller roten är egentligen inte utsagan ”$x = \frac{35}{2}$” utan talet $\frac{35}{2}$ självt). Mängden av alla lösningar, d.v.s. den mängd som innehåller alla element som gör ekvationen till en sann utsaga, men inga andra element, kallas för ekvationens *lösningsmängd*.

Att *lösa* en ekvation är samma sak som att bestämma dess lösningsmängd.

En ekvation i den obekanta variabeln x är alltid ekvivalent med utsagan att x tillhör ekvationens lösningsmängd. Det följer att två ekvationer är ekvivalenta om och endast om de har samma lösningsmängd.

Lösningsmängden till ekvationen (1) innehåller alltså talet $7/2$. Finns det fler lösningar? Hur kommer man fram till den här lösningen, utan att pröva ett element i taget (vilket tar väldigt lång tid med tanke på att $\mathbb{R}$ innehåller *oändligt* många element som alla i sådana fall måste prövas)? Jo, vi manipulerar ekvationen ett steg i taget och får på så sätt en följd av ekvationer, som förhoppningsvis slutar i en så pass enkel ekvation att vi på den direkt ser vilken lösningsmängden just *den* ekvationen har. Om varje manipulation ger upphov till en *ekvivalent* ekvation, så har ursprungsekvationen samma lösningsmängd som den sista ekvationen, och vi är klara. Så har vi t.ex.

$$5x = 3x + 7 \qquad \Longleftrightarrow \qquad 2x = 7 \qquad \Longleftrightarrow \qquad x = \frac{7}{2}.$$

(I första steget adderar vi $-3x$ till båda leden och i andra multiplicerar vi båda leden med $1/2$; för att gå åt andra hållet multiplicerar vi med 2 och adderar med $3x$. Hur man i praktiken manipulerar enkla ekvationer bör vara mycket välbekant från tidigare studier på gymnasium eller i grundskolan.)

Den sista ekvationen är uppenbarligen sann om och endast om x är talet $7/2$ så dess lösningsmängd är $\{7/2\}$. Eftersom den sista ekvationen är ekvivalent med den första är $\{7/2\}$ också lösningsmängden för den ursprungliga ekvationen.

1.4.1.1 Tolkning av en redovisad ekvationslösning

Betrakta följande redovisning av en ekvationslösning:

$$5x = 3x + 7 \qquad \Longleftrightarrow \qquad 2x = 7 \qquad \Longleftrightarrow \qquad x = \frac{7}{2}.$$

Första ekvationen är vår ursprungliga ekvation och sista ekvationen är den vi använder för att utläsa lösningsmängden. Dessa är ekvivalenta, d.v.s.

$$5x = 3x + 7 \qquad \Longleftrightarrow \qquad x = \frac{7}{2}.$$

Liksom alla ekvivalenser består denna utsaga av två utsagor, som båda gäller. Först har vi

$$5x = 3x + 7 \qquad \Longrightarrow \qquad x = \frac{7}{2}$$

som säger att *om* $5x = 3x + 7$ *så* måste $x = 7/2$, d.v.s. *om* ekvationen har en lösning, *så* måste lösningen vara $7/2$. Det kan alltså inte finnas *några andra* lösningar. Däremot säger inte den här implikationen något om huruvida $7/2$ faktiskt *är* en lösning, eller ens om ekvationen över huvud taget har några lösningar. Men vi har också

$$5x = 3x + 7 \qquad \Longleftarrow \qquad x = \frac{7}{2}$$

som säger att *om* $x = 7/2$, *så* är $5x = 3x + 7$, d.v.s. $7/2$ *är* en lösning till ekvationen. Låt oss sammanfatta våra observationer:

Observation 1. Låt $EQ(x)$ vara en ekvation i den obekanta variabeln x och låt S vara lösningsmängden. Om A är en mängd så innebär påståendet

$$EQ(x) \qquad \Longrightarrow \qquad x \in A$$

att elementen i A är de enda möjliga kandidaterna till rötter, d.v.s. $S \subset A$. Det kan inte finnas några rötter till ekvationen som *inte* finns i A. Den omvända implikationen

$$EQ(x) \qquad \Longleftarrow \qquad x \in A$$

betyder å andra sidan att varje element i A faktiskt löser ekvationen, d.v.s. $A \subset S$. Om båda implikationerna är sanna, d.v.s. om

$$EQ(x) \qquad \Longleftrightarrow \qquad x \in A$$

så är $A = S$, d.v.s. A är ekvationens lösningsmängd: varje lösning till ekvationen finns i A, och mängden A innehåller inga andra element än just lösningarna.

1.4.2 Andragradsekvationer

Låt oss värma upp genom att lösa den mycket okomplicerade andragradsekvationen $x^2 = 9$. Syftet är enbart att diskutera notationen. Det är klart att

$$x^2 = 9 \qquad \Longleftrightarrow \qquad x = -3 \quad \text{eller} \quad x = 3.$$

Ofta brukar högra ledet förkortas genom att man skriver $x = \pm 3$:

$$x^2 = 9 \qquad \Longleftrightarrow \qquad x = \pm 3.$$

Tolkningen är fortfarande densamma. Man kan också skriva

$$x^2 = 9 \qquad \Longleftrightarrow \qquad x \in \{-3, 3\}.$$

Fördelen med det här skrivsättet är att lösningsmängden $\{-3, 3\}$ till ekvationen *bokstavligt talat* framgår i slutet av lösningsgången, vilket onekligen medför ett visst mått av elegans till redovisningen.

En allmän andragradsekvation $ax^2 + bx + c = 0$ kan alltid lösas medelst kvadratkomplettering.

Exempel

Lös ekvationen $6x^2 + x - 2 = 0$.

Lösning:

$$\begin{aligned}
6x^2 + x - 2 = 0 \qquad &\Longleftrightarrow \qquad x^2 + \frac{x}{6} - \frac{1}{3} = 0 \qquad \Longleftrightarrow \\
\Longleftrightarrow \qquad &\left(x + \frac{1}{12}\right)^2 - \left(\frac{1}{12}\right)^2 - \frac{1}{3} = 0 \qquad \Longleftrightarrow \\
\Longleftrightarrow \qquad &\left(x + \frac{1}{12}\right)^2 = \frac{49}{144} \qquad \Longleftrightarrow \\
\Longleftrightarrow \qquad &x + \frac{1}{12} = \pm\frac{7}{12} \qquad \Longleftrightarrow \\
\Longleftrightarrow \qquad &x = -\frac{1}{12} \pm \frac{7}{12} \qquad \Longleftrightarrow \\
\Longleftrightarrow \qquad &x = -\frac{2}{3} \quad \text{eller} \quad x = \frac{1}{2}.
\end{aligned}$$

Svar: $6x^2 + x - 2 = 0 \Leftrightarrow x = -\frac{2}{3}$ eller $x = \frac{1}{2}$.

Följande svar är också tänkbart, och kanske även mer lättläst:

Svar: $6x^2 + x - 2 = 0 \Leftrightarrow x \in \left\{-\frac{2}{3}, \frac{1}{2}\right\}$.

Exempel

Lös ekvationen $x^2 + 3x - 1 = 0$.

Lösning:

$$x^2+3x-1=0 \iff \left(x+\frac{3}{2}\right)^2-\left(\frac{3}{2}\right)^2-1=0 \iff$$
$$\iff \left(x+\frac{3}{2}\right)^2=\frac{13}{4} \iff x+\frac{3}{2}=\pm\frac{\sqrt{13}}{2} \iff$$
$$\iff x=-\frac{3}{2}\pm\frac{\sqrt{13}}{2}.$$

Svar: $x^2+3x-1 \Leftrightarrow x=-\frac{3}{2}\pm\frac{\sqrt{13}}{2}$.

Exempel

Lös ekvationen $x^2+x-2=0$.

Lösning:

$$x^2+x-2=0 \iff \left(x+\frac{1}{2}\right)^2-\frac{1}{4}-2=0 \iff$$
$$\iff \left(x+\frac{1}{2}\right)^2=\frac{9}{4} \iff x+\frac{1}{2}=\pm\frac{3}{2} \iff$$
$$\iff x=-\frac{1}{2}\pm\frac{3}{2} \iff x\in\{-2,1\}.$$

Svar: $x^2+x-2=0 \Leftrightarrow x\in\{-2,1\}$.

Det finns ett bättre sätt att lösa föregående uppgift. Man kan nämligen direkt se att $x^2+x-2=(x-1)(x+2)$. Hur gör man det? Jo, man ser onekligen direkt att $x=1$ är ett nollställe till polynomet. Enligt den s.k. faktorsatsen, som vi kommer att introducera noggrannare senare, gäller då att $x-1$ måste vara en faktor i polynomet, d.v.s. $x^2+x-2=(x-1)(x+c)$ för något $c\in\mathbb{R}$. Man inser direkt att $-2=-1\cdot c$ så $c=2$. Följande redovisning är alltså *mycket bättre*:

Exempel

Lös ekvationen $x^2+x-2=0$.

Lösning:

$$x^2+x-2=0 \iff (x-1)(x+2)=0 \iff x\in\{-2,1\}.$$

Svar: $x^2+x-2=0 \Leftrightarrow x\in\{-2,1\}$.

Kroppen $\mathbb{R}$ har den egenskapen, att det finns polynom med reella tal som koefficienter vilka saknar reella rötter.

Exempel

Bestäm alla reella rötter till $9x^2+6x+10=0$.

Lösning:

$$9x^2+6x+10=0 \iff x^2+\frac{2}{3}x+\frac{10}{9}=0 \iff$$
$$\iff \left(x+\frac{1}{3}\right)^2-\frac{1}{9}+\frac{10}{9}=0 \iff$$
$$\iff \left(x+\frac{1}{3}\right)^2=-1$$

som är omöjligt om $x\in\mathbb{R}$ (d.v.s. utsagan är falsk för varje $x\in\mathbb{R}$). Alltså saknar ekvationen reella lösningar.

Svar: Ekvationen saknar reella lösningar.

Vi återkommer till andragradsekvationer lite senare, efter att vi bekantat oss med de komplexa talen. Då kommer vi även att ge exempel på ekvationslösning av andragradsekvationer med icke-reella rötter.

1.4.3 Rotekvationer

Om $a=b$ så är det självklart så att $a^2=b^2$:

$$a=b \implies a^2=b^2.$$

Däremot gäller inte omvändningen: Bara för att $a^2=b^2$ måste inte $a=b$, utan $a=-b$ är också tänkbart. Till exempel, om $x=3$ så är det klart att $x^2=9$, men bara för att $x^2=9$ måste inte $x=3$ – vi kan också ha $x=-3$.

Om vi kvadrerar båda leden i en ekvation har vi därför bara implikation åt höger, inte åt vänster. Enligt Observation 1 får vi då en ny ekvation vars lösningsmängd kan vara *större* än den ursprungliga ekvationens; vi kan med andra ord inte vara säkra på att alla lösningarna till den nya ekvationen verkligen löser den ursprungliga. Vi måste därför *pröva* de erhållna rötterna i ursprungsekvationen (eller någon ekvation som är ekvivalent med denna), vilket i praktiken brukar låta sig göras, eftersom de oftast är få till antalet.

Ett vanligt exempel där vi förlorar ekvivalensen genom kvadrering är vid lösning av rotekvationer.

Exempel

Lös ekvationen $\sqrt{x+3}=x+1$.

Lösning:

$$\sqrt{x+3}=x+1 \implies x+3=(x+1)^2 \iff$$
$$\iff x+3=x^2+2x+1 \iff$$
$$\iff x^2+x-2=0 \iff$$
$$\iff (x-1)(x+2)=0 \iff$$
$$\iff x\in\{-2,1\}.$$

Prövning i ursprungsekvationen ger

$$x=-2 \implies 1=-1$$
$$x=1 \implies 2=2$$

så att endast $x = 1$ är en rot.

Svar: $\sqrt{x+3} = x + 1 \Leftrightarrow x = 1.$

Ett alternativt förfarande är att se till att ekvivalens råder i varje steg. Man måste då göra något åt den informationsförlust som inträffar vid kvadreringen, genom att lägga till något villkor på högra sidan. Vi har

$$a = b \qquad \Longrightarrow \qquad a^2 = b^2. \tag{2}$$

Om vi vet att $a^2 = b^2$ så är det klart att $\sqrt{a^2} = \sqrt{b^2}$, d.v.s. $|a| = |b|$, d.v.s. $a = \pm b$. För att kunna dra slutsatsen $a = b$ räcker det tydligen med att känna till att a och b har samma tecken. Låt oss pröva det:

Exempel

Lös ekvationen $\sqrt{x+3} = x + 1$.

Lösning:

$$\begin{aligned}
\sqrt{x+3} = x + 1 \quad &\Longleftrightarrow \quad \begin{cases} x + 3 = (x+1)^2 \\ x + 1 \geq 0 \end{cases} \quad \Longleftrightarrow \\
&\Longleftrightarrow \quad \begin{cases} x + 3 = x^2 + 2x + 1 \\ x \geq -1 \end{cases} \quad \Longleftrightarrow \\
&\Longleftrightarrow \quad \begin{cases} x^2 + x - 2 = 0 \\ x \geq -1 \end{cases} \quad \Longleftrightarrow \\
&\Longleftrightarrow \quad \begin{cases} (x-1)(x+2) = 0 \\ x \geq -1 \end{cases} \quad \Longleftrightarrow \\
&\Longleftrightarrow \quad \begin{cases} x \in \{-2, 1\} \\ x \geq -1 \end{cases} \quad \Longleftrightarrow \quad x = 1.
\end{aligned}$$

Svar: $\sqrt{x+3} = x + 1 \Leftrightarrow x = 1.$

Den här metoden är väldigt elegant, men kanske lite svårare att förstå och få rätt. Notera i synnerhet att kravet vi lägger till är $x + 1 \geq 0$, INTE $x + 3 \geq 0$. I ekvationen $\sqrt{x+3} = x + 1$ framgår båda utsagorna: $x + 3 \geq 0$ ty $D_{\sqrt{}} = [0, \infty[$ och $x + 1 \geq 0$ ty $V_{\sqrt{}} = [0, \infty[$. I $x + 3 = (x+1)^2$ framgår fortfarande att $x + 3 \geq 0$ ty $(x+1)^2 \geq 0$; däremot framgår inte att $x + 1 \geq 0$.

Hur vet vi då att denna tillagda information är tillräcklig för att åstadkomma implikationen åt vänster? Jo, om $x + 3 = (x+1)^2$ så är det, som sagt, klart att $x + 3 \geq 0$, så vi har en likhet mellan två icke-negativa tal. Alltså måste deras kvadratrötter vara lika:

$$\sqrt{x+3} = \sqrt{(x+1)^2} = |x+1|.$$

Om vi nu lägger till pusselbiten $x + 1 \geq 0$ får vi därför

$$\sqrt{x+3} = \sqrt{(x+1)^2} = |x+1| = x + 1,$$

precis som önskat.

En alternativ förklaring, som också synliggör kopplingen till den allmänna observationen vid (2) ovan, är som följer:

$$\sqrt{x+3} = x+1 \qquad \Longrightarrow \qquad \left(\sqrt{x+3}\right)^2 = (x+1)^2.$$

Vi får ekvivalens om vi också, till höger, lägger till upplysningen att $\sqrt{x+3}$ och $x+1$ har samma tecken. Eftersom $V_{\sqrt{}} = [0, \infty[$ är denna pusselbit precis utsagan $x+1 \geq 0$. Tycker man att den här är invecklat går det naturligtvis bra att skippa ekvivalensen och sedan pröva rötterna i ursprungsekvationen.

Exempel

Lös ekvationen $\sqrt{x+1} = \sqrt{x+3} - 1$.

Lösning:

$$\begin{aligned} &\sqrt{x+1} = \sqrt{x+3} - 1 \quad \Longrightarrow \quad x+1 = \left(\sqrt{x+3} - 1\right)^2 \quad \Longleftrightarrow \\ &\Longleftrightarrow \quad x+1 = (x+3) - 2\sqrt{x+3} + 1 \quad \Longleftrightarrow \\ &\Longleftrightarrow \quad \sqrt{x+3} = \frac{3}{2} \quad \Longrightarrow \quad x+3 = \frac{9}{4} \quad \Longleftrightarrow \\ &\Longleftrightarrow \quad x = -\frac{3}{4}. \end{aligned}$$

Med $x = -3/4$ lyder ursprungsekvationens respektive led

$$VL = \sqrt{x+1} = \sqrt{\left(-\frac{3}{4}\right) + 1} = \sqrt{\frac{1}{4}} = \frac{1}{2}$$

$$HL = \sqrt{x+3} - 1 = \sqrt{\left(-\frac{3}{4}\right) + 3} - 1 = \sqrt{\frac{9}{4}} - 1 = \frac{3}{2} - 1 = \frac{1}{2}$$

så $x = -3/4$ är en rot till ursprungsekvationen.

Svar: $\sqrt{x+1} = \sqrt{x+3} - 1 \Leftrightarrow x = -3/4$.

Exempel

Lös ekvationen $\sqrt{x+1} = \sqrt{x+3} + 1$.

Lösning: Vi har

$$\begin{aligned} &\sqrt{x+1} = \sqrt{x+3} + 1 \quad \Longrightarrow \quad x+1 = \left(\sqrt{x+3} + 1\right)^2 \quad \Longleftrightarrow \\ &\Longleftrightarrow \quad x+1 = (x+3) + 2\sqrt{x+3} + 1 \quad \Longleftrightarrow \\ &\Longleftrightarrow \quad \sqrt{x+3} = -\frac{3}{2} \end{aligned}$$

som inte är sant för något $x \in \mathbb{R}$ (eftersom $V_{\sqrt{}} = [0, \infty[$).

Svar: Ekvationen saknar lösningar.

1.4.4 Ekvationer med beloppstecken

Om en ekvation innehåller beloppstecken brukar man dela upp $\mathbb{R}$ i disjunkta intervall, inom vilka man känner till tecknet hos varje beloppsfunktions argument. I varje sådant intervall kan man därför enkelt bli av med beloppstecknen, och därmed finna de rötter som finns i intervallet. Vi illustrerar förfarandet med ett exempel.

Exempel

Lös ekvationen $|x+1| + |x-2| = 2|x+3|$.

Lösning: Vi har

$$|x+1| = \begin{cases} x+1 & \text{om } x \geq -1 \\ -x-1 & \text{om } x < -1, \end{cases}$$

$$|x-2| = \begin{cases} x-2 & \text{om } x \geq 2 \\ -x+2 & \text{om } x < 2 \end{cases}$$

samt

$$|x+3| = \begin{cases} x+3 & \text{om } x \geq -3 \\ -x-3 & \text{om } x < -3. \end{cases}$$

Därför, om $x < -3$ lyder ekvationen

$$-x-1+(-x+2) = 2(-x-3) \iff -x-1-x+2 = -2x-6 \iff 1 = -6$$

vilket är absurt. Det finns alltså inget $x < -3$ som löser ekvationen. I $[-3,-1[$ lyder ekvationen

$$-x-1+(-x+2) = 2(x+3) \iff -x-1-x+2 = 2x+6 \iff 4x = -5 \iff x = -\frac{5}{4}$$

som är en lösning (notera att $-\frac{5}{4} \in [-3,-1[$). Om $x \in [-1,2[$ lyder ekvationen

$$x+1+(-x+2) = 2(x+3) \iff x+1-x+2 = 2x+6 \iff 2x = -3 \iff x = -\frac{3}{2} (= -1.5)$$

vilket är absurt eftersom $x \in [-1,2[$. Slutligen, om $x \geq 2$ så lyder ekvationen

$$x+1+x-2 = 2(x+3) \iff x+1+x-2 = 2x+6 \iff -1 = 6$$

vilket också är absurt, så inga $x \geq 2$ löser ekvationen.

Svar: $|x+1| + |x-2| = 2|x+3| \Leftrightarrow x = -5/4$.

Lägg märke till hur de falldefinierade beloppsuttrycken ändrar karaktär vid $x = -3, -1$ och 2, men samtliga ges av *samma* uttryck mellan dessa punkter. Vi löser alltså ekvationen i intervallen $]-\infty,-3[$, $[-3,-1[$, $[-1,2[$ samt $[2,\infty[$ var för sig. Vi får då samtliga lösningar, ty varje $x \in \mathbb{R}$ finns i något av dessa fyra intervall.

1.4.5 Olikheter

Olikhetssymbolerna $<, \leq, >$ och $\geq$ kan använda mellan reella tal på ungefär samma sätt som likhetstecknet $=$ och dess negation $\neq$. Lösning av olikheter fungerar på liknande sätt, och vi kan tala om en olikhets lösningsmängd. Till exempel har vi

$$2x + 4 > x + 10 \qquad \Longleftrightarrow \qquad x > 6$$

så att olikhetens lösningsmängd är $]6, \infty[$ – alla tal i denna mängd, men inga andra, gör så att olikheten blir en sann utsaga. Som ytterligare exempel inser vi att de två olikheterna $(x+6)^2 \geq 0$ och $x^2 < -5$ har $\mathbb{R}$ respektive $\emptyset$ som sina lösningsmängder. Den kanske största skillnaden mellan mekanisk manipulation av likheter och av olikheter är att multiplikation med negativt tal vänder på olikhetstecknet. Till exempel är

$$x > 2 \qquad \Longleftrightarrow \qquad -x < -2$$

(multiplikation av båda leden med -1). Det är därför i allmänhet mycket olämpligt att multiplicera olikheter med okända tal, eftersom man då inte vet om tecknet skall vändas eller inte.

Exempel

Lös olikheten $x^2 - 5x \geq \frac{24}{x+2} - 12$.

Lösning:

$$\begin{aligned}
& x^2 - 5x \geq \frac{24}{x+2} - 12 \quad \Longleftrightarrow \quad x^2 - 5x - \frac{24}{x+2} + 12 \geq 0 \quad \Longleftrightarrow \\
\Longleftrightarrow \quad & \frac{(x^2-5x)(x+2)}{x+2} - \frac{24}{x+2} + \frac{12(x+2)}{x+2} \geq 0 \quad \Longleftrightarrow \\
\Longleftrightarrow \quad & \frac{(x^2-5x)(x+2) - 24 + 12(x+2)}{x+2} \geq 0 \quad \Longleftrightarrow \\
\Longleftrightarrow \quad & \frac{x^3 - 3x^2 + 2x}{x+2} \geq 0 \quad \Longleftrightarrow \\
\Longleftrightarrow \quad & \frac{x(x-1)(x-2)}{x+2} \geq 0 \quad \Longleftrightarrow \\
\Longleftrightarrow \quad & x \in \,]-\infty, -2[\, \cup [0, 1] \cup [2, \infty[
\end{aligned}$$

där sista steget framgår av teckentabellen

		-2		0		1		2	
x	$-$		$-$	0	$+$		$+$		$+$
$x-1$	$-$		$-$		$-$	0	$+$		$+$
$x-2$	$-$		$-$		$-$		$-$	0	$+$
$x+2$	$-$	0	$+$		$+$		$+$		$+$
$\frac{x(x-1)(x-2)}{x+2}$	$+$	$\nexists$	$-$	0	$+$	0	$-$	0	$+$

.

Svar: $x^2 - 5x \geq \frac{24}{x+2} - 12 \quad \Longleftrightarrow \quad x \in \,]-\infty, -2[\, \cup [0, 1] \cup [2, \infty[$.

Är man allergisk mot mängdlärans symboler kan man förstås också svara

Svar: $x^2 - 5x \geq \frac{24}{x+2} - 12 \quad \Longleftrightarrow \quad x < -2$ eller $0 \leq x \leq 1$ eller $x \geq 2.$

Hade det rört sig om vanlig ekvationslösning, hade det varit frestande att multiplicerade båda leden med det nollskilda talet $x + 2$ som första steg i lösningen. Nu vet vi emellertid inte vilket tecken $x + 2$ har, så den metoden är inte gångbar med tanke på att vi arbetar med en olikhet. I stället flyttar vi över alla termer till samma sida, skriver dem på gemensamt bråkstreck och faktoriserar täljare och nämnare. Det blir då en enkel uppgift att bestämma tecknet hos kvoten – i praktiken behöver vi bara ”räkna minustecken” bland faktorerna: udda antal blir minus, jämnt antal blir plus.[11]

Om (i en liknande uppgift) en av faktorerna i det resulterande bråket alltid är positiv, t.ex. $x^2 + 2$, kan man förstås helt avstå från att ta med den i teckentabellen. [Däremot kan inte *icke-negativa* termer, såsom x^2, tas bort, eftersom de kan bli noll, om de inte är positiva.]

Får man en faktor -1 i bråket, kan man antingen (1) ta bort den och vända olikhetstecknet (man ”gångrar” olikheten med ett negativt tal), (2) ta med den i teckentabellen (-1 har alltid tecknet $-$) eller (3) ändra sin regel så att *udda* antal minustecken i en kolumn blir till ett plustecken i summaraden och tvärtom. Alla dessa tre varianter är fullt godtagbara, men man kan misstänka att den sistnämnda medför störst risk för slarvfel.

[11] Om inte någon faktor är noll, i vilket fall antingen hela bråket har värdet noll eller är odefinierat, beroende på om faktorn finns i täljaren eller i nämnaren.

1.5 Funktioner, igen

Det här avsnittet är en fortsättning på avsnittet om funktioner (s. 57). Nu när vi bekantat oss med ekvationer och olikheter kan vi nämligen ge lite mer intressanta exempel på bestämning av naturliga definitionsmängder och inverser.

Exempel

Bestäm definitionsmängden för funktionen

$$f: x \mapsto \frac{1}{x} + \sqrt{x^2 - x^3}.$$

Lösning: Det är två saker som kan gå fel: dels måste $x \neq 0$ för att $1/x$ skall vara definierat, dels måste $x^2 - x^3 \geq 0$ eftersom $D_{\sqrt{}} = [0, \infty[$. På grund av att

$$x^2 - x^3 = x^2(1 - x) \geq 0 \qquad \Longleftrightarrow \qquad x \leq 1$$

betyder detta att

$$D_f =]-\infty, 1] \setminus \{0\}.$$

Exempel

Bestäm definitionsmängden för funktionen

$$g: x \mapsto \ln\left(\frac{x-1}{x+1}\right).$$

Lösning: Eftersom $D_{\ln} = \mathbb{R}^+ (=]0, \infty[)$ måste

$$\frac{x-1}{x+1} > 0$$

vilket inträffar precis då $x \notin [-1, 1]$. Sålunda är

$$D_g = [-1,1]^C.$$

Exempel

Bestäm definitionsmängden för funktionen

$$h: x \mapsto \ln\left(\frac{1-x}{1+x}\right)$$

och, om möjligt, den inversa funktionen.

Lösning: Här måste

$$\frac{1-x}{1+x} > 0$$

vilket inträffar precis då $x \in \,]-1,1[$, så $D_h = \,]-1,1[$. Vi försöker nu bestämma den inversa funktionen (om den finns). Låt $y \in V_h$ vara godtycklig. Då finns det ett $x \in D_h$ sådant att $y = h(x)$. Eftersom

$$\begin{aligned} y = \ln\frac{1-x}{1+x} \quad &\Longrightarrow \quad e^y = \frac{1-x}{1+x} \quad \Longrightarrow \\ &\Longrightarrow \quad (1+x)e^y = 1-x \quad \Longrightarrow \\ &\Longrightarrow \quad e^y + xe^y = 1-x \quad \Longrightarrow \\ &\Longrightarrow \quad x(e^y+1) = 1-e^y \quad \Longrightarrow \\ &\Longrightarrow \quad x = \frac{1-e^y}{1+e^y} \end{aligned}$$

finns det bara *ett* $x \in D_h$ som avbildas på y (nämligen $(1-e^y)/(1+e^y)$); h är alltså injektiv, och har en invers. Vi har dessutom funnit en formel för den.

Svar: $D_h = \,]-1,1[$ och $h^{-1}(x) = \frac{1-e^x}{1+e^x}$.

Exempel

Bestäm definitionsmängden för funktionen

$$\tilde{h}: x \mapsto \ln\left(\frac{1-x^2}{1+x^2}\right)$$

och, om möjligt, den inversa funktionen.

Lösning: Eftersom

$$\frac{1-x^2}{1+x^2} = \frac{(1-x)(1+x)}{1+x^2} > 0 \quad \Longleftrightarrow \quad x \in \,]-1,1[$$

har vi $D_{\tilde{h}} = \,]-1,1[$. Det är emellertid uppenbart att $\tilde{h}$ inte är injektiv, ty $\tilde{h}$ är en jämn funktion definierad kring origo: $\tilde{h}(x) = \tilde{h}(-x)$ för varje $x \in D_{\tilde{h}}$. Till exempel är $\tilde{h}\left(\frac{1}{2}\right) = \tilde{h}\left(-\frac{1}{2}\right)$.

Svar: $D_{\tilde{h}} = \,]-1,1[$. $\tilde{h}$ saknar invers.

Ett mycket vanligt fel

I båda de två senaste fallen har vi definitionsmängden $D_h = \,]-1,1[$. Förhållandevis ofta redovisas sådana fakta på följande sätt av studenter:

$$D_h = -1 < x < 1 \qquad \text{(absurt!!)}$$

vilket förstås är mycket suspekt. Vänsterledet i ekvationen, D_h, är ju en *mängd*, medan högerledet, $-1 < x < 1$ är ett *påstående*. Det rör sig alltså om två olika *typer* av objekt. Därför kan de inte vara lika. I det korrekta skrivsättet, $D_h = \,]-1,1[$, är ju D_h *mängden* av alla tal man kan stoppa in i funktionen, och $]-1,1[$ är *mängden* av alla reella tal mellan -1 och 1 (exkl. gränserna). Dessa två mängder är identiska, så det är mycket naturligt (nästintill oundvikligt!) att sätta likhetstecken mellan dem.

I exemplen $D_f =]-\infty, 1] \setminus \{0\}$ och $D_g = [-1,1]^C$ fordras viss kunskap om elementär mängdlära för att kunna skriva på det mest naturliga sättet. Men även utan sådan kunskap kan man skriva på ett *korrekt* sätt, genom att helt enkelt skriva

$$D_f = \{x \in \mathbb{R}: \quad x \leq 1 \text{ och } x \neq 0\}$$

respektive

$$D_g = \{x \in \mathbb{R}: \quad x < -1 \text{ eller } x > 1\}.$$

Om man är totalt allergisk mot allt som har med mängdlära att göra, är det minst dåliga sättet att skriva något i stil med

$$\begin{aligned} &D_h: \quad -1 < x < 1, \\ &D_f: \ x \leq 1 \text{ och } x \neq 0, \\ &D_g: \ x < -1 \text{ eller } x > 1 \end{aligned}$$

(i våra tre respektive exempel) även om det inte är särskilt snyggt eller tydligt – det är i varje fall bättre än direkta tokigheter som ”$D_h = -1 < x < 1$”.

Ett till vanligt fel

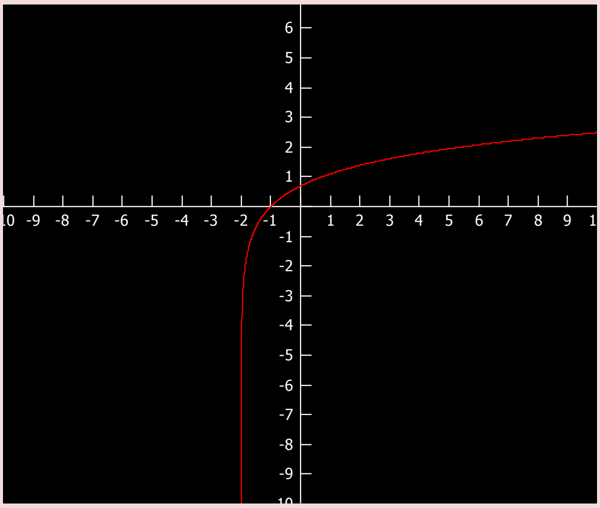
Studenter blandar ofta ihop det man stoppar in i en funktion med det man får ut ur den.

Säg att vi skall bestämma definitionsmängden till funktionen $f: x \mapsto \ln(x+2)$. Vi vet att definitionsmängden $D_{\ln} = \mathbb{R}^+$, d.v.s. vi kan bara stoppa in positiva tal i ln-funktionen. Alltså måste $x + 2 > 0$, så $x > -2$. Vi har därför $D_f =]-2, \infty[$, d.v.s. vi kan bara stoppa in tal större än -2 i f.

När man rättar skrivningar ser man däremot ofta följande:

”Eftersom $\ln(x+2) > 0$ måste $x > -2$.”

vilket är mycket konstigt. Först är det rent felaktigt att säga att $\ln(x+2) > 0$, eftersom $\ln(x+2)$ närmar sig minus oändligheten när x närmar sig -2. I grafen nedan syns klart och tydligt att $\ln(x+2)$ antar negativa värden.



Felet studenten begår är att han/hon blandar ihop det man stoppar in i funktionen med det man får ut ur den. I uppgiften frågas efter vad man kan stoppa in för tal i f, och vi vet att man bara kan stoppa in positiva tal i ln, så $x + 2$ måste vara positiv. Det studenten menar är alltså

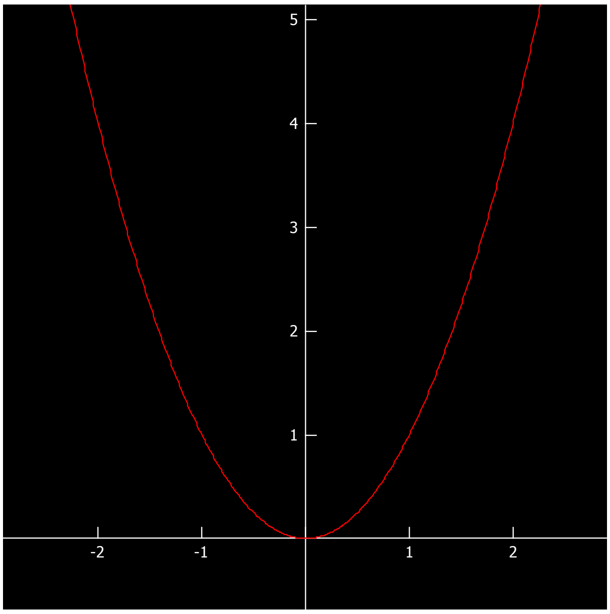
”Eftersom vi måste ha $x + 2 > 0$ så måste $x > -2$”.

1.5.1 Restriktioner

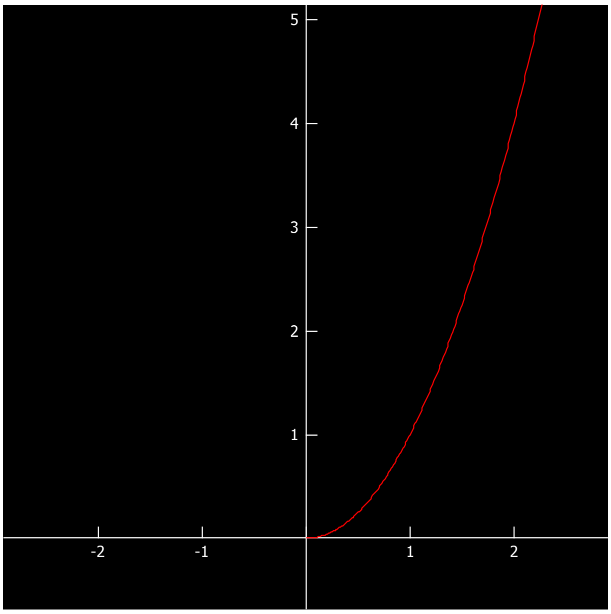
Om $f: D_f \to Y$ är en funktion och $S \subset D_f$ är en *äkta* delmängd av D_f, kan man införa funktionen $\tilde{f}: S \to Y$ definierad av $\tilde{f}(x) \coloneqq f(x)$ för varje $x \in S$. Funktionen $\tilde{f}$ är alltså *identisk* med funktionen f, så när som på att $\tilde{f}$ är definierad på en (äkta) *delmängd* av D_f. Trots det är förstås $f \neq \tilde{f}$.

Funktionen $\tilde{f}$ säges vara en *restriktion* till f. Det här handlar egentligen bara om formalia – man ignorerar helt enkelt alla element utanför S när man betraktar funktionen $\tilde{f}$.

Som exempel, betrakta kvadreringsfunktionen $f: \mathbb{R} \to \mathbb{R}, x \mapsto x^2$, med den välkända parabeln som graf. Funktionen $g: [0, \infty[\to \mathbb{R}, x \mapsto x^2$ är då restriktionen av kvadreringsfunktionen till de icke-negativa talen. Nedan visas graferna för f och dess restriktion g.



$y = f(x)$



$y = g(x)$

Man kan notera att g är strängt växande – och därmed injektiv. f saknar båda dessa egenskaper. I själva verket kan vi nu formulera den mycket sympatiska insikten

Observation. Kvadratrotsfunktionen $x \mapsto \sqrt{x}$ är inversen till restriktionen av kvadreringsfunktionen $x \mapsto x^2$ till de icke-negativa talen $[0, \infty[$.

Vi har ju också, som sig bör, identiteterna $\sqrt{x^2} = \sqrt{x}^2 = x$ för alla $x \geq 0$.

De följande tre exemplen bygger vidare på samma idé. Rita graferna för f i de tre fallen!

Exempel

Bestäm, om möjligt, inversen till funktionen $f: \mathbb{R} \to \mathbb{R}$ definierad av

$$f(x) \coloneqq 1 + 4x^2.$$

Lösning: Eftersom f är jämn (och definierad på hela $\mathbb{R}$) är den inte injektiv. Till exempel är $f(-1) = f(1) = 5$. f har därför ingen invers.

Exempel

Bestäm, om möjligt, inversen till funktionen $f: [0, \infty[\to \mathbb{R}$ definierad av

$$f(x) \coloneqq 1 + 4x^2.$$

Lösning: Välj ett godtyckligt $y \in V_f$ och låt $y = f(x)$ för något $x \in D_f$. Då är

$$y = 1 + 4x^2 \implies y - 1 = 4x^2 \implies x^2 = \frac{1}{4}(y-1) \implies$$
$$\overset{*}{\implies} x = \frac{1}{2}\sqrt{y-1}$$

där (*) beror på att $x \geq 0$ eftersom $x \in D_f$. Alltså är $f^{-1}(x) = \frac{1}{2}\sqrt{x-1}$.

Exempel

Bestäm, om möjligt, inversen till funktionen $f:]-\infty, 0] \to \mathbb{R}$ definierad av

$$f(x) \coloneqq 1 + 4x^2.$$

Lösning: Välj ett godtyckligt $y \in V_f$ och låt $y = f(x)$ för något $x \in D_f$. Då är

$$y = 1 + 4x^2 \implies y - 1 = 4x^2 \implies x^2 = \frac{1}{4}(y-1) \implies$$
$$\overset{*}{\implies} x = -\frac{1}{2}\sqrt{y-1}$$

där (*) beror på att $x \leq 0$ eftersom $x \in D_f$. Alltså är $f^{-1}(x) = -\frac{1}{2}\sqrt{x-1}$.

1.6 Trigonometri

1.6.1 Trigonometriska funktioner

I det här avsnittet ger vi en introduktion till de trigonometriska funktionerna.

1.6.1.1 Definitionen av sinus och cosinus

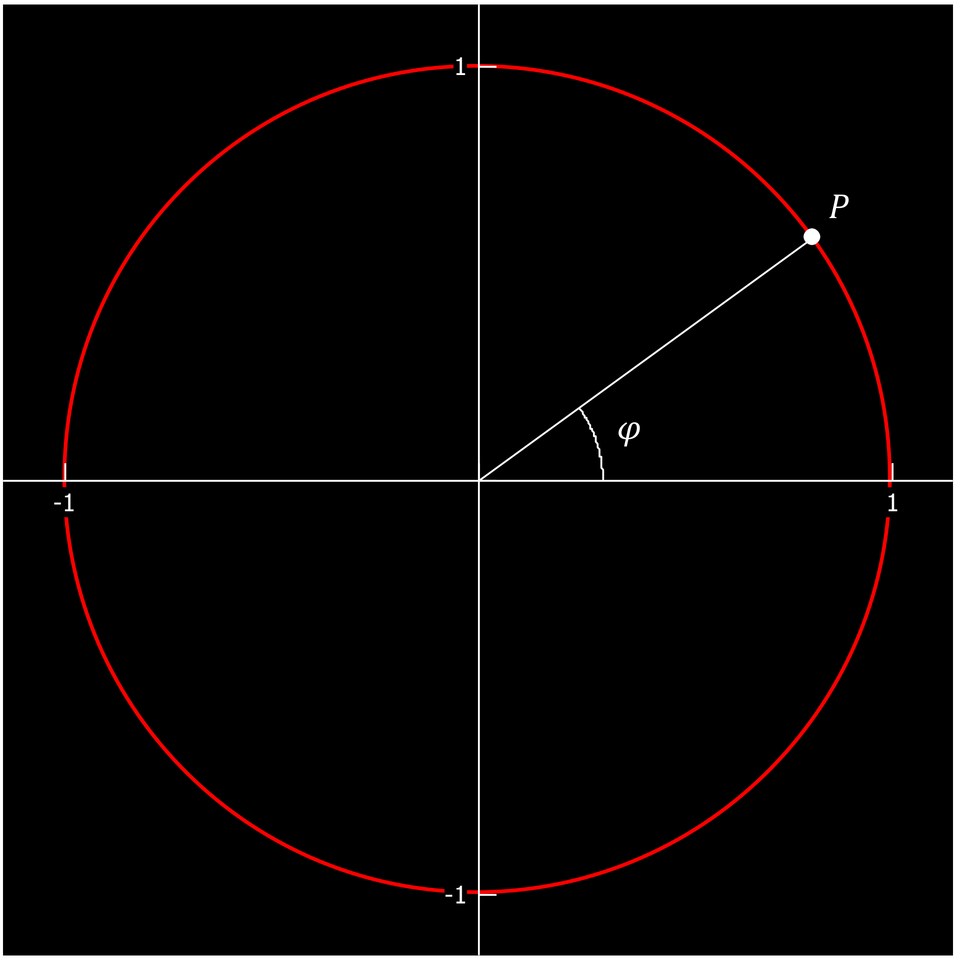
Vi börjar med att definiera vad som menas med ”sinus” och ”cosinus”. Sinus och cosinus är båda funktioner $\mathbb{R} \to [-1, 1]$. För att på ett smidigt sätt definiera $\sin\varphi$ och $\cos\varphi$ inför vi först en hjälpfunktion $\mathcal{P}: \mathbb{R} \to S^1$ som tar in ett reellt tal $\varphi \in \mathbb{R}$ och ger ifrån sig en punkt $P = (x, y) \in S^1$ på enhetscirkeln

$$S^1 = \{(x, y) \in \mathbb{R}^2: \quad x^2 + y^2 = 1\}.$$

Välj ett $\varphi \in \mathbb{R}$ godtyckligt. Om $\varphi \geq 0$ är $\mathcal{P}(\varphi)$ den punkt på enhetscirkeln som du kommer till om du börjar vid $(1, 0)$ och går sträckan φ moturs längs cirkeln. Om $\varphi < 0$ går du i stället medurs sträckan $|\varphi| = -\varphi$. Eftersom enhetscirkeln har omkrets 2π följer det att funktionen $\mathcal{P}$ är *periodisk* med *perioden* 2π, d.v.s.

$$\mathcal{P}(\varphi + n \cdot 2\pi) = \mathcal{P}(\varphi), \qquad \forall n \in \mathbb{Z}.$$

Emedan enhetscirkeln har radie 1 kan vi tolka $\mathcal{P}(\varphi)$ på ett alternativt sätt: Låt ℓ vara den stråle (halvlinje) som börjar i origo och bildar **vinkeln** φ uppåt (om $\varphi \geq 0$) eller vinkeln $-\varphi$ nedåt (om $\varphi < 0$) från positiva φ-axeln. Då är $\{\mathcal{P}(\varphi)\} = \ell \cap S^1$. I bilden nedan visas $P \coloneqq \mathcal{P}(\varphi)$.



Låt nu $\pi_x, \pi_y: \mathbb{R}^2 \to \mathbb{R}$ vara *projektionsfunktionerna* i planet, d.v.s. $\pi_x(x, y) = x$ och $\pi_y(x, y) = y$. π_x är alltså den funktion som tar in en punkt i planet och ger ifrån sig punktens x-koordinat, och motsvarande för π_y. Då definierar vi funktionerna *cosinus* respektive *sinus* som

$$\cos \coloneqq \pi_x \circ \mathcal{P}$$
$$\sin \coloneqq \pi_y \circ \mathcal{P}.$$

$\cos(\varphi)$ är med andra ord x-koordiaten för punkten $\mathcal{P}(\varphi)$ medan $\sin(\varphi)$ är y-koordinaten. Eftersom $\mathcal{P}$ är periodisk med perioden 2π gäller detsamma för sinus och cosinus:

$$\sin(\varphi + n \cdot 2\pi) = \sin \varphi$$
$$\cos(\varphi + n \cdot 2\pi) = \cos \varphi$$

för varje $n \in \mathbb{Z}$. Vi inser också direkt att

$$\sin 0 = \sin \pi = 0, \qquad \sin \frac{\pi}{2} = 1, \qquad \sin\left(-\frac{\pi}{2}\right) = -1.$$

De två första likheterna är specialfall av

$$\sin n\pi = 0, \qquad \forall n \in \mathbb{Z}.$$

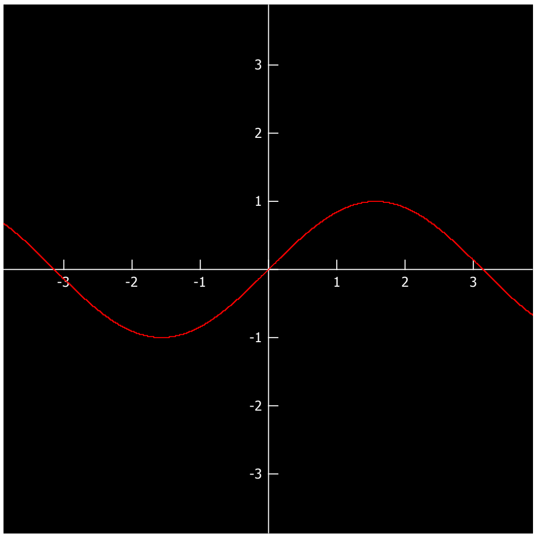
Analogt har vi

$$\cos 0 = 1, \qquad \cos \pi = -1, \qquad \cos\left(\pm\frac{\pi}{2}\right) = 0$$

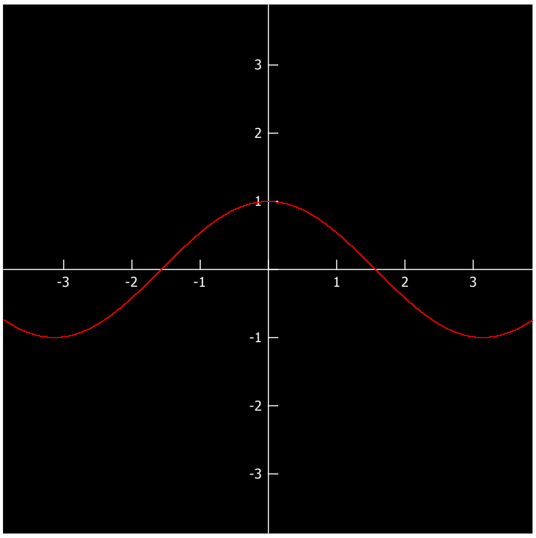
där de två första likheterna är specialfall av

$$\cos n\pi = (-1)^n, \qquad \forall n \in \mathbb{Z}.$$

Graferna till sinus och cosinus är troligtvis redan bekanta, inte bara från sidan 67, utan också från tidigare matematikstudier (på gymnasienivå):



$y = \sin x$



$y = \cos x$

Förutom ovannämnda fullkomligt uppenbara funktionsvärden måste man memorera följande *standardvärden*:

v	$\pi/6$	$\pi/4$	$\pi/3$
$\sin v$	$\frac{1}{2}$	$\frac{1}{\sqrt{2}}$	$\frac{\sqrt{3}}{2}$
$\cos v$	$\frac{\sqrt{3}}{2}$	$\frac{1}{\sqrt{2}}$	$\frac{1}{2}$

Värdena för $\pi/4$ är mycket lätta att inse. Strålen ℓ från origo som bildar vinkeln $\pi/4$ från positiva x-axeln är ju en bisektris till den räta vinkeln mellan de positiva koordinataxlarna; strålen har sålunda ekvationen $y = x$. Eftersom $P = (x, y) \in \ell \cap S^1$ har vi

$$\begin{cases} x = y \\ x^2 + y^2 = 1 \end{cases} \implies x^2 + x^2 = 1 \implies 2x^2 = 1 \implies$$
$$\implies x^2 = \frac{1}{2} \implies x = \frac{1}{\sqrt{2}}$$

där sista implikationen kommer från att $x > 0$. Med bara lite mer arbete kan man resonera sig fram till de två andra kolumnerna i tabellen.

Notera att $V_{\sin} = V_{\cos} = [-1, 1]$. Detta innebär speciellt att både sinus och cosinus är begränsade funktioner.

Övning

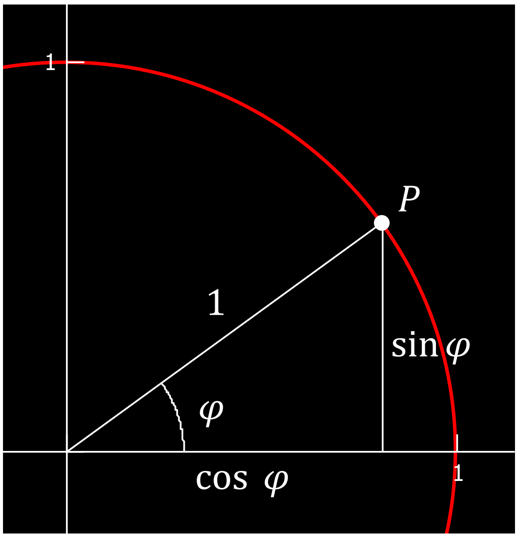
1. I den här övningen härleder vi de återstående fyra standardvärdena.
   (1) Rita enhetscirkeln.
   (2) Markera de två strålarna från origo vilka bildar vinkeln $\frac{\pi}{6}$ mot den positiva x-axeln.
   (3) Dessa två strålar skär enhetscirkeln i två punkter, P uppe ($y > 0$) och P' nere ($y < 0$); förbind dessa med en rät linje.
   (4) Du har nu erhållit en triangel. Markera triangelns samtliga vinklar, och notera att triangeln är liksidig. Markera triangelns samtliga sidlängder.
   (5) Dra slutsaten att $\sin\frac{\pi}{6} = \frac{1}{2}$ och (därefter) att $\cos\frac{\pi}{6} = \frac{\sqrt{3}}{2}$.
   (6) Dra nu den stråle från origo i övre halvplanet $y > 0$ vilken bildar vinkeln $\frac{\pi}{3}$ mot positiva x-axeln. Låt Q vara skärningspunkten med enhetscirkeln.
   (7) Dra ett horisontellt linjestycke mellan Q och y-axeln.
   (8) En ny triangel uppkommer. Notera att dess spetsvinkel är $\frac{\pi}{6}$ och att triangeln är kongruent med (=har exakt samma form och storlek som) den första triangeln.
   (9) Dra slutsatsen att $\sin\frac{\pi}{3} = \frac{\sqrt{3}}{2}$ och $\cos\frac{\pi}{3} = \frac{1}{2}$.

1.6.1.2 Identiteter med sinus och cosinus

Eftersom $(\cos x, \sin x) \in S^1$ har vi

$$\cos^2 x + \sin^2 x = 1, \qquad \forall x \in \mathbb{R}.$$

Denna identitet[12] brukar kallas för ”trigonometriska ettan”. Man kan förstås också ”härleda” den genom att applicera Pythagoras sats i enhetscirkeln.[13] Om t.ex. $x \in]0, \pi/2[$ har vi ju följande figur:



Liknande trianglar[14] kan skapas för alla φ (förutom heltalsmultipler av $\pi/2$) med samma resultat. Utöver trigonometriska ettan är följande identiteter uppenbara:

$$\sin(x \pm \pi) = -\sin x\,, \qquad \cos(x \pm \pi) = -\cos x.$$

Det är heller inte alltför svårt att inse att

$$\sin\left(\frac{\pi}{2} - x\right) = \cos x\,, \qquad \cos\left(\frac{\pi}{2} - x\right) = \sin x.$$

Dessa fyra identiteter är också specialfall av de mer allmänna additionsformlerna som vi kommer till inom kort. Vidare är cos jämn medan sin är udda:

$$\cos(-x) = \cos x\,, \qquad \sin(-x) = -\sin x.$$

Vi har också

$$\cos(\pi - x) = -\cos x\,, \qquad \sin(\pi - x) = \sin x.$$

De fyra senaste identiteterna ser man direkt från definitionen i enhetscirkeln. Utöver trigonometriska ettan är det fyra identiteter (som inte är uppenbara) som man måste memorera. Dessa är *additions-formlerna* för sinus och cosinus, samt *formlerna för dubbla vinkeln*. Nedan listas alla fem identiteter man måste memorera:

1. $\sin^2 x + \cos^2 x = 1$ (trigonometriska ettan)
2. $\sin(u \pm v) = \sin u \cos v \pm \cos u \sin v$ (additionsformeln för sinus)

[12] En ”identitet” är en ekvation som är en sann utsaga för alla värden på de variabler som ingår i den.
[13] Hur inser man att enhetscirkelns ekvation är just $x^2 + y^2 = 1$? Jo, med Pythagoras sats! De båda sätten att ”härleda” trigonometriska ettan är alltså väsentligen identiska.
[14] Man får bara se upp med tecknen: $\sin x$ är ju negativt [för vinklar x sådana att P ligger] i undre halvplanet och $\cos x$ är negativt [för vinklar x sådana att P ligger] i vänstra halvplanet. Om man vill rita trianglar i andra kvadranter än den första gäller alltså att deras sidlängder är $|\sin x|$ och $|\cos x|$, med beloppstecknen. Men detta påverkar förstås inte kvadraterna $\sin^2 x$ och $\cos^2 x$.

3. $\cos(u \pm v) = \cos u \cos v \mp \sin u \sin v$ (additionsformeln för cosinus)
4. $\sin 2x = 2 \sin x \cos x$ (sinus för dubbla vinkeln)
5. $\cos 2x = \cos^2 x - \sin^2 x$ (cosinus för dubbla vinkeln).

Dubbla vinkeln-formlerna följer förstås direkt från additionsformlerna[15], men de bör ändå memoreras, så att man känner igen högerleden när man ser dem. För att visa de fyra formlerna (2) och (3) räcker det med att visa en av dem. De återstående tre följer sedan direkt av denna och de uppenbara identiteterna vi listade tidigare.

Exempel

Visa att $\sin\left(x + \frac{\pi}{2}\right) = \cos x$ och $\cos\left(x - \frac{\pi}{2}\right) = \sin x$.

Lösning: Vi har

$$\sin\left(x + \frac{\pi}{2}\right) = \sin x \underbrace{\cos\frac{\pi}{2}}_{=0} + \cos x \underbrace{\sin\frac{\pi}{2}}_{=1} = \cos x$$

och

$$\cos\left(x - \frac{\pi}{2}\right) = \cos x \underbrace{\cos\frac{\pi}{2}}_{=0} + \sin x \underbrace{\sin\frac{\pi}{2}}_{=1} = \sin x.$$

I praktiska sammanhang, speciellt integration (som vi kommer till om några kapitel) är det **mycket viktigt** att kunna skriva om en *kvadrat* av en sinus- eller cosinusterm till något enklare. Detta låter sig göras med hjälp av cosinus för dubbla vinkeln.

Säg att vi vill skriva om $\sin^2 x$. Vi noterar då att

$$\cos 2x = \cos^2 x - \sin^2 x = (1 - \sin^2 x) - \sin^2 x = 1 - 2\sin^2 x$$

som ger

$$\sin^2 x = \frac{1}{2} - \frac{1}{2}\cos 2x.$$

Denna omskrivning är mycket viktig! Glöm inte bort den‼

Övning

1. Skriv om $\cos^2 x$ på liknande sätt.

[15] Additionsformlerna är egentligen *fyra* olika identiteter, en för varje val av tecknen: man kan antingen välja översta eller understa tecknet i symbolerna $\pm$ och $\mp$, men man måste göra *samma* val (*översta* eller *understa*) i båda leden. I sinus-fallen blir det alltså *samma* tecken i båda leden, medan cosinusidentiteterna har *olika* tecken i de två leden.

1.6.1.3 Tangens och cotangens (och sekant och cosekant)

Funktionen *tangens* definieras av

$$\tan x := \frac{\sin x}{\cos x}$$

och är därför en udda funktion med definitionsmängden

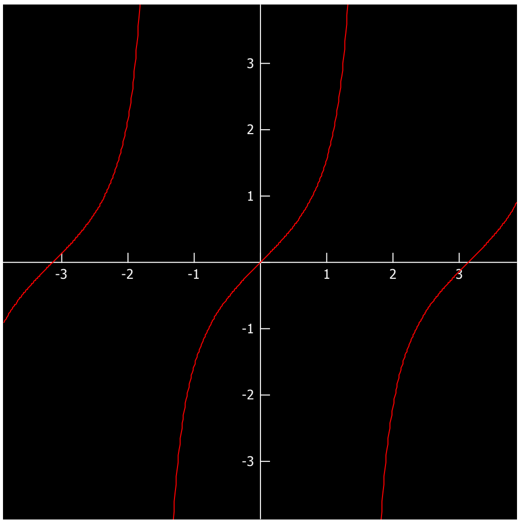
$$D_{\tan} = \{x \in \mathbb{R}: \quad \cos x \neq 0\} = \left\{x \in \mathbb{R}: \quad x \neq \frac{\pi}{2} + n\pi \ \text{ för något } \ n \in \mathbb{Z}\right\}.$$

Det är klart att $\tan\frac{\pi}{4} = \tan\left(-\frac{3\pi}{4}\right) = 1$, eftersom sinus och cosinus är lika här (vi är på linjen $x = y$). Vidare är $\tan\left(-\frac{\pi}{4}\right) = \tan\frac{3\pi}{4} = -1$, eftersom sinus och cosinus är lika till beloppet men har olika tecken här (vi är på linjen $x = -y$). Vidare är det klart att $\tan n\pi = 0$ eftersom $\sin n\pi = 0$ ($n \in \mathbb{Z}$).

Eftersom sin och cos båda har perioden 2π är det självklart att även tan har perioden 2π. Men eftersom

$$\tan(x + n\pi) = \frac{\sin(x + n\pi)}{\cos(x + n\pi)} = \frac{\sin x \cos n\pi + \cos x \sin n\pi}{\cos x \cos n\pi - \sin x \sin n\pi} = \frac{\sin x \cos n\pi}{\cos x \cos n\pi} = \frac{\sin x}{\cos x} = \tan x$$

för varje $n \in \mathbb{Z}$ ser vi att tan till och med är periodisk med perioden π. Grafen till tangens visas nedan. Notera att tan är obegränsad och går mot oändligheten till beloppet då cos går mot noll.



$y = \tan x$

Vi kan nu utöka vår tabell med standardvärden:

v	$\pi/6$	$\pi/4$	$\pi/3$
$\sin v$	$\frac{1}{2}$	$\frac{1}{\sqrt{2}}$	$\frac{\sqrt{3}}{2}$
$\cos v$	$\frac{\sqrt{3}}{2}$	$\frac{1}{\sqrt{2}}$	$\frac{1}{2}$
$\tan v$	$\frac{1}{\sqrt{3}}$	1	$\sqrt{3}$

Man brukar också införa *cotangens* genom

$$\cot x \coloneqq \frac{\cos x}{\sin x}$$

som har definitionsmängden

$$D_{\cot} = \{x \in \mathbb{R}: \quad \sin x \neq 0\} = \{x \in \mathbb{R}: \quad x \neq n\pi \ \text{ för något } n \in \mathbb{Z}\}.$$

Vi har alltså $\tan x \cot x = 1$, $\tan x = 1/\cot x$, $\cot x = 1/\tan x$. Det följer att även cotangens är periodisk med perioden π. Slutligen kan man införa *sekanten* och *cosekanten* via

$$\sec x = \frac{1}{\cos x}, \qquad \csc x = \frac{1}{\sin x}.$$

Det är uppenbart att $D_{\sec} = D_{\tan}$ och $D_{\csc} = D_{\cot}$.

Exempel

Beräkna sinus, cosinus och tangens för vinklarna i $\left\{\frac{5\pi}{6}, -\frac{\pi}{4}, -\frac{2\pi}{3}, \frac{25\pi}{6}\right\}$.

Lösning: Vi ser med hjälp av enhetscirkeln (eller räknelagar) och standardvärdena att

$$\begin{aligned}
\sin\frac{5\pi}{6} &= \sin\left(\pi - \frac{5\pi}{6}\right) = \sin\frac{\pi}{6} = \frac{1}{2} \\
\cos\frac{5\pi}{6} &= -\cos\left(\pi - \frac{5\pi}{6}\right) = -\cos\frac{\pi}{6} = -\frac{\sqrt{3}}{2} \\
\sin\left(-\frac{\pi}{4}\right) &= -\sin\frac{\pi}{4} = -\frac{1}{\sqrt{2}} \\
\cos\left(-\frac{\pi}{4}\right) &= \cos\frac{\pi}{4} = \frac{1}{\sqrt{2}} \\
\sin\left(-\frac{2\pi}{3}\right) &= -\sin\left(\pi - \frac{2\pi}{3}\right) = -\sin\frac{\pi}{3} = -\frac{\sqrt{3}}{2} \\
\cos\left(-\frac{2\pi}{3}\right) &= -\cos\left(\pi - \frac{2\pi}{3}\right) = -\cos\frac{\pi}{3} = -\frac{1}{2} \\
\sin\frac{25\pi}{6} &= \sin\left(4\pi + \frac{\pi}{6}\right) = \sin\frac{\pi}{6} = \frac{1}{2} \\
\cos\frac{25\pi}{6} &= \cos\left(4\pi + \frac{\pi}{6}\right) = \cos\frac{\pi}{6} = \frac{\sqrt{3}}{2}.
\end{aligned}$$

Därför är

$$\tan\frac{5\pi}{6} = -\frac{1}{\sqrt{3}}, \qquad \tan\left(-\frac{\pi}{4}\right) = -1, \qquad \tan\left(-\frac{2\pi}{3}\right) = \sqrt{3}, \qquad \tan\frac{25\pi}{6} = \frac{1}{\sqrt{3}}.$$

I praktiken tar man inte fram dylika funktionsvärden genom algebraiska manipulationer, utan man *ser direkt* i enhetscirkeln vad värdet blir, genom att utnyttja symmetrier och standardvärdena. Mellanleden i redovisningarna ovan bör alltså snarast tolkas som ledtrådar till vilken sorts symmetri som kan användas i respektive fall (t.ex. plus ett halvt varv, spegling i y-axeln eller plus några hela varv).

1.6.1.4 Identiteter med tangens

Vi härleder här tangens additionsformel. Om u och v är reella tal sådana att $u \pm v$ *inte* är ett nollställe till cosinus, så är $\tan(u \pm v)$ definierat, och vi har

$$\tan(u \pm v) = \frac{\sin(u \pm v)}{\cos(u \pm v)} = \frac{\sin u \cos v \pm \cos u \sin v}{\cos u \cos v \mp \sin u \sin v}.$$

Om det *dessutom* är så att varken u eller v i sig är ett nollställe till cosinus, så kan vi förkorta bråket med $\cos u \cos v$, varvid vi erhåller

$$\tan(u \pm v) = \frac{\tan u \pm \tan v}{1 \mp \tan u \tan v}.$$

Identiteten är alltså giltig för alla reella u och v sådana att $\tan(u \pm v)$, $\tan u$ och $\tan v$ är definierade. Om vänsterledet är definierat, så måste inte högerledet vara det. Ett motexempel är $u = v = \pi/2$ då vänsterledet är 0 och högerledet är odefinierat. Däremot, om högerledet är definierat, då är också vänsterledet definierat (varför?). Vi har alltså

$$y = \frac{\tan u \pm \tan v}{1 \mp \tan u \tan v} \qquad \Longrightarrow \qquad y = \tan(u \pm v),$$

men bara ”nästan alltid” omvändningen. Mer precist,

$$y = \frac{\tan u \pm \tan v}{1 \mp \tan u \tan v} \qquad \Longleftrightarrow \qquad \begin{cases} y = \tan(u \pm v) \\ \cos u \cos v \neq 0. \end{cases}$$

Vi kan också erhålla tangens för dubbla vinkeln, genom att sätta $u = v$:

$$\tan 2u = \frac{2 \tan u}{1 - \tan^2 u}$$

som ger ekvivalensen

$$y = \frac{2 \tan u}{1 - \tan^2 u} \qquad \Longleftrightarrow \qquad \begin{cases} y = \tan 2u \\ \cos u \neq 0. \end{cases}$$

Slutligen noterar vi (mest för framtida referens) identiteten

$$\frac{1}{\cos^2 x} = \frac{\sin^2 x + \cos^2 x}{\cos^2 x} = \tan^2 x + 1$$

som vi kommer stöta på i samband med derivering samt identiteterna (där $\cos\frac{x}{2} \neq 0$, d.v.s. $x \neq \pi + n \cdot 2\pi$)

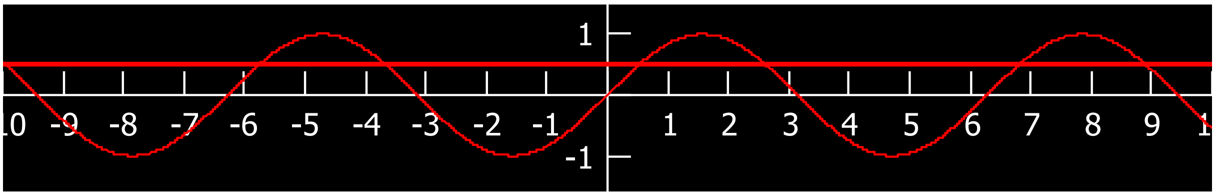
$$\sin x = \frac{\sin\left(2 \cdot \frac{x}{2}\right)}{1} = \frac{2 \sin\frac{x}{2} \cos\frac{x}{2}}{\sin^2\frac{x}{2} + \cos^2\frac{x}{2}} = \frac{2 \tan\frac{x}{2}}{1 + \tan^2\frac{x}{2}}$$

$$\cos x = \frac{\cos\left(2 \cdot \frac{x}{2}\right)}{1} = \frac{\cos^2\frac{x}{2} - \sin^2\frac{x}{2}}{\sin^2\frac{x}{2} + \cos^2\frac{x}{2}} = \frac{1 - \tan^2\frac{x}{2}}{1 + \tan^2\frac{x}{2}}$$

som vi kommer stöta på i samband med motsatsen, d.v.s. integration.

1.6.1.5 Trigonometriska ekvationer

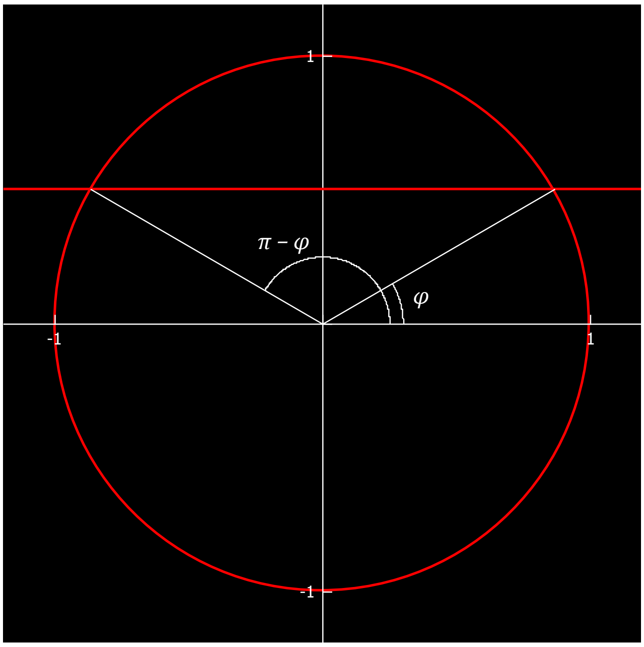
Välj något $a \in [-1, 1] = V_{\sin}$. Om $\sin x = a$, vad kan då sägas om x? Det är klart att ekvationen $\sin x = a$ har oändligt många lösningar, vilket illustreras med grafen $y = \sin x$ och den räta linjen $y = a$ i fallet $a = 1/2$.



Några specialfall är enkla och inses direkt med definitionen av sinus i enhetscirkeln:

$$\sin x = 0 \quad \Longleftrightarrow \quad x = n\pi \text{ för något } n \in \mathbb{Z}$$
$$\sin x = 1 \quad \Longleftrightarrow \quad x = \frac{\pi}{2} + n \cdot 2\pi \text{ för något } n \in \mathbb{Z}$$
$$\sin x = -1 \quad \Longleftrightarrow \quad x = -\frac{\pi}{2} + n \cdot 2\pi \text{ för något } n \in \mathbb{Z}.$$

I allmänhet, om φ är *en* lösning till ekvationen, så är också $\pi - \varphi$ en lösning, eftersom $\sin(\pi - \varphi) = \sin \varphi \overset{\text{def}}{=} a$. Detta illustreras i bilden nedan (där $a = \frac{1}{2}$ och $\varphi = \frac{\pi}{6}, \pi - \varphi = \frac{5\pi}{6}$):



Det är uppenbart att dessa är de *enda* lösningarna, så när som på periodiciteten (d.v.s. så när som på addition av heltalsmultipler av 2π). Alltså kan *samtliga* lösningar till ekvationen $\sin x = a$ uttryckas med hjälp av *en* av lösningarna, φ:

$$\sin x = a \quad \Longleftrightarrow \quad \begin{cases} x = \varphi + n \cdot 2\pi \\ \text{eller} \\ x = \pi - \varphi + n \cdot 2\pi \end{cases}$$

där n är ett heltal. Vi får med andra ord två familjer av lösningar, i alla fall förutom $\sin x = \pm 1$, då de två familjerna sammanfaller. (Jämför med de tre enkla specialfallen ovan.)

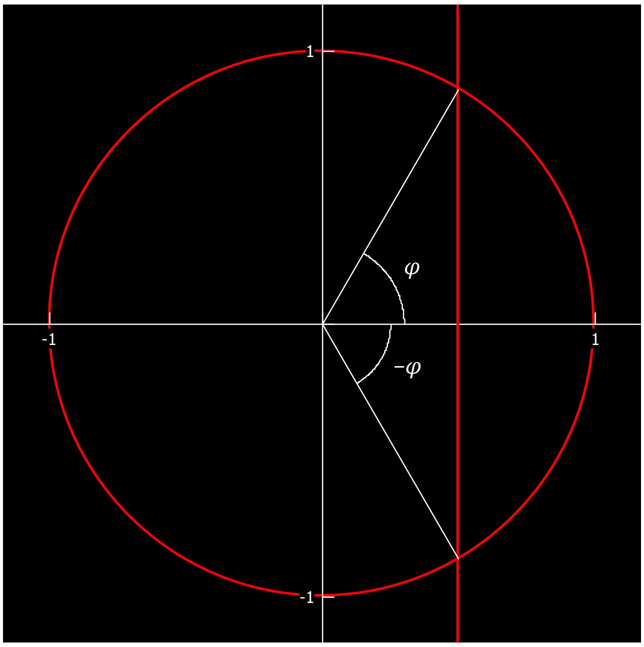
Exempel

Lös ekvationen $\sin x = \frac{1}{2}$.

Lösning:

$$\sin x = \frac{1}{2} \qquad \Longleftrightarrow \qquad \begin{cases} x = \dfrac{\pi}{6} + n \cdot 2\pi \\ \text{eller} \\ x = \dfrac{5\pi}{6} + n \cdot 2\pi. \end{cases}$$

Betrakta nu ekvationen $\cos x = a$ för något $a \in [-1, 1] = V_{\cos}$. Återigen har vi oändligt många lösningar. Om φ är *en* av dem, är det klart att $-\varphi$ är en annan (titta i enhetscirkeln!).



Det är också uppenbart att dessa är de *enda* lösningarna, så när som på hela varv. Alltså är

$$\cos x = a \qquad \Longleftrightarrow \qquad x = \pm\varphi + n \cdot 2\pi.$$

Enkla specialfall är

$$\cos x = 0 \qquad \Longleftrightarrow \qquad x = \frac{\pi}{2} + n\pi$$
$$\cos x = 1 \qquad \Longleftrightarrow \qquad x = n \cdot 2\pi$$
$$\cos x = -1 \qquad \Longleftrightarrow \qquad x = \pi + n \cdot 2\pi.$$

Exempel

Lös ekvationen $\cos x = \frac{1}{2}$.

Lösning:

$$\cos x = \frac{1}{2} \qquad \Longleftrightarrow \qquad x = \pm\frac{\pi}{3} + n \cdot 2\pi.$$

Exempel

Lös ekvationen $\cos x = -\frac{\sqrt{3}}{2}$.

Lösning:

$$\cos x = -\frac{\sqrt{3}}{2} \qquad \Longleftrightarrow \qquad x = \pm\frac{5\pi}{6} + n \cdot 2\pi.$$

Slutligen betraktar vi $\tan x = a$ för något $a \in \mathbb{R} = V_{\tan}$. Eftersom tangens är periodisk med period π och strängt växande inom varje period gäller att om φ är *en* lösning till ekvationen, så har vi

$$\tan x = a \qquad \Longleftrightarrow \qquad x = \varphi + n\pi.$$

Exempel

Lös ekvationen $\tan x = -\sqrt{3}$.

Lösning:

$$\tan x = -\sqrt{3} \qquad \Longleftrightarrow \qquad x = -\frac{\pi}{3} + n\pi.$$

1.6.1.5.1 Olika vinklar, samma funktionsvärde

Ett mycket närbesläktat problem är när två vinklar har samma trigonometriskt funktionsvärde, t.ex. om $\sin a = \sin b$. *En* möjlighet är förstås att vinklarna är lika, $a = b$. En annan är att de är varandras spegelbilder i y-axeln, d.v.s. att $b = \pi - a$ (tänk $\sin\frac{\pi}{6} = \sin\frac{5\pi}{6}$). Dessa två möjligheter är de enda, så när som på att det givetvis även kan skilja hela varv mellan vinklarna.

Motsvarande situation gäller för cosinus: Om $\cos a = \cos b$ är en möjlighet att $a = b$, en annan att $a = -b$; utöver detta kan det skilja hela varv mellan vinklarna.

Exempel

Lös ekvationen $\sin(2x + 1) = \sin(x + 1)$.

Lösning: Vi har

$$\sin(2x+1) = \sin(x+1) \qquad \Longleftrightarrow \qquad \begin{cases} 2x + 1 = x + 1 + n \cdot 2\pi \\ \text{eller} \\ 2x + 1 = \pi - (x + 1) + n \cdot 2\pi \end{cases} \qquad \Longleftrightarrow$$

$$\Longleftrightarrow \qquad \begin{cases} x = n \cdot 2\pi \\ \text{eller} \\ 3x + 2 = \pi + n \cdot 2\pi \end{cases} \qquad \Longleftrightarrow$$

$$\Longleftrightarrow \qquad \begin{cases} x = n \cdot 2\pi \\ \text{eller} \\ x = \dfrac{\pi - 2}{3} + n \cdot \dfrac{2\pi}{3}. \end{cases}$$

1.6.1.5.2 Exempel

Vi kan nu ge en samling enkla exempel på trigonometriska ekvationer. Uppgifter av den här sorten är typiska i inledande matematikkurser på högskolenivå.

Exempel

Lös ekvationen $\sin 2x = \sin x$.

Lösning (metod 1): Vi har

$$\sin 2x = \sin x \iff \begin{cases} 2x = x + n \cdot 2\pi \\ \text{eller} \\ 2x = \pi - x + n \cdot 2\pi \end{cases} \iff \begin{cases} x = n \cdot 2\pi \\ \text{eller} \\ 3x = \pi + n \cdot 2\pi \end{cases} \iff \begin{cases} x = n \cdot 2\pi \\ \text{eller} \\ x = \frac{\pi}{3} + n \cdot \frac{2\pi}{3}. \end{cases}$$

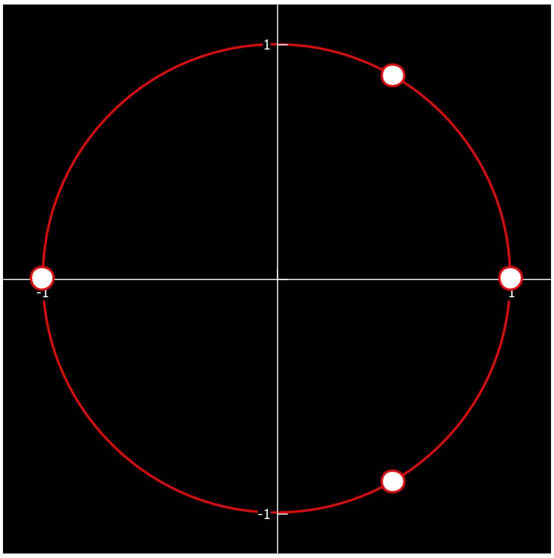
Svar: $\sin 2x = \sin x \iff x = n \cdot 2\pi$ eller $x = \frac{\pi}{3} + n \cdot \frac{2\pi}{3}$ för något $n \in \mathbb{Z}$.

Lösning (metod 2): Vi har

$$\sin 2x = \sin x \iff 2\sin x \cos x = \sin x \iff \begin{cases} \sin x = 0 \\ \text{eller} \\ 2\cos x = 1 \end{cases} \iff \begin{cases} \sin x = 0 \\ \text{eller} \\ \cos x = \frac{1}{2} \end{cases} \iff \begin{cases} x = n\pi \\ \text{eller} \\ x = \pm\frac{\pi}{3} + n \cdot 2\pi. \end{cases}$$

Svar: $\sin 2x = \sin x \iff x = n\pi$ eller $x = \pm\frac{\pi}{3} + n \cdot 2\pi$ för något $n \in \mathbb{Z}$.

Lägg märke till hur de två metoderna ger olika (men ekvivalenta) beskrivningar av lösningsmängden. Oavsett vilken av de två beskrivningarna man använder, får man nämligen följande vinklar (modulo 2π): 0, π, $\pi/3$ samt $-\pi/3$. Lösningsmängden (modulo 2π) illustreras i enhetscirkeln till höger.



Exempel

Lös ekvationen $\frac{1}{2}\sin 2x + \cos^2 x = 1$.

Lösning:

$$\frac{1}{2}\sin 2x + \cos^2 x = 1 \iff \sin 2x = 2(1-\cos^2 x) \iff \sin 2x = 2\sin^2 x \iff 2\sin x\cos x = 2\sin^2 x \iff \sin x\cos x = \sin^2 x \iff \begin{cases} \sin x = 0 \\ \text{eller} \\ \cos x = \sin x \end{cases} \iff \begin{cases} x = n\pi \\ \text{eller} \\ x = \frac{\pi}{4} + n\pi. \end{cases}$$

Svar: $\frac{1}{2}\sin 2x + \cos^2 x \iff x = n\pi$ eller $x = \frac{\pi}{4} + n\pi$ för något $n \in \mathbb{Z}$.

Exempel

Lös ekvationen $\sin^2 x + \sin x = \frac{3}{4}$.

Lösning: Kvadratkomplettering ger

$$\sin^2 x + \sin x = \left(\sin x + \frac{1}{2}\right)^2 - \frac{1}{4}$$

varför ekvationen är ekvivalent med

$$\left(\sin x + \frac{1}{2}\right)^2 - \frac{1}{4} = \frac{3}{4} \iff \left(\sin x + \frac{1}{2}\right)^2 = 1 \iff \sin x + \frac{1}{2} = \pm 1 \iff \sin x \in \left\{\frac{1}{2}, -\frac{3}{2}\right\}.$$

Eftersom $V_{\sin} = [-1, 1]$ är $\sin x = -3/2 < -1$ absurt, så vi måste ha $\sin x = 1/2$. Men

$$\sin x = \frac{1}{2} \iff \begin{cases} x = \frac{\pi}{6} + n\cdot 2\pi \\ \text{eller} \\ x = \frac{5\pi}{6} + n\cdot 2\pi. \end{cases}$$

Svar: $\sin^2 x + \sin x = \frac{3}{4} \iff x = \frac{\pi}{6} + n\cdot 2\pi$ eller $x = \frac{5\pi}{6} + n\cdot 2\pi$ för något $n \in \mathbb{Z}$.

Exempel (*)

Lös ekvationen $\sin^2 x = \frac{\sqrt{3}+2}{4}$.

Lösning: Det är klart att

$$\sin x = \pm \frac{\sqrt{\sqrt{3}+2}}{2},$$

men det hjälper oss knappast med att bestämma x. Å andra sidan är

$$\begin{aligned}
&\sin^2 x = \frac{\sqrt{3}}{4} + \frac{1}{2} \quad \Longleftrightarrow \quad 2\sin^2 x - 1 = \frac{\sqrt{3}}{2} \quad \Longleftrightarrow \\
&\Longleftrightarrow \quad 2\sin^2 x - (\sin^2 x + \cos^2 x) = \frac{\sqrt{3}}{2} \quad \Longleftrightarrow \\
&\Longleftrightarrow \quad \sin^2 x - \cos^2 x = \frac{\sqrt{3}}{2} \quad \Longleftrightarrow \\
&\Longleftrightarrow \quad -\cos 2x = \frac{\sqrt{3}}{2} \quad \Longleftrightarrow \quad \cos 2x = -\frac{\sqrt{3}}{2} \quad \Longleftrightarrow \\
&\Longleftrightarrow \quad 2x = \pm\frac{5\pi}{6} + n \cdot 2\pi \quad \Longleftrightarrow \quad x = \pm\frac{5\pi}{12} + n\pi.
\end{aligned}$$

Svar: $\sin^2 x = \frac{\sqrt{3}+2}{4} \quad \Longleftrightarrow \quad x = \pm\frac{5\pi}{12} + n\pi$ för något $n \in \mathbb{Z}$.

Exempel (*)

Speciellt är

$$\sin^2 \frac{5\pi}{12} = \frac{\sqrt{3}+2}{4}.$$

Eftersom $5\pi/12 \in [0, \pi]$ är $\sin 5\pi/12 \geq 0$ så

$$\sqrt{\sin^2 \frac{5\pi}{12}} = \left|\sin\frac{5\pi}{12}\right| = \sin\frac{5\pi}{12};$$

om vi tar kvadratroten av båda (de positiva) leden i ekvationen $\sin^2 \frac{5\pi}{12} = \frac{\sqrt{3}+2}{4}$ får vi därför

$$\sin\frac{5\pi}{12} = \frac{\sqrt{\sqrt{3}+2}}{2}.$$

Det är alltså fullt möjligt att utöka tabellen med standardfunktionsvärden för sinus, cosinus och tangens till fler vinklar än bara $\frac{\pi}{6}$, $\frac{\pi}{4}$ och $\frac{\pi}{3}$ (och de triviala, som 0, $\frac{\pi}{2}$ och π). Däremot är resultat som det nyligen härledda inte något man går runt och kan i huvudet.

Som en bonusövning i allmän räknefärdighet visar vi här att vi kan skriva det nyhärledda funktionsvärdet på tre alternativa sätt:

$$\sin\frac{5\pi}{12} = \frac{\sqrt{\sqrt{3}+2}}{2} = \frac{1+\sqrt{3}}{2\sqrt{2}} = \frac{\sqrt{2}+\sqrt{6}}{4} = \frac{1}{\sqrt{6}-\sqrt{2}}.$$

Det första och det andra uttrycket för sinusvärdet är uppenbarligen positiva, så för att visa att de verkligen är lika, räcker det med att visa att de har samma kvadrat. Och mycket riktigt är

$$\left(\frac{\sqrt{\sqrt{3}+2}}{2}\right)^2=\frac{\sqrt{3}+2}{4}=\frac{2\sqrt{3}+4}{8}=\frac{1+2\sqrt{3}+3}{8}=\left(\frac{1+\sqrt{3}}{\sqrt{8}}\right)^2=\left(\frac{1+\sqrt{3}}{2\sqrt{2}}\right)^2.$$

Förlängning med $\sqrt{2}$ ger den tredje formen. Genom att i stället förlänga andra formen med täljarens konjugat erhåller vi den fjärde formen:

$$\frac{1+\sqrt{3}}{2\sqrt{2}}=\frac{(1+\sqrt{3})(1-\sqrt{3})}{2\sqrt{2}(1-\sqrt{3})}=\frac{1-3}{2\sqrt{2}(1-\sqrt{3})}=-\frac{2}{2\sqrt{2}(1-\sqrt{3})}=\frac{1}{\sqrt{2}(\sqrt{3}-1)}=\frac{1}{\sqrt{6}-\sqrt{2}}.$$

1.6.1.6 Ekvationslösning med villkor

Vi har sett exempel på ekvationer i sinus, cosinus och tangens där vi, på grund av periodiciteten, i allmänhet får oändligt många lösningar. Med extra villkor på den oberoende variabeln kan man förstås få färre eller rent av unika lösningar.

Exempel

Bestäm alla vinklar $v \in [0, \pi/2]$ sådana att $\sin^2 x = 3\cos^2 x$.

Lösning:

$$\sin^2 x = 3\cos^2 x \iff 1-\cos^2 x = 3\cos^2 x \iff 1 = 4\cos^2 x \iff \cos^2 x = \frac{1}{4} \iff \cos x = \pm\frac{1}{2}.$$

Eftersom $v \in \left[0, \frac{\pi}{2}\right]$ är $\cos x \geq 0$ så vi måste ha $\cos x = 1/2$. Därför är

$$x = \pm\frac{\pi}{3} + n \cdot 2\pi.$$

Av dessa vinklar är det endast $x = \pi/3$ som tillhör $[0, \pi/2]$.

Svar: $v \in [0, \pi/2]$ och $\sin^2 x = 3\cos^2 x \iff x = \pi/3$.

1.6.1.7 Samband mellan olika funktioners värden för samma vinkel

Exempel

Bestäm $\sin x$ om $\cos x = \frac{1}{\sqrt{7}}$ och $x \in \left[-\frac{\pi}{2}, 0\right]$.

Lösning: Vi har

$$\sin^2 x + \cos^2 x = 1 \iff \sin^2 x = 1 - \cos^2 x \iff \sin x = \pm\sqrt{1-\cos^2 x}.$$

Eftersom $x \in [-\pi/2, 0]$ är $\sin x \leq 0$ så att

$$\sin x = -\sqrt{1-\cos^2 x} = -\sqrt{1-\frac{1}{7}} = -\sqrt{\frac{6}{7}}.$$

Exempel

Bestäm $\tan x$ om $\sin x = \frac{1}{5}$ och $x \in \left[\frac{\pi}{2}, \pi\right]$.

Lösning: Vi har

$$\sin^2 x + \cos^2 x = 1 \iff \cos^2 x = 1 - \sin^2 x \iff \cos x = \pm\sqrt{1 - \sin^2 x}.$$

Eftersom $x \in [\pi/2, \pi]$ är $\cos x \leq 0$ så att

$$\cos x = -\sqrt{1 - \sin^2 x} = -\sqrt{1 - \frac{1}{25}} = -\sqrt{\frac{24}{25}} = -\frac{\sqrt{24}}{5} = -\frac{2\sqrt{6}}{5}$$

och

$$\tan x = \frac{\sin x}{\cos x} = \frac{\frac{1}{5}}{-\frac{2\sqrt{6}}{5}} = -\frac{1}{2\sqrt{6}}.$$

Exempel

Bestäm $\sin x$ om $\tan x = 100$ och $x \in [0, \pi/2]$.

Lösning: Det är uppenbart att $x \in \,]0, \pi/2[$ (varför?). Därför är $\cos x > 0$ och $\sin x >$ [att $\tan x > 0$ medför ju bara att $\sin x$ och $\cos x$ har *samma* tecken]. Med detta i åtanke (framför allt $\cos x > 0$) finner vi att

$$\tan x = \frac{\sin x}{\cos x} = \frac{\sin x}{+\sqrt{1 - \sin^2 x}} = 100.$$

Med $s \coloneqq \sin x \in \,]0,1[$ har vi därför

$$\frac{s}{\sqrt{1 - s^2}} = 100 \iff s = 100\sqrt{1 - s^2} \iff s^2 = 10000(1 - s^2) \iff 10001 s^2 = 10000 \iff s^2 = \frac{10000}{10001} \iff s = \frac{100}{\sqrt{10001}}.$$

Alltså har vi funnit

$$\sin x = \frac{100}{\sqrt{10001}} (= 0.99995000375\ldots).$$

En genväg till att bestämma $\sin x$ eller $\cos x$ – framför allt den senare – när $\tan x$ är känt är att utnyttja identiteten

$$1 + \tan^2 x = \frac{1}{\cos^2 x}$$

som vi enkelt härledde på sidan 98.

Exempel

Bestäm $\cos x$ om $\tan x = 4$ och $x \in \left[0, \frac{\pi}{2}\right]$.

Lösning: Vi har

$$1 + \tan^2 x = \frac{1}{\cos^2 x} \iff \cos^2 x = \frac{1}{1 + \tan^2 x} = \frac{1}{17} \iff \cos x = \pm\frac{1}{\sqrt{17}} \iff \cos x = \frac{1}{\sqrt{17}}$$

eftersom $x \in \left[0, \frac{\pi}{2}\right]$ medför att $\cos x \geq 0$.

1.6.1.8 Hjälpvinkelmetoden

Ett uttryck på formen $A\sin x + B\cos x$ kan alltid skrivas på formen $C\sin(x + \delta)$ med ett par konstanter C och δ som beror på A och B. Vi har nämligen (om inte både A och B är noll i vilket fall allt är trivialt)

$$A\sin x + B\cos x = \sqrt{A^2 + B^2}\left(\frac{A}{\sqrt{A^2 + B^2}}\sin x + \frac{B}{\sqrt{A^2 + B^2}}\cos x\right).$$

Eftersom

$$\left(\frac{A}{\sqrt{A^2 + B^2}}\right)^2 + \left(\frac{B}{\sqrt{A^2 + B^2}}\right)^2 = 1$$

ligger punkten $\left(\frac{A}{\sqrt{A^2+B^2}}, \frac{B}{\sqrt{A^2+B^2}}\right)$ på enhetscirkeln. Alltså är

$$\left(\frac{A}{\sqrt{A^2 + B^2}}, \frac{B}{\sqrt{A^2 + B^2}}\right) = (\cos\delta, \sin\delta)$$

för någon vinkel $\delta \in \mathbb{R}$. Vi har därför

$$\begin{aligned} A\sin x + B\cos x &= \sqrt{A^2 + B^2}\left(\frac{A}{\sqrt{A^2 + B^2}}\sin x + \frac{B}{\sqrt{A^2 + B^2}}\cos x\right) = \\ &= \sqrt{A^2 + B^2}(\cos\delta\sin x + \sin\delta\cos x) = \sqrt{A^2 + B^2}\sin(x + \delta) \end{aligned}$$

som utlovat.

Exempel

Lös ekvationen $\sin x + \sqrt{3}\cos x = 1$.

Lösning: Hjälpvinkelmetoden ger

$$\sin x + \sqrt{3}\cos x = \sqrt{1^2 + \left(\sqrt{3}\right)^2}\sin(x+\delta) = 2\sin(x+\delta)$$

där enda kravet på δ är att $\cos\delta = 1/2$ och $\sin\delta = \sqrt{3}/2$. En fullt godtagbar hjälpvinkel är sålunda $\delta = \pi/3$ varför

$$\sin x + \sqrt{3}\cos x = 2\sin\left(x + \frac{\pi}{3}\right), \qquad \forall x \in \mathbb{R}$$

och vi har

$$\sin x + \sqrt{3}\cos x = 1 \iff 2\sin\left(x+\frac{\pi}{3}\right) = 1 \iff$$

$$\iff \sin\left(x+\frac{\pi}{3}\right) = \frac{1}{2} \iff \begin{cases} x + \frac{\pi}{3} = \frac{\pi}{6} + n\cdot 2\pi \\ \text{eller} \\ x + \frac{\pi}{3} = \frac{5\pi}{6} + n\cdot 2\pi \end{cases} \iff$$

$$\iff \begin{cases} x = -\frac{\pi}{6} + n\cdot 2\pi \\ \text{eller} \\ x = \frac{\pi}{2} + n\cdot 2\pi. \end{cases}$$

Svar: $\sin x + \sqrt{3}\cos x = 1 \iff x = -\frac{\pi}{6} + n\cdot 2\pi$ eller $x = \frac{\pi}{2} + n\cdot 2\pi$ för något $n \in \mathbb{Z}$.

Exempel

Lös ekvationen $\sin x + \cos x = 1$.

Lösning: Uppgiften *kan* mycket väl lösas med hjälpvinkelmetoden, men det behövs inte. $(\cos x, \sin x)$ är ju en punkt på enhetscirkeln, så $\cos^2 x + \sin^2 x = 1$. Men nu är också $\sin x + \cos x = 1$, d.v.s. $(\cos x, \sin x)$ ligger också på den räta linjen $y = 1 - x$. Snittet mellan denna linje och enhetscirkeln undersökte vi i avsnittet ”Samspelet mellan algebra och geometri” på sidan 55. Vi får alltså direkt

$$(\sin x, \cos x) \in \{(1,0),(0,1)\}$$

d.v.s.

$$x = n\cdot 2\pi \quad \text{eller} \quad x = \frac{\pi}{2} + n\cdot 2\pi$$

i båda fallen för något heltal n.

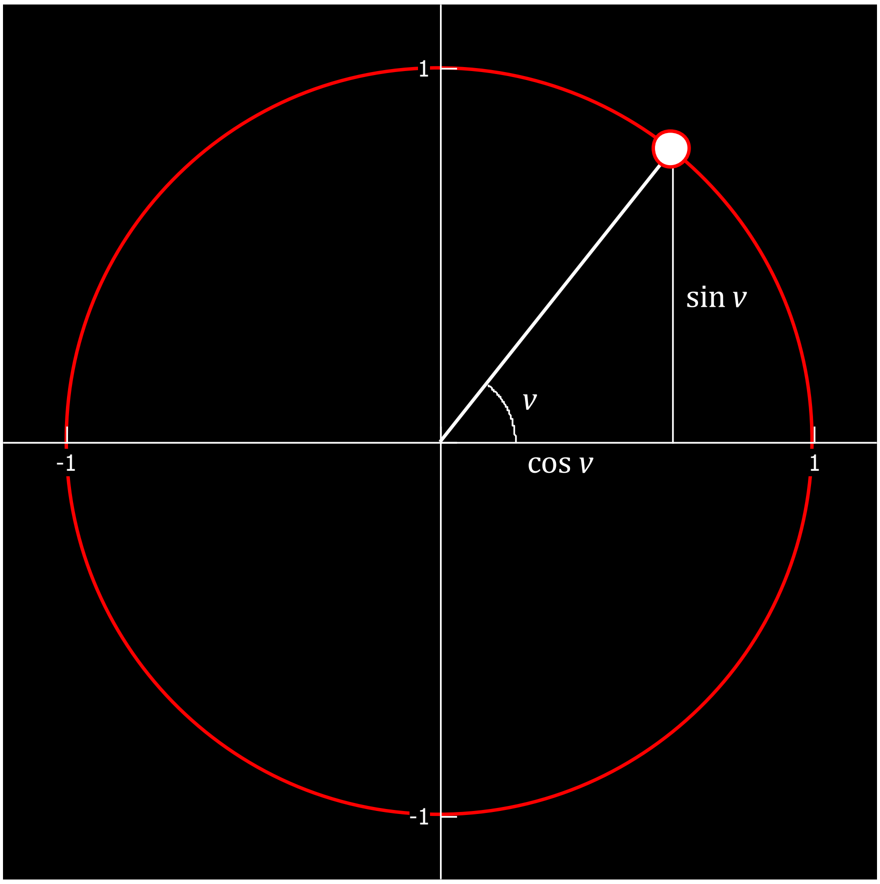
1.6.2 Trigonometri och trianglar

Trigonometri förknippas med trianglar, och vi skall se att det inte alls är konstigt. Man hör på själva ordet *trigonometri* att det handlar om *mätning* (-metri) i figurer med tre (tri-) vinklar (-gon-), d.v.s. i trianglar (jfr engelskan *triangle*). Emellertid, som vi sett, kan de trigonometriska funktionerna i sig introduceras utan att ens nämna ordet ”triangel”. Dessutom är dessa elementära matematiska funktioner fundamentala för mycket matematik, och i många tillämpningar lyser triangelkopplingen med sin frånvaro. Till exempel används de trigonometriska funktionerna

- för att beskriva *rotationer* inom geometrin,
- för att beskriva vågor, t.ex. elektromagnetiska vågor, samt
- för att beskriva utslaget hos en fjäder eller pendel som funktion av tiden.

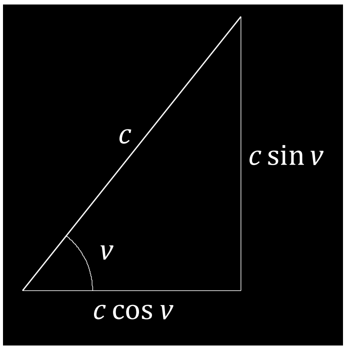
Icke desto mindre stöter svenska skolelever oftast på de trigonometriska funktionerna först i samband med just trianglar, närmare bestämt rätvinkliga sådana. I vilket fall som helst är beskrivningar av och beräkningar i trianglar *en* av alla tillämpningar av de trigonometriska funktionerna. I det här avsnittet ger vi kopplingen.

Välj en vinkel $v \in \left]0, \frac{\pi}{2}\right[$ och titta i enhetscirkeln:

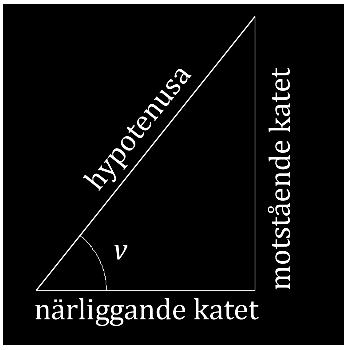


Kopplingen till rätvinkliga trianglar torde vara uppenbar! Genom att dra en vertikal linje från $P = (\cos v, \sin v)$ ned till x-axeln [d.v.s. till $(\cos v, 0)$] erhålles en rätvinklig triangel som avgränsas av linjen ℓ från origo till $P = (\cos v, \sin v)$, x-axelns segment från origo till $(\cos v, 0)$ och den vertikala linjen.

Triangelns bas är $\cos v$ och dess höjd är $\sin v$; hypotenusans längd är 1 (enhetscirkelns radie). Om vi skalar triangeln likformigt med faktorn $c > 0$ erhåller vi en ny triangel som är likformig med triangeln i enhetscirkeln. Dess bas, höjd och hypotenusa är $c \cos v$, $c \sin v$ respektive c. Vinkeln nere till vänster är fortfarande v:



Omvänt, vilken rätvinklig triangel som helst med hypotenusa c kan skalas med faktorn c^{-1} varvid en triangel i första kvadranten innanför enhetscirkeln erhålles. (Eventuellt måste förstås triangeln roteras och translateras så att den hamnar stående på x-axeln med en spets i origo och andra spetsen på enhetscirkeln.) Om detta förfarande utförs med triangeln



så får den skalade triangels sidor längderna $\frac{\text{närliggande katet}}{\text{hypotenusa}}$, $\frac{\text{motstående katet}}{\text{hypotenusa}}$, respektive 1 (hypotenusan). Men eftersom vi befinner oss i enhetscirkeln gäller

$$\sin v = \frac{\text{motstående katet}}{\text{hypotenusa}}, \qquad \cos v = \frac{\text{närliggande katet}}{\text{hypotenusa}}.$$

Av detta följer dessutom att

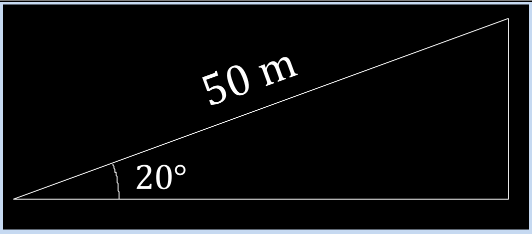
$$\tan v = \frac{\sin v}{\cos v} = \frac{\text{motstående katet}}{\text{närliggande katet}}.$$

Det är på det här sättet som de trigonometriska funktionerna brukar *definieras* i skolmatematiken (på gymnasiet). Men i dessa fall blir funktionerna bara definierade för talen[16] $]0, \pi/2[$, vilket bara är en fjärdedels period av vad man får om definierar sin, cos och tan med enhetscirkeln som vi gjorde i förra avsnittet.

Exempel

Bestäm alla sidlängder i triangeln nedan.

[16] En vinkel i en rätvinklig triangel är alltid mindre än 90°. Om en rätvinklig triangel har vinklarna u, v och $90°$ så är ju vinkelsumman $u + v + 90° = 180°$ varför $u + v = 90°$. Eftersom $u > 0$ är $v < 90°$; eftersom $v > 0$ är $u < 90°$.



Lösning (med räknare/dator): Basen är $50 \cos 20° \approx 47$ och höjden är $50 \sin 20° \approx 17$.

Svar: Basen är 47 m och höjden är 17 m.

1.6.2.1 Om att använda trianglar för att lösa trigonometriska problem (pedagogiska funderingar)

Som nämnt i inledningen till det här avsnittet så har de trigonometriska funktionerna $\mathbb{R} \to \mathbb{R}$ egentligen inget särskilt med trianglar att göra. Däremot har vi sett att funktionernas restriktioner till rätvinkeltriangelvinklarna (mitt nya favoritord!) $\left]0, \frac{\pi}{2}\right[$ kan beskrivas fullständigt med hjälp av rätvinkliga trianglar. Vissa trigonometriska problem, framför allt dem vi diskuterade i avsnittet *Samband mellan olika funktioners värden för samma vinkel* på sidan 105, kan därför lösas med hjälp av rätvinkliga trianglar, som ett *alternativ* till de metoder vi diskuterade i avsnittet om de trigonometriska funktionerna. Däremot fungerar metoden inte alltid. Till exempel, problemet

Exempel

Bestäm $\sin x$ om $\cos x = \frac{1}{\sqrt{7}}$ och $x \in \left[-\frac{\pi}{2}, 0\right]$.

som vi löste i det ovan nämnda avsnittet, kan *inte* på ett ”direkt” sätt lösas med hjälp av rätvinkliga trianglar, eftersom talen i $\left[-\frac{\pi}{2}, 0\right]$ knappast kan fungera som vinklar i någon triangel. Följande problem låter sig däremot lösas både med standardmetoden och triangelmetoden:

Exempel

Bestäm $\sin x$ om $\cos x = \frac{1}{\sqrt{7}}$ och $x \in \left[0, \frac{\pi}{2}\right]$.

Lösning (standardmetoden): Vi har

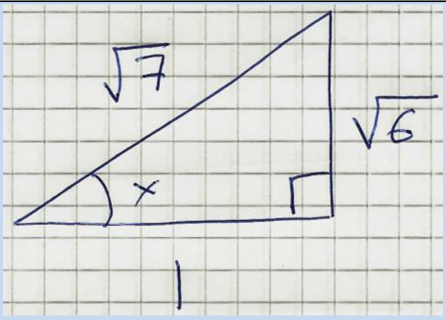
$$\sin^2 x + \cos^2 x = 1 \quad \Longleftrightarrow \quad \sin^2 x = 1 - \cos^2 x \quad \Longleftrightarrow \quad \sin x = \pm\sqrt{1 - \cos^2 x}.$$

Eftersom $x \in \left[0, \frac{\pi}{2}\right]$ är $\sin x \geq 0$ så att

$$\sin x = \sqrt{1 - \cos^2 x} = \sqrt{1 - \frac{1}{7}} = \sqrt{\frac{6}{7}}.$$

* * *

Lösning (triangelmetoden): Det är klart att $x \in \left]0, \frac{\pi}{2}\right[$ så x är en vinkel i en rätvinklig triangel, enligt illustration.



(Här skriver vi först ut x, 1 och $\sqrt{7}$, så att $\cos x = \frac{1}{\sqrt{7}}$ gäller. Sedan använder vi Pythagoras sats för att sätta dit etiketten $\sqrt{6}$.) I triangeln ser man att

$$\sin x = \frac{\sqrt{6}}{\sqrt{7}}.$$

Om i ett dylikt problem det givna värdet är tangensvärdet, blir lösningen till och med marginellt lättare med ”triangelmetoden”. Låt oss illustrera med ett tidigare exempel (som vi här upprepar):

Exempel

Bestäm $\sin x$ om $\tan x = 100$ och $x \in [0, \pi/2]$.

Lösning (standardmetoden): Det är uppenbart att $x \in]0, \pi/2[$. Därför är $\cos x > 0$ och $\sin x > 0$. Med detta i åtanke (framförallt $\cos x > 0$) finner vi att

$$\tan x = \frac{\sin x}{\cos x} = \frac{\sin x}{+\sqrt{1 - \sin^2 x}} = 100.$$

Med $s \coloneqq \sin x \in]0,1[$ har vi därför

$$\frac{s}{\sqrt{1 - s^2}} = 100 \iff s = 100\sqrt{1 - s^2} \iff$$
$$\iff s^2 = 10000(1 - s^2) \iff$$
$$\iff 10001 s^2 = 10000 \iff$$
$$\iff s^2 = \frac{10000}{10001} \iff s = \frac{100}{\sqrt{10001}}.$$

Alltså har vi funnit

$$\sin x = \frac{100}{\sqrt{10001}} (= 0.99995000375\ldots).$$

* * *

Lösning (standardmetoden nummer 2): Vi har

$$1 + \tan^2 x = \frac{1}{\cos^2 x} = \frac{1}{1 - \sin^2 x} \iff$$
$$\iff 1 - \sin^2 x = \frac{1}{1 + \tan^2 x} \iff$$

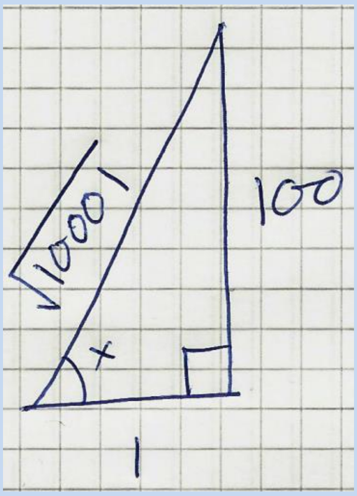
$$\iff \qquad \sin^2 x = 1 - \frac{1}{1+\tan^2 x} = 1 - \frac{1}{10001} = \frac{10000}{10001}.$$

Eftersom $x \in [0, \pi/2]$ är $\sin x \geq 0$ så att

$$\sin x = +\sqrt{\frac{10000}{10001}} = \frac{100}{\sqrt{10001}}.$$

* * *

Lösning (triangelmetoden): Det är klart att $x \in \left]0, \frac{\pi}{2}\right[$ så x är en vinkel i en rätvinklig triangel, enligt illustration.



(Här skriver vi först ut x, 100 och 1 så att $\tan x = 100$. Sedan använder vi Pythagoras sats för att sätta dit etiketten $\sqrt{10001}$.) I triangeln ser man att

$$\sin x = \frac{100}{\sqrt{10001}}.$$

En nackdel med ”triangelmetoden” är att den inte är lika ”ren” som standardmetoden. I standardmetoden använder vi ju i princip enbart definitionerna av de trigonometriska funktionerna i enhetscirkeln. I ”triangelmetoden” använder vi (implicit) kopplingen till rätvinkliga trianglar. En annan nackdel är att det blir lite knöligare att redovisa lösningsgången, eftersom man måste hänvisa till en illustration; dessutom måste läsaren förstå kopplingen mellan det aktuella problemet och triangeln. Det blir lite bökigare att *serialisera* tankegången i text.

Vidare fungerar ”triangelmetoden” i sin enklaste form förstås bara när argumenten till de trigonometriska funktionerna ligger i intervallet $\left]0, \frac{\pi}{2}\right[$. Det är onekligen svårt (här en eufemism för ”omöjligt”) att rita en rätvinklig triangel där vissa sidlängder är negativa, eller där någon inre vinkel är negativ eller större än 90°.

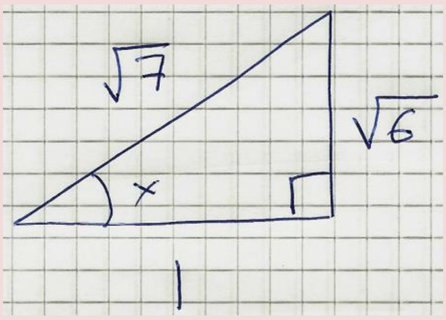
Mindre lyckat bruk av ”triangelmetoden”

Det hindrar emellertid inte studenter från att försöka lösa allmänna trigonometriska problem med hjälp av rätvinkliga trianglar. Till exempel ser man ibland följande försök till lösning på vårt tidigare exempelproblem:

Exempel

Bestäm $\sin x$ om $\cos x = \frac{1}{\sqrt{7}}$ och $x \in \left[-\frac{\pi}{2}, 0\right]$.

”Lösning”:

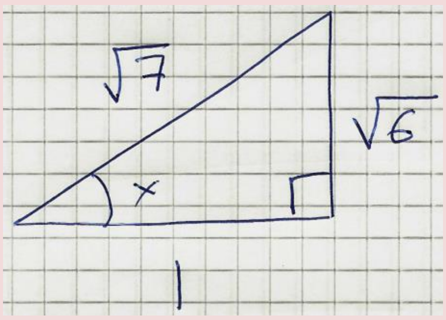


I triangeln ser vi att $\sin x = \frac{\sqrt{6}}{\sqrt{7}}$.

Här blev svaret fel; jämför med den tidigare givna korrekta lösningen som ger det korrekta svaret $-\sqrt{\frac{6}{7}}$. Faktum är att svaret inte bara är fel, det är dessutom orimligt: eftersom $x \in \left[-\frac{\pi}{2}, 0\right]$ är det uppenbart att $\sin x < 0$. Vidare är vinkeln x i triangeln negativ, vilket är omöjligt. Kopplingen mellan det aktuella problemet och den uppritade (omöjliga) triangeln är också (uppenbarligen!) oklar.

Inte sällan får studenten rätt svar även i dessa fall, fast med en något tveksam redovisning:

”Lösning”: Vi ritar en triangel:



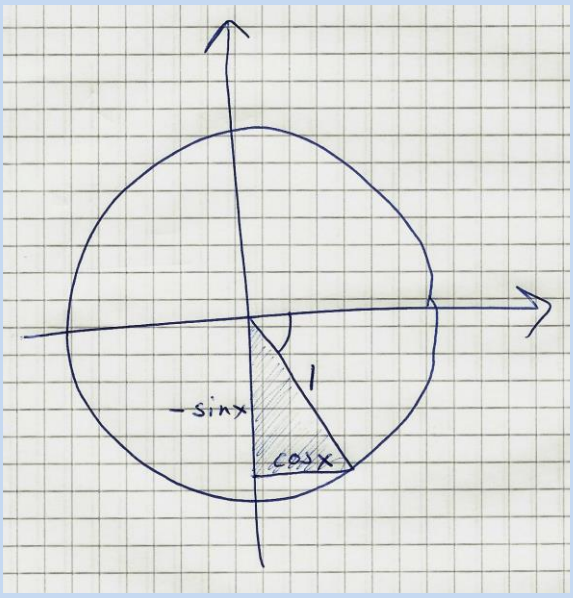
Vi ser att $\sin x = \frac{\sqrt{6}}{\sqrt{7}}$, men eftersom $x \in \left[-\frac{\pi}{2}, 0\right]$ är $\sin x < 0$.

Svar: $\sin x = -\frac{\sqrt{6}}{\sqrt{7}}$

Här är svaret rätt, men lösningen är mycket tveksam (högst tveksamt om det blir någon poäng på uppgiften, trots rätt svar). Studenten hävdar ju (felaktigt!) att $\sin x = \frac{\sqrt{6}}{\sqrt{7}}$. I och med att studenten sedan hävdar att $\sin x = -\frac{\sqrt{6}}{\sqrt{7}}$, så hävdar han ju dessutom att $\frac{\sqrt{6}}{\sqrt{7}} = -\frac{\sqrt{6}}{\sqrt{7}}$, vilket inte heller är sant (milt uttryckt). Även bortsett från dessa tokigheter, så är det inte uppenbart vilken koppling (den omöjliga) triangeln har till det aktuella problemet. Man kan åtminstone konstatera att det inte går att följa redovisningen av lösningen – den är obegriplig. I och för sig är det inte en slump att svaret blir rätt, men matematik handlar ju om att resonera, och att göra det rätt, och att kunna *formulera* resonemanget på ett begripligt sätt. Det kan man inte säga är gjort i det här fallet.

Det går förstås att lösa problemet korrekt även "med hjälp av" en rätvinklig triangel. Man kan då förslagsvis rita ut triangeln i fjärde kvadranten innanför enhetscirkeln. Men i sådana fall använder man ju egentligen bara definitionerna av de trigonometriska funktionerna och "härleder" trigonometriska ettan själv.

Lösning (otymplig): Vi har följande situation:



Pythagoras sats ger

$$-\sin x = \sqrt{1-\cos^2 x} = \sqrt{1-\frac{1}{7}} = \sqrt{\frac{6}{7}}$$

så att $\sin x = -\sqrt{\frac{6}{7}}$.

I det här fallet har studenten i praktiken använt enbart definitonen av de trigonometriska funktionerna i enhetscirkeln. Studenten har implicit använt $x \in \left[-\frac{\pi}{2}, 0\right]$ genom att rita vinkeln i den fjärde kvadranten. Den rätvinkliga triangelns höjd är $0 - \sin x = -\sin x > 0$ och dess bas är $\cos x > 0$. Tillämpning av Pythagoras sats i enhetscirkeln är däremot precis vad som används för att "härleda" trigonometriska ettan [eg. enhetscirkelns ekvation], så den här lösningen är i själva verket bara en tillkrånglad (och mindre lättläst/lättförstådd) version av exempellösningen.

För att sammanfatta så fungerar ”triangelmetoden” bara för vinklar mellan 0 och 90° (*exklusive* gränserna). Men även i dessa fall är standardmetoderna kanske att föredra, eftersom de är ”renare” och ger upphov till lösningsgångar som är lättare att följa. Dessutom är det en viss fördel att alltid använda en och samma metod.

Om vinklarna inte är låsta till intervallet $\left]0, \frac{\pi}{2}\right[$ bör ingen variant av ”triangelmetoden” användas.

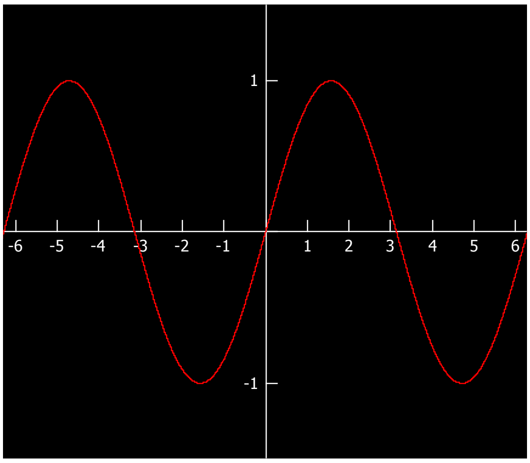
I fallet med vinklar i $\left]0, \frac{\pi}{2}\right[$ är det emellertid aldrig ”fel” att använda triangelmetoden, och den kan dessutom ge marginellt enklare lösningsgång om det givna funktionsvärdet är tangens.

1.6.3 Arcusfunktioner

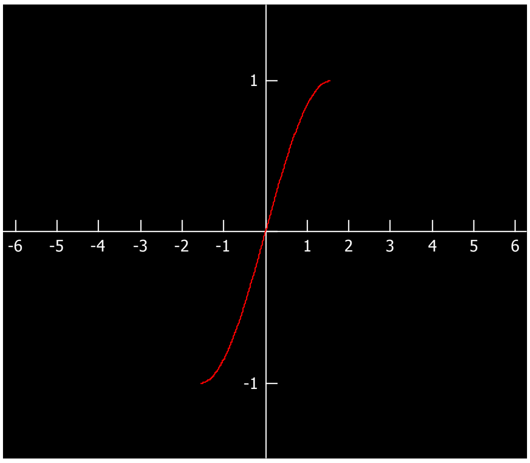
De trigonometriska funktionerna (här betraktar vi främst sin, cos och tan) är *inte* injektiva eftersom de är periodiska. I fallet med sin och cos är de inte ens injektiva inom en period (säg, på $[0, 2\pi[$). De trigonometriska funktionerna har därför *inte* några inversa funktioner.

Samma sak gäller för kvadreringsfunktionen $f: \mathbb{R} \to \mathbb{R}, x \mapsto x^2$, men där införde vi restriktionen $\tilde{f}: [0, \infty[\to \mathbb{R}, x \mapsto x^2$. $\tilde{f}$ är strängt växande och därför injektiv, och vi kallar dess invers för kvadratrotsfunktionen $\sqrt{\ }$. För varje $y \in V_{\tilde{f}} = V_f = [0, \infty[$ är med andra ord $\sqrt{y} \in D_{\tilde{f}} \subset D_f$ det (entydigt bestämda) icke-negativa tal vars kvadrat är y. Vi gör i det här avsnittet motsvarande sak för de trigonometriska funktionerna.

Om vi tittar på sinus definition i enhetscirkeln, eller på sinuskurvan $y = \sin x$, inser vi att sinus är strängt växande, och därmed injektiv, på intervallet $[-\frac{\pi}{2}, \frac{\pi}{2}]$ och att alla värden i $V_{\sin}$ antas i intervallet. Vi inför därför funktionen $f: \left[-\frac{\pi}{2}, \frac{\pi}{2}\right] \to [-1, 1], x \mapsto \sin x$ som vår restriktion av sinus. Graferna $y = \sin x$ och $y = f(x)$ visas nedan.

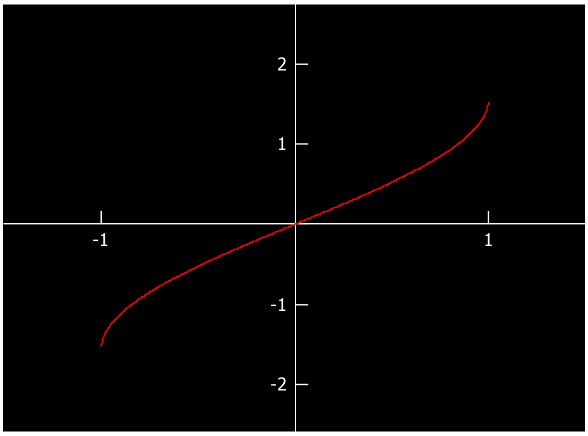


$y = \sin x$



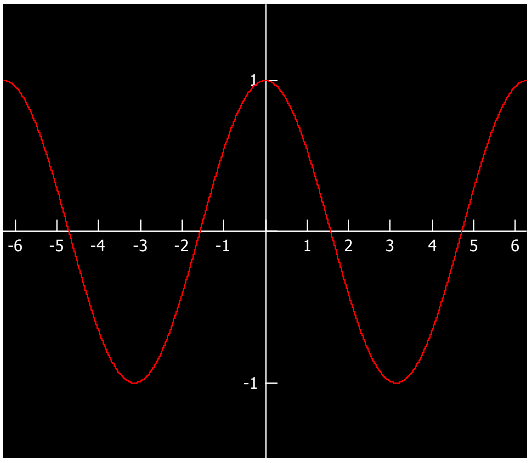
$y = f(x)$

Eftersom f är injektiv har den en invers, och det är den som kallas för arcsin, eller ”arcus sinus”. Emedan $D_f = \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$ och $V_f = [-1, 1]$ har vi $D_{\arcsin} = [-1, 1]$ och $V_{\arcsin} = \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$. Grafen till arcsin visas nedan.

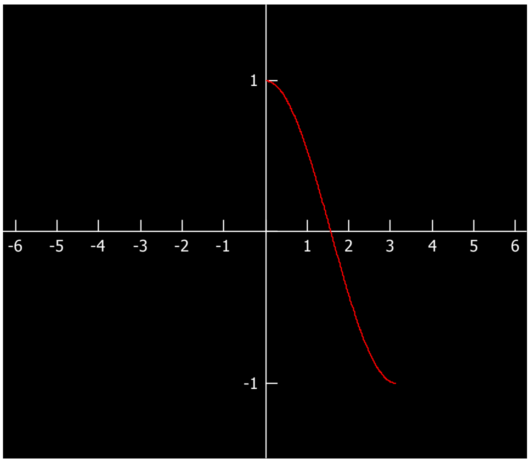


$y = \arcsin x$

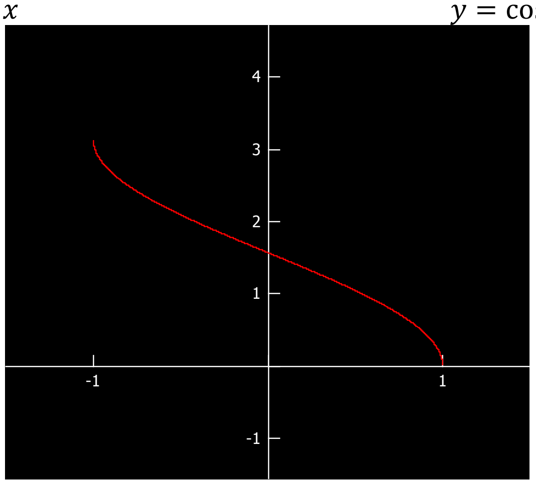
Betrakta nu cosinus. Cosinus är *inte* injektiv på $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$, men däremot på $[0, \pi]$, där cosinus är strängt avtagande, och dessutom antar alla värden i sin värdemängd. Funktionen arccos, eller ”arcus cosinus”, definieras som inversen till restriktionen av cosinus till detta intervall, $[0, \pi]$. Vi har sålunda $D_{\text{arccos}} = [-1, 1]$ och $V_{\text{arccos}} = [0, \pi]$. Nedan visas graferna till cosinus, dess restriktion och arccos.



$y = \cos x$

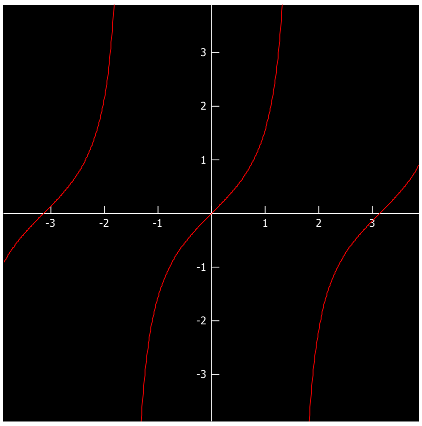


$y = \cos x, x \in [0, \pi]$

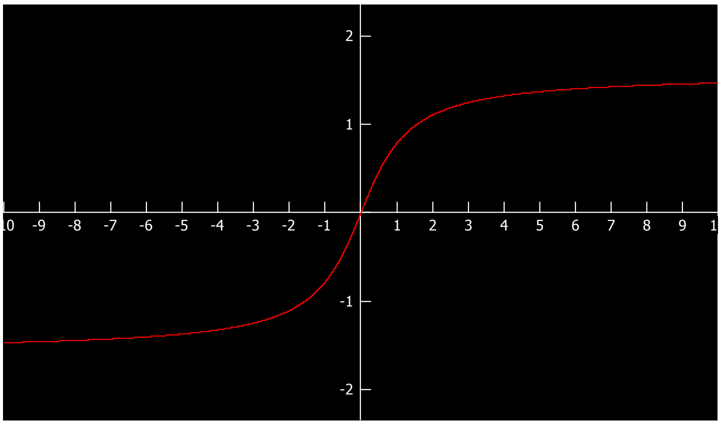


$y = \arccos x$

Nu kan vi det här, så när det gäller tangens konstaterar vi kort och gott att arctan, ”arcus tangens” är inversen till restriktionen av tangens till intervallet $\left]-\frac{\pi}{2}, \frac{\pi}{2}\right[$. Alltså är $D_{\text{arctan}} = \mathbb{R}$ och $V_{\text{arctan}} = \left]-\frac{\pi}{2}, \frac{\pi}{2}\right[$. Nedan visas graferna till tan och arctan.



$y = \tan x$



$y = \arctan x$

Graferna till sinus, cosinus, tangens och arcustangens bör man kunna i huvudet (d.v.s. man bör kunna rita dem utan att ens behöva fundera).

1.6.3.1 På ren svenska …

- … är $\arcsin x$ *den* vinkel i intervallet $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$ för vilken sinus är x.
- … är $\arccos x$ *den* vinkel i intervallet $[0, \pi]$ för vilken cosinus är x.
- … är $\arctan x$ *den* vinkel i intervallet $\left]-\frac{\pi}{2}, \frac{\pi}{2}\right[$ för vilken tangens är x.

1.6.3.2 Observationer

Följande observationer, som följer omedelbart av definitionerna, bör man tycka är självklara.

- $\arcsin x$ är alltid ett tal i intervallet $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$.
- $\arccos x$ är alltid ett tal i intervallet $[0, \pi]$.
- $\arctan x$ är alltid ett tal i intervallet $\left]-\frac{\pi}{2}, \frac{\pi}{2}\right[$.

Notera att

$$\sin(\arcsin x) = x, \qquad \forall x \in [-1,1].$$

Per definition är ju $\arcsin x$ *en* av alla de vinklar, för vilka sinus antar värdet x. Vilket värde antar sinus för $\arcsin x$? Jo, x, så klart![17]

Däremot gäller INTE att

$$\arcsin(\sin x) = x \quad \text{(FEL i allmänhet!)}$$

för alla $x \in \mathbb{R}$. Om t.ex. $x = 16\pi$ så är ju $\sin x = 0$. *Den* vinkel i intervallet $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$ för vilken sinus är noll är förstås 0, så $\arcsin 0 = 0$. Alltså är $\arcsin(\sin 16\pi)$ inte 16π, utan 0.

Vi har alltså (ta sinus av de båda lika talen):

$$x = \arcsin y \qquad \Longrightarrow \qquad \sin x = y$$

[17] Vardagsanalogi: Välj ut *en* av alla svenskar som är 50 år. Hur gammal är hon du valde ut?

men INTE omvändningen. Sista ekvationen har ju (för t.ex. $y = \frac{1}{2}$) *oändligt* många lösningar, medan x i första ekvationen måste vara exakt $\arcsin y$ (d.v.s. $\frac{\pi}{6}$ om $y = \frac{1}{2}$).

Motsvarande observationer kan göras för cos/arccos och för tan/arctan.

Exempel

Bestäm $\arcsin \sin \frac{35\pi}{6}$.

Lösning (metod 1): Vi har

$$\arcsin \sin \frac{35\pi}{6} = \arcsin \sin \left(\frac{36\pi}{6} - \frac{\pi}{6}\right) = \arcsin \sin \left(6\pi - \frac{\pi}{6}\right) = \arcsin \sin \left(-\frac{\pi}{6}\right) = -\frac{\pi}{6}$$

eftersom $-\frac{\pi}{6} \in [-\frac{\pi}{2}, \frac{\pi}{2}]$. [Alternativt kan man förstås på valfritt sätt inse att sinusvärdet är $-\frac{1}{2}$.]

1.6.3.3 Trigonometriska ekvationer, igen

Med hjälp av arcusfunktionerna kan vi mycket snyggare än tidigare konstatera att

$$\sin x = a \quad\Longleftrightarrow\quad \begin{cases} x = \arcsin a + n \cdot 2\pi \\ \text{eller} \\ x = \pi - \arcsin a + n \cdot 2\pi, \end{cases}$$

$$\cos x = a \quad\Longleftrightarrow\quad x = \pm \arccos a + n \cdot 2\pi$$

samt

$$\tan x = a \quad\Longleftrightarrow\quad x = \arctan a + n\pi.$$

Jämför med de klumpigare formuleringarna i avsnittet *Trigonometriska ekvationer* på sidan 99. Den stora vinsten, emellertid, är att vi nu kan använda symboler i stil med $\arcsin \frac{8}{11}$ när vi inte kan eller behöver bestämma ett exakt uttryck för värdet [samma sak gäller ju t.ex. symbolen $\sqrt{2}$, som är lättare att skriva än ”det entydigt bestämda icke-negativa tal vars kvadrat är lika med 2”]:

Exempel

Lös ekvationen $\tan^2 x - 3 \tan x = 1$.

Lösning: Kvadratkomplettering ger

$$\begin{aligned} \tan^2 x - 3\tan x = \left(\tan x - \frac{3}{2}\right)^2 - \frac{9}{4} = 1 \quad &\Longleftrightarrow \\ \Longleftrightarrow \quad \left(\tan x - \frac{3}{2}\right)^2 = \frac{13}{4} \quad &\Longleftrightarrow \\ \Longleftrightarrow \quad \tan x - \frac{3}{2} = \pm\frac{\sqrt{13}}{2} \quad &\Longleftrightarrow \\ \Longleftrightarrow \quad \tan x = \frac{3+\sqrt{13}}{2} \quad \text{eller} \quad \tan x = \frac{3-\sqrt{13}}{2}&. \end{aligned}$$

I första fallet har vi

$$x = \arctan\frac{3+\sqrt{13}}{2} + n\pi$$

för något $n \in \mathbb{Z}$ och i andra fallet har vi

$$x = \arctan\frac{3-\sqrt{13}}{2} + n\pi$$

för något $n \in \mathbb{Z}$.

Svar: $\tan^2 x - 3\tan x = 1 \Leftrightarrow x = \arctan\frac{3+\sqrt{13}}{2} + n\pi$ eller $x = \arctan\frac{3-\sqrt{13}}{2} + n\pi$ för något $n \in \mathbb{Z}$.

1.6.4 Fler trigonometriska exempel

Exempel

Beräkna $\arctan 2 + \arctan 3$.

Lösning: Vi bestämmer först

$$\tan(\arctan 2 + \arctan 3) = \frac{\tan\arctan 2 + \tan\arctan 3}{1 - \tan\arctan 2\tan\arctan 3} = \frac{2+3}{1-2\cdot 3} = -1.$$

Därför är

$$\arctan 2 + \arctan 3 = -\frac{\pi}{4} + n\pi$$

för något $n \in \mathbb{Z}$. Eftersom $\arctan 2$ och $\arctan 3$ båda ligger mellan $\frac{\pi}{4}$ och $\frac{\pi}{2}$ (varför?) är det klart att deras summa ligger mellan $\frac{\pi}{2}$ och π. Därför måste $n = 1$ och

$$\arctan 2 + \arctan 3 = -\frac{\pi}{4} + \pi = \frac{3\pi}{4}.$$

Svar: $\arctan 2 + \arctan 3 = \frac{3\pi}{4}$.

Ett vanligt fel

Det händer förhållandevis ofta att studenter anger

$$\arctan 2 + \arctan 3 = -\frac{\pi}{4} + n\pi \quad (n \text{ heltal})$$

som *svar* på en dylik uppgift. Det kan tolkas på två sätt. Antingen vet studenten att summan $\arctan 2 + \arctan 3$ är *ett* av talen $-\frac{\pi}{4} + n\pi$ (d.v.s. det som erhålles för *något* heltal n), men vet inte hur man bestämmer *vilket*, och hoppas alltså på något halvt poäng.

Den andra tolkningen är att studenten tror att $\arctan 2 + \arctan 3$ faktiskt betyder *alla* dessa tal, vilket är värre. Arctan är ju en funktion, som man stoppar in *ett* tal i, varvid man får ut *ett* tal. Så $\arctan 2$ och $\arctan 3$ är båda bestämda tal (1.10714871779… resp. 1.2490457724…). Alltså är deras summa *ett* bestämt tal, i det här fallet talet $\frac{3\pi}{4} = 2.35619449019\ldots$.

Exempel

Lös ekvationen $\tan 2x \tan x = 1$.

Lösning: Notera att $\tan x$ är definierat om ekvationen är uppfylld. Därför har vi

$$\tan 2x \tan x = 1 \iff \frac{2\tan x}{1-\tan^2 x}\tan x = 1 \iff \frac{2\tan^2 x}{1-\tan^2 x} = 1 \iff 2\tan^2 x = 1-\tan^2 x \iff 3\tan^2 x = 1 \iff \tan^2 x = \frac{1}{3} \iff \tan x = \pm\frac{1}{\sqrt{3}} \iff \left\{\begin{array}{c}\tan x = \frac{1}{\sqrt{3}} \\ \text{eller} \\ \tan x = -\frac{1}{\sqrt{3}}\end{array}\right. \iff \left\{\begin{array}{c} x = \frac{\pi}{6} + n\pi \\ \text{eller} \\ x = -\frac{\pi}{6} + n\pi. \end{array}\right.$$

Svar: $\tan 2x \tan x = 1 \iff x = \pm\frac{\pi}{6} + n\pi$ för något $n \in \mathbb{Z}$.

Exempel

Lös ekvationen $\sin^2 x + \tan^2 x = 1$.

Lösning: Eftersom

$$\tan^2 x = \frac{\sin^2 x}{\cos^2 x} = \frac{\sin^2 x}{1-\sin^2 x}$$

lyder ekvationen

$$\sin^2 x + \frac{\sin^2 x}{1-\sin^2 x} = 1.$$

Med $s := \sin x$ har vi därför

$$s^2 + \frac{s^2}{1-s^2} = 1 \iff \frac{s^2}{1-s^2} = 1 - s^2 \iff s^2 = 1 - 2s^2 + s^4 \iff s^4 - 3s^2 + 1 = 0 \iff \left(s^2 - \frac{3}{2}\right)^2 - \frac{9}{4} + 1 = 0 \iff \left(s^2 - \frac{3}{2}\right)^2 = \frac{5}{4} \iff s^2 - \frac{3}{2} = \pm\frac{\sqrt{5}}{2} \iff s^2 = \frac{3 \pm \sqrt{5}}{2} \iff s^2 = \frac{3-\sqrt{5}}{2}$$

där sista steget kommer av att $s \in [-1, 1] \Rightarrow s^2 \in [0, 1]$. Slutligen har vi

$$s^2 = \frac{3-\sqrt{5}}{2} \qquad \Longleftrightarrow \qquad s = \pm\frac{\sqrt{3-\sqrt{5}}}{\sqrt{2}} \in [-1, 1].$$

Om vi inför beteckningen

$$\xi \coloneqq \frac{\sqrt{3-\sqrt{5}}}{\sqrt{2}} \in \,]0, 1[$$

har vi sålunda (där sista ledet kommer av att arcsin är udda)

$$s = \sin x = \pm\xi \qquad \Longleftrightarrow \qquad \begin{cases} \sin x = \xi \\ \text{eller} \\ \sin x = -\xi \end{cases} \qquad \Longleftrightarrow$$

$$\Longleftrightarrow \qquad \begin{cases} x = \arcsin\xi + n\cdot 2\pi \\ \text{eller} \\ x = \pi - \arcsin\xi + n\cdot 2\pi \\ \text{eller} \\ x = \arcsin(-\xi) + n\cdot 2\pi \\ \text{eller} \\ x = \pi - \arcsin(-\xi) + n\cdot 2\pi \end{cases} \qquad \Longleftrightarrow$$

$$\Longleftrightarrow \qquad \begin{cases} x = \pm\arcsin\xi + n\cdot 2\pi \\ \text{eller} \\ x = \pi \pm \arcsin\xi + n\cdot 2\pi\,. \end{cases}$$

Vi får därför fyra familjer av lösningar.

Svar: $\sin^2 x + \tan^2 x = 1 \qquad \Longleftrightarrow \qquad \begin{cases} x = \pm\arcsin\xi + n\cdot 2\pi \quad \text{för något} \quad \in \mathbb{Z} \\ \text{eller} \\ x = \pi \pm \arcsin\xi + n\cdot 2\pi \quad \text{för något} \quad n \in \mathbb{Z} \end{cases}$

där $\xi \coloneqq \sqrt{3-\sqrt{5}}/\sqrt{2}$.

Ekvationen $\sin x = \pm\xi$ ger förstås de vinklar x vilka erhålles då enhetscirkeln skärs med de två linjerna $y = \pm\xi$, där $\xi \in \,]0,1[$. Fyra familjer av lösningar är alltså att vänta.

Följande **viktiga** skolboksräkning kan ses som ett specialfall av exemplen i avsnittet *Samband mellan olika funktioners värden för samma vinkel* på sidan 105.

Exempel

Förenkla $\cos\arcsin x$.

Lösning: Eftersom $\arcsin x \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$ är $\cos\arcsin x \geq 0$ så att

$$\cos\arcsin x = \sqrt{1 - \sin^2\arcsin x} = \sqrt{1 - x^2}.$$

1.7 Potenser och logaritmer

Vad betyder a^b, d.v.s. vad innebär ”upphöjt till”? Det är en fråga som man måhända tycker borde vara självklar för alla som studerar matematik på högskolan. Men frågan är knepigare än vad den vid en första anblick verkar, och det har kanske främst att göra med att de reella talens egenskaper utgör en pedagogisk svårighet i inledande matematiklitteratur.

Visst, *om* $b \in \mathbb{Z}^+$ är ett positivt heltal så är det klart att

$$a^b = \underbrace{a \cdot a \cdot \dots \cdot a}_{b \text{ faktorer}}.$$

Däremot är ju ”nästan inga” av elementen i $\mathbb{R}$ positiva heltal! Låt oss först repetera den klassiska metoden för att definiera a^b för varje $a > 0$ och $b \in \mathbb{R}$.

1. Låt $a^n = \underbrace{a \cdot a \cdot \dots \cdot a}_{n \text{ faktorer}}$ för varje $n \in \mathbb{Z}^+$.
2. Låt, för varje $n \in \mathbb{Z}^+$, $a^{1/n}$ vara det entydigt bestämda icke-negativa tal som upphöjt till n är a.
3. Sätt, för varje $p, q \in \mathbb{Z}^+$, $a^{p/q} = (a^p)^{\frac{1}{q}}$. Vi har nu definierat a^b för varje $b \in \mathbb{Q} \cap \mathbb{R}^+$.
4. Sätt $a^0 = 1$ och $a^{-r} = \frac{1}{a^r}$ för varje $r \in \mathbb{Q} \cap \mathbb{R}^+$. Vi har nu definierat a^b för varje $b \in \mathbb{Q}$.

Konstruktionen ovan bör vara bekant för studenter med gymnasiematematiken färskt i minnet.

Men vad menas då med a^b om $b \notin \mathbb{Q}$; vad menas t.ex. med 2^π? Som bekant är

$$\pi = 3.14159265358979323846 26\ldots$$

och vi kan approximera π godtyckligt nära med rationella tal genom att ta med godtyckligt många decimaler från denna oändliga följd av decimaler. Elementen i (den växande) följden

$$3.1, 3.14, 3.141, 3.1415, 3.14159, 3.141592, 3.1415926, \ldots$$

kommer ju allt närmare π i den meningen att differensen mellan π och elementen i följden blir mindre och mindre. Vi får då också en följd (också växande)

$$2^{3.1}, 2^{3.14}, 2^{3.141}, 2^{3.1415}, 2^{3.14159}, 2^{3.141592}, 2^{3.1415926}, \ldots$$

som blir en bättre och bättre approximation av vad vi rimligtvis *bör* mena med symbolen ”2^π”. I själva verket definieras 2^π som det s.k. *gränsvärdet* av följden. Att det ens existerar ett sådant har med de reella talens egenskaper att göra, vilka vi inte går in på här i detalj (det s.k. *supremumaxiomet* är väsentligt).

Nu har vi sålunda definierat a^b för varje $a > 0$ och $b \in \mathbb{R}$.[18] Speciellt kan man införa den matematiska konstanten $e \in \mathbb{R} \setminus \mathbb{Q}$ (som torde vara bekant sedan tidigare: $e = 2.71828182846\ldots$) och låta $\exp: \mathbb{R} \to \mathbb{R}^+, x \mapsto e^x$ vara den s.k. *exponentialfunktionen* som är strängt växande med den s.k. *naturliga logaritmfunktionen* $\ln: \mathbb{R}^+ \to \mathbb{R}$ som invers funktion. Det visar sig då dels att $\ln(t)$ kan tolkas

[18] Man brukar också sätta $0^b = 0$ för varje $b \in \mathbb{R}^+$ samt – självklart – tillåta godtyckliga baser $a \in \mathbb{R}$ om $b \in \mathbb{Z}$, med det enda undantaget att 0^0 är odefinierat.

som ”arean” under kurvan $1/x$ från $x = 1$ till $x = t$, dels att vi har identiteten $a^b = e^{b \ln a}$ (om $a > 0$).

Det sista steget i den klassiska definitionen av a^b, från $b \in \mathbb{Q}$ till $b \in \mathbb{R}$, är alltså något bökigt, och sällan något man ägnar någon större tid åt t.ex. i gymnasieskolan, eller ens i inledande kurser på hög-skolenivå. Det finns dock ett alternativt sätt att definiera ”upphöjt till”, som ger upphov till precis samma begrepp:

1. Definiera den naturliga logaritmfunktionen $\ln: \mathbb{R}^+ \to \mathbb{R}$ genom att låta $\ln(t)$ vara ”arean” under kurvan $y = \frac{1}{x}$ från $x = 1$ till $x = t$. Om $t < 1$ låter man i stället $\ln(t)$ vara *minus* arean under kurvan från $x = t$ till $x = 1$.
2. Det följer omedelbart att ln är strängt växande. Låt exp: $\mathbb{R} \to \mathbb{R}^+$, exponentialfunktionen, vara dess invers. Låt $e \coloneqq \exp(1)$.
3. Definiera nu, för varje $a > 0$ och $b \in \mathbb{R}$, $a^b \coloneqq e^{b \ln a}$.

(Sedan lägger man till ”utökningarna”, t.ex. $a \notin \mathbb{R}^+$ men $b \in \mathbb{Z}^+$, på samma sätt som tidigare.) Här ligger egenskaperna hos de reella talen i stället gömda i första steget, nämligen i begreppet ”area”, eller, egentligen ”integral”. På sätt och vis har vi alltså bara flyttat problemet, men i praktiken är den här alternativa metoden ändå tilltalande, om inte annat för att den ser lättare ut (den består åt-minstone av färre steg). Dessutom är den här metoden mycket praktisk i många situationer, eftersom uttrycket $e^{b \ln a}$ på många sätt är lättare att arbeta med än a^b. Till viss del beror nog detta på att $e^{b \ln a} = \exp(b \ln a)$ är en sammansättning av funktioner av *en* variabel, medan uttrycket a^b i sig är en funktion av *två* variabler.

Naturligtvis går inte fördelarna med uttrycket $e^{b \ln a}$ förlorade om vi i stället baserar vår teori på den första definitionen av ”upphöjt till”, eftersom definitionerna är ekvivalenta. I första metoden gäller ändå $a^b = e^{b \ln a}$, även om formeln i det fallet är en *sats*, och inte en *definition*, som i andra meto-den. Denna distinktion, den mellan sats och definition, utgör emellertid en fördel i elegans för den andra metoden: Att gå till ett objekts definition för att visa dess egenskaper är *uppenbart*, att gå till en sats bara *tänkbart*.

Följande logaritmlagar måste man kunna (de två första kräver $a, b \in \mathbb{R}^+$; den senare kräver $a \in \mathbb{R}^+$):

- $\ln ab = \ln a + \ln b$
- $\ln \frac{a}{b} = \ln a - \ln b$
- $\ln a^b \left(= \ln\left(a^b\right)\right) = b \ln a.$

Dessa är enkla att inse om man använder den ”klassiska” approachen. Till exempel har vi ju att

$$e^{\ln a + \ln b} = e^{\ln a} e^{\ln b} = ab = e^{\ln ab}$$

så att $\ln a + \ln b = \ln ab$ (ty exponentialfunktionen är injektiv).

Övning

1. Visa de två andra logaritmlagarna.

Ett vanligt fel

Det händer inte sällan att studenter hittar på egna räknelagar, vilka inte är sanna. Exempel på vanligt förekommande, men falska, identiteter är $\ln(a+b) = \ln a \cdot \ln b$, $\ln(a+b) = \ln a + \ln b$ och $\ln ab = \ln a \cdot \ln b$. Att de två sista identiteterna *inte* gäller betyder att logaritmfunktionen *inte* är "additiv" eller "multiplikativ". Även den tämligen uppenbara (varför?) tokigheten $(\ln a)^b = b \ln a$ ses ibland.

Det är inte bara nya "logaritmlagar" som hittas på. Tokigheter som $\sqrt{a+b} = \sqrt{a} + \sqrt{b}$, $(a+b)^2 = a^2 + b^2$ och $\sin(a+b) = \sin a + \sin b$ ses ofta. Studenter tror att allt är additivt och linjärt!

Man måste vara försiktig med logaritmidentiteterna. Uttrycket $\ln ab$ är ju definierat när $ab > 0$, d.v.s. när ab har samma (nollskilda) tecken, d.v.s. omm

$$\begin{cases} a > 0 \\ b > 0 \end{cases} \quad \text{eller} \quad \begin{cases} a < 0 \\ b < 0. \end{cases}$$

Högerledet $\ln a + \ln b$ är däremot bara definierat i första fallet. Liknande förhållanden gäller för andra och tredje identiteten: $\ln \frac{-2}{-3} = \ln \frac{2}{3}$ är förstås definierat, men inte $\ln(-2) - \ln(-3)$; $\ln(-5)^2 = \ln 25$ är förstås definierat, men inte $2\ln(-5)$. Däremot, om högerleden är definierade, så är också vänsterleden det: om $a > 0$ och $b > 0$ så är förstås $ab > 0$ och $\frac{a}{b} > 0$, och om $a > 0$ och $b \in \mathbb{R}$ så är $a^b > 0$.

Exempel

Bestäm definitionsmängden till funktionen

$$f : x \mapsto \ln\left(\frac{1-x}{2-x}\right).$$

Lösning: Funktionsuttrycket är definierat omm

$$\frac{1-x}{2-x} > 0 \qquad \Longleftrightarrow \qquad x \notin [1,2].$$

Svar: $D_f = [1,2]^C$.

Följande lösning är däremot uppenbart felaktig, eftersom svaret är felaktigt:

"Lösning": Eftersom

$$\ln\left(\frac{1-x}{2-x}\right) = \ln(1-x) - \ln(2-x)$$

ser vi att uttrycket är definierat precis då $1 - x > 0$ och $2 - x > 0$, d.v.s. precis då $x < 1$ och $x < 2$, d.v.s. precis då $x < 1$.

Svar: $D_f =]-\infty, 1[$.

När man använder logaritmidentiteterna måste man alltså vara observant på att deras vänsterled är definierade för många fler värden på a och b än högerleden. Ett specialfall av detta, som vi just såg, är att man vid bestämning av definitionsmängden av ett logaritmuttryck inte kan använda dessa lagar för omskrivning. Även vid ekvationslösning måste man vara försiktig. Till exempel är ekvationen $y = \ln ab$ inte ekvivalent med $y = \ln a + \ln b$; den förra kan ju ha lösningar där både a och $b < 0$.

1.7.1 Ekvationslösning

Exempel

Lös ekvationen $2^x = 5$.

Lösning: Vi har

$$2^x = 5 \quad \Longleftrightarrow \quad \ln 2^x = \ln 5 \quad \Longleftrightarrow \quad x \ln 2 = \ln 5 \quad \Longleftrightarrow \quad x = \frac{\ln 5}{\ln 2}.$$

Svar: $2^x = 5 \quad \Longleftrightarrow \quad x = \ln 5 / \ln 2.$

Ett alternativt – och kanske mer naturligt – sätt att lösa föregående ekvation på är att ställa sig frågan vad ”2^x” egentligen betyder. Jo, symbolen 2^x betyder ju exakt $e^{x \ln 2}$, så ekvationen lyder egentligen $e^{x \ln 2} = 5 \Leftrightarrow x \ln 2 = \ln 5 \Leftrightarrow x = \ln 5 / \ln 2$.

Exempel

Lös ekvationen $2^{2x+1} - 2^{x-1} + 4^x = 1$.

Lösning: Med $t \coloneqq 2^x$ har vi

$$2^{2x+1} - 2^{x-1} + 4^x = 2^{2x} \cdot 2 - \frac{2^x}{2} + (2^2)^x = (2^x)^2 \cdot 2 - \frac{2^x}{2} + (2^x)^2 = 2t^2 - \frac{1}{2}t + t^2 = 3t^2 - \frac{1}{2}t$$

så att ekvationen är ekvivalent med

$$\begin{aligned}
3t^2 - \frac{1}{2}t = 1 \quad &\Longleftrightarrow \quad t^2 - \frac{1}{6}t = \frac{1}{3} \quad \Longleftrightarrow \\
&\Longleftrightarrow \quad \left(t - \frac{1}{12}\right)^2 - \frac{1}{144} = \frac{1}{3} \quad \Longleftrightarrow \\
&\Longleftrightarrow \quad \left(t - \frac{1}{12}\right)^2 = \frac{49}{144} \quad \Longleftrightarrow \\
&\Longleftrightarrow \quad t - \frac{1}{12} = \pm \frac{7}{12} \quad \Longleftrightarrow \\
&\Longleftrightarrow \quad t = \frac{2}{3} \quad \text{eller} \quad t = -2 \quad \Longleftrightarrow \quad t = \frac{2}{3}
\end{aligned}$$

eftersom $t = 2^x > 0$. Ekvationen är alltså ekvivalent med

$$\begin{aligned}
2^x = \frac{2}{3} \quad &\Longleftrightarrow \quad \ln 2^x = \ln \frac{2}{3} \quad \Longleftrightarrow \\
&\Longleftrightarrow \quad x \ln 2 = \ln 2 - \ln 3 \quad \Longleftrightarrow \\
&\Longleftrightarrow \quad x = \frac{\ln 2 - \ln 3}{\ln 2} = 1 - \frac{\ln 3}{\ln 2}.
\end{aligned}$$

Svar: $2^{2x+1} - 2^{x-1} + 4^x = 1 \quad \Longleftrightarrow \quad x = 1 - \frac{\ln 3}{\ln 2}.$

Exempel

Bestäm de $x > 0$ sådana att $x^x = x$.

Lösning: Per definition är $x^x = e^{x \ln x}$ så att ekvationen lyder

$$e^{x \ln x} = x \iff x \ln x = \ln x \iff (x-1)\ln x = 0 \iff x = 1 \text{ eller } \ln x = 0 \iff x = 1.$$

Svar: Om $x > 0$ är $x^x = x \iff x = 1$.

Exempel (*)

Bestäm de talpar $(x, y) \in \mathbb{R}^+ \times \mathbb{R}$ sådana att $x^y = 1$.

Lösning: Per definition är $x^y = e^{y \ln x}$ så att ekvationen lyder $e^{y \ln x} = 1$ vilket inträffar precis då $y \ln x = 0$, d.v.s. precis då $y = 0$ eller $\ln x = 0$, d.v.s. precis då $y = 0$ eller $x = 1$.

Svar: Om $(x, y) \in \mathbb{R}^+ \times \mathbb{R}$ så gäller $x^y = 1 \iff x = 1$ eller $y = 0$.

Exempel

Lös ekvationen $\ln(x+2) - \ln(x-2) = 10$.

Lösning: Vänsterledet är bara definierat då $x > 2$ och för dessa x gäller

$$\ln(x+2) - \ln(x-2) = 10 \iff \ln\frac{x+2}{x-2} = 10 \iff \frac{x+2}{x-2} = e^{10} \iff x + 2 = e^{10}(x-2) \iff (1-e^{10})x = -2 - 2e^{10} \iff x = \frac{2+2e^{10}}{e^{10}-1} (> 2).$$

Svar: $\ln(x+2) - \ln(x-2) = 10 \iff x = \frac{2+2e^{10}}{e^{10}-1}$.

Eftersom den erhållna lösningen ligger i $]-2, \infty[$ är den korrekt. Ett mer pedantiskt, men egentligen tydligare, sätt att redovisa lösningen på är som följer:

Lösning:

$$\ln(x+2) - \ln(x-2) = 10 \iff \begin{cases} \ln\frac{x+2}{x-2} = 10 \\ x > 2 \end{cases} \iff \begin{cases} \frac{x+2}{x-2} = e^{10} \\ x > 2 \end{cases} \iff \begin{cases} x + 2 = e^{10}(x-2) \\ x > 2 \end{cases} \iff \begin{cases} (1-e^{10})x = -2 - 2e^{10} \\ x > 2 \end{cases} \iff \begin{cases} x = \frac{2+2e^{10}}{e^{10}-1} \\ x > 2 \end{cases} \iff$$

$$\iff \qquad x = \frac{2 + 2e^{10}}{e^{10} - 1}.$$

Påståendet $x > 2$ är ju en följd av $x = \frac{2+2e^{10}}{e^{10}-1}$, så konjunktionen (OCH:et) mellan dessa utsagor är ekvivalent med bara $x = \frac{2+2e^{10}}{e^{10}-1}$, på samma sätt som påståendet ”Han är i Sverige och han är i Stockholm” är ekvivalent med ”Han är i Stockholm”.

Exempel

Lös ekvationen $\ln(x+2) - 2\ln x = -1$.

Lösning: Vänsterledet är definierat för $x > 0$ och för dessa x är

$$\ln(x+2) - 2\ln x = -1 \iff \ln\frac{x+2}{x^2} = -1 \iff$$

$$\iff \frac{x+2}{x^2} = e^{-1} \iff x + 2 = e^{-1}x^2 \iff$$

$$\iff x^2 - ex = 2e \iff \left(x - \frac{e}{2}\right)^2 - \frac{e^2}{4} = 2e \iff$$

$$\iff \left(x - \frac{e}{2}\right)^2 = \frac{1}{4}e(8+e) \iff x - \frac{e}{2} = \pm\frac{1}{2}\sqrt{8e + e^2} \iff$$

$$\iff x = \frac{e}{2} \pm \frac{1}{2}\sqrt{8e + e^2} \iff x = \frac{e}{2} + \frac{1}{2}\sqrt{8e + e^2}$$

med tanke på att $x > 0$ [notera att $\sqrt{8e + e^2} > \sqrt{e^2} = e$].

Svar: $\ln(x+2) - 2\ln x = -1 \iff x = \frac{1}{2}\left(e + \sqrt{8e + e^2}\right)$.

Precis som i föregående exempel bör redovisningen ovan tolkas som en förkortad variant av följande, mer explicita (och eleganta) lösningsgång:

Lösning:

$$\ln(x+2) - 2\ln x = -1 \iff \begin{cases} \ln\dfrac{x+2}{x^2} = -1 \\ x > 0 \end{cases} \iff$$

$$\iff \begin{cases} \dfrac{x+2}{x^2} = e^{-1} \\ x > 0 \end{cases} \iff \begin{cases} x + 2 = e^{-1}x^2 \\ x > 0 \end{cases} \iff$$

$$\iff \begin{cases} x^2 - ex = 2e \\ x > 0 \end{cases} \iff \begin{cases} \left(x - \dfrac{e}{2}\right)^2 - \dfrac{e^2}{4} = 2e \\ x > 0 \end{cases} \iff$$

$$\iff \begin{cases} \left(x - \dfrac{e}{2}\right)^2 = \dfrac{1}{4}e(8+e) \\ x > 0 \end{cases} \iff \begin{cases} x - \dfrac{e}{2} = \pm\dfrac{1}{2}\sqrt{8e + e^2} \\ x > 0 \end{cases} \iff$$

$$\iff \begin{cases} x = \dfrac{e}{2} \pm \dfrac{1}{2}\sqrt{8e + e^2} \\ x > 0 \end{cases} \iff x = \frac{e}{2} + \frac{1}{2}\sqrt{8e + e^2}.$$

Svar: $\ln(x+2) - 2\ln x = -1 \iff x = \frac{1}{2}\left(e + \sqrt{8e + e^2}\right)$.

1.8 Abstrakt algebra*

Det här avsnittet har inte skrivits än. Här kommer jag att på ett mycket enkelt och tydligt sätt introducera följande centrala begrepp:

- Kompositionsregel
- Algebraisk struktur
- Magma, semigrupp, monoid
- Grupp, ring, kropp
- Modul, vektorrum
- Relationer, ekvivalensrelationer

Efter att jag skrivit det här avsnittet kommer jag att skriva om den asteriskmarkerade introduktionen till de komplexa talen. Den nya versionen av det avsnittet kommer att bli mycket kortare (och mer överskådlig).

1.9 Komplexa tal

I det här kapitlet introducerar vi de komplexa talen. Först ger vi en informell introduktion där vi mest säger vad det hela går ut på och sedan ger vi en rigorös introduktion där inga detaljer utelämnas. Lägg märke till att den rigorösa introduktionen är *-märkt.

1.9.1 Informell introduktion av de komplexa talen

1.9.1.1 Algebraiska strukturer

Vi är väl förtrogna med de naturliga talen, d.v.s. elementen i

$$\mathbb{N} = \{0, 1, 2, 3, \dots\}.$$

Detta är *inte* bara en mängd av symboler, utan vi har dessutom två regler, addition och multiplikation, som till varje *par* av naturliga tal ordnar ett nytt naturligt tal. Till exempel är $5 + 7 = 12$ och $4 \cdot 6 = 24$. Vi säger då att $(\mathbb{N}, +, \cdot)$ är en *algebraisk struktur*. En algebraisk struktur är i allmänhet en mängd (som inte nödvändigtvis måste bestå av "tal") tillsammans med en eller flera räkneoperationer, s.k. *kompositionsregler*. Om mängden heter X är en kompositionsregel en funktion $X \times X \to X$.

En större mängd är mängden av alla heltal,

$$\mathbb{Z} = \{\dots, -3, -2, -1, 0, 1, 2, 3, \dots\}$$

som också är en algebraisk struktur med addition och multiplikation. Ännu större är mängden av alla rationella tal, $\mathbb{Q}$, d.v.s. mängden av alla tal på formen $\frac{p}{q}$ där p och q är heltal och $q \neq 0$. Även här har vi addition och multiplikation. Slutligen har vi mängden $\mathbb{R}$ av alla reella tal, som utöver de rationella talen också inkluderar de irrationella talen på tallinjen, t.ex. $\sqrt{2}$ och π. Även $(\mathbb{R}, +, \cdot)$ är förstås en algebraisk struktur.

Ett exempel på en mängd som *inte* (åtminstone inte *a priori*) är en algebraisk struktur är mängden

$$D = \{\text{hund}, \text{katt}, \text{råtta}\}$$

bestående av tre trevliga tamdjur: vi har ju inte definierat någon kompositionsregel på mängden (t.ex. har vi inte sagt vad hund $+$ katt är – är det hund, katt eller råtta?).

Som *mängder* gäller som bekant att

$$\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R}.$$

Däremot gäller även att t.ex. $\mathbb{N}$ är en *understruktur* till $\mathbb{Z}$. Det betyder bara att kompositionsreglerna $+$ och $\cdot$ i $\mathbb{N}$ är restriktionerna till $\mathbb{N}$ av motsvarande kompositionsregler i $\mathbb{Z}$. Till exempel är ju summan av de två *heltalen* 7 och 5 lika med *heltalet* 12, och summan av de två *naturliga talen* 7 och 5 är lika med det *naturliga talet* 12. Samma sak gäller för varje par av strukturer i inklusionskedjan ovan.

De komplexa talen är i princip en superstruktur till $\mathbb{R}$, d.v.s. en algebraisk struktur $(\mathbb{C}, +, \cdot)$ som har $\mathbb{R}$ som en delstruktur. Som mängder gäller alltså $\mathbb{R} \subset \mathbb{C}$, och dessutom finns kompositionsregler kallade "addition" och "multiplikation" i $\mathbb{C}$, och deras restriktioner till $\mathbb{R}$ är den vanliga additionen resp. multiplikationen i $\mathbb{R}$.

1.9.1.2 Exakt vad är $\mathbb{C}$?

Nu när vi vet vad för ”sorts” objekt $\mathbb{C}$ är, är vi mogna att ta reda på *exakt* vad det är.

Om vi bara kan arbeta i $\mathbb{N}$ stöter vi snart på problem: ekvationen $x + 3 = 1$ saknar nämligen lösning. I $\mathbb{Z}$, däremot, har ekvationen lösningen -2. Men även i $\mathbb{Z}$ finns liknande problem, t.ex. saknar ekvationen $2x = 1$ lösning. I $\mathbb{Q}$ har samma ekvation lösningen $1/2$. Men även $\mathbb{Q}$ har sina fläckar, t.ex. beträffande ekvationen $x^2 = 2$. I $\mathbb{R}$ har denna ekvation lösningen $\sqrt{2}$. Inte ens i denna stora mängd har emellertid *varje* polynomekvation en lösning. Till exempel saknar ekvationen $x^2 = -1$ reella lösningar. I $\mathbb{C}$, däremot, har även denna ekvation en lösning. Faktum är att *varje* icke-konstant polynom med komplexa koefficienter har minst ett nollställe i $\mathbb{C}$. Så man kan säga att

$$\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R} \subset \mathbb{C}$$

är så långt man ”behöver” gå om man är intresserad av polynomekvationer.

Vi vill alltså att varje polynomekvation skall ha en lösning i $\mathbb{C}$, så t.ex. $z^2 = -1$ skall gälla för något $z \in \mathbb{C}$. Om vi kallar *en* av ekvationens lösningar för i så har vi $i^2 = -1$. Eftersom $(\mathbb{C}, +, \cdot)$ skall vara en superstruktur till $\mathbb{R}$ måste vi också kunna beräkna $b \cdot i$ för varje $b \in \mathbb{R}$ och dessutom måste vi kunna bilda $a + b \cdot i$ för varje $a, b \in \mathbb{R}$. Alla tal på formen $a + bi$, med $a, b \in \mathbb{R}$ måste med andra ord vara komplexa tal. Om vi dessutom vill att samma räknelagar[19] skall gälla i $\mathbb{C}$ som i $\mathbb{R}$ (beträffande kompositionsreglerna $+$ och $\cdot$) måste

$$(a_1 + b_1 i) + (a_2 + b_2 i) = (a_1 + a_2) + (b_1 + b_2)i$$

och

$$\begin{aligned}(a_1 + b_1 i) \cdot (a_2 + b_2 i) &= a_1 a_2 + a_1 b_2 i + b_1 a_2 i + b_1 b_2 i^2 = a_1 a_2 + a_1 b_2 i + b_1 a_2 i + b_1 b_2 (-1) = \\ &= (a_1 a_2 - b_1 b_2) + (a_1 b_2 + a_2 b_1)i.\end{aligned}$$

Det visar sig att om vi *definierar* ett komplext tal som ett formellt uttryck på formen $a + bi$ där $a, b \in \mathbb{R}$, d.v.s. i princip som ett ***ordnat par*** (a, b) av två reella tal, och sedan *definierar* kompositionsreglerna $+ : \mathbb{C} \times \mathbb{C} \to \mathbb{C}$ och $\cdot : \mathbb{C} \times \mathbb{C} \to \mathbb{C}$ enligt ovan så får vi precis vad som inom matematiken kallas för *de komplexa talen* $(\mathbb{C}, +, \cdot)$. Man kan visa att denna algebraiska struktur uppfyller samma räknelagar beträffande addition och multiplikation som gäller för $(\mathbb{R}, +, \cdot)$. Dessutom framgår av de två formlerna för addition och multiplikation i $\mathbb{C}$ att de komplexa talen på formen $a + 0i$ fungerar precis som de reella talen, ty för sådana tal har vi

$$\begin{aligned}(a_1 + 0i) + (a_2 + 0i) &= (a_1 + a_2) + 0i \\ (a_1 + 0i) \cdot (a_2 + 0i) &= (a_1 a_2) + 0i\end{aligned}$$

så vi skriver då kort och gott a även när vi menar det komplexa talet $a + 0i$. Vi får då

$$\begin{aligned}a_1 \oplus a_2 &= a_1 + a_2 \\ a_1 \odot a_2 &= a_1 \cdot a_2\end{aligned}$$

[19] Exempel: Additionen och multiplikationen skall vara kommutativa och associativa, och multiplikationen skall distribuera över additionen.

där vi för tydlighetens skull använde symbolerna $\oplus$ och $\odot$ för kompositionsreglerna i $\mathbb{C}$. Här framgår att restriktionen av kompositionsreglerna i $\mathbb{C}$ till de ”reella talen” är de vanliga kompositionsreglerna i $\mathbb{R}$. De reella talen $(\mathbb{R},+,\cdot)$ är med andra ord en delstruktur till $(\mathbb{C},+,\cdot)$ med den här identifikationen.

Slutligen kan man visa att *varje* polynomekvation med komplexa koefficienter har minst en lösning, som vi lovade.

1.9.1.3 What's the point?

Vad är då poängen med att utvidga $\mathbb{R}$ till $\mathbb{C}$? Om man är intresserad av abstrakt algebra och polynomekvationer är det klart att $\mathbb{C}$ är av intresse, men annars? Det är ju inte särdeles ofta man i tekniska eller vetenskapliga tillämpningar stöter på storheter som uppenbart naturligt beskrivs med icke-reella tal; t.ex. är det sällan man ser en stång av längd $2+3i$ eller en termos med temperaturen $1-i$. Nej, poängen är faktiskt till stor del den att många problem av *reell* natur kan lösas med metoder från komplex algebra och analys i mellanstegen. Vi kommer att se några exempel på det framöver.

1.9.1.4 Ett par räkneexempel

Eftersom samma räknelagar gäller i $(\mathbb{C},+,\cdot)$ som i $(\mathbb{R},+,\cdot)$ kan läsaren redan räkna med komplexa tal! Det enda han behöver komma ihåg är att $i^2=-1$.

Exempel

$$i+(2+3i)(5-i)=i+10-2i+15i-3i^2=i+10-2i+15i+3=13+14i.$$

Vid division av komplexa tal är förlängning med nämnarens konjugat den sedvanliga vägen att gå:

Exempel

$$\frac{5+2i}{3-i}=\frac{5+2i}{3-i}\cdot\frac{3+i}{3+i}=\frac{(5+2i)(3+i)}{10}=\frac{15+5i+6i-2}{10}=\frac{13+11i}{10}=1.3+1.1i$$

1.9.1.5 Real- och imaginärdel

Om $z=a+bi$ är ett komplext tal (underförstått att $a,b\in\mathbb{R}$) så kallas a och b för z:s *realdel* respektive *imaginärdel*. Dessa betecknas $\operatorname{Re} z$ respektive $\operatorname{Im} z$, så

$$\begin{aligned}\operatorname{Re} z&=a\\ \operatorname{Im} z&=b.\end{aligned}$$

Notera att både realdelen och imaginärdelen är ett reellt tal! (I själva verket är det ju dessa två reella tal som det komplexa talet ”består av”.)

1.9.2 Rigorös introduktion av de komplexa talen (*)

Som bekant används beteckningarna

$$\begin{aligned}&\mathbb{N}=\{0,1,2,\dots\},\\ &\mathbb{Z}=\{0,\pm1,\pm2,\dots\},\\ &\mathbb{Q}=\left\{\frac{p}{q}\in\mathbb{R}\quad p,q\in\mathbb{Z},\quad q\neq 0\right\},\\ &\mathbb{R}\end{aligned}$$

för mängden av de naturliga talen, heltalen, de rationella talen respektive de reella talen. Vi har förstås

$$\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R}$$

om alla element betraktas som reella tal. Nu är emellertid $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$ och $\mathbb{R}$ inte ”bara” mängder av objekt, utan vi har också två operationer, $+$ och $\cdot$, som verkar i dem, s.k. *kompositionsregler*. Vi kan ta två tal och beräkna deras summa och produkt: $2 + 5 = 7$, $2 \cdot 3 = 6$.

En mängd X med en kompositionsregel $X \times X \to X$ kallas för en *grupp* om kompositionsregeln uppfyller vissa villkor, som t.ex. associativitet. En mängd med *två* kompositionsregler kallas för en *ring* om dessa två regler uppfyller vissa villkor; bland annat skall mängden vara en kommutativ grupp under en av reglerna. Exempel på ringar är $(\mathbb{Z}, +,\cdot)$. $(\mathbb{Q}, +,\cdot)$ och $(\mathbb{R}, +,\cdot)$, d.v.s. heltalen, de rationella talen och de reella talen under vanlig addition och multiplikation. Dessa ringar är dessutom kommutativa, d.v.s. multiplikationen $\cdot$ är kommutativ.

En egenskap som $(\mathbb{Q}, +,\cdot)$ och $(\mathbb{R}, +,\cdot)$ har, men som inte $(\mathbb{Z}, +,\cdot)$ har, är att varje element, förutom det neutrala elementet med avseende på addition (d.v.s. 0), har en invers med avseende på multiplikation. Till exempel är den multiplikativa inversen till 5 lika med $\frac{1}{5}$ som inte är ett heltal. En kommutativ ring där varje element, förutom additiva neutrala elementet, har en multiplikativ invers kallas för en *kropp*. $(\mathbb{Q}, +,\cdot)$ och $(\mathbb{R}, +,\cdot)$ är alltså exempel på kroppar.

Jag tänker här inte ge någon fullständig introduktion till de algebraiska begreppen grupp, ring och kropp; för en fullständig introduktion till dylika algebraiska strukturer hänvisas till kurser eller böcker i ämnet (abstrakt) algebra. Här nöjer jag mig i stället med att lista de egenskaper som gör att $\mathbb{Q}$ och $\mathbb{R}$ faktiskt *är* kroppar under vanlig addition och multiplikation. Listan nedan talar för $(\mathbb{R}, +,\cdot)$, men exakt samma egenskaper gäller för $(\mathbb{Q}, +,\cdot)$.

$(\mathbb{R}, +,\cdot)$ är en *kropp*, eftersom…

$(\mathbb{R}, +)$ är en kommutativ grupp:

Associativitet:	$(x + y) + z = x + (y + z)$
Existens av neutralt element:	$0 + x = x + 0 = x$
Existens av inverser:	$x + (-x) = (-x) + x = 0$
Kommutativitet:	$x + y = y + x$

$(\mathbb{R}, +,\cdot)$ är en kommutativ ring eftersom vi *dessutom* har:

Associativitet:	$(x \cdot y) \cdot z = x \cdot (y \cdot z)$
Distributivitet:	$x \cdot (y + z) = x \cdot y + x \cdot z$
	$(x + y) \cdot z = x \cdot z + y \cdot z$
Existens av neutralt element[20]*:*	$1 \cdot x = x \cdot 1 = x$
Kommutativitet:	$x \cdot y = y \cdot x$

$(\mathbb{R}, +,\cdot)$ är en kropp eftersom vi *dessutom* har:

Existens av inverser:	För varje $x \neq 0$ är $x \cdot \frac{1}{x} = 1$.

[20] En del författare kräver inte det axiomet av en ring. Däremot har en *kropp* alltid ett neutralt element med avseende på multiplikation, en s.k. ”etta”.

Jag kunde inte låta bli att dela in listan i tre delar för att antyda definitionerna för de tre algebraiska begreppen *grupp*, *ring* och *kropp*, trots att jag egentligen har för avsikt att undvika en orgie i abstrakt algebra i det här avsnittet. Om man ogärna vill gå ner sig i algebran kan man alltså med fördel ignorera rubrikerna och grupperingarna i listan. Det viktiga är att listan anger de egenskaper som operationerna plus och gånger för reella (och rationella) tal har. Den anger explicit de räknelagar som gäller för sådana tal, och inför dessutom begreppet ”kropp” för en algebraisk struktur som besitter dessa egenskaper. Detta kommer vi att ha stor nytta av alldeles strax, när vi introducerat de komplexa talen: vi kommer då att visa att dessa också, precis som $\mathbb{Q}$ och $\mathbb{R}$, utgör en kropp, så att vi kan räkna på ”som vanligt” med komplexa tal, åtminstone när det kommer till addition och multiplikation.

En mängd X tillsammans med två kompositionsregler $+$ och $\cdot$ kallas alltså för en *kropp* om de räknelagar som listas ovan gäller. Från dessa räknelagar kan man härleda ytterligare räknelagar, som följaktligen också gäller i varje kropp. Nedan ges några sådana exempel.

Lemma. Om $\cdot$ är en kompositionsregel på en mängd X så finns det max ett neutralt element, d.v.s. max ett element e sådant att $e \cdot x = x \cdot e = x$ för varje $x \in X$.

Bevis. Antag att e och $e' \in X$ båda är neutrala element. I sådana fall är $e = e \cdot e' = e'$ där första likheten kommer av att e' är ett neutralt element och andra av att e är det. Sålunda är $e = e'$. ■

Hos en kropp $(X, +, \cdot)$ gäller därför att både $+$ och $\cdot$ har *exakt* ett neutralt element. Hos $\mathbb{R}$ är dessa 0 respektive 1.

Lemma. Om $\cdot$ är en **associativ** kompositionsregel på en mängd X med neutralt element e så har varje $x \in X$ högst en invers, d.v.s. det finns högst ett element $y \in X$ sådant att $x \cdot y = y \cdot x = e$.

Bevis. Antag att både y och $y' \in X$ är inverser till x. Då är $y = y \cdot e = y \cdot (x \cdot y') = (y \cdot x) \cdot y' = e \cdot y' = y'$. ■

Hos en kropp $(X, +, \cdot)$ är båda kompositionsreglerna associativa, så varje $x \in X$ har *exakt* en additiv invers (vilken betecknas $-x$) och varje $x \in X \setminus \{0\}$ har *exakt* en multiplikativ invers (vilken betecknas x^{-1}).

Lemma. Om $\cdot$ är en associativ kompositionsregel på mängden X med neutralt element e så gäller att inversen till inversen till x är lika med x för varje $x \in X$.

Bevis. Låt y vara inversen till x och låt z vara inversen till y; vi skall då visa att $z = x$. Men $x = x \cdot e = x \cdot (y \cdot z) = (x \cdot y) \cdot z = e \cdot z = z$. ■

Hos en kropp $(X, +, \cdot)$ gäller alltså att $-(-x) = x$ och $(x^{-1})^{-1} = x$ för varje $x \in X$ respektive $x \in X \setminus \{0\}$.

Sats. Om $(X, +, \cdot)$ är en kropp med additivt neutralt element 0 så är $0 \cdot x = 0$ för alla $x \in X$.

Bevis. $0 \cdot x + 0 \cdot x = (0 + 0) \cdot x = 0 \cdot x$. Eftersom varje element i en kropp har en additiv invers, så finns också en additiv invers till elementet $0 \cdot x \in X$. Addition av denna ger $0 \cdot x = 0$. ■

Sats. Om $(X,+,\cdot)$ är en kropp så är $(-x)\cdot y = -(x\cdot y)$ och $(-x)\cdot(-y) = x\cdot y$ för varje $x,y \in X$.

Bevis. $(x\cdot y) + [(-x)\cdot y] = \big(x + (-x)\big)\cdot y = 0\cdot y = 0$ vilket visar att $(-x)\cdot y$ är den additiva inversen till $x\cdot y$, och vi har visat första delen. Användande av denna ger $(-x)\cdot(-y) = -\big(x\cdot(-y)\big) = -\big((-y)\cdot x\big) = -\big(-(y\cdot x)\big) = y\cdot x = x\cdot y$. ■

Sats. Om $(X,+,\cdot)$ är en kropp och $x,y \in X$ så gäller

$$x\cdot y = 0 \qquad \Longleftrightarrow \qquad \begin{cases} x = 0 \\ \text{eller} \\ y = 0. \end{cases}$$

Bevis. $\Rightarrow$) Antag att $x\cdot y = 0$. Om $x = 0$ är vi klara. Om $x \neq 0$ så har x en multiplikativ invers x^{-1}. Multiplikation av denna ger $x^{-1}\cdot(x\cdot y) = (x^{-1}\cdot x)\cdot y = e\cdot y = y = x^{-1}\cdot 0 = 0$. $\Leftarrow$) Omvändningen är trivial med tanke på en tidigare sats på den här sidan. ■

Den stora poängen är sålunda den, att om en mängd tillsammans med två kompositionsregler uppfyller kraven för en kropp, så uppfyller de väsentligen *samtliga* trevliga egenskaper vi är vana vid hos t.ex. de reella talen.

1.9.2.1 Delmängder och ärvda operationer (forts.)*

$(\mathbb{R},+,\cdot)$ och $(\mathbb{Q},+,\cdot)$ är alltså två exempel på kroppar. Notera att pluset i respektive kropp inte är samma. I $(\mathbb{R},+,\cdot)$ är pluset en funktion $\mathbb{R}\times\mathbb{R}\to\mathbb{R}$, medan den i $(\mathbb{Q},+,\cdot)$ är en funktion $\mathbb{Q}\times\mathbb{Q}\to\mathbb{Q}$, och samma sak gäller för multiplikationen. Vi använder däremot samma symbol, av den anledning som ges i nästa stycke.

Eftersom varje rationellt tal också är ett reellt tal, kan två rationella tal dels fungera som operander till additionen och multiplikationen i $\mathbb{R}$, dels som operander till additionen och multiplikationen i $\mathbb{Q}$. Naturligtvis sammanfaller dessa, så att, t.ex., additionen av de två *reella* talen 2 och 5 med plusfunktionen i $\mathbb{R}$ ger samma resultat som additionen av de två *rationella* talen 2 och 5 med plusfunktionen i $\mathbb{Q}$. Additionen i $\mathbb{Q}$ är med andra ord restriktionen av additionen i $\mathbb{R}$ till par av rationella tal, och liknande för multiplikationen. Man säger därför att $\mathbb{Q}$ är en *underkropp* till $\mathbb{R}$, eller att $\mathbb{R}$ är en *kroppsutvidgning* av $\mathbb{Q}$.

Kroppen $\mathbb{C}$ av komplexa tal som vi kommer att införa kommer att vara en kroppsutvidgning av $\mathbb{R}$, d.v.s. $\mathbb{R}$ kommer att vara en underkropp till $\mathbb{C}$. Det betyder dels att $\mathbb{R}$ är en delmängd av $\mathbb{C}$, dels att additionen och multiplikationen i $\mathbb{R}$ bara är restriktionerna av motsvarande kompositionsregler i $\mathbb{C}$ till par av reella tal.

1.9.2.2 Motivering (forts.)*

Vad är då poängen med att utöka $\mathbb{R}$ till en ännu större kropp? Med andra ord, vad är poängen med att införa $\mathbb{C}$ som ännu ett steg i följande lista:

$$\mathbb{N}\subset\mathbb{Z}\subset\mathbb{Q}\subset\mathbb{R}\subset\mathbb{C}?$$

En teoretisk förklaring, som ofta nämns i det här sammanhanget, är att det finns polynom med reella koefficienter som saknar reella nollställen. Till exempel finns det inget $x\in\mathbb{R}$ sådant att $x^2+1=0$. I kroppen $\mathbb{Q}$ är problemet ännu större: det finns t.ex. inte något $x\in\mathbb{Q}$ sådant att $x^2=2$. I ringen $\mathbb{Z}$ är problemet, återigen, *ännu* större: här har inte ens ekvationen $2x = 5$ någon lösning. I $\mathbb{N}$, som inte ens är en grupp, får vi problem redan vid $x + 5 = 3$.

Däremot kan man visa att kroppen $\mathbb{C}$ är *algebraiskt sluten*, d.v.s. *varje* icke-konstant polynom med koefficienter från $\mathbb{C}$ har ett nollställe i $\mathbb{C}$. Det följer (av den s.k. faktorsatsen) att ett polynom av grad n i komplexa koefficienter alltid har precis n komplexa nollställen, med multiplicitet räknat.

En kanske mer övertygande motivering är att många problem, som till synes inte har någonting alls med komplexa tal att göra, faktiskt kan lösas med metoder från komplex algebra och analys. Till exempel kan komplex analys användas till att beräkna vissa reella integraler. Listan över tillämpningar av komplex analys kan göras mycket lång, men vi kommer inte att ge särskilt många fler exempel här. I stället ger vi oss nu i kast med konstruktionen av $\mathbb{C}$.

1.9.2.3 Konstruktion av kroppen $\mathbb{C}$ (forts.)*

Att definiera den algebraiska strukturen $(\mathbb{C}, +, \cdot)$ är i det närmaste barnsligt enkelt:

Definition. Låt $\mathbb{C}$ vara mängden av alla ordnade par (x, y) av reella tal. Elementen i $\mathbb{C}$ kallas för *komplexa tal*. Om $z = (z_1, z_2)$ och $w = (w_1, w_2)$ är två komplexa tal, så är deras summa

$$z + w = (z_1 + w_1, \quad z_2 + w_2)$$

och produkt

$$zw = (z_1w_1 - z_2w_2, \quad z_1w_2 + z_2w_1).$$

Eftersom ett komplext tal är ett par (x, y) av reella tal kan det åskådliggöras som en punkt i planet $\mathbb{R}^2$. När planet föreställs som mängden av alla komplexa tal brukar det kallas *komplexa talplanet*.

Som *mängd* är alltså $\mathbb{C} = \mathbb{R}^2$, men man brukar ändå inte skriva just "$\mathbb{C} = \mathbb{R}^2$". Anledningen är att man med $\mathbb{C}$ nästan alltid menar *kroppen* $(\mathbb{C}, +, \cdot)$ medan man med $\mathbb{R}^2$ nästan alltid menar *vektorrummet* $(\mathbb{R}^2, +, \cdot)$ [operationerna $+$ och $\cdot$ är där vektoraddition respektive multiplikation med skalär – den senare är med andra ord inte någon multiplikation mellan vektorer], och dessa är definitivt inte lika – de är ju t.o.m. två olika *sorters* algebraiska strukturer.

Vi visar nu att man kan räkna "som vanligt" i $\mathbb{C}$, d.v.s. att de räknelagar som gäller för addition och multiplikation av reella (eller rationella) tal också gäller för addition och multiplikation av komplexa tal.

Sats. $(\mathbb{C}, +, \cdot)$ är en kropp.

Bevis. Vi behöver visa alla egenskaperna som en kropp har, d.v.s. de som listades ovan. Beviset är mycket enkelt, i de flesta fall trivialt, men för pedanteriets skull redovisar vi det ändå i detalj.

Låt $x = (x_1, x_2)$, $y = (y_1, y_2)$ och $z = (z_1, z_2)$ vara komplexa tal. Då har vi *associativitet för additionen*, ty

$$(x + y) + z = (x_1 + y_1, x_2 + y_2) + z = ([x_1 + y_1] + z_1, [x_2 + y_2] + z_2) = (x_1 + [y_1 + z_1], x_2 + [y_2 + z_2]) =$$
$$= x + (y_1 + z_1, y_2 + z_2) = x + (y + z)$$

eftersom additionen i $\mathbb{R}$ är associativ. Det komplexa talet $(0, 0)$ är ett *neutralt element med avseende på addition*, ty

$$(0, 0) + x = (0 + x_1, 0 + x_2) = (x_1, x_2) = x$$
$$x + (0, 0) = (x_1 + 0, x_2 + 0) = (x_1, x_2) = x$$

eftersom 0 är ett neutralt element med avseende på addition i $\mathbb{R}$. Vidare har (x_1, x_2) den *additiva inversen* $(-x_1, -x_2)$ ty

$$(x_1, x_2) + (-x_1, -x_2) = \big(x_1 + (-x_1), x_2 + (-x_2)\big) = (0, 0)$$
$$(-x_1, -x_2) + (x_1, x_2) = \big((-x_1) + x_1, (-x_2) + x_2\big) = (0, 0)$$

eftersom de reella talen x_1 och x_2 har additiva inverser $-x_1$ resp. $-x_2$ med avseende på additionen i $\mathbb{R}$. $(\mathbb{C}, +)$ är sålunda en grupp, och den är dessutom kommutativ, ty

$$x + y = (x_1, x_2) + (y_1, y_2) = (x_1 + y_1, x_2 + y_2) = (y_1 + x_1, y_2 + x_2) = (y_1, y_2) + (x_1, x_2) = y + x$$

på grund av att additionen i $\mathbb{R}$ är kommutativ. Beträffande multiplikationen i $\mathbb{C}$ har vi *associativitet*, ty

$$\begin{aligned}(x \cdot y) \cdot z &= (x_1y_1 - x_2y_2, \quad x_1y_2 + x_2y_1) \cdot (z_1, z_2) = \\ &= (x_1y_1z_1 - x_2y_2z_1 - x_1y_2z_2 - x_2y_1z_2, \quad x_1y_1z_2 - x_2y_2z_2 + x_1y_2z_1 + x_2y_1z_1)\end{aligned}$$

och

$$\begin{aligned}x \cdot (y \cdot z) &= (x_1, x_2) \cdot (y_1z_1 - y_2z_2, \quad y_1z_2 + y_2z_1) = \\ &= (x_1y_1z_1 - x_1y_2z_2 - x_2y_1z_2 - x_2y_2z_1, \quad x_1y_1z_2 + x_1y_2z_1 + x_2y_1z_1 - x_2y_2z_2)\end{aligned}$$

som är identiska. Vi har även *distribution* från vänster:

$$\begin{aligned}x \cdot (y + z) &= (x_1, x_2) \cdot (y_1 + z_1, y_2 + z_2) = \big(x_1(y_1 + z_1) - x_2(y_2 + z_2), \quad x_1(y_2 + z_2) + x_2(y_1 + z_1)\big) = \\ &= (x_1y_1 + x_1z_1 - x_2y_2 - x_2z_2, \quad x_1y_2 + x_1z_2 + x_2y_1 + x_2z_1) = \\ &= (x_1y_1 - x_2y_2, \quad x_1y_2 + x_2y_1) + (x_1z_1 - x_2z_2, \quad x_1z_2 + x_2z_1) = x \cdot y + x \cdot z.\end{aligned}$$

Distribution från höger visas på motsvarande sätt. Vi har ett *neutralt element*, nämligen $(1, 0)$, ty

$$(1,0) \cdot x = (1, 0) \cdot (x_1, x_2) = (1 \cdot x_1 - 0 \cdot x_2, \quad 1 \cdot x_2 + 0 \cdot x_1) = (x_1, x_2) = x$$
$$x \cdot (1,0) = (x_1, x_2) \cdot (1, 0) = (x_1 \cdot 1 - x_2 \cdot 0, \quad x_1 \cdot 0 + x_2 \cdot 1) = (x_1, x_2) = x.$$

Vi har också *kommutativitet*:

$$x \cdot y = (x_1, x_2) \cdot (y_1, y_2) = (x_1y_1 - x_2y_2, \quad x_1y_2 + x_2y_1) = (y_1x_1 - y_2x_2, \quad y_1x_2 + y_2x_1) = (y_1, y_2) \cdot (x_1, x_2) = y \cdot x.$$

Nu kommer vi till den enda icke-triviala delen av beviset. Vi skall nämligen visa att varje komplext tal förutom det additiva neutrala elementet $(0, 0)$ har en multiplikativ invers. Detta åstadkommer vi genom att visa att det komplexa talet $(x_1, x_2) \neq (0, 0)$ har $\left(\frac{x_1}{x_1^2+x_2^2}, -\frac{x_2}{x_1^2+x_2^2}\right)$ som invers [nämnarna är nollskilda eftersom det komplexa talet *inte* är $(0, 0)$]. Detta är däremot trivialt:

$$(x_1, x_2) \cdot \left(\frac{x_1}{x_1^2 + x_2^2}, -\frac{x_2}{x_1^2 + x_2^2}\right) = \left(\frac{x_1^2}{x_1^2 + x_2^2} - \frac{-x_2^2}{x_1^2 + x_2^2}, \quad -\frac{x_1x_2}{x_1^2 + x_2^2} + \frac{x_1x_2}{x_1^2 + x_2^2}\right) = (1, 0).$$

Beviset är klart. ■

Anmärkning. Valet av uttrycket för den multiplikativa inversen kan förefalla godtyckligt, och läsaren kanske undrar om det finns *fler* inverser till (x_1, x_2). Svaret är ”nej”, eftersom vi just visat att $(\mathbb{C}, +, \cdot)$ är en *kropp* och i en kropp är den multiplikativa inversen alltid entydigt bestämd.

1.9.2.4 $\mathbb{R}$ är en underkropp till $\mathbb{C}$ (* forts.)

Vi visar härnäst att $\mathbb{R}$ är en *underkropp* till $\mathbb{C}$, så att varje reellt tal är ett komplext tal. Den noggranne läsaren har nog redan ställt sig frågande till hur detta ens är möjligt, med tanke på att varje komplext tal är ett *par* av reella tal, och inget sådant *par* kan ju vara exakt samma sak som *ett* reellt tal, per definition. Poängen är i stället att de komplexa talen på formen $(x, 0)$ *beter* sig exakt som de reella talen; mer precist gäller detta om vi identifierar varje reellt tal x med det komplexa talet $(x, 0)$. För att inse detta, notera bara att

$$(x_1, 0) + (x_2, 0) = (x_1 + x_2, 0)$$
$$(x_1, 0) \cdot (x_2, 0) = (x_1x_2, 0).$$

så att den *komplexa* summan av x_1 och x_2 är lika med den vanliga *reella* summan $x_1 + x_2$, och analogt för produkterna.

Låt $\tilde{\mathbb{R}}$ vara mängden av de komplexa talen på formen $(x_1, 0)$. Vi ser att restriktionen av additionen och multiplikationen i $\mathbb{C}$ till par av element i $\tilde{\mathbb{R}}$ är funktioner $\tilde{\mathbb{R}} \times \tilde{\mathbb{R}} \to \tilde{\mathbb{R}}$, och det är också lätt att se att t.ex. de två neutrala elementen i $\mathbb{C}$ samt alla additiva och multiplikativa inverser till elementen i $\tilde{\mathbb{R}}$ också finns i $\tilde{\mathbb{R}}$. Det följer därför att $\tilde{\mathbb{R}}$ är en **underkropp** till $\mathbb{C}$.

Som nämnt ovan beter sig $\tilde{\mathbb{R}}$ precis som $\mathbb{R}$ – enda skillnaden är att elementen skrivs $(x, 0)$ i stället för x. Med ett fint ord är kropparna $\tilde{\mathbb{R}}$ och $\mathbb{R}$ *isomorfa*, vilket i princip betyder att de är identiska så när som på att elementen betecknas[21] olika. Det är i den meningen som $\mathbb{R}$ är en ”underkropp” till $\mathbb{C}$.

Om $x \in \mathbb{R}$ kommer vi hädanefter att skriva kort och gott x även när vi betraktar det som ett komplext tal (och därför egentligen menar $(x, 0)$). I det komplexa talplanet hittar vi sålunda de reella talen längs x-axeln.

1.9.2.5 Notationen $a + bi$ (* forts.)

Vi inför beteckningen

$$i \coloneqq (0, 1)$$

för det speciella komplexa talet $(0, 1)$ och noterar att

$$i^2 = i \cdot i = (0,1) \cdot (0,1) = (-1, 0) = -1.$$

Det komplexa talet i, som således har den exotiska egenskapen att dess kvadrat är lika med det reella talet -1, kallas för *den imaginära enheten*. Vi noterar också att ett godtyckligt komplext tal $z = (z_1, z_2)$ på ett unikt sätt kan skrivas

$$z = (z_1, z_2) = (z_1, 0) + (0, z_2) = (z_1, 0) + (z_2, 0) \cdot (0, 1) = z_1 + z_2 \cdot i.$$

Vi kommer i fortsättningen alltid att skriva komplexa tal på det sättet. z_1 kallas för talets *realdel*; z_2 kallas för talets *imaginärdel*. Lägg märke till att båda är reella tal. I fortsättningen kommer vi att utelämna multiplikationstecknet vid multiplikation mellan komplexa tal (precis som vi gör för reella tal). Vi skriver alltså $z_1 + z_2 i$. Vi inför också beteckningarna

$$\begin{aligned} \operatorname{Re} z &= z_1 \\ \operatorname{Im} z &= z_2; \end{aligned}$$

Re och Im är som synes funktioner $\mathbb{C} \to \mathbb{R}$. Precis som de reella talen finns längs x-axeln i det komplexa talplanet, så hittar vi de s.k. *rent imaginära talen* bi ($b \in \mathbb{R}$) längs y-axeln. Ett komplext tal z är reellt omm $\operatorname{Im} z = 0$, det är rent imaginärt omm $\operatorname{Re} z = 0$.

1.9.2.6 Konjugering och belopp (* forts.)

Om $z = a + bi$ (med $a, b \in \mathbb{R}$) så definieras vi *konjugatet* $\bar{z} = \overline{a + bi} = a - bi$. Konjugering är alltså en funktion $\mathbb{C} \to \mathbb{C}$, tydligen en *involution* (d.v.s. den är sin egen invers). Notera att

$$\operatorname{Re} z = \frac{z + \bar{z}}{2}, \qquad \operatorname{Im} z = \frac{z - \bar{z}}{2i}.$$

[21] Datoranalogi: Olika teckensnitt men samma kodpunkt?!

Det är vidare en enkel övning att visa att

$$\overline{z+w} = \bar{z} + \bar{w}, \qquad \overline{zw} = \bar{z} \cdot \bar{w}$$

och

$$\bar{z}z = (\operatorname{Re} z)^2 + (\operatorname{Im} z)^2.$$

Vi inför slutligen *beloppet*

$$|z| \coloneqq \sqrt{\bar{z}z} = \sqrt{(\operatorname{Re} z)^2 + (\operatorname{Im} z)^2}$$

som är en funktion $\mathbb{C} \to [0, \infty[$ med $|z| = 0 \Leftrightarrow z = 0$. Om $z = a + bi = (a, b)$ åskådliggörs i det komplexa talplanet är $|z|$ alltså avståndet till origo. Vi har

$$|zw| = |z|\,|w|, \qquad |z+w| \leq |z| + |w|$$

där olikheten går under namnet *triangelolikheten*; namnet kommer förstås från den geometriska bilden (vilken?).

1.9.2.7 En smak av komplexitet (forts.)*

Låt oss nu äntligen pröva vår nya kropp. Precis som för reella tal inför vi den binära operatorn $-$ (minustecken) som en kortform för addition med den additiva inversen, och division (och bråkstreck) som en kortform för multiplikation med den multiplikativa inversen. Eftersom $\mathbb{C}$ är en kropp kan vi räkna ”som vanligt” med de komplexa talen. Om vi dessutom använder notationen $a + bi$ för det komplexa talet (a, b), så är egentligen det enda vi behöver komma ihåg att $i^2 = -1$.

Exempel

$$i + (2+3i)(5-i) = i + 10 - 2i + 15i - 3i^2 = i + 10 - 2i + 15i + 3 = 13 + 14i.$$

Vid division av komplexa tal är förlängning med nämnarens konjugat den sedvanliga vägen att gå:

Exempel

$$\frac{5+2i}{3-i} = \frac{5+2i}{3-i} \cdot \frac{3+i}{3+i} = \frac{(5+2i)(3+i)}{10} = \frac{15+5i+6i-2}{10} = \frac{13+11i}{10} = 1.3 + 1.1i$$

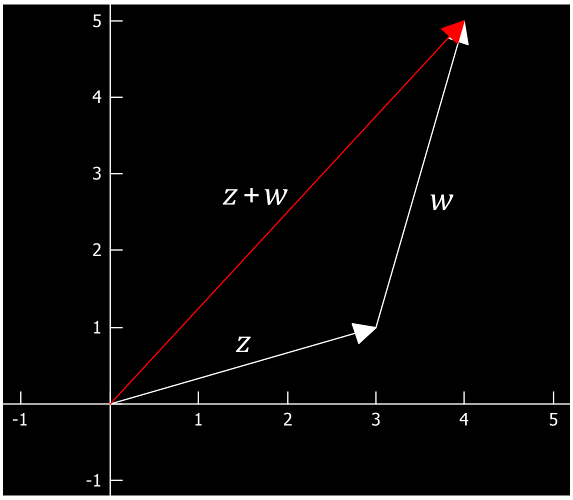
där en rad egenskaper hos kroppen $\mathbb{C}$ användes implicit (precis som vi använder dem utan att tänka på dem när vi räknar i $\mathbb{R}$). Som övning kan läsaren med fördel verifiera i detalj att räkningarna ovan är korrekta och helt i enighet med konstruktionen av kroppen $\mathbb{C}$.

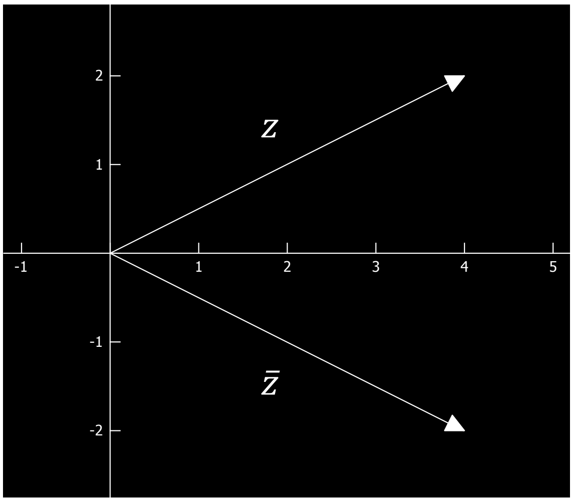
Detta avslutar vår ~~orgie i abstrakt algebra~~ formella introduktion till de komplexa talen. Härnäst skall vi studera de geometriska tolkningarna av de komplexa talen och de två operationerna $+$ och $\cdot$ på dem.

1.9.3 Geometriska tolkningar

Ett komplext tal $a + bi$ kan som bekant illustreras med en punkt i planet $\mathbb{R}^2$. I själva verket råder ju ett 1–1-förhållande mellan de komplexa talen $a + bi$ och punkterna (a, b) i planet. Ett komplext tal $a + bi$ kan därför också representeras som den vektor (pil) som börjar i origo och slutar i punkten

(a, b). Med det åskådliggörandet följer det direkt av definitionen av kompositionsregeln $+$ att summan av två komplexa tal erhålles genom vektoraddition (se vänstra bilden):





I bilden ovan, till vänster, är $z = 3 + i$, $w = 1 + 4i$ och summan $z + w = 4 + 5i$. Notera att vektorerna för z och $z + w$ startar i origo och därmed verkligen pekar på talen z respektive $z + w$ i det komplexa talplanet, medan vektorn w är förflyttad för att illustrera (vektor)additionen, och alltså inte pekar på talet w.

Per definition är $|z|$ avståndet från origo till det komplexa talet $z \in \mathbb{C}$, d.v.s. längden av vektorn från origo till z. I bilden ovan illustreras mycket tydligt triangelolikheten $|z + w| \leq |z| + |w|$ samtidigt som det torde framgå varifrån namnet på olikheten kommer. Om $z = a + bi$ $(a, b \in \mathbb{R})$ definieras vidare det komplexa *konjugatet* som talet $\bar{z} = \overline{a + bi} = a - bi$. Vektorn $\bar{z}$ erhålles tydligen från z genom spegling i den reella axeln, se bilden ovan till höger.

Övningar

Låt $z = a + bi$ och $w = c + di$ vara komplexa tal $(a, b, c, d \in \mathbb{R})$.

1. Visa att
   a. $\operatorname{Re} z = \frac{z+\bar{z}}{2}$
   b. $\operatorname{Im} z = \frac{z-\bar{z}}{2i}$.
2. Visa att
   a. $\overline{z + w} = \bar{z} + \bar{w}$
   b. $\overline{zw} = \bar{z} \cdot \bar{w}$.
3. Visa att
   a. $\bar{\bar{z}} = z$
   b. $|\bar{z}| = |z|$
4. Visa att $z\bar{z} = |z|^2$.
5. Visa att
   a. $|zw| = |z||w|$
   b. $\left|\frac{z}{w}\right| = \frac{|z|}{|w|}$.

1.9.3.1 Polär form

För att ange en punkt i planet kan man förstås ange punktens kartesiska koordinater x och y. Men man kan också ange avståndet till origo samt den vinkel som punktens ortsvektor bildar med positiva x-axeln (rent av modulo 2π). På detta sätt introduceras också de *polära koordinaterna* (ρ, φ) för punkten med de (vanliga) kartesiska koordinaterna (x, y). Mer precist är sambandet

$$x = \rho\cos\varphi$$
$$y = \rho\sin\varphi.$$

Detta samband kan (bör) tolkas som en funktion $[0, \infty[\times\mathbb{R}\to\mathbb{R}^2, (\rho,\varphi)\mapsto(x,y)$. Den är förstås inte injektiv: dels svarar alla koordinater (ρ, φ) med $\rho = 0$ mot origo, dels är φ bara bestämd av $(x, y) \neq (0,0)$ så när som på heltalsmultipler av 2π. Restriktionen av avbildningen till $\mathbb{R}^+ \times [0,2\pi[$, säg, är injektiv och ”nästan” surjektiv: värdemängden är $\mathbb{R}^2 \setminus \{(0,0)\}$. I samtliga fall är

$$x^2 + y^2 = \rho^2\cos^2\varphi + \rho^2\sin^2\varphi = \rho^2$$

så att avståndet till origo

$$\rho = \sqrt{x^2 + y^2}$$

(hör och häpna!). Eftersom komplexa tal identifieras med punkterna i planet, kan ett komplext tal $z = a + bi$ också anges på polär form. Avståndet till origo, ρ, är i sådana fall inget annat än beloppet $|z| = \sqrt{a^2 + b^2}$. Vinkeln φ [eller, rättare sagt, *någon* sådan vinkel] kallas för talets *argument*, och är bara bestämt av talet modulo 2π. Vi har förstås sambandet

$$z = a + bi = \rho(\cos\varphi + i\sin\varphi).$$

1.9.3.2 Den komplexa exponentialfunktionen

Exponentialfunktionen $\exp: x \mapsto e^x$ är bekant sedan tidigare. Det är en funktion som tar in och ger ifrån sig reella tal: $\exp: \mathbb{R} \to \mathbb{R}^+$.

Inom ämnet som kallas *komplex analys* brukar man utvidga de elementära funktionerna (sin, cos, exp, ln, arcsin, …) så att de även blir definierade för icke-reella komplexa tal. Man utvidgar alltså funktionernas definitionsmängder, så att de blir större (men för tillåtna reella tal beter sig dessa nya funktioner precis som förut). Det rör sig med andra ord om rena definitioner: man bestämmer vad som menas med t.ex. sinus för ett icke-reellt komplext tal, så att man ger mening åt uttryck såsom $\sin 2i$ och $\sin(5 - i)$.

När det gäller definitioner finns det strängt taget inget ”rätt” eller ”fel”. Man *skulle* t.ex. kunna definiera $\sin z$, för $z \notin \mathbb{R}$, som enhetsmatrisen av storlek 27×27 om z:s imaginärdel är en heltalsmultipel av 5, och som Petersengrafen i annat fall. Detta vore emellertid av föga nytta. I stället definierar man $\sin z$ för komplext z på ett sådant sätt att den erhållna komplexa funktionen blir en ”naturlig” (och användbar) generalisering av den reella sinusfunktionen.

Nästan inget av detta kommer att behandlas i den här boken, utan läsaren hänvisas till sympatisk litteratur i just ämnet komplex analys. Det enda vi kommer att bekymra oss för här, är hur just exponentialfunktionen $\exp: \mathbb{R} \to \mathbb{R}^+$ kan utvidgas till $\mathbb{C}$. [Detta kommer också att vara grunden för hur *alla* de andra elementära funktionerna utvidgas, men – återigen – det är inget vi behandlar i den här texten.]

Det visar sig att det i särklass mest ”naturliga” sättet att göra denna utvidgning på är att först definiera

$$e^{ix} := \cos x + i \sin x$$

för varje $x \in \mathbb{R}$. Notera att högerledet, ett komplext tal, är väldefinierat eftersom både real- och imaginärdelen är definierade sedan tidigare: det rör sig ju om de gamla hederliga *reella* trigonometriska funktionerna, eftersom $x \in \mathbb{R}$. Vi har nu utvidgat exp till att också vara definierat för rent imaginära tal. För ett godtyckligt komplext tal $z = x + iy$ sätter vi slutligen

$$e^z = e^{x+iy} := e^x(\cos y + i \sin y)$$

där e^x i högerledet är den gamla vanliga, reella exponentialfunktionen; notera att detta stämmer överens med definitionen av e^{iy}, $y \in \mathbb{R}$, tidigare (sätt $x = 0$).

Vi har nu utvidgat exp till hela $\mathbb{C}$. Den ”nya” definitionen av e^z överensstämmer med den gamla om $z \in \mathbb{R}$, för i sådana fall är ju $z = x + 0i$ med $x \in \mathbb{R}$ så att

$$e^z = e^{x+0i} = e^x(\cos 0 + i \sin 0) = e^x(1 + i \cdot 0) = e^x$$

där högerledet är den vanliga reella exponentialfunktionen. Det rör sig alltså om en *utvidgning* av den reella exponentialfunktionen, och alltså inte en ”helt ny” funktion. [Om så inte vore fallet, skulle det vara förfärligt dumt att använda samma symboler (exp samt $x \mapsto e^x$) för båda funktionerna.]

Den naturliga frågan är nu, förstås, *varför* vi utvidgar på just det här sättet; det finns naturligtvis många andra funktioner definierade på $\mathbb{C}$, sådana att deras restriktioner till $\mathbb{R}$ sammanfaller med den reella exponentialfunktionen. Svaret är att den erhållna komplexa funktionen, definierad på hela $\mathbb{C}$, delar många egenskaper som den reella exponentialfunktionen, definierad på $\mathbb{R}$, besitter. Vi kommer att stöta på flera av dessa framöver.

1.9.3.2.1 Inledande egenskaper hos den komplexa exponentialfunktionen

Den komplexa exponentialfunktionen är en funktion $\mathbb{C} \to \mathbb{C} \setminus \{(0,0)\}$. För att inse detta, notera helt enkelt att om $z = x + iy$ så är

$$e^z = e^x(\cos y + i \sin y)$$

inget annat än det komplexa tal som har belopp $e^x > 0$ och argument $y \in \mathbb{R}$. Vi kan redan här se den första stora användningspotentialen för den komplexa exponentialfunktionen: den kan användas för att mycket kompakt ange det komplexa tal $z = re^{i\varphi}$ som har belopp $r \in [0, \infty[$ och argument φ.

För rent imaginära tal $z = 0 + iy$, $y \in \mathbb{R}$, har vi förstås $|e^z| = \left|e^{iy}\right| = |\cos y + i \sin y| = 1$, eftersom $\cos y + i \sin y$ är en punkt på enhetscirkeln.

1.9.3.2.2 Algebraiska egenskaper hos den komplexa exponentialfunktionen

Nu skall vi ge första exemplet på egenskaper hos den reella exponentialfunktionen vilka också gäller för den komplexa. Vi har nämligen

Sats. Om z och $w \in \mathbb{C}$ så är $e^z e^w = e^{z+w}$.

Bevis. Om $z = a + bi$ och $w = c + di$ $[a, b, c, d \in \mathbb{R}]$ så är

$$\begin{aligned}
e^z e^w = e^{a+bi} e^{c+di} = e^a(\cos b + i \sin b) e^c(\cos d + i \sin d) = \\
= e^{a+c}(\cos b \cos d + i \cos b \sin d + i \sin b \cos d - \sin b \sin d) = \\
= e^{a+c}\big(\cos b \cos d - \sin b \sin d + i(\sin b \cos d + \cos b \sin d)\big) = \\
= e^{a+c}(\cos(b+d) + i \sin(b+d)) = e^{a+c+(b+d)i} = e^{z+w}
\end{aligned}$$

som önskat. ■

Speciellt får vi

$$(e^z)^2 = e^z e^z = e^{z+z} = e^{2z}$$

och mer allmänt (med induktion) har vi

Följdsats. Om $z \in \mathbb{C}$ och $n \in \mathbb{Z}^+$ så är $(e^z)^n = e^{nz}$.

Även för nollskilda komplexa tal är det naturligt att definiera $z^0 = 1$ och $z^{-n} = 1/z^n$ för $n \in \mathbb{Z}^+$. Eftersom

$$e^z e^{-z} = e^{z+(-z)} = e^0 = 1$$

följer det att

$$e^{-z} = \frac{1}{e^z}.$$

De två senaste observationerna medför att vi i själva verket har

Följdsats. Om $z \in \mathbb{C}$ och $n \in \mathbb{Z}$ så är $(e^z)^n = e^{nz}$.

Observera emellertid att satsen *inte* gäller för exponenter som *inte* är heltal. Vi har ju inte ens definierat vad som menas med ett (allmänt) komplext tal upphöjt till något som inte är ett heltal! Vi kan inte heller använda t.ex. $z^w := e^{w \ln z}$ eftersom vi inte har definierat $\ln z$ för $z \in \mathbb{C} \setminus \mathbb{R}^+$!

1.9.3.3 Geometrisk tolkning av multiplikation

Vi kan nu också ge en geometrisk tolkning av kompositionsregeln $\cdot$ (multiplikationen) på $\mathbb{C}$. Om $z = re^{\varphi i}$ och $w = qe^{\psi i}$ [med $r, q \geq 0$ och $\varphi, \psi \in \mathbb{R}$] så är ju

$$zw = re^{\varphi i} qe^{\psi i} = rqe^{\varphi i} e^{\psi i} = rq\, e^{(\varphi+\psi)i},$$

d.v.s. produkten av två komplexa tal z och w är det komplexa tal vars belopp är *produkten* mellan beloppen hos z och w och vars argument är *summan* av argumenten hos z och w: gångra beloppen och addera argumenten.

Speciellt inser vi att produkten av två komplexa tal som båda ligger på enhetscirkeln, också ligger på enhetscirkeln.

Vi har också

$$\frac{z}{w} = \frac{re^{\varphi i}}{qe^{\psi i}} = \frac{r}{q} e^{\varphi i} \cdot \frac{1}{e^{\psi i}} = \frac{r}{q} e^{\varphi i} \cdot e^{-\psi i} = e^{(\varphi-\psi)i},$$

d.v.s. kvoten mellan två komplexa tal z och w är det komplexa tal vars belopp är *kvoten* mellan beloppen av z och w och vars argument är *differensen* mellan argumenten hos z och w: dela beloppen och subtrahera argumenten.

1.9.4 Eulers formler

Från definitionen av den komplexa exponentialfunktionen har vi

$$e^{ix} = \cos x + i \sin x$$

för varje $x \in \mathbb{R}$. Vi har förstås också

$$e^{-ix} = \cos x - i \sin x$$

eftersom cosinus är jämn medan sinus är udda. Genom addition resp. subtraktion av dessa ekvationer ser vi att

$$\cos x = \frac{e^{ix} + e^{-ix}}{2}, \qquad \sin x = \frac{e^{ix} - e^{-ix}}{2i}$$

som kallas för *Eulers formler*. Notera att högerleden är reellvärda trots att de involverar icke-reella exponentialfunktioner. Vi kan redan nu ge ett första exempel på hur man med *komplex* algebra och analys kan lösa *reella* problem. Även om problemet nedan mycket väl *kan* lösas utan inblandning av komplexa tal, så är lösningen som presenteras här anmärkningsvärt kort och smidig.

Exempel

Skriv $\sin^5 x$ som en summa av (rena) sinus- och cosinustermer.

Lösning: Med Eulers formler, binomialsatsen och Pascals triangel har vi

$$\sin^5 x = \left(\frac{e^{ix} - e^{-ix}}{2i}\right)^5 = \frac{e^{5ix} - 5e^{3ix} + 10e^{ix} - 10e^{-ix} + 5e^{-3ix} - e^{-5ix}}{32i} =$$
$$= \frac{1}{16} \cdot \frac{e^{5ix} - e^{-5ix}}{2i} - \frac{5}{16} \cdot \frac{e^{3ix} - e^{3ix}}{2i} + \frac{5}{8} \cdot \frac{e^{ix} - e^{-ix}}{2i} =$$
$$= \frac{1}{16} \sin 5x - \frac{5}{16} \sin 3x + \frac{5}{8} \sin x.$$

1.9.5 Polynomekvationer med komplexa rötter

Algebrans fundamentalsats säger att ett polynom av grad $n \in \mathbb{Z}^+$ med komplexa koefficienter har minst ett komplext nollställe. Med hjälp av den s.k. faktorsatsen följer att ett sådant polynom i själva verket har precis n nollställen, räknat med multiplicitet. Lösning av polynomekvationer med icke-reella koefficienter eller rötter fungerar i stort sett likadant som lösning av motsvarande reella problem, så när som på vissa kritiska steg. Låt oss först bekanta oss med de tre icke-triviala typfall som erhålles efter kvadratkompletteringen vid lösning av andragradsekvationer.

Exempel (positivt högerled)

Lös ekvationen $x^2 = 9$.

Lösning: $x^2 = 9 \Leftrightarrow x = \pm 3$.

Exempel (negativt högerled)

Lös ekvationen $z^2 = -9$.

Lösning: $z^2 = -9 \Leftrightarrow z = \pm 3i$.

(I det här fallet är det mycket lätt att inse rent *geometriskt* att $\pm 3i$ är de enda rötterna till ekvationen. Tänk bara på den geometriska tolkningen av multiplikation.)

Exempel (icke-reellt högerled)

Lös ekvationen $z^2 = 2 + i$.

Lösning: Sätt $z = a + bi$ $[a, b \in \mathbb{R}]$, så att $z^2 = (a + bi)^2 = a^2 + 2abi - b^2 = a^2 - b^2 + 2abi$.

Vi har då

$$
\begin{aligned}
z^2 = 2 + i \quad &\Longleftrightarrow \quad a^2 - b^2 + 2abi = 2 + i \quad \Longleftrightarrow \\
&\Longleftrightarrow \begin{cases} a^2 - b^2 = 2 \\ 2ab = 1 \end{cases} \quad \Longleftrightarrow \\
&\overset{b \neq 0}{\Longleftrightarrow} \begin{cases} a^2 - b^2 = 2 \\ a = \dfrac{1}{2b} \end{cases} \quad \Longleftrightarrow \quad \begin{cases} \dfrac{1}{4b^2} - b^2 = 2 \\ a = \dfrac{1}{2b} \end{cases} \quad \Longleftrightarrow \\
&\Longleftrightarrow \begin{cases} 1/4 - b^4 = 2b^2 \\ a = \dfrac{1}{2b} \end{cases} \quad \Longleftrightarrow \quad \begin{cases} b^4 + 2b^2 = 1/4 \\ a = \dfrac{1}{2b} \end{cases} \quad \Longleftrightarrow \\
&\Longleftrightarrow \begin{cases} (b^2 + 1)^2 - 1 = 1/4 \\ a = \dfrac{1}{2b} \end{cases} \quad \Longleftrightarrow \quad \begin{cases} b^2 = -1 \pm \sqrt{5}/2 \\ a = \dfrac{1}{2b} \end{cases} \quad \Longleftrightarrow \\
&\overset{b \in \mathbb{R}}{\Longleftrightarrow} \begin{cases} b^2 = -1 + \sqrt{5}/2 \\ a = \dfrac{1}{2b} \end{cases} \quad \Longleftrightarrow \quad \begin{cases} b = \pm\sqrt{\sqrt{5}/2 - 1} \\ a = \pm\dfrac{1}{2\sqrt{\sqrt{5}/2 - 1}}. \end{cases}
\end{aligned}
$$

Svar: $z^2 = 2 + i \quad \Longleftrightarrow \quad z = \pm\left(\frac{1}{2\sqrt{\sqrt{5}/2-1}} + \sqrt{\sqrt{5}/2 - 1}\, i\right)$.

I sista steget i lösningen är det underförstått att samma tecken ($+$ eller $-$) väljs på båda raderna (d.v.s. för både b och för a). Notera också att det av den andra ursprungsekvationen $2ab = 1$ följer att både $a \neq 0$ och $b \neq 0$. Den här metoden, där vi ansätter att $z = a + bi$, kan förstås använda i *samtliga* av de tre exemplen ovan. Däremot är det bara i sista exemplet som det verkligen behövs (de två tidigare är triviala).

Ett vanligt fel

Ett mycket vanligt förekommande fel är att studenter skriver roten ur negativa eller icke-reella tal. Betrakta t.ex. följande lösning på det andra typexemplet ovan (negativt högerled):

Lös ekvationen $z^2 = -9$.

Lösning (inte bra): $z^2 = -9 \iff z = \pm\sqrt{-9} = \pm 3i$.

Problemet med den här lösningen är att symbolen ”$\sqrt{-9}$” inte är definierad. Ingenstans i dokumentet ovan, eller i de grundläggande universitetskurser i matematik i vilka författaren har undervisat, definieras vad som menas med roten ur ett negativt reellt tal. Och om det inte är känt vad en symbol *betyder*, så kan man förstås inte resonera med den.

Det är i och för sig möjligt att utvidga kvadratrotsfunktionen till hela $\mathbb{R}$ genom att *definiera* $\sqrt{-x} = xi$ för varje $x > 0$. På så sätt blir alltså $\sqrt{a}$, för varje negativt a, det entydigt bestämda komplexa tal på den *positiva* imaginära axeln vars kvadrat är a. Men eftersom detta inte har gjorts tidigare i det här dokumentet, eller i någon av de nämnda kurserna, måste det i *dessa* sammanhang räknas som ett fel att använda symbolen $\sqrt{a}$ om $a \notin [0, \infty[$.

Ett i någon mening värre fel är när studenter skriver roten ur ett icke-reellt tal, som i nedan ”lösning”:

Lös ekvationen $z^2 = 2 + i$.

”Lösning”: $z^2 = 2 + i \iff z = \pm\sqrt{2 + i}$.

Ett problem, återigen, är att vi inte definierat vad som menas med roten ur ett icke-reellt tal. Och här är det faktiskt mindre uppenbart hur det skulle göras. Det finns förstås exakt två olika komplexa tal, vars kvadrat är $2 + i$. *Vilken* av dessa skulle symbolen $\sqrt{2 + i}$ representera? Vidare, även om man gör något slags rimligt val av vad $\sqrt{z}$ betyder för $z \in \mathbb{C}$ i allmänhet, så är det ganska uppenbart att ”svaret” $z = \pm\sqrt{2 + i}$ på uppgiften är väldigt implicit: i princip säger man ju då bara att z är ett av de två tal som kvadreras till $2 + i$, vilket ju står redan i uppgiften. I den riktiga lösningen ovan får vi i stället fram två tal på formen $z = a + bi$, vilka är explicita bestämningar av de två lösningarna.

Nu har vi sett hur vi behandlar fallen $z^2 = a$ för alla $a \in \mathbb{C}$. Vi kan då ge exempel på allmänna andragradsekvationer, vilka på vanligt sätt överförs till formen $z^2 = a$ genom den ädla konsten kvadratkomplettering. Först ger vi ett exempel på en polynomekvation av grad två med reella koefficienter, fast icke-reella rötter. Sedan ger vi oss i kast med en som dessutom har icke-reella koefficienter.

Exempel

Lös ekvationen $x^2 - 2x + 8 = 0$.

Lösning:

$$x^2 - 2x + 8 = 0 \iff (x-1)^2 + 7 = 0 \iff$$
$$\iff (x-1)^2 = -7 \iff x - 1 = \pm\sqrt{7}\,i \iff$$
$$\iff x = 1 \pm \sqrt{7}\,i.$$

Svar: $x^2 - 2x + 8 = 0 \iff x = 1 \pm \sqrt{7}\,i$.

Exempel

Lös ekvationen $x^2 + (1+i)x = 2+i$.

Lösning:

$$x^2 + (1+i)x = 2+i \quad \Longleftrightarrow \quad \left(x + \left(\frac{1}{2} + \frac{i}{2}\right)\right)^2 - \left(\frac{1}{2} + \frac{i}{2}\right)^2 = 2+i \quad \Longleftrightarrow$$

$$\Longleftrightarrow \quad \left(x + \left(\frac{1}{2} + \frac{i}{2}\right)\right)^2 = 2 + \frac{3}{2}i.$$

Med

$$w \coloneqq x + \frac{1}{2} + \frac{i}{2} \eqqcolon a + bi$$

har vi sålunda ekvationen

$$w^2 = 2 + \frac{3}{2}i \quad \Longleftrightarrow \quad a^2 - b^2 + 2abi = 2 + \frac{3}{2}i \quad \Longleftrightarrow$$

$$\Longleftrightarrow \quad \begin{cases} a^2 - b^2 = 2 \\ 2ab = \frac{3}{2} \end{cases} \quad \Longleftrightarrow \quad \begin{cases} a^2 - \frac{9}{16a^2} = 2 \\ b = \frac{3}{4a} \end{cases} \quad \Longleftrightarrow$$

$$\Longleftrightarrow \quad \begin{cases} a^4 - \frac{9}{16} = 2a^2 \\ b = \frac{3}{4a} \end{cases} \quad \Longleftrightarrow \quad \begin{cases} a^4 - 2a^2 - \frac{9}{16} = 0 \\ b = \frac{3}{4a} \end{cases} \quad \Longleftrightarrow$$

$$\Longleftrightarrow \quad \begin{cases} (a^2 - 1)^2 - 1 - \frac{9}{16} = 0 \\ b = \frac{3}{4a} \end{cases} \quad \Longleftrightarrow \quad \begin{cases} (a^2 - 1)^2 = \frac{25}{16} \\ b = \frac{3}{4a} \end{cases} \quad \Longleftrightarrow$$

$$\Longleftrightarrow \quad \begin{cases} a^2 - 1 = \pm\frac{5}{4} \\ b = \frac{3}{4a} \end{cases} \quad \Longleftrightarrow \quad \begin{cases} a^2 = 1 \pm \frac{5}{4} \\ b = \frac{3}{4a} \end{cases} \quad \Longleftrightarrow$$

$$\Longleftrightarrow \quad \begin{cases} a^2 = \frac{9}{4} \\ b = \frac{3}{4a} \end{cases} \quad \Longleftrightarrow \quad \begin{cases} a = \pm\frac{3}{2} \\ b = \pm\frac{1}{2}. \end{cases}$$

Vi har därför

$$w = \pm\left(\frac{3}{2} + \frac{1}{2}i\right)$$

och därmed

$$x + \frac{1}{2} + \frac{i}{2} = \pm\left(\frac{3}{2} + \frac{1}{2}i\right)$$

som är ekvivalent med

$$x = 1 \quad \text{eller} \quad x = -2 - i.$$

Svar: $x^2 + (1+i)x = 2+i \quad \Longleftrightarrow \quad x \in \{1, -2-i\}$.

1.9.6 Ekvationer av typen $z^n = w$

En ekvation av typen $z^n = w$ där $n \in \mathbb{Z}^+$ och $w \in \mathbb{C}$ är givna löses med fördel på polär form.

Exempel

Lös ekvationen $z^5 = 1 + \sqrt{3}i$.

Lösning: På polär form är $1 + \sqrt{3}i = 2e^{\pi i/3}$. Om $z = re^{i\varphi}$ har vi därför

$$z^5 = 1 + \sqrt{3}i \iff r^5 e^{5i\varphi} = 2e^{\pi i/3} \iff \begin{cases} r = \sqrt[5]{2} \\ 5\varphi = \dfrac{\pi}{3} + n \cdot 2\pi \end{cases} \iff \begin{cases} r = \sqrt[5]{2} \\ \varphi = \dfrac{\pi}{15} + n \cdot \dfrac{2\pi}{5} \end{cases}$$

Ekvationens fem lösningar ges tydligen för $n = 0, 1, \dots, 4$ (varför?) och dessa är (se vänstra bilden)

$$\sqrt[5]{2}e^{\pi i/15}, \quad \sqrt[5]{2}e^{7\pi i/15}, \quad \sqrt[5]{2}e^{13\pi i/15}, \quad \sqrt[5]{2}e^{19\pi i/15}, \quad \sqrt[5]{2}e^{25\pi i/15}.$$

Exempel

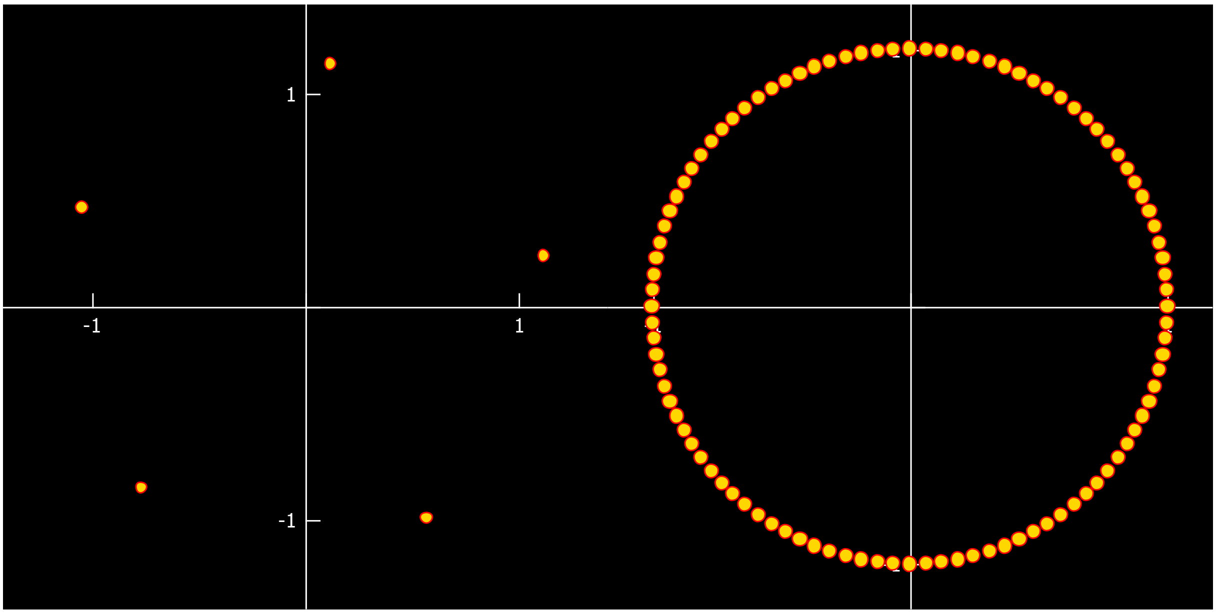
Lös ekvationen $z^{100} = 1$.

Lösning: Med $z = re^{i\varphi}$ har vi

$$z^{100} = 1 \iff r^{100} e^{100 i\varphi} = 1e^{0i} \iff \begin{cases} r^{100} = 1 \\ 100\varphi = 0 + n \cdot 2\pi \end{cases} \iff \begin{cases} r = 1 \\ \varphi = \dfrac{n}{100} \cdot 2\pi. \end{cases}$$

Svar: $z^{100} = 1 \iff z = e^{\frac{n}{100} \cdot 2\pi i}$ för något $n = 0, 1, \dots, 99$ (högra bilden).

De n rötterna till ekvationen $z^n = 1$ kallas för de n:te *enhetsrötterna*. De är alltid jämnt utspridda på enhetscirkeln: följden $(\varphi_k)_{k=0}^{n-1}$ definierad av $\varphi_k = (k/n) \cdot 2\pi$ är ju aritmetisk med differens $2\pi/n$.



1.10 Mer om polynom

1.10.1 Polynomdivision

Om p är ett polynom så är

$$p(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0$$

för något $n \in \mathbb{N}$ och någon följd $a_0, a_1, \dots, a_n$ av koefficienter. Om $a_n \neq 0$ säges $n =: \deg p$ vara *graden* av polynomet p. Om $a_n = 1$, d.v.s. om koefficienten framför högsta termen är 1, säges polynomet vara *moniskt*. Man kan visa att, givet två polynom p och q med $\deg p \geq \deg q$ [speciellt är båda nollskilda] finns det alltid entydigt bestämda polynom k och r sådana att

$$p(x) = k(x)q(x) + r(x), \qquad \deg r < \deg q \quad \text{eller} \quad r(x) = 0.$$

För nästan alla $x \in \mathbb{R}$ kan denna ekvation skrivas

$$\frac{p(x)}{q(x)} = k(x) + \frac{r(x)}{q(x)}$$

vilket torde motivera det faktum att polynomen k och r kallas för *kvoten* respektive *resten* när p divideras med q. Polynom fungerar alltså mycket som *heltal*.

Exempel

Vi har

$$\frac{52}{16} = 3 + \frac{4}{16}$$

d.v.s. 3 är kvoten och 4 är resten då 52 delas med 16. Notera att $4 < 16$.

Vi har också

$$\frac{x^4 + x^3 - 3x^2 + 2}{x^2 + 2x + 1} = x^2 - x - 2 + \frac{5x + 4}{x^2 + 2x + 1}$$

d.v.s. $x^2 - x - 2$ är kvoten och $5x + 4$ är resten då $x^2 + x^3 - 3x^2 + 2$ delas med $x^2 + 2x + 1$. Notera att $\deg(5x + 4) = 1 < 2 = \deg(x^2 + 2x + 1)$.

* Inom den abstrakta algebran säger man att ringen $\mathbb{Z}$ av alla heltal delar många egenskaper med ringen $K[x]$ av alla polynom med koefficienter från en godtycklig kropp K. Speciellt delar ringen $\mathbb{R}[x]$ (eller $\mathbb{C}[x]$) många egenskaper med ringen $\mathbb{Z}$.

Divisionsalgoritmen som används vid division av heltal (och som ofta kallas för ”liggande stolen” eller ”trappan”) fungerar också nästan oförändrad för division av polynom; *siffrorna* i heltalet svarar då mot *termerna* i polynomet.

Om $p(x) = 0$ för något $x \in \mathbb{R}$ [eller $\mathbb{C}$] så säges x vara ett *nollställe* till polynomet. *Faktorsatsen* säger att a är ett nollställe till ett icke-konstant polynom p om och endast om $x - a$ är en *faktor* i p, d.v.s. om och endast om det finns ett polynom q sådant att $p(x) = (x - a)q(x)$. I sådana fall är förstås $\deg q = \deg p - 1$ och polynomdivision ger att kvoten mellan p och $x - a$ är q med rest lika med nollpolynomet. Faktorsatsen och polynomdivision kan därför användas för att lösa många polynomekvationer med polynom av högre grad än 2, om man kan *gissa* någon rot.

Exempel

Lös ekvationen $x^3 + 2x^2 - 5x + 2 = 0$.

Lösning: Låt $p(x)$ beteckna polynomet i vänsterledet. Vi ser direkt att $p(1) = 0$, så $x - 1$ måste vara en faktor i $p(x)$. Polynomdivision ger

$$x^3 + 2x^2 - 5x + 2 = (x-1)(x^2 + 3x - 2)$$

så de övriga nollställena till $p(x)$ måste vara nollställena till $x^2 + 3x - 2$. Men

$$x^2 + 3x - 2 = 0 \iff \left(x + \frac{3}{2}\right)^2 - \frac{9}{4} - 2 = 0 \iff \left(x + \frac{3}{2}\right)^2 = \frac{17}{4} \iff x + \frac{3}{2} = \pm\frac{\sqrt{17}}{2} \iff x = -\frac{3}{2} \pm \frac{\sqrt{17}}{2}.$$

Svar: $x^3 + 2x^2 - 5x + 2 = 0 \iff x \in \left\{1, -\frac{3}{2} - \frac{\sqrt{17}}{2}, -\frac{3}{2} + \frac{\sqrt{17}}{2}\right\}$.

Antag att p är ett polynom och att $p(a) = p(b) = 0$ för $a \neq b$. Eftersom $p(a) = 0$ ger faktorsatsen att $p(x) = (x-a)q(x)$ för något polynom q. Eftersom $p(b) = 0$ har vi därför $(b-a)q(b) = 0$ och med tanke på att $a \neq b$ så måste $q(b) = 0$. Faktorsatsen ger därmed att $q(x) = (x-b)w(x)$ för något polynom w. Sålunda är $p(x) = (x-a)q(x) = (x-a)(x-b)w(x)$ så att i själva verket $(x-a)(x-b)$ är en faktor i p.

Exempel

Lös ekvationen $x^4 + x^3 - 2x^2 - x + 1 = 0$.

Lösning: Vi ser direkt att $x = 1$ och $x = -1$ är lösningar till ekvationen, så både $x - 1$ och $x + 1$ måste vara faktorer i polynomet i vänsterledet. Alltså måste $(x-1)(x+1) = x^2 - 1$ vara en faktor. Polynomdivision ger

$$x^4 + x^3 - 2x^2 - x + 1 = (x^2 - 1)(x^2 + x - 1)$$

där nollställena till kvoten $x^2 + x - 1$ ges av

$$x^2 + x - 1 = 0 \iff \left(x + \frac{1}{2}\right)^2 - \frac{1}{4} - 1 = 0 \iff \left(x + \frac{1}{2}\right)^2 = \frac{5}{4} \iff x + \frac{1}{2} = \pm\frac{\sqrt{5}}{2} \iff x = -\frac{1}{2} \pm \frac{\sqrt{5}}{2}.$$

Svar: $x^4 + x^3 - 2x^2 - x + 1 = 0 \iff x \in \left\{-1, 1, -\frac{1}{2} - \frac{\sqrt{5}}{2}, -\frac{1}{2} + \frac{\sqrt{5}}{2}\right\}$.

Exempel

Lös ekvationen $x^4 + x^3 - 12x^2 + 17x - 7 = 0$.

Lösning: Prövning ger att $x = 1$ är en rot, så $x - 1$ är en faktor. Polynomdivision ger

$$x^4 + x^3 - 12x^2 + 17x - 7 = (x-1)(x^3 + 2x^2 - 10x + 7).$$

Prövning ger att $x = 1$ är ett nollställe även till kvoten, som alltså också innehåller faktorn $x - 1$. Polynomdivision ger

$$x^3 + 2x^2 - 10x + 7 = (x-1)(x^2 + 3x - 7).$$

Därför är

$$x^4 + x^3 - 12x^2 + 17x - 7 = (x-1)^2(x^2 + 3x - 7).$$

För att finna nollställena till den sista kvoten använder vi kvadratkomplettering, som vanligt:

$$x^2 + 3x - 7 = 0 \iff \left(x + \frac{3}{2}\right)^2 - \frac{9}{4} - 7 = 0 \iff \left(x + \frac{3}{2}\right)^2 = \frac{37}{4} \iff x + \frac{3}{2} = \pm\frac{\sqrt{37}}{2} \iff x = -\frac{3}{2} \pm \frac{\sqrt{37}}{2}.$$

Svar: $x^4 + x^3 - 12x^2 + 17x - 7 = 0 \iff x \in \left\{1, -\frac{3}{2} - \frac{\sqrt{37}}{2}, -\frac{3}{2} + \frac{\sqrt{37}}{2}\right\}$.

1.10.2 Polynom med reella koefficienter

Ett polynom med enbart reella koefficienter kan som bekant mycket väl sakna reella lösningar. Ett exempel är $x^2 = 1$ med lösningarna $x = \pm i$. Här får vi i stället ett *par* av komplexkonjugerade tal, d.v.s. det enda talet $(-i)$ är det komplexa konjugatet av det andra $(+i)$ och tvärtom. I ett tidigare exempel såg vi att

$$x^2 - 2x + 8 = 0 \iff x = 1 \pm \sqrt{7}i.$$

Även där har vi ett par av komplexkonjugerade tal som lösningar. Å andra sidan har vi sett att

$$x^2 + (1+i)x = 2 + i \iff x \in \{1, -2 - i\}.$$

Här förekommer lösningen $-2 - i$ *inte* tillsammans med sin komplexkonjugerade partner $-2 + i$, men polynomet hade ju här inte heller reella koefficienter. Om ett polynom har reella koefficienter är det emellertid *alltid* så att icke-reella lösningar förekommer i komplexkonjugerade par, vilket följande sats visar.

Sats. Om p är ett polynom med reella koefficienter och $p(\alpha) = 0$ så är också $p(\bar{\alpha}) = 0$.

Bevis. Låt $p(z) = \sum_{k=0}^{n} a_k z^k$ där varje $a_k \in \mathbb{R}$. Vi har då

$$p(\bar{\alpha}) = \sum_{k=0}^{n} a_k (\bar{\alpha})^k = \sum_{k=0}^{n} a_k \overline{\alpha^k} = \sum_{k=0}^{n} \overline{a_k}\, \overline{\alpha^k} = \sum_{k=0}^{n} \overline{a_k \alpha^k} = \overline{\sum_{k=0}^{n} a_k \alpha^k} = \overline{p(\alpha)} = \bar{0} = 0$$

där vi använt att $a_k = \overline{a_k}$ ty $a_k \in \mathbb{R}$. ■

Exempel

Lös ekvationen $x^4 - 4x^3 + 5x^2 - 4x + 4 = 0$.

Lösning: Prövning ger att i är ett nollställe till polynomet (med reella koefficienter) i vänsterledet. Därför måste också $-i$ vara ett nollställe. Sålunda måste polynomet innehålla båda faktorerna $x - i$ och $x + i$, och därmed även $(x - i)(x + i) = x^2 + 1$.

Polynomdivision ger

$$x^4 - 4x^3 + 5x^2 - 4x + 4 = (x^2 + 1)(x^2 - 4x - 4).$$

Nollställena till kvoten ges av

$$x^2 - 4x - 4 = 0 \quad\Longleftrightarrow\quad (x - 2)^2 = 0 \quad\Longleftrightarrow\quad x = 2.$$

Svar: $x^4 - 4x^3 + 5x^2 - 4x + 4 = 0 \quad\Longleftrightarrow\quad x \in \{2, i, -i\}$.

Mer allmänt, om ett polynom $p(x)$ med reella koefficienter har $\alpha \in \mathbb{C}$ som ett nollställe, så har det även $\bar{\alpha}$ som nollställe. Det innehåller då dels faktorn $x - \alpha$, dels faktorn $x - \bar{\alpha}$. Det innehåller alltså faktorn $(x - \alpha)(x - \bar{\alpha})$. Denna faktor är faktiskt alltid ett andragradspolynom med reella koefficienter, ty

$$(x - \alpha)(x - \bar{\alpha}) = x^2 - (\alpha + \bar{\alpha})x + \alpha\bar{\alpha} = x^2 - (2 \operatorname{Re} \alpha)x + |\alpha|^2$$

som är ett polynom med reella koefficienter 1, $-2 \operatorname{Re} \alpha$ och $|\alpha|^2$.

1.10.3 Faktorisering av polynom

Algebrans fundamentalsats och faktorsatsen säger som bekant att ett polynom

$$p(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0$$

av grad $n \in \mathbb{Z}^+$ med komplexa koefficienter har precis n komplexa nollställen, räknat med multiplicitet. Om vi betecknar dessa med $\alpha_1, \alpha_2, \dots, \alpha_n$ [där alltså två tal i följden tillåts vara lika] har vi dessutom

$$p(x) = a_n (x - \alpha_1)(x - \alpha_2) \cdots (x - \alpha_n),$$

d.v.s. polynomet kan faktoriseras i n stycken moniska förstagradspolynom. Denna faktorisering är vidare i unik så när som på ordningen mellan faktorerna. Man kan jämföra med fullständig faktorisering av positiva heltal i primtalsfaktorer: den faktoriseringen är också unik så när som på ordningen mellan faktorerna. (Detta är *aritmetikens* fundamentalsats.)

Ett polynom med reella koefficienter kan inte alltid skrivas som en produkt av reella förstagradspolynom (t.ex. $x^2 + 1$). Däremot följer det (hur?) av resultaten i förra avsnittet att ett reellt polynom alltid kan skrivas som en produkt av reella faktorer av grad ≤ 2.

Exempel

Tidigare i kapitlet har vi insett att

$$x^2 - 2x + 8 = \left(x - 1 - \sqrt{7}i\right)\left(x - 1 + \sqrt{7}i\right)$$
$$x^2 + (1 + i)x - 2 - i = (x - 1)(x + 2 + i)$$
$$x^3 + 2x^2 - 5x + 2 = (x - 1)\left(x + \frac{3}{2} + \frac{\sqrt{17}}{2}\right)\left(x + \frac{3}{2} - \frac{\sqrt{17}}{2}\right)$$
$$x^4 + x^3 - 2x^2 - x + 1 = (x - 1)(x + 1)\left(x + \frac{1}{2} + \frac{\sqrt{5}}{2}\right)\left(x + \frac{1}{2} - \frac{\sqrt{5}}{2}\right)$$
$$x^4 + x^3 - 12x^2 + 17x - 7 = (x - 1)^2\left(x + \frac{3}{2} + \frac{\sqrt{37}}{2}\right)\left(x + \frac{3}{2} - \frac{\sqrt{37}}{2}\right)$$
$$x^4 - 4x^3 + 5x^2 - 4x + 4 = (x^2 + 1)(x - 2)^2 = (x - i)(x + i)(x - 2)^2.$$

Faktorisering är i allmänhet ett *svårt* problem (trots att vi inte märkt av det, tack vare väl valda exempelpolynom). Till exempel är det knappast en enkel sak att inse att

$$x^4 + 2x^3 - 4x^2 - 6x + 3 = \left(x + \sqrt{3}\right)\left(x - \sqrt{3}\right)\left(x + 1 + \sqrt{2}\right)\left(x + 1 - \sqrt{2}\right).$$

Ännu svårare att faktorisera är t.ex.

$$x^5 + 2x^4 - 3x^3 + x^2 + 2x - 1.$$

1.11 Hyperboliska funktioner

Vi har sett att Eulers formler för sinus och cosinus följer av definitionen av den *komplexa* exponentialfunktionen $\exp\colon \mathbb{C} \to \mathbb{C} \setminus \{0\}$. Eulers formler lyder, som bekant,

$$\sin x = \frac{e^{ix} - e^{-ix}}{2i}, \qquad \cos x = \frac{e^{ix} + e^{-ix}}{2}.$$

Trots att den komplexa exponentialfunktionen ger ifrån sig allmänna komplexa tal, så blir linjärkombinationerna i högerleden inget annat än de vanliga reella trigonometriska funktionerna. Om man tar bort alla förekomster av den imaginära enheten i de två uttrycken ovan erhåller man

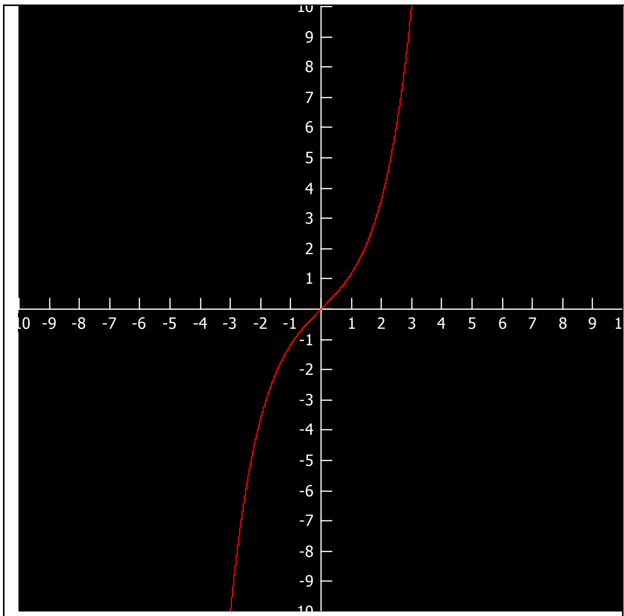
$$\frac{e^{x} - e^{-x}}{2}, \qquad \frac{e^{x} + e^{-x}}{2}.$$

Här är det alltså den vanliga, *reella* exponentialfunktionen $\exp\colon \mathbb{R} \to \mathbb{R}^+$ som figurerar. Dessa två uttryck förekommer så pass ofta både inom ren matematik och inom tekniska och vetenskapliga tillämpningar att de har fått speciella namn.

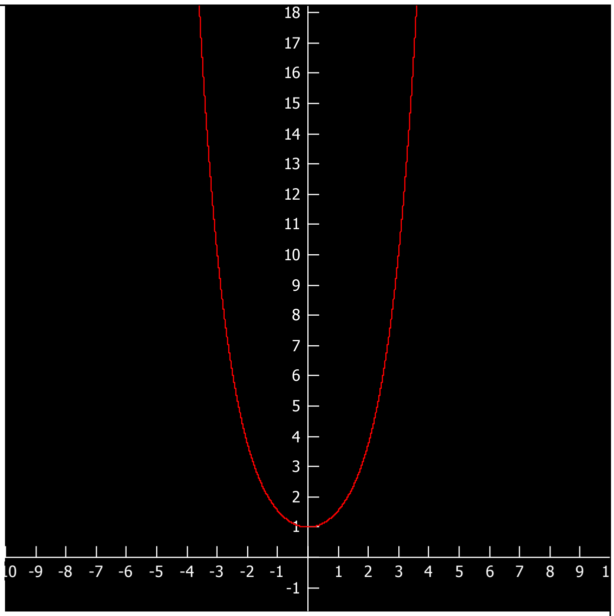
Definition. *Sinus* och *cosinus hyperbolicus*, sinh respektive cosh, är de funktioner $\mathbb{R} \to \mathbb{R}$ som definieras av

$$\sinh x = \frac{e^{x} - e^{-x}}{2}, \qquad \cosh x = \frac{e^{x} + e^{-x}}{2}.$$

Sina namn har funktionerna förstås fått av likheten med de trigonometriska funktionernas uttryck i Eulers formler. Nedan visas graferna till cosh och sinh:



$y = \sinh x$



$y = \cosh x$

Grafen till cosinus hyperbolicus bör vara bekant. Det är nämligen den form som en kedja får när den hänger fritt i ett konstant tyngdkraftsfält (t.ex.). Direkt från definitionerna ser man att cosh och sinh är obegränsade funktioner, att sinh är udda och strängt växande och att cosh är jämn och nedåt begränsad (rent av positiv).

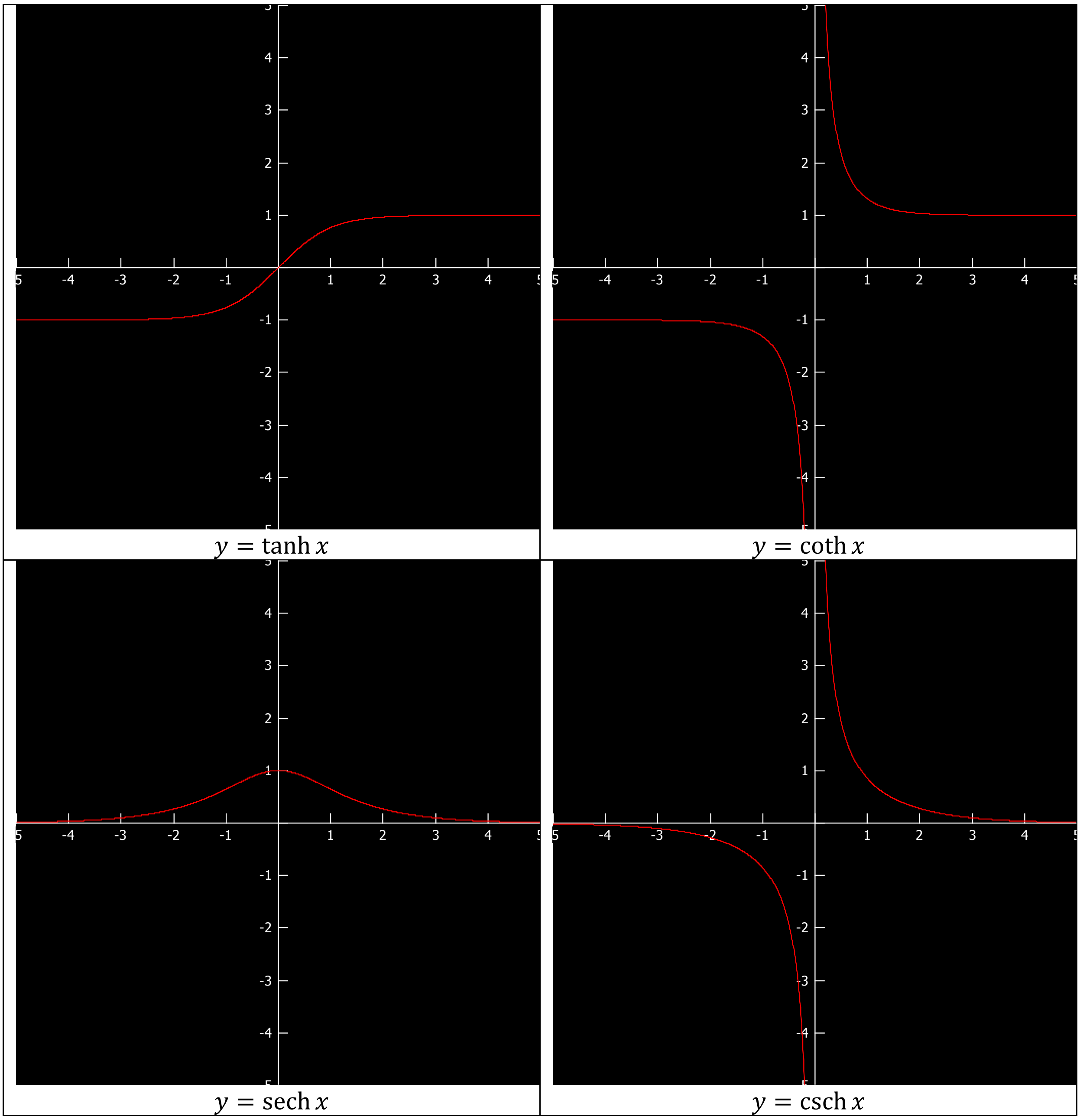
Vi inför också *tangens* och *cotangens hyperbolicus* via

$$\tanh x = \frac{\sinh x}{\cosh x}, \qquad \coth x = \frac{\cosh x}{\sinh x}$$

samt *sekant* och *cosekant hyperbolicus* via

$$\operatorname{sech} x = \frac{1}{\cosh x}, \qquad \operatorname{csch} x = \frac{1}{\sinh x}.$$

Eftersom $\cosh x$ aldrig är noll är tanh definierad på hela $\mathbb{R}$, precis som sech. Däremot är coth och csch tydligen definierade bara på $\mathbb{R} \setminus \{0\}$. Nedan visas graferna till tanh, coth, sech och csch.



$y = \tanh x$ $\quad$ $y = \coth x$

$y = \operatorname{sech} x$ $\quad$ $y = \operatorname{csch} x$

Det är lätt att se varför dessa fyra grafer har de kvalitativa egenskaper de har. Till exempel är det trivialt att tanh, coth och csch är udda medan sech är jämn. Det är också klart att tanh (och på samma sätt coth) snabbt närmar sig ± 1 till höger/vänster eftersom $\sinh x \approx \pm \cosh x$ bara en liten bit till höger/vänster. Vidare är det uppenbart att både sech och csch snabbt närmar sig noll utanför origo, med tanke på hur snabbt beloppet av cosh respektive sinh växer iväg från $x = 0$. Slutligen är det klart att coth och csch växer obegränsat när vi närmar oss origo, eftersom divisionen med sinus hyperbolicus då ger division med mindre och mindre tal.

1.11.1 Hyperboliska identiteter

De hyperboliska funktionerna uppfyller, precis som de vanliga trigonometriska funktionerna, en rad identiteter. Med tanke på likheterna mellan Eulers formler för sinus och cosinus och definitionerna av sinus och cosinus hyperbolicus är det kanske inte så konstigt att de två uppsättningarna med identiteter påminner mycket om varandra.

Sats. För varje x och $y \in \mathbb{R}$ gäller

$$\begin{aligned}
&\cosh^2 x - \sinh^2 x = 1 \\
&\sinh(x \pm y) = \sinh x \cosh y \pm \cosh x \sinh y \\
&\cosh(x \pm y) = \cosh x \cosh y \pm \sinh x \sinh y \\
&\sinh 2x = 2 \sinh x \cosh x \\
&\cosh 2x = \cosh^2 x + \sinh^2 x.
\end{aligned}$$

Bevis. Vi har

$$\cosh^2 x - \sinh^2 x = \left(\frac{e^x + e^{-x}}{2}\right)^2 - \left(\frac{e^x - e^{-x}}{2}\right)^2 = \frac{e^{2x} + 2 + e^{-2x}}{4} - \frac{e^{2x} - 2 + e^{-2x}}{4} = 1$$

och

$$\begin{aligned}
\sinh x \cosh y + \cosh x \sinh y &= \frac{e^x - e^{-x}}{2} \cdot \frac{e^y + e^{-y}}{2} + \frac{e^x + e^{-x}}{2} \cdot \frac{e^y - e^{-y}}{2} = \\
&= \frac{e^{x+y} + e^{x-y} - e^{-x+y} - e^{-x-y}}{4} + \frac{e^{x+y} - e^{x-y} + e^{-x+y} - e^{-x-y}}{4} = \\
&= \frac{2e^{x+y} - 2e^{-x-y}}{4} = \frac{e^{x+y} - e^{-x-y}}{2} = \sinh(x + y)
\end{aligned}$$

och liknande för minustecknet. På samma sätt visas cosh:s additionsformel. Dubbla vinkel-formlerna följer sedan direkt av respektive additionsformel. ∎

Övningar

1. Jämför de hyperboliska identiteterna med motsvarande trigonometriska identiteter (s. 93).
2. Härled additionsformeln för tangens hyperbolicus och jämför resultatet med den trigonometriska motsvarigheten (s. 98).
3. Visa att $\cosh x \pm \sinh x = e^{\pm x}$.
4. Visa att $1 - \tanh^2 x = \operatorname{sech}^2 x$. Vad är motsvarande trigonometriska formel?

1.11.2 Inversa hyperboliska funktioner

Sinus, tangens, cotangens och cosekant hyperbolicus är injektiva, men det är inte cosinus och sekant hyperbolicus; de injektiva funktionernas inverser betecknas arcsinh, arctanh, arccoth och arccsch. Det är emellertid klart att restriktionerna av cosinus och sekant hyperbolicus till $[0, \infty[$ är injektiva, och inverserna till *dessa* restriktioner betecknas arccosh respektive arcsech.

Sats.

$$\begin{aligned}
\operatorname{arcsinh} x &= \ln\left(x + \sqrt{x^2+1}\right), \qquad \forall x \in \mathbb{R} \\
\operatorname{arccosh} x &= \ln\left(x + \sqrt{x^2-1}\right), \qquad \forall x \geq 1 \\
\operatorname{arctanh} x &= \frac{1}{2}\ln\frac{1+x}{1-x}, \qquad \forall x \in \left]-1, 1\right[.
\end{aligned}$$

Bevis. Välj något $y \in V_{\sinh}$. Då är

$$y = \sinh x = \frac{e^x - e^{-x}}{2}$$

för något $x \in D_{\sinh}$. Med $t \coloneqq e^x$ har vi alltså

$$\begin{aligned}
y = \frac{t - t^{-1}}{2} \quad &\Longrightarrow \quad 2y = t - t^{-1} \quad \Longrightarrow \quad 2yt = t^2 - 1 \quad \Longrightarrow \\
&\Longrightarrow \quad t^2 - 2yt - 1 = 0 \quad \Longrightarrow \quad (t-y)^2 - y^2 - 1 = 0 \quad \Longrightarrow \\
&\Longrightarrow \quad (t-y)^2 = 1 + y^2 \quad \overset{*}{\Longrightarrow} \quad t - y = \sqrt{1+y^2} \quad \Longrightarrow \\
&\Longrightarrow \quad t = y + \sqrt{1+y^2}
\end{aligned}$$

där $*$ följer av att $t - y = e^x - \sinh x = \cosh x \geq 0$. Vi har sålunda $e^x = y + \sqrt{1+y^2}$ så att $x = \ln\left(y + \sqrt{1+y^2}\right)$. Välj nu ett $y \in V_{\cosh}$. Vi vill hitta det $x \in [0, \infty[\subset D_{\cosh}$ för vilket

$$y = \cosh(x) = \frac{e^x + e^{-x}}{2}.$$

Med $t \coloneqq e^x$ har vi då

$$\begin{aligned}
y = \frac{t + t^{-1}}{2} \quad &\Longrightarrow \quad 2y = t + t^{-1} \quad \Longrightarrow \quad 2yt = t^2 + 1 \quad \Longrightarrow \\
&\Longrightarrow \quad t^2 - 2yt + 1 = 0 \quad \Longrightarrow \quad (t-y)^2 - y^2 + 1 = 0 \quad \Longrightarrow \\
&\Longrightarrow \quad (t-y)^2 = y^2 - 1 \quad \overset{*}{\Longrightarrow} \quad t - y = \sqrt{y^2-1} \quad \Longrightarrow \\
&\Longrightarrow \quad t = y + \sqrt{y^2-1}
\end{aligned}$$

där $*$ följer av att $t - y = e^x - \cosh x = \sinh x \geq 0$ ty $x \geq 0$. Sålunda är $e^x = y + \sqrt{y^2-1}$ varför $x = \ln\left(y + \sqrt{y^2-1}\right)$. Slutligen väljer vi något $y \in V_{\tanh}$ så att

$$y = \tanh x = \frac{\sinh x}{\cosh x} = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

för något $x \in D_{\tanh}$. Med $t \coloneqq e^x$ har vi

$$y = \frac{t - t^{-1}}{t + t^{-1}} \quad \Longrightarrow \quad ty + t^{-1}y = t - t^{-1} \quad \Longrightarrow$$
$$\Longrightarrow \quad t^2 y + y = t^2 - 1 \quad \Longrightarrow \quad (1 - y)t^2 = 1 + y \quad \Longrightarrow$$
$$\Longrightarrow \quad t^2 = \frac{1 + y}{1 - y} \quad \Longrightarrow \quad t = \sqrt{\frac{1 + y}{1 - y}}$$

där de två sista stegen följer av att $y < 1$ och att $t = e^x > 0$. Således är

$$x = \ln \sqrt{\frac{1 + y}{1 - y}} = \frac{1}{2} \ln \frac{1 + y}{1 - y}$$

och satsen är bevisad. ■

Övning

1. Bestäm explicita formler för $\operatorname{arccoth} x$, $\operatorname{arcsech} x$ och $\operatorname{arccsch} x$.
2. Rita graferna till de inversa hyperboliska funktionerna.

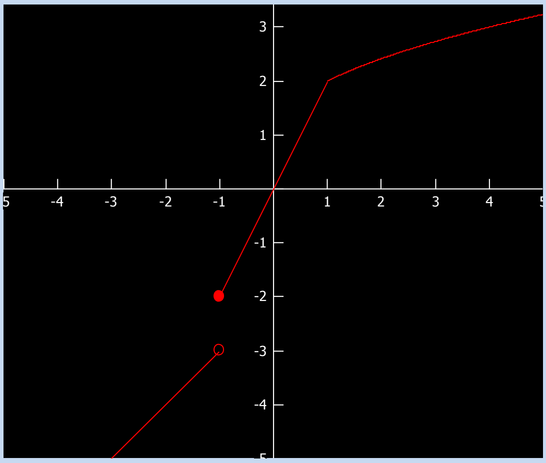
1.12 Blandade exempel

Exempel

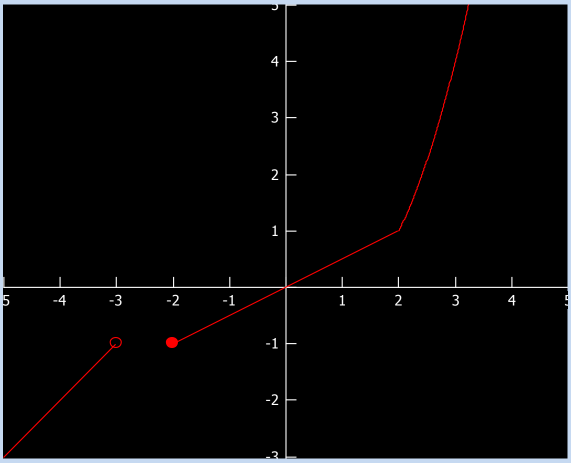
Bestäm, om möjligt, inversen till funktionen $f\colon \mathbb{R} \to \mathbb{R}$ definierad av

$$f(x) := \begin{cases} x - 2 & \text{om } x < -1 \\ 2x & \text{om } x \in [-1, 1[\\ \sqrt{x} + 1 & \text{om } x \geq 1. \end{cases}$$

Lösning: Funktionen är falldefinierad, och i varje enskilt intervall är den strängt växande. Dessutom ser vi (t.ex. i grafen) att hela funktionen är strängt växande, och därmed injektiv. Grafen till funktionen visas f nedan till vänster.



$y = f(x)$



$y = f^{-1}(x)$

Om $y = f(x)$ gäller att (1) om $y < -3$ så är $x = y + 2$, (2) om $y \in [-2, 2[$ så är $x = \frac{1}{2}y$ och (3) om $y \geq 2$ så är $x = (y-1)^2$. Alltså är inversen $f^{-1}\colon]-\infty, -3[\cup [-2, \infty[$ definierad av

$$f^{-1}(x) = \begin{cases} x + 2 & \text{om } x < -3 \\ x/2 & \text{om } x \in [-2, 2[\\ (x-1)^2 & \text{om } x \geq 2. \end{cases}$$

Grafen till inversen f^{-1} visas ovan till höger.

Exempel

Är funktionen $x \mapsto \sin x + \cos x$ injektiv på $\left[0, \frac{\pi}{4}\right]$?

Lösning: Sinustermen växer och cosinustermen avtar på intervallet, så det är inte helt uppenbart hur summan beter sig. Men med hjälpvinkelmetoden ser vi att

$$\sin x + \cos x = \sqrt{2}\sin(x + \delta)$$

där $\cos\delta = \sin\delta = 1/\sqrt{2}$, förslagsvis $\delta = \pi/4$. Sålunda är $\sin x + \cos x = \sqrt{2}\sin\left(x + \frac{\pi}{4}\right)$ som är strängt växande på $\left[0, \frac{\pi}{4}\right]$. Funktionen är därför injektiv på intervallet.

Exempel

Lös ekvationen $\cos^3 x \sin x - \cos x \sin^3 x = \frac{\sqrt{3}}{8}$.

Lösning: Eftersom

$$\cos^3 x \sin x - \cos x \sin^3 x = \sin x \cos x\,(\cos^2 x - \sin^2 x) = \frac{1}{2}\sin 2x \cos 2x = \frac{1}{4}\sin 4x$$

är ekvationen ekvivalent med

$$\sin 4x = \frac{\sqrt{3}}{2}$$

som i sin tur är ekvivalent med

$$\begin{cases} 4x = \dfrac{\pi}{3} + n \cdot 2\pi \\ \text{eller} \\ 4x = \dfrac{2\pi}{3} + n \cdot 2\pi. \end{cases}$$

Svar: $\cos^3 x \sin x - \cos x \sin^3 x = \frac{\sqrt{3}}{8} \iff x = \frac{\pi}{12} + \frac{n\pi}{2}$ eller $x = \frac{\pi}{6} + \frac{n\pi}{2}$ för något $n \in \mathbb{Z}$.

Alternativt kan Eulers formler användas för att mer mekaniskt skriva om en produkt av sinus- och cosinustermer som en summa av enkla sådana:

Lösning: Vi har

$$\cos^3 x \sin x = \left(\frac{e^{ix} + e^{-ix}}{2}\right)^3 \left(\frac{e^{ix} - e^{-ix}}{2i}\right) = \frac{e^{3ix} + 3e^{ix} + 3e^{-ix} + e^{-3ix}}{8} \cdot \frac{e^{ix} - e^{-ix}}{2i} =$$
$$= \frac{e^{4ix} + 3e^{2ix} + 3 + e^{-2ix} - e^{2ix} - 3 - 3e^{-2ix} - e^{-4ix}}{16i}$$

och

$$\cos x \sin^3 x = \left(\frac{e^{ix} + e^{-ix}}{2}\right)\left(\frac{e^{ix} - e^{-ix}}{2i}\right)^3 = \frac{e^{ix} + e^{-ix}}{2} \frac{e^{3ix} - 3e^{ix} + 3e^{-ix} - e^{-3ix}}{-8i} =$$
$$= \frac{-e^{4ix} + 3e^{2ix} - 3 + e^{-2ix} - e^{2ix} + 3 - 3e^{-2ix} + e^{-4ix}}{16i}$$

så att

$$\cos^3 x \sin x - \cos x \sin^3 x = \frac{2e^{4ix} - 2e^{-4ix}}{16i} = \frac{e^{4ix} - e^{-4ix}}{8i} = \frac{1}{4}\sin 4x$$

Därför är ekvationen ekvivalent med …

Exempel

Lös ekvationen $\cos^4 x + \sin^2 x = 1$.

Lösning: Eftersom $\cos^4 x = (\cos^2 x)^2 = (1 - \sin^2 x)^2$ kan ekvationen skrivas

$$(1 - t^2)^2 + t^2 = 1$$

där $t \coloneqq \sin x$. Eftersom

$$(1 - t^2)^2 + t^2 = 1 \iff 1 - 2t^2 + t^4 + t^2 = 1 \iff t^4 - t^2 = 0 \iff t^2(t^2 - 1) = 0 \iff t \in \{-1, 0, 1\}$$

är ekvationen ekvivalent med $\sin x \in \{-1, 0, 1\}$ d.v.s. $x = \frac{n\pi}{2}$ för något $n \in \mathbb{Z}$.

Svar: $\cos^4 x + \sin^2 x = 1 \iff x = \frac{n\pi}{2}$ för något $n \in \mathbb{Z}$.

Exempel

Lös ekvationen $\sin^2 x - 3\sin 2x = -8\cos^2 x$.

Lösning: Eftersom $\cos x \neq 0$ har vi

$$\sin^2 x - 3\sin 2x = -8\cos^2 x \iff \sin^2 x - 6\sin x\cos x = -8\cos^2 x \iff \tan^2 x - 6\tan x = -8 \iff \tan^2 x - 6\tan x + 8 = 0 \iff (\tan x - 2)(\tan x - 4) = 0 \iff \begin{cases} x = \arctan 2 + n\pi \\ \text{eller} \\ x = \arctan 4 + n\pi. \end{cases}$$

Svar: $\sin^2 x - 3\sin 2x = 8\cos^2 x \iff x = \arctan 2 + n\pi$ eller $x = \arctan 4 + n\pi$ för ngt. $n \in \mathbb{Z}$.

Exempel

Bestäm alla komplexa lösningar till ekvationen $e^z = 1$.

Lösning: Sätt $z = a + bi$ där $a, b \in \mathbb{R}$. Per definition är $e^z = e^a(\cos b + i\sin b)$ så att ekvationen lyder

$$e^a \cos b + i\, e^a \sin b = 1$$

som är ekvivalent med

$$\begin{cases} e^a \cos b = 1 \\ e^a \sin b = 0 \end{cases}$$

som först ger $b = n \cdot 2\pi$ (motivera!) för något $n \in \mathbb{Z}$ och sedan $e^a = 1$, d.v.s. $a = 0$.

Svar: $e^z = 1 \iff z = n \cdot 2\pi i$ för något $n \in \mathbb{Z}$.

Exempel

Beräkna summan $1-\frac{1}{2}+\frac{1}{4}-\frac{1}{8}-\cdots+\frac{1}{4096}$.

Lösning: Det rör sig tydligen om en geometrisk summa

$$\sum_{k=0}^{N}\left(-\frac{1}{2}\right)^k$$

för något $N \in \mathbb{Z}$. Vilket? Jo, sista termen är

$$\left(-\frac{1}{2}\right)^N = \frac{1}{4096}$$

som ger $N = 12$. Alltså är

$$1-\frac{1}{2}+\frac{1}{4}-\frac{1}{8}-\cdots+\frac{1}{4096} = \sum_{k=0}^{12}\left(-\frac{1}{2}\right)^k = \frac{1-\left(-\frac{1}{2}\right)^{13}}{1-\left(-\frac{1}{2}\right)} = \frac{1+\frac{1}{2^{13}}}{\frac{3}{2}} = \frac{2}{3}+\frac{2}{3\cdot 2^{13}} = \frac{2}{3}+\frac{1}{3\cdot 2^{12}} =$$
$$= \frac{2}{3}+\frac{1}{12288}\left(=\frac{2731}{4096} = 0.666748046875\right).$$

Exempel

Lös ekvationen $1+x+x^2+x^3+\cdots+x^{10} = 0$.

Lösning: Vänsterledet är en geometrisk summa med kvot x, första term 1 och antalet termer 11. Sålunda är, med tanke på att ekvationen medför att $x \neq 1$,

$$1+x+x^2+x^3+\cdots+x^{10} = \frac{1-x^{11}}{1-x}$$

varför

$$1+x+x^2+x^3+\cdots+x^{10} = 0 \iff \frac{1-x^{11}}{1-x} = 0 \iff$$
$$\iff \begin{cases} 1-x^{11} = 0 \\ x \neq 1 \end{cases} \iff \begin{cases} x^{11} = 1 \\ x \neq 1. \end{cases}$$

Med $x = re^{i\varphi}$ har vi

$$x^{11} = 1 \iff r^{11}e^{11i\varphi} = 1e^{0i} \iff$$
$$\iff \begin{cases} r^{11} = 1 \\ 11\varphi = 0 + n\cdot 2\pi \end{cases} \iff \begin{cases} r = 1 \\ \varphi = n\cdot\dfrac{2\pi}{11}. \end{cases}$$

Med andra ord är $x^{11} = 1$ ekvivalent med att $x = e^{n\cdot\frac{2\pi}{11}i}$ för något $n = 0, \dots, 10$. Men *ursprungs-ekvationen* är ju ekvivalent med att $x^{11} = 1$ *och* $x \neq 1$, så $n = 0$ måste tydligen exkluderas.

Svar: $1+x+x^2+x^3+\cdots+x^{10} = 0 \iff x = e^{n\cdot\frac{2\pi}{11}i}$ för något $n = 1, \dots, 10$.

Speciellt ser vi att ekvationen *saknar* reell lösning. Vad händer om vi tar med en term till i summan, så att den slutar med x^{11}?

Exempel

Välj ett fixt tal $a \in \mathbb{R}$. Visa att resten som erhålles då polynomet $p(x)$ delas med det moniska förstagradspolynomet $x - a$ är lika med $p(a)$.

Lösning: Divisionsalgoritmen ger att

$$p(x) = (x - a)k(x) + r(x), \qquad \forall x \in \mathbb{R}$$

där $k(x)$ är kvoten och $r(x)$ är resten. Eftersom nämnaren $x - a$ har grad 1 måste resten $r(x)$ vara en konstant, så vi kan skriva $r(x) =: r$, så

$$p(x) = (x - a)k(x) + r, \qquad \forall x \in \mathbb{R}.$$

Speciellt gäller, för $x = a$, att

$$p(a) = r$$

som önskat. ■

Exempel

Lös ekvationen $\operatorname{sech} x = \frac{1}{2}$.

Lösning (naiv): Vi har

$$\begin{aligned}
\operatorname{sech} x = \frac{1}{2} \quad &\Longleftrightarrow \quad \frac{1}{\cosh x} = \frac{1}{2} \quad \Longleftrightarrow \quad \frac{2}{e^x + e^{-x}} = \frac{1}{2} \quad \Longleftrightarrow \\
&\Longleftrightarrow \quad e^x + e^{-x} = 4 \quad \Longleftrightarrow \quad e^{2x} + 1 = 4e^x \quad \Longleftrightarrow \\
&\Longleftrightarrow \quad e^{2x} - 4e^x + 1 = 0 \quad \Longleftrightarrow \quad (e^x - 2)^2 - 3 = 0 \quad \Longleftrightarrow \\
&\Longleftrightarrow \quad (e^x - 2)^2 = 3 \quad \Longleftrightarrow \quad e^x - 2 = \pm\sqrt{3} \quad \Longleftrightarrow \\
&\Longleftrightarrow \quad e^x = 2 \pm \sqrt{3} \quad \Longleftrightarrow \quad x = \ln\left(2 \pm \sqrt{3}\right).
\end{aligned}$$

Svar: $\operatorname{sech} x = \frac{1}{2} \quad \Longleftrightarrow \quad x = \ln\left(2 \pm \sqrt{3}\right)$.

Vi kan förstås också använda oss av den explicita formeln för arccosh:

Lösning: Vi har

$$\begin{aligned}
\operatorname{sech} x = \frac{1}{2} \quad &\Longleftrightarrow \quad \frac{1}{\cosh x} = \frac{1}{2} \quad \Longleftrightarrow \quad \cosh x = 2 \quad \Longleftrightarrow \\
&\Longleftrightarrow \quad x = \pm \operatorname{arccosh} 2 = \pm \ln\left(2 + \sqrt{2^2 - 1}\right) = \pm \ln\left(2 + \sqrt{3}\right).
\end{aligned}$$

Notera att $-\ln\left(2 + \sqrt{3}\right) = \ln\left(2 + \sqrt{3}\right)^{-1} = \ln\frac{1}{2+\sqrt{3}} = \ln\frac{2-\sqrt{3}}{(2+\sqrt{3})(2-\sqrt{3})} = \ln\frac{2-\sqrt{3}}{4-3} = \ln\left(2 - \sqrt{3}\right)$ så de två svaren stämmer överens.