



STOCKHOLMS MATEMATISKA CIRKEL

HYPERBOLISK GEOMETRI

DAN PETERSEN

ALAN SOLA

REVIDERAT AV

DANIEL ERIKSSON

HAMPUS NYBERG

INSTITUTIONEN FÖR MATEMATIK, KTH OCH
MATEMATISKA INSTITUTIONEN, STOCKHOLMS UNIVERSITET
2025–2026

STOCKHOLMS MATEMATISKA CIRKEL genom tiderna
(tidigare KTH:S MATEMATISKA CIRKEL)

2025-2026	Hyperbolisk geometri
2024-2025	Felrättande koder
2023-2024	Elliptiska Kurvor
2022-2023	Variationskalkyl
2021-2022	Matematik och AI
2020-2021	Musik och matematik
2019-2020	Datorernas matematik
2018-2019	Grafteori med inriktning på färgläggning
2017-2018	Geometriska konstruktioner
2016-2017	Vad är ett tal?
2015-2016	Fraktaler
2014-2015	Polytoper
2013-2014	Grupper, mönster och symmetrier
2012-2013	Den matematiska analysens grunder
2011-2012	Diofantiska ekvationer
2010-2011	Polynom
2009-2010	Hyperbolisk geometri
2008-2009	Talteori
2007-2008	Sannolikhetsteori
2006-2007	Gruppteori
2005-2006	Vad är ett tal?
2004-2005	Integraler
2003-2004	Linjär algebra och bioinformatik
2002-2003	Algebra och kryptografi
2001-2002	Analysens grunder
2000-2001	Talföljder, rekursioner och iterationer
1999-2000	Linjära avbildningar

Innehåll

Lista över symboler	v
Några ord på vägen	vi
1 Mängdlära	1
1.1 Definition, axiom, sats och bevis	1
1.2 Bevistekniker	2
1.3 Olika satser och hur man bevisar dem	4
1.4 Mängder	5
1.5 Funktioner	8
1.6 Mängdoperationer	11
1.7 Kartor	14
2 Euklidisk geometri	19
2.1 Definitioner	19
2.2 Automorfier	21
2.3 Axiom	25
2.4 Historik över parallellpostulatet	27
3 Komplexa tal	31
3.1 Operationer med punkter i planet	31
3.2 Real- och imaginärdel av komplexa tal	33
3.3 Konjugat, längd och vinkel.	35
3.4 Komplexa tal på polär form	35
3.5 Komplexa funktioner	37
3.6 Komplexa polynom	39
3.7 Mer allmänna funktioner av komplexa tal	39
3.8 Komplex derivata	39

4	Båglängd	43
4.1	Kurvor i komplexa talplanet	43
4.2	Linjeintegraler	46
4.3	Variabelbyten och automorfier	49
4.4	Avstånd och geodeter	52
5	Generella automorfier och Riemannsfären	56
5.1	Automorfier för sfären	56
5.2	Kartor	56
5.3	Riemannsfären	57
5.4	Automorfier	60
5.5	Möbiusavbildningar	62
6	Vidare egenskaper hos $\text{Möb}(\mathbb{C})$.	68
6.1	Cirklar och linjer	68
6.2	3-transitivitet	70
6.3	Konforma avbildningar	73
7	Delgruppen $\text{Aut}(\mathbb{H})$.	78
7.1	Möbiusavbildningar som bevarar övre halvplanet	78
7.2	Hyperboliska linjer	82
8	Hyperbolisk längd och hyperboliskt avstånd	86
8.1	Hyperbolisk längd	86
8.2	Invariants under $\text{Möb}(\mathbb{H})$	88
8.3	Hyperboliskt avstånd och hyperboliska geodeter	89
8.4	Hyperbolisk geometri	92
8.5	Sfärisk geometri	93
8.6	Areor, vinklar och Gauss-Bonnets sats	93
	Lösningar till udda övningsuppgifter	97
A	Appendix	111
A.1	Derivator	111
A.2	Integraler	112
	Referenser och förslag till vidare läsning	115

Lista över grekiska alfabetet

A	α	alfa
B	β	beta
Γ	γ	gamma
Δ	δ	delta
E	ε	epsilon
Z	ζ	zeta
H	η	eta
Θ	θ	theta
I	ι	iota
K	κ	kappa
Λ	λ	lambda
M	μ	my
N	ν	ny
Ξ	ξ	xi
O	o	omikron
Π	π	pi
P	ρ	rho
Σ	σ	sigma
T	τ	tau
Υ	υ	ypsilon
Φ	ϕ	fi
X	χ	chi
Ψ	ψ	psi
Ω	ω	omega

Lista över mängdsymboler

$x \in X$	x är ett element i mängden X .
$X \subset Y$	X är en delmängd till Y .
$\forall x \in X$	för varje x som är ett element i mängden X .
$\exists x \in X \dots$	existerar ett element x i mängden X sådant att $\dots$
$f: X \rightarrow Y$	f är en funktionen med definitionsmängd X och målmängd Y .

Några ord på vägen

Detta kompendium är skrivet för att användas som kurslitteratur till STOCKHOLMS MATEMATISKA CIRKEL under läsåret 2025–2026 och består av åtta kapitel. Det är en omarbetning av materialet till cirkeln given 2009–2010 skriven av Dan Peterson och Alan Sola.

Kompendiet är inte tänkt att läsas enbart på egen hand, utan ska ses som ett skriftligt komplement till undervisningen. Alla elever rekommenderas att läsa igenom varje kapitel själv innan föreläsningen. Det är inte nödvändigt att förstå alla detaljer vid den första genomläsningen.

Som de flesta matematiska skrifter på högre nivå är kompendiet kompakt skrivet. Detta innebär att man i allmänhet inte kan läsa det som en vanlig bok. Istället bör man pröva nya satser och definitioner genom att på egen hand exemplifiera. Därmed uppnår man oftast en mycket bättre förståelse av vad dessa satser och deras bevis går ut på.

Till varje kapitel finns ett antal övningsuppgifter. De udda övningarna har lösningar längst bak i kompendiet. Syftet med dessa är att eleverna ska kunna lösa dem och på egen hand kontrollera att de förstått materialet. Övningar med jämna nummer saknar facit och kan användas som examination. Det rekommenderas dock att man försöker lösa dessa uppgifter även om man inte examineras på dem.

Om man kör fast kan man alltid fråga en kompis, en lärare på sin skola eller någon av författarna. Under årets gång kommer det att finnas övningstillfällen där eleverna kan jobba med uppgifterna, själva eller i grupp, och få hjälp av oss.

Vissa övningar kan ha flera lösningar och det som står i facit bör i detta fall endast ses som ett förslag.

Några ord om cirkeln

STOCKHOLMS MATEMATISKA CIRKEL är en kurs för matematikintresserade gymnasieelever, som arrangeras av Kungliga Tekniska högskolan och Stockholms universitet. Cirkeln startade 1999. Vid starten hade den namnet KTH:s MATEMATISKA CIRKEL och hölls i KTH:s ensamma regi. Ambitionen med cirkeln är att sprida kunskap om matematiken och dess användningsområden utöver vad eleverna får genom gymnasiekurser, och att etablera ett närmare samarbete mellan gymnasieskolan och högskolan. Cirkeln ska särskilt stimulera elevernas matematikintresse och inspirera dem till fortsatta naturvetenskapliga och matematiska studier.

Till varje kurs skrivs ett kompendium som distribueras gratis till eleverna. Detta material, föreläsningsschema och övrig information om STOCKHOLMS MATEMATISKA CIRKEL finns tillgängligt på

<https://www.math-stockholm.se/cirkel>.

Cirkeln godkänns ofta som en gymnasiekurs eller som matematisk breddning på gymnasieskolorna. Det är upp till varje skola att godkänna cirkeln som en kurs och det är lärarna från varje skola som sätter betyg på kursen. Lärarna är självklart också välkomna till cirkeln och många har kommit överens med sin egen skola om att få cirkeln godkänd som fortbildning eller som undervisning.

Vi vill gärna understryka att föreläsningarna är öppna för alla gymnasieelever, lärare eller andra matematikintresserade.

Vi har avsiktligt valt materialet för att ge eleverna en inblick i matematisk teori och tankesätt och presenterar därför både några huvudsatser inom varje område och bevisen för dessa resultat. Vi har också som målsättning att bevisa alla satser som används om de inte kan förutsättas bekanta av elever från gymnasiet.

Årets tema är hyperbolisk geometri. Hyperbolisk geometri skiljer sig både från vanlig euklidisk geometri (geometri i planet) och från sfärisk geometri. Det hyperboliska planet 'buktar inåt' överallt och vinkelsummor i trianglar har mindre än 180 grader i vinkelsumma. Hyperbolisk geometri används i fysiken för att beskriva universum. Andrew Wiles använde också funktioner på hyperboliska planet i sitt kända bevis av Fermats stora sats.

Kapitel 1 är en grundläggande introduktion till mängdlära. För att förstå likheter och skillnad mellan hyperbolisk och euklidisk geometri beskriver vi euklidisk geometri i kapitel 2. I kapitel 3 går vi igenom komplexa tal som vi använder för att kartlägga det hyperboliska planet. Båglängd av kurvor i komplexa talplanet diskuteras i kapitel 4. I kapitel 5 definierar vi automorfier och Riemannsfären. Automorfierna är en "välskapt" familj av funktioner som bevarar avstånd och vinklar. Riemannsfären är en kartläggning av sfären med hjälp av komplexa tal. Vi definierar även Möbiusgruppen som generaliserar de automorfierna vi stöter på. Fler egenskaper hos Möbiusgruppen behandlas i kapitel 6. I kapitel 7 begränsar vi oss till övre halvplanet, där kommer Möbiusgruppen ge oss en kartläggning för det hyperboliska planet. Vi avslutar med att introducera hy-

perbolisk längd och hyperboliskt avstånd i kapitel 8 för att, slutligen, beskriva det hyperboliska planet.

Författarna, sommaren 2025

1 Mängdlära

Temat för årets matematiska cirkel är Hyperbolisk geometri. Detta kapitel är en introduktion i den matematiska metoden och ett antal grundbegrepp som vi kommer använda oss av i kursen.

1.1 Definition, axiom, sats och bevis

I detta avsnitt ska vi beskriva den matematiska metoden utifrån fyra begrepp: *definition*, *sats*, *bevis* och *axiom*.

En *definition* bestämmer vad en term betyder så att man kan arbeta matematiskt med den. Till exempel kan vi definiera udda och jämna tal på följande sätt.

Definition 1.1.1. Ett heltal n är *udda* om det finns ett heltal k som uppfyller att $n = 2k + 1$.

Definition 1.1.2. Ett heltal n är *jämnt* om det finns ett heltal k som uppfyller att $n = 2k$.

Ofta har man en intuition om vad en term betyder redan innan man definierar den. Läsaren hade till exempel säkert en uppfattning om vad udda och jämna tal är innan vi definierade dem. Syftet med en definition är att precisera detta.

När definitionen är gjord, så överger man sina tidigare uppfattningar om vad termen betyder och utgår endast ifrån definitionen. Man säger att definitionen är *stipulativ*. En definition är alltså inte rätt eller fel, utan bara mer eller mindre användbar och intuitiv.

Definitioner bygger ofta på begrepp som läsaren är bekant med. Till exempel utgår Definition 1.1.1 och 1.1.2 från att läsaren redan vet vad ett heltal är.

En *sats* är ett påstående som bevisats vara sant. Varje sats hör samman med ett *bevis*: ett argument för att påståendet är sant.

Sats 1.1.3. Om n är udda, så är $n + 1$ jämnt.

Bevis. Om n är udda så finns det ett heltal k så att $n = 2k + 1$. Då gäller att

$$n + 1 = 2k + 1 + 1 = 2k + 2 = 2(k + 1).$$

Eftersom $k + 1$ är ett heltal, så är $n + 1$ ett jämnt tal. □

Bevisen kombinerar definitioner och olika logiska slutledningsregler för att nå den önskade slutsatsen. Sats 1.1.3 har en syskonsats. Beviset är mer eller mindre identiskt, och lämnas som övning.

Sats 1.1.4. Om n är jämnt, så är $n + 1$ udda.

En sats vars främsta syfte är att användas i beviset av en annan sats kallas för en *hjälpssats* eller ett *lemma*. En sats som följer omedelbart ur en annan sats, till exempel som ett specialfall, kallas för en *följdsats* eller ett *korollarium*.

Ett påstående måste vara bevisat för att få kallas för en sats. Om man har goda skäl att tro att ett påstående är sant men inte formellt bevisat det kallas påståendet för en *förmodan*, eller *hypotes*. Två exempel är *Riemannhypotesen* och *primtalstvillingsförmodan*.

En förmodan kan förbli obevisad i hundratals år. Ett berömt exempel är *Fermats stora sats*, som formulerades av Pierre de Fermat (1607–1665) år 1637 men bevisades först av Andrew Wiles år 1995. Riemannhypotesen, som ännu är obevisad, formulerades 1859 av Bernard Riemann (1826–1866).

Eftersom bevisen utgår ifrån definitionen, och inte vår intuition, så behöver man ibland bevisa saker som känns uppenbara. Läsaren vet till exempel att

- (i) alla tal är antingen udda eller jämna, och
- (ii) ett tal kan inte vara udda och jämnt samtidigt.

Men om man läser Definition 1.1.1 och 1.1.2 så ingår inte dessa påståenden. Kan man inte tänka sig tal som varken är udda eller jämnt? Eller tal som är både och?

Bevis utgår ifrån antaganden och tidigare kända satser. Dessa tidigare satser måste också bevisas innan de kan anses giltiga. Men dessa bevis måste också bygga på antaganden och satser, som också måste bevisas, och så vidare.

För att undvika en oändlig kedja av bevis, eller ett cirkulärt bevis (ett bevis som använder sig av det man försöker bevisa) så måste man göra grundantaganden som inte behöver bevisa. Dessa kallas för *axiom*. Exempel på axiom är att mängden av heltal existerar och att addition uppfyller

- (Associativitet) För alla heltal n, m, p gäller $(n + m) + p = n + (m + p)$.
- (Identitet) Det finns ett element 0, som vi kallar *nollan*, sådan att för alla heltal n så gäller $0 + n = n + 0 = n$.
- (Inverser) För varje heltal n existerar ett heltal $(-n)$, som vi kallar för *minus n*, sådan att $0 = n + (-n) = (-n) + n$.
- (Kommutativitet) För alla heltal n, m så gäller $n + m = m + n$.

1.2 Bevistekniker

Ett bevis för en sats är ett argument som förklarar varför satsen är sann. Vi har redan sett ett exempel när vi bevisade Sats 1.1.3. I detta avsnitt ska vi gå igenom två metoder för att bevisa matematiska satser: direkta bevis och motsägelsebevis.¹

Ett *direkt bevis* utgår ifrån satsens antaganden och definitioner och bevisar satsen rakt på, så att säga. Beviset av Sats 1.1.3 är ett exempel på direkt bevis. Ett annat är följande sats.

¹Ibland förekommer termen *indirekt bevis*. Vissa använder det som synonym till motsägelsebevis, andra som en synonym till bevisregeln *modus tollens*. Vi undviker den helt.

Sats 1.2.1. Antalet funktioner från en mängd A med n element till en mängd B med m element är m^n .

Bevis. Varje funktion från A till B kan beskrivas som en tabell där varje element i A motsvaras av precis ett element i B . Listan innehåller totalt n platser, och på varje plats kan vi välja bland m element att välja bland. Alltså finns det totalt

$$\underbrace{m \cdot m \cdots m \cdot m}_{n \text{ stycken}} = m^n$$

olika funktioner. □

Ibland går det inte att använda direkta bevis, till exempel när man ska bevisa att något inte är fallet. Då kan det vara enklare att anta att det man vill bevisa är falskt, och visa att detta leder till en motsägelse. Om alla steg i beviset är korrekta så måste det ursprungliga antagandet vara fel. Detta kallas för ett *motsägelsebevis*.

Sats 1.2.2. Ett tal kan inte vara udda och jämnt samtidigt.

Bevis. Antag att n är ett tal som är både udda och jämnt. Då finns det två heltal, k och l , så att $n = 2k$ och $n = 2l + 1$. Då gäller att

$$2k = n = 2l + 1 \implies 2k - 2l = 1 \implies 2(k - l) = 1.$$

Med andra ord finns det heltal $m = k - l$ så att $2m = 1$. Kan det finnas ett sådant tal? Det finns två fall.

(i) Om $m \leq 0$, så är $1 = 2m \leq 0$. Motsägelse!

(ii) Om $m \geq 1$ så är $2m \geq 2 > 1$. Motsägelse! □

Ett berömt motsägelsebevis är följande.

Sats 1.2.3. Talet $\sqrt{2}$ är irrationellt.

Bevis. Antag motsatsen, det vill säga att $\sqrt{2} = a/b$ för några heltal a och b . Antag att a och b är förkortade så långt som möjligt. Då kan endast en av a eller b vara jämn, eftersom om båda är jämna kan vi skriva

$$\sqrt{2} = \frac{a}{b} = \frac{2c}{2d} = \frac{c}{d}$$

och då var inte a och b förkortade så långt som möjligt.

Av definitionen av $\sqrt{2}$ får vi att

$$\sqrt{2}^2 = 2 = \frac{a^2}{b^2} \implies 2b^2 = a^2.$$

Den sista ekvationen säger att a^2 är jämn. Eftersom kvadrater av udda tal är udda (se Övning 1.22), så måste a vara ett jämnt tal, det vill säga $a = 2k$ för något heltal k . Då får vi att

$$2b^2 = (2k)^2 = 4k^2 \implies b^2 = 2k^2.$$

Eftersom b^2 är jämnt, så måste b vara jämnt. Men nu har vi bevisat att både a och b är jämna, vilket var omöjligt eftersom vi hade förkortat bråket så långt som möjligt. Detta är en motsägelse. $\square$

Att bevis inkluderar antaganden och små hjälpsatser som inte nämns är snarare regel än undantag, till exempel lägger vi nästan aldrig tid på att visa $2 > 0$ i mitten av ett bevis. Ifall man bevisade precis vartenda antagande och påstående utifrån axiomen varje gång skulle bevisen bli väldigt långa och komplicerade. Läsaren förväntas själv fylla i de luckor som uppstår.

Det händer dock att uppenbara antaganden är mycket svåra, till och med omöjliga, att bevisa utifrån definitionerna. Historien är fylld av matematiker som gjort till synes självklara antaganden som sedan visat sig vara svåra att bevisa.

Beviset av Sats 1.2.3 är ett exempel på det. Vi antar att ett bråk kan förkortas så långt som möjligt. Detta är inte självklart, utan bygger i själva verket på aritmetikens fundamentalsats som vi inte kommer att behandla i kursen.

Den tredje bevistekniken som finns kallas för *induktionsbevis* men den kommer vi inte behöva i den här kursen.

1.3 Olika satser och hur man bevisar dem

I föregående avsnitt diskuterade vi olika bevistekniker. Men vilka tekniker är lämpliga för vilka typer av satser?

- **Implikation:** Man säger att P implicerar Q om Q är sant när P är det. Ett exempel är Sats 1.1.4, som säger att om ett tal n är jämnt, så är talet $n + 1$ udda. Man brukar beteckna implikationer med en tjock pil $\implies$, så att P medför Q skrivs

$$P \implies Q.$$

En implikation kan bevisas med ett direkt bevis. Då antar man att P är sant, och sedan visar man att Q också måste vara sant (det är så vi bevisar Sats 1.1.4). Man kan också använda ett motsägelsebevis. Då antar man att P är sann och att Q är falsk, och bevisar en motsägelse.

Ett tredje sätt att bevisa att P implicerar Q är att bevisa att om Q är falsk, så är P falsk. Detta kallas för *omvändningen* av en implikation.

- **Ekvivalens:** En ekvivalens är när två påståenden P och Q implicerar varandra, alltså att om P så Q , och om Q så P . Man brukar använda frasen P om och endast om Q . Man använder tjocka dubbelpilar för att beteckna ekvivalenser, så att P om och endast om Q skrivs som

$$P \iff Q.$$

Ekvivalenser bevisas genom att första visa att P implicerar Q , och sedan att Q implicerar P .

- **Universalsats:** En universalsats säger att alla n i en mängd M uppfyller något villkor P . Universalsatser kan bevisas som implikationer, genom att omformulera universalsatsen som att om n ligger i mängden M , så uppfyller n villkoret P , det vill säga

$$n \in M \implies n \text{ uppfyller } P$$

Man kan även bevisa en universalsats genom ett motsägelsebevis. Då antar man att det finns ett n i M som inte uppfyller P , och bevisar att det är omöjligt.

- **Existenssats:** En existenssats säger att finns ett objekt n som har egenskapen P . Den typiska existenssatsen är ekvationslösning. Att $x^2 = 3$ har en lösning är en existenssats, och kan omformuleras som att det finns ett tal x så att $x^2 = 3$.

Ett sätt att bevisa en existenssats är att konstruera det sökta objektet utifrån objekt man redan vet finns. Till exempel så kan man bevisa att det finns ett udda kvadrattal genom att notera att $3^2 = 9$ är udda och ett kvadrattal.

Man kan också använda ett motsägelsebevis. Då antar man att det inte existerar någon objekt med egenskapen P och visar att det leder till en motsägelse. Dessa bevis har fördelen att vi inte behöver beskriva hur objektet konstrueras. I gengäld kan bevisen vara mycket komplicerade.

En variant på universalsatsen är att inget n i M uppfyller P . Den kan omformuleras som att alla n i M saknar egenskapen P . För dessa typer av satser är motsägelsebevis ofta smidiga: man antar att det finns ett n i M som uppfyller P och härleder en motsägelse.

Universal- och existenssatser är duala till varandra, i bemärkelsen att om du ska bevisa en existenssats med hjälp av ett motsägelsebevis så antar du en universalsats, och vice versa, se bevisen av 1.2.2 och 1.2.3.

1.4 Mängder

En *mängd* är en samling objekt. Man kan samla nästan vad man vill i en mängd: tal, katter, och andra mängder.² Det viktiga är att man alltid kan avgöra ifall ett objekt tillhör mängden eller inte. De objekt som ligger i mängden kallas för *element*.

Det lättaste sättet att beskriva en mängd är att räkna upp elementen som ingår i den. För att markera att objekten ligger i en mängd, så omger man listan med *mängdklamrar* $\{$ och $\}$. Mängden som innehåller 1, 2 och 3 skrivs alltså som

$$\{1, 2, 3\}.$$

²Vi skriver *nästan* av en anledning. Det finns samlingar av objekt som kan beskrivas men som inte utgör en mängd. Detta kallas *Russells paradox*, efter Bertrand Russell (1872–1970). Russells exempel är samlingen av alla mängder som inte innehåller sig själv.

Två mängder A och B är lika om de innehåller samma element, vilket skrivs $A = B$. Det spelar ingen roll i vilken ordning man skriver elementen eller hur många gånger de listas. Därför gäller att

$$\{1, 1, 2, 3\} = \{1, 2, 3\} = \{2, 3, 1\}.$$

Om ett element x tillhör en mängd A brukar man skriva $x \in A$, vilket uttalas som x *tillhör* A . Om x inte tillhör A skriver man $x \notin A$. Antalet element som tillhör en mängd A brukar betecknas med $|A|$.

Mängden på formen $\{\}$ innehåller inte några element alls och kallas den *tomma mängden*. Den brukar betecknas med $\emptyset$ och är unik i aspekten att den saknar element, vi skriver alltså $|\emptyset| = 0$.

En mängd kan innehålla andra mängder som element. Mängden

$$A = \{\{1, 2\}, 3\}$$

har två element vilket skrivs som $|A| = 2$. Dess element är: mängden $\{1, 2\}$ och talet 3. Mängden $\{1, 2\}$ innehåller i sin tur elementen 1 och 2. Däremot innehåller A varken 1 eller 2, det vill säga

$$\{1, 2\} \in A \quad \text{men } 1 \notin A.$$

Att mängder kan innehålla andra mängder kan ha paradoxala konsekvenser. Till exempel kan vi lägga den tomma mängden i en mängd, och bilda mängden av den tomma mängden.

$$A = \{\emptyset\} = \{\{\}\}$$

Mängden A innehåller ett element, den tomma mängden, och är därför inte tom. Mängden av den tomma mängden är alltså inte lika med den tomma mängden.

Detta verkar motsägelsefullt. Den tomma mängden är ju tom, så mängden av den tomma mängden borde ju också vara tom? Tricket är att skilja på mängden och elementen i mängden. Den tomma mängden är ju ett element i sig, även om den inte innehåller några element, precis som att 0 är ett tal, trots att representerar ett antal som inte finns. Man kan tänka sig att en påse som innehåller en annan tom påse, inte är tom.

Det finns ingen begränsning på hur stor en mängd kan vara, och de flesta mängder man studerar innehåller oändligt många element. Dessa mängder kan naturligtvis inte skrivas ut som en lista. Istället beskriver man dem med *mängdbyggaren*, som har följande allmänna form $\{x \mid \text{villkor på } x\}$. Den här mängden består av alla element som uppfyller villkoret. Ett exempel är mängden

$$\{n \mid n \text{ är jämnt}\} = \{\dots, -4, -2, 0, 2, 4, \dots\}$$

som innehåller alla jämna tal.

En mängd B är en *delmängd* av en mängd A om alla element som tillhör B även tillhör A . Man skriver detta som $B \subset A$. Till exempel så är $\{1, 2\}$ en delmängd av $\{1, 2, 3\}$, eftersom 1 och 2 är element i båda mängderna. Om två mängder är delmängder av varandra så är de lika.

En icke-tom mängd har alltid minst två delmängder: sig själv och den tomma mängden. En delmängd B av A är *äkta* om $B \neq A$.

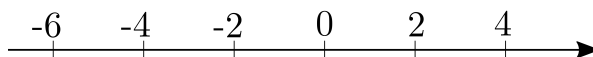
Det är lätt att blanda ihop element och delmängder. Det beror på att mängder kan innehålla andra mängder, så att en delmängd av en mängd kan vara ett element i mängden. Mängden $A = \{\emptyset\}$ är ett bra exempel. Den tomma mängden är både ett element i och en delmängd av A .

I mängden $A = \{1, 2, \{1, 2\}\}$ är $\{1, 2\}$ både en delmängd och ett element. Däremot så är $\{1\}$ enbart en delmängd av A , medan 1 enbart är ett element.

De olika talsystemen kan ses som mängder av tal, och har fått egna beteckningar. De *naturliga talen* betecknas med $\mathbb{N}$ och består av talen 0, 1, 2, 3, och så vidare.³

Naturliga tal kan adderas och multipliceras utan problem. Resultatet är alltid ett nytt naturligt tal. För att subtrahera behöver vi införa de negativa talen -1 , -2 , och så vidare. De naturliga talen tillsammans med de negativa talen kallas för *heltalen*, och betecknas med $\mathbb{Z}$ (av tyskans *Zahl* = tal).

Heltal kan adderas, subtraheras och multipliceras. Man kan däremot inte dividera dem med varandra. För detta krävs *rationella tal*. De definieras som alla kvoter a/b , där a och b är heltal och b är skilt från 0. Mängden av alla rationella tal betecknas med $\mathbb{Q}$.

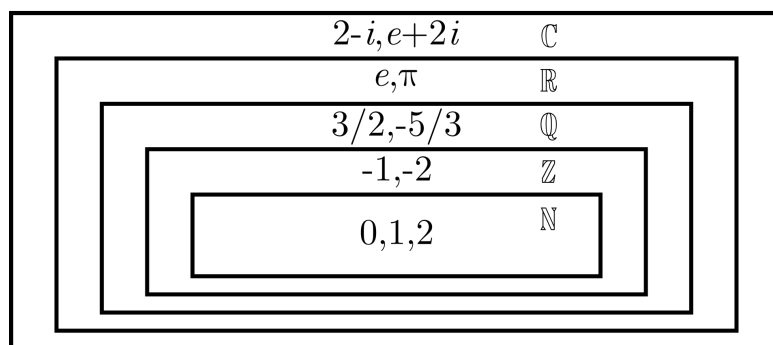


Figur 1.1: Tallinjen runt 0.

De rationella talen ligger på den så kallade *tallinjen*, som går från negativa tal till vänster och till positiva tal till höger (se Figur 1.1). Det finns dock tal som inte är rationella, men som ändå ligger på tallinjen. Ett exempel är $\sqrt{2}$, som är längden på diagonalen i en kvadrat med sidan 1. Läger man till dessa tal får de *reella talen*, som betecknas med $\mathbb{R}$.⁴ Reella tal som inte är rationella kallas för *irrationella*.

³Vissa exkluderar 0 från de naturliga talen. Att inkludera 0 har dock fördelar. Om man börjar räkna från 0 och går ett steg i taget kommer man ha gått n steg när man räknat till n . Exempel: om vi räknar till 3 från 0 så får vi $0 \rightarrow 1 \rightarrow 2 \rightarrow 3$, vilket är 3 steg. Om vi börjar från 1 får vi istället $1 \rightarrow 2 \rightarrow 3$, vilket är 2 steg.

⁴Reella tal är mycket mystiska. Den matematiska cirkeln 2016–2017, *Vad är ett tal?*, handlade om hur man kan definiera dem i termer av rationella tal. Den intresserade läsaren uppmanas att söka upp kompendiet på Cirkelns hemsida: <https://www.math-stockholm.se/samverkan/cirkel/>



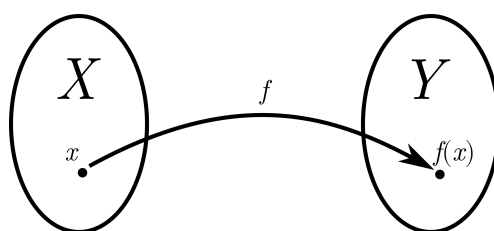
Figur 1.2: De olika talsystemen från $\mathbb{N}$ till $\mathbb{C}$.

De reella talen kan utvidgas ytterligare till de *komplexa talen*, dessa kommer vi beskriva i Kapitel 3.

1.5 Funktioner

En funktion f från en mängd X till en mängd Y beskriver ett sätt att para ihop element i en mängd X med element i en mängd Y . Det brukar skrivas som $f: X \rightarrow Y$. Mängden X kallas för *definitions mängd* och mängden Y kallas för *målmängd*. Man kan se f som en process som tar ett element i mängden X och avger ett element som ligger i mängden Y . När man tillämpar en funktion på ett element x i X så kallas x för funktionens *argument*. Mängden av alla värden en funktion i praktiken antar kallas för funktionens *värdemängd*, och denna betecknas ofta med V_f . Värdemängden är då alltså en delmängd av målmängden, och kan beskrivas som $V_f = \{f(x) \mid x \in X\}$. Till exempel är $\sin(x)$ en funktion från $\mathbb{R}$ till $\mathbb{R}$, så funktionens målmängd är alltså $\mathbb{R}$, men värdemängden är $[-1, 1]$.

Två funktioner är lika när de har samma definitions mängd, samma målmängd och de är lika på alla element i definitions mängden. Definitions- och målmängden är alltså en del av funktionen.



Figur 1.3: En funktion f från X till Y .

Funktioner beskrivs ofta med formler. Exempelvis så kan funktionen $f: \mathbb{N} \rightarrow \mathbb{N}$ som tar ett naturligt tal och returnerar dess kvadrat beskrivas som $f(n) = n^2$. Alla polynom kan ses som en funktion från $\mathbb{R}$ till $\mathbb{R}$, som beräknas genom att man sätter in talet x i uttrycket. En funktion måste dock inte ges av en formel. Det enda som krävs är att funktionen är definierad för alla element i definitions mängden, och att den alltid ger samma svar. Vi ger nu ett par

exempel på detta och hur man istället kan beskriva en funktion.

Exempel 1.5.1. Vårt första exempel är absolutbeloppet $|x|$ av ett reellt tal x , som definieras som avståndet på tallinjen från x till origo. Detta är en funktion vars definitionsmängd och målmängd är $\mathbb{R}$. Man kan beräkna den genom att man tar bort eventuella minustecken framför talet, det vill säga

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

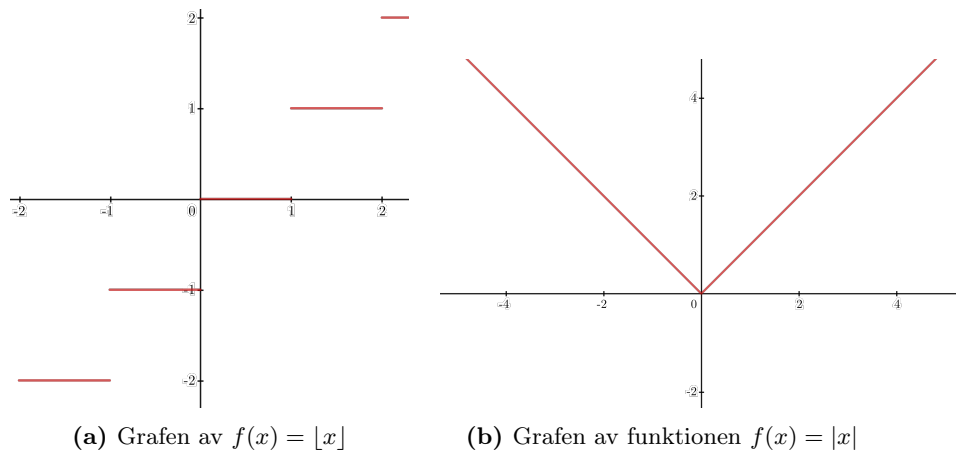
Exempelvis så gäller $|-3| = -(-3) = 3$ och $|2| = 2$.

Exempel 1.5.2. *Golvfunktionen* är funktionen som avbildar reella tal x på heltalet man får när man avrundar x nedåt. Vi benämner den funktionen med $\lfloor \dots \rfloor$. För att förtydliga hur denna och andra funktioner används brukar vi skriva på följande sätt

$$\begin{aligned} \lfloor \dots \rfloor : \mathbb{R} &\rightarrow \mathbb{Z} \\ x &\mapsto \lfloor x \rfloor. \end{aligned}$$

eftersom det kanske inte är uppenbart för läsaren att man ska skriva $\lfloor x \rfloor$ när man vill referera till funktionens värde vid punkten x istället för $\lfloor \dots \rfloor_x$ eller $\lfloor \dots \rfloor(x)$. Exempelvis så gäller $\lfloor \pi \rfloor = 3$.

En funktion $f: \mathbb{R} \rightarrow \mathbb{R}$ kan beskrivas genom sin *graf*, som definieras som mängden av punkter i planet på formen $(x, f(x))$.



Figur 1.4: Golv- och absolutbeloppsfunktionen.

Om man istället betraktar en funktion med ändlig definitionsmängd kan man beskriva den med en tabell.

Exempel 1.5.3. Betrakta funktionen från mängden $\{1, 2, 3\}$ till mängden $\{2, 3, 4, 5\}$ där $f(1) = 3$ och $f(2) = 4$ och $f(3) = 3$. Vi kan beskriva denna funktion med en tabell eftersom vi bara har ändligt många element i vår

definitions mängd,

x	f(x)
1	3
2	4
3	3

Vi kan också beskriva den med en formel eftersom definitions mängden och målmängden är delmängder till de reella talen $\mathbb{R}$ så till exempel gäller

$$f(x) = 4 - (x - 2)^2.$$

Exempel 1.5.4. Betrakta funktionen från mängden {hund, katt} till mängden $\{0, 1\}$ där $f(\text{hund}) = 0$ och $f(\text{katt}) = 1$. Vi kan beskriva denna funktion med en tabell eftersom vi bara har ändligt många element i vår definitions mängd,

x	f(x)
hund	0
katt	1

Det finns dock inte någon vettig formel som beskriver f eftersom det inte finns något välkänt sätt att 'addera' eller 'multiplicera' orden hund och katt på ett sätt som ger ett tal.

Definition 1.5.5. Givet två funktioner sådana att den enas definitions mängd är den andras målmängd, $f: Y \rightarrow Z$ och $g: X \rightarrow Y$ kan vi definiera deras *sammansättning* $f \circ g: X \rightarrow Z$ enligt regeln

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

Definition 1.5.6. Givet en funktion $f: X \rightarrow Y$ och en delmängd $A \subset X$ så kallar vi mängden $\{f(x) \mid x \in A\}$ för *bilden av A* och den betecknas $f(A)$. Om x är ett element i X brukar vi även kalla $f(x)$ för *bilden av x*.

Definition 1.5.7. Givet en funktion $f: X \rightarrow Y$ och en delmängd $B \subset Y$ så kallar vi mängden $\{x \mid f(x) \in B\}$ för *urbilden av B*. Om y är ett element i Y brukar vi kalla urbilden $\{x \mid f(x) = y\}$ för *fibern av y*. Givet ett element x i fibern av y kommer vi i den här kursen beteckna fibern av y med $[x]$.

Definition 1.5.8. Vi säger att en funktion $f: X \rightarrow Y$ är *surjektiv* om alla element i Y är bilden av något element i X , det vill säga om värdemängden överensstämmer med målmängden. En funktion sägs vara *injektiv* om varje element i värdemängden är bilden av exakt ett element i definitions mängden. Om en funktion är både surjektiv och injektiv säger vi att funktionen är *bijektiv*.

Notera att en bijektiv funktion är *inverterbar*: Det vill säga, det existerar en funktion som kallas *den inversa funktionen* eller f^{-1} med beskrivningen $f^{-1}: Y \rightarrow X$ och $f^{-1}(f(x)) = x$ för alla $x \in X$ och $f(f^{-1}(y)) = y$ för alla $y \in Y$. Den inversa funktionen, om den finns, är alltså den funktion som för varje element $y \in Y$ ger det unika elementet $x \in X$ som har egenskapen att $f(x) = y$. I övning 1.14 ser vi exempelvis att en strikt växande funktion $f: \mathbb{R} \rightarrow \mathbb{R}$ är injektiv. Det är även värt att notera att en injektiv funktion blir bijektiv om vi byter ut målmängden mot värdemängden.

Huruvida en funktion är inverterbar eller ej beror inte bara på regeln som beskriver funktionen, utan på definitions- och målmängden.

Exempel 1.5.9. Betrakta funktionen $f(x) = x^2$ som tar ett tal och kvadrerar det.

Betraktat som funktion från $\mathbb{R}$ till $\mathbb{R}$ är den inte inverterbar. Den är inte surjektiv eftersom det till exempel inte finns något $x \in \mathbb{R}$ så att $x^2 = -1$. Problemet här är alltså att funktionens målmängd inte överensstämmer med dess värdemängd, det vill säga att funktionen inte är surjektiv.

Detta problem kan lösas genom att begränsa målmängden till värdemängden, men även om vi betraktar f som en funktion från $\mathbb{R}$ till dess värdemängd, det vill säga mängden av icke-negativa tal $[0, \infty)$, så är funktionen inte inverterbar. Vi har fortfarande problemet att funktionen inte är injektiv. Till exempel så har vi $(-1)^2 = 1^2 = 1$ och eftersom det finns två element som avbildas på 1, kan vi inte entydigt definiera en invers funktion.

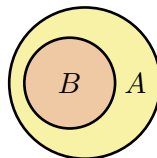
Om vi däremot betraktar f som en funktion från $[0, \infty)$ till $[0, \infty)$ så är den både injektiv och surjektiv, det vill säga bijektiv, och således har funktionen en väldefinierad invers, nämligen $f^{-1}(x) = \sqrt{x}$. Till exempel så är i detta fall $f^{-1}(1) = 1$ eftersom vi bara får välja den positiva roten.

Vi har en liknande situation med $\sin(x)$ och $\cos(x)$. Betraktade som funktioner från $\mathbb{R}$ till $\mathbb{R}$ är dessa funktioner varken injektiva eller surjektiva, men genom att betrakta $\sin(x)$ som en funktion från $[-\pi/2, \pi/2]$ till $[-1, 1]$ och $\cos(x)$ som en funktion från $[0, \pi]$ till $[-1, 1]$ erhåller vi bijektiva, och således inverterbara funktioner.

1.6 Mängdoperationer

I detta avsnitt ska vi beskriva ett antal olika sätt att från en eller flera mängder skapa en ny mängd. Vi kallar dessa för *mängdoperationer*. I nästa kapitel kommer vi även att se hur vissa av dessa mängdoperationer går att generalisera till andra situationer.

För att illustrera mängder använder man ibland *Vennndiagram*, efter matematikern John Venn (1834–1923). Där representeras mängder som enkla former, oftast cirklar, och formernas förhållanden till varandra motsvarar mängdernas. Till exempel kan man illustrera att B är en delmängd av A genom att rita dem som två cirklar, där B ligger inuti A .



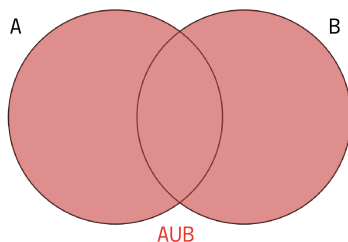
Figur 1.5: Vennndiagram för $B \subset A$.

Unionen

Unionen av mängderna A och B är mängden som består av alla element som ligger i A *eller* i B . Den betecknas med $A \cup B$ och definieras som

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

Ett exempel är $\{1, 2, 3\} \cup \{2, 3, 4\} = \{1, 2, 3, 4\}$.



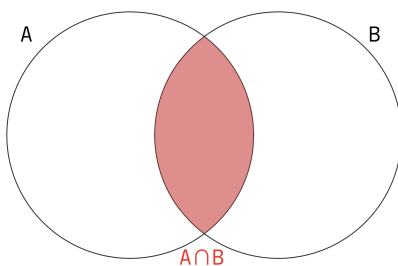
Figur 1.6: Venndiagram för $A \cup B$.

Snittet

Snittet av mängderna A och B är mängden som består av alla element som ligger i A *och* i B . Den betecknas med $A \cap B$ och definieras som

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

Ett exempel är $\{1, 2, 3\} \cap \{2, 3, 4\} = \{2, 3\}$.



Figur 1.7: Venndiagram för $A \cap B$.

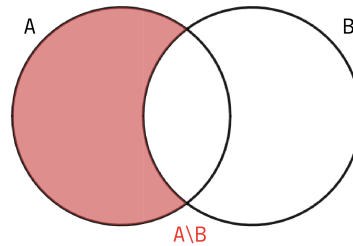
Två mängder A och B sägs vara *disjunkta* om de inte har några gemensamma element, det vill säga om $A \cap B = \emptyset$.

Differensen och komplementet

Differensen av en mängd A och B är mängden som består av alla element som ligger i A men inte i B . Den betecknas med $A \setminus B$ och definieras som

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

Ett exempel är $\{1, 2, 3\} \setminus \{2, 3, 4\} = \{1\}$.



Figur 1.8: Venndiagram för $A \setminus B$.

Notera att $A \setminus B$ inte är lika med $B \setminus A$, exempelvis gäller

$$\{2, 3, 4\} \setminus \{1, 2, 3\} = \{4\} \neq \{1\} = \{1, 2, 3\} \setminus \{2, 3, 4\}.$$

Om alla uppträdande mängder är delmängder av en viss mer eller mindre underförstådd grundmängd M talar man ofta om $M \setminus A$ som *komplementet* till A (med avseende på M). Vi kommer att beteckna komplementet till A med A^C . Om vi till exempel pratar om mängder av heltal, och vi till exempel betraktar mängden $A = \{1, 2, 3\}$, så avser komplementet till A mängden av alla heltal *förutom* 1, 2 och 3.

Den kartesiska produkten

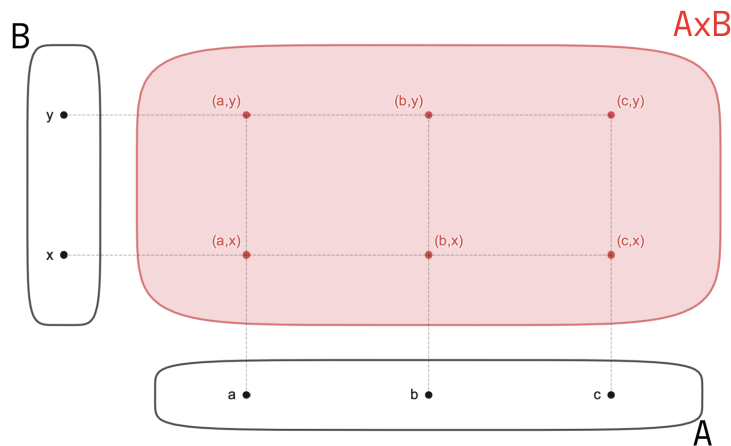
Den kartesiska produkten av två mängder A och B är mängden som består av alla par av element så att det första ligger i A och det andra i B . Den betecknas med $A \times B$ och definieras som

$$A \times B = \{(x, y) \mid x \in A \text{ och } y \in B\}.$$

Två olika exempel är

$$\begin{aligned} \{a, b, c\} \times \{x, y\} &= \{(a, x), (a, y), (b, x), (b, y), (c, x), (c, y)\} \\ \{1, 2, 3\} \times \{2, 3, 4\} &= \{(1, 2), (1, 3), (1, 4), (2, 2), (2, 3), (2, 4), (3, 2), (3, 3), (3, 4)\}. \end{aligned}$$

Notera att ordningen är viktig, exempelvis gäller $(1, 2) \in \{1, 2, 3\} \times \{2, 3, 4\}$ men $(2, 1) \notin \{1, 2, 3\} \times \{2, 3, 4\}$.



Figur 1.9: Venn diagram för $A \times B$.

När vi tar den kartesiska produkten av en mängd med sig själv brukar vi använda notationen $A^2 = A \times A$. Mer allmänt skriver vi

$$A^n = \underbrace{A \times \dots \times A}_{n \text{ gånger}} = \{(a_1, \dots, a_n) \mid \text{alla } a_i \in A\}.$$

Till exempel är det reella talplanet $\mathbb{R} \times \mathbb{R}$ vilket vi brukar beteckna med $\mathbb{R}^2$.

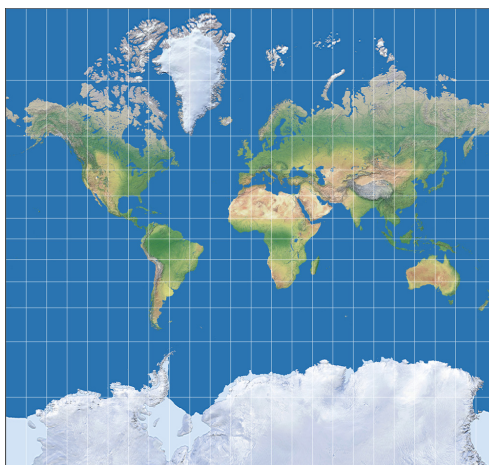
1.7 Kartor

En karta över jorden kan också ses som en funktion. Om vi tänker oss att kartan själv är en delmängd $R \subset \mathbb{R}^2$, och jorden är en delmängd $J \subset \mathbb{R}^3$ kan vi beskriva kartan som en funktion $f : R \rightarrow J$, där vi parar i hop en punkt i R med den punkten på jorden som kartan syftar på, se Figur 1.10. Om f är surjektiv betyder det att kartan kartlägger hela jorden. Om f är injektiv betyder det att en punkt på jorden inte ska synas på två ställen på kartan.

Övningar

Övning 1.1. Lista elementen i följande mängder.

- (i) $A = \{n \in \mathbb{N} \mid n < 5\}$.
- (ii) $B = \{1, 2, \{2, 3\}\}$.
- (iii) $C = \{k \in \mathbb{Z} \mid k^2 < 16\}$.
- (iv) $A \cap B$.
- (v) $(C \setminus A) \cup B$.



Figur 1.10: Karta över världen med Mercators projektion. Som en funktion är den injektiv men inte surjektiv, då hela nordpolen och sydpolen inte finns med.

Bild av Tobias Jung [CC BY-SA 4.0]

(<http://creativecommons.org/licenses/by-sa/4.0>), via <http://map-projections.net>

Övning 1.2. Lista elementen i följande mängder.

- (i) $A = \{x \in \mathbb{Q} \mid x^2 = 2\}$.
- (ii) $B = \{0, 1, 2, 3\}$.
- (iii) $C = \{p/q \mid p, q \in \mathbb{N}, 0 \leq p < 3 \text{ och } 1 \leq q < 3\}$.
- (iv) $(B \cup A) \cap C$.
- (v) $(C \cap B) \setminus A$.

Övning 1.3. För nedanstående par av mängder A och B , avgör om A och B är lika, disjunkta, någon av dem är en äkta delmängd av den andra eller ingetdera.

- (i) $A = \{1, 2, 3\}$ och $B = \{1, 1, 2\}$.
- (ii) $A = \{0, 1, 2\}$ och $B = \{n \in \mathbb{N} \mid n^2 < 9\}$.
- (iii) $A = \{\{\}\}$ och $B = \{x \in \mathbb{N} \mid 2x = -2\}$.
- (iv) $A = \{x \in \mathbb{R} \mid |x| < 1\}$ och $B = \{x \in \mathbb{R} \mid |x - 1| < 1\}$.
- (v) $A = \{x \in \mathbb{Q} \mid x^2 = 2\}$ och $B = \{x \in \mathbb{R} \mid x^2 = 2\}$.

Övning 1.4. För nedanstående par av mängder A och B , avgör om A och B är lika, disjunkta eller någon av dem är en äkta delmängd av den andra.

- (i) $A = \{-2, 0, 2\}$ och $B = \{x \in \mathbb{Z} \mid |x| < 3 \text{ och } x \text{ är jämnt}\}$.
- (ii) $A = \{x \in \mathbb{R} \mid x^2 < 2\}$ och $B = \{x \in \mathbb{Q} \mid x^2 \leq 2\}$.
- (iii) $A = \{x \in \mathbb{Z} \mid x \text{ är jämnt}\}$ och $B = \{x \in \mathbb{Z} \mid x \text{ är kvadrattal}\}$.

(iv) $A = \{x \in \mathbb{Z} \mid 2x = -2\}$ och $B = \{x \in \mathbb{N} \mid 2x = 2\}$.

(v) $A = \{\emptyset, \{\emptyset\}\}$ och $B = \{\emptyset\}$.

Övning 1.5. Använd mängdbyggaren för att definiera följande mängder.

(i) Mängden av jämna, positiva heltal.

(ii) Mängden av rationella tal r så att $2r$ är ett heltal.

(iii) Mängden av irrationella tal som ligger inom avstånd 1 från origo.

Övning 1.6. Använd mängdbyggaren för att definiera följande mängder.

(i) Mängden av alla kvadrattal som är större än 2.

(ii) Mängden av rationella lösningar till $x^4 + x^2 - 1 = 0$.

(iii) Mängden av rationella tal som är volymen av en kub med rationella sidor.

Övning 1.7. Ange möjlig definitionsmängd och målmängd för följande funktioner.

(i) Funktionen som ger det n :te kvadrattalet.

(ii) Funktionen som beräknar arean av triangel.

(iii) Funktionen beräknar derivatan av ett andragradspolynom.

Övning 1.8. Ange möjlig definitionsmängd och målmängd för följande funktioner.

(i) Funktionen som ger arean av cirkel med radie r .

(ii) Funktionen som ger avståndet mellan 1 och ett tal r på tallinjen.

(iii) Funktionen som ger de rationella nollställena till ett förstgradspolynom med rationella koefficienter.

Övning 1.9. Är följande funktioner eller inte? Om inte, motivera varför.

(i) $f : \mathbb{R} \rightarrow \mathbb{R}$ där

$$f(x) = \begin{cases} 1 & \text{om } x \in \mathbb{Q} \\ 0 & \text{om } x \notin \mathbb{Q}. \end{cases}$$

(ii) $f : \mathbb{N} \rightarrow \mathbb{Q}$ där $f(n) = \sqrt{n}$.

(iii) $f : \mathbb{R} \rightarrow \mathbb{R}$ så att $f(x) = 0$ med sannolikhet $1/2$ och $f(x) = 1$ med sannolikhet $1/2$.

(iv) $f : \{0\} \rightarrow \mathbb{R}$ där $f(0) = 1$ om ordet Balkong börjar på B.

Övning 1.10. Är följande funktioner eller inte? Om inte, motivera varför.

- (i) $f : \mathbb{Z} \rightarrow \mathbb{N}$, där $f(n)$ är siffersumman i det vanliga (decimala) talsystemet. Obs, alla *siffror* är icke-negativa.
- (ii) $f : \mathbb{R} \rightarrow \mathbb{Q}$, där $f(x) = x/2$.
- (iii) $f : \mathbb{N} \rightarrow \mathbb{R}$, där $f(n) = \sqrt[n+1]{n+1}$.
- (iv) $f : \mathbb{Q} \rightarrow \mathbb{Z}$, där $f(p/q) = p$.

Övning 1.11. Avgör om följande funktioner är lika eller inte? Motivera varför.

- (i) $f : \mathbb{R} \rightarrow \mathbb{R}$, $f(x) = |x|$ och $g : \mathbb{R} \rightarrow \mathbb{R}$, $g(x) = \sqrt{x^2}$.
- (ii) $f : \mathbb{Z} \rightarrow \mathbb{Q}$, $f(n) = 1/n$ och $g : \mathbb{N} \rightarrow \mathbb{Q}$, $g(m) = 1/m$.
- (iii) $f : \mathbb{N} \rightarrow \mathbb{Q}$, $f(n) = n/(n+1)$ och $g : \mathbb{N} \rightarrow \mathbb{R}$, $g(z) = z/(z+1)$.

Övning 1.12. Avgör om följande funktioner är lika eller inte? Om inte, motivera varför.

- (i) $f : \mathbb{R} \rightarrow \mathbb{R}$, $f(x) = x^2 + 2x + 1$ och $g : \mathbb{R} \rightarrow \mathbb{R}$, $g(x) = (x+1)^2$.
- (ii) $f : \mathbb{Z} \rightarrow \mathbb{Z}$, $f(x) = x^2$ och $g : \mathbb{Z} \rightarrow \mathbb{Z}$, $g(x) = |x|^2$.

Övning 1.13. Vad är värdemängden i exemplen 1.5.3 och 1.5.4? Avgör i vilka av de två exemplen som f är injektiv och i vilka som f är surjektiv.

Övning 1.14. Bevisa att en strängt växande funktion, alltså en funktion med egenskapen

$$x_1 < x_2 \implies f(x_1) < f(x_2)$$

är injektiv.

Övning 1.15. Visa att A och B är disjunkta om och endast om $A \setminus B = A$.

Övning 1.16. Visa att för två mängder A, B så uppfyller operationerna $\cup, \cap, A^C$ (där A^C betecknar mängdkomplement) följande räkneregler,

$$(A \cup B)^C = A^C \cap B^C \quad (A \cap B)^C = A^C \cup B^C.$$

Övning 1.17. Visa att följande räkneregler håller

$$A \setminus B = A \setminus (A \cap B).$$

Övning 1.18. Visa att sammansättningen av två bijektioner är en bijektion.

Övning 1.19. Visa att antalet element i $A \times B$ är produkten av antalet element i A och antalet element i B .

Övning 1.20. Visa att den kartesiska produkten av två delmängder är delmängd till den kartesiska produkten av grundmängderna. Alltså om $B \subset A$ och $Y \subset X$ gäller det att

$$B \times Y \subset A \times X.$$

Övning 1.21 (*). Använd ett direkt bevis för att bevisa att om n är ett jämnt tal så är $n+1$ ett udda tal (detta är Sats 1.1.4).

Övning 1.22 (★). Använd ett direkt bevis för att bevisa att om n är udda så är n^2 udda.

Övning 1.23 (★). Använd ett motsägelsebevis för att bevisa att summan av ett irrationellt tal och ett rationellt tal är irrationellt.

Övning 1.24 (★). Använd ett motsägelsebevis för att bevisa att om $a + b \geq c$ så är antingen $a \geq c/2$ eller $b \geq c/2$

Övning 1.25 (★★). Bevisa att om $ab = c$ så är $a \geq \sqrt{c}$ och $b \leq \sqrt{c}$, eller tvärtom.

Övning 1.26 (★). Bevisa att det finns två irrationella tal a och b så att a^b är rationellt. Tips: använd att $\sqrt{2}$ är irrationellt.

Övning 1.27 (★★). Bevisa att alla rationella tal kan skrivas som en produkt av två irrationella tal.

Övning 1.28 (★). Visa följande påståenden för heltalen (du bör använda det du bevisat i ena för att bevisa nästa).

- (i) Multiplikation med ett positivt tal bevarar olikheter. Alltså för $m > 0$ gäller

$$r \leq n \implies mr \leq mn.$$

- (ii) Ett positivt heltal kan inte dela ett mindre positivt heltal. Alltså

$$0 < r < n \implies n \text{ delar inte } r.$$

- (iii) Differensen av två heltal delbara med n är också delbar med n .

2 Euklidisk geometri

I detta kapitel kommer vi att snabbt ge en introduktion till euklidisk geometri. Euklidisk geometri är helt enkelt vanlig geometri i planet — namnet kommer från den grekiske matematikern Euklides (aktiv ca 300 f.Kr.) och används för att skilja det från den icke-euklidiska geometrin som först introducerades under 1800-talet.

Målet i detta kapitel kommer dock inte att vara att läsaren skall lära sig mycket euklidisk geometri, eftersom det inte är detta kursen handlar om. Det vi kommer att göra är att presentera de grundläggande definitionerna och ramverket för euklidisk geometri. Utöver definitioner av saker som kan förväntas vara kända sedan tidigare, diskuterar vi också begreppet *automorfi* av planet, och beskriver hur mängden av automorfier till det euklidiska planet ser ut. Målet i resten av kursen kommer att vara att ge en beskrivning av den hyperboliska geometrin liknande denna.

Vi kommer också att beskriva Euklides ursprungliga axiomatiska beskrivning av geometri, och ge en kort historik över det så kallade parallellpostulatet. Den fascinerande historien kring detta postulat var den stora drivkraften bakom varför den icke-euklidiska geometrin från början över huvud taget infördes.

2.1 Definitioner

Euklides arbetade systematiskt med ett litet antal begrepp: han använde sig av punkter, linjer, avstånd och vinklar. Det Euklides gjorde var att ställa upp ett antal *axiom* som dessa begrepp skulle uppfylla (till exempel: genom två punkter passerar exakt en linje); det vi kommer göra är att konstruera en *modell* för euklidisk geometri. Med detta menar vi att Euklides aldrig faktiskt definierade vad en punkt eller en linje är för något, bara vilka egenskaper dessa skall ha. Vi kommer i stället att direkt definiera vad en punkt, en linje och ett avstånd är för något, på ett sådant sätt att Euklides axiom alla är uppfyllda.

Låt oss definiera alla dessa begrepp ett i taget.

Definition 2.1.1. Mängden $\mathbb{R}^2$ består av alla ordnade par (x, y) av tal, där x och y tillhör de reella talen $\mathbb{R}$. När vi talar om det euklidiska planet, eller helt enkelt planet, så menar vi denna mängd.

Anmärkning 2.1.2. Vi visualiserar $\mathbb{R}^2$ som ett xy -plan, där första talet anger koordinaten i x -led och andra talet koordinaten i y -led.

Definition 2.1.3. En *punkt* p i planet är ett element i mängden $\mathbb{R}^2$.

Definition 2.1.4. Vi säger att λ är en *linje* i planet om den är en delmängd som har följande form:

$$\lambda = \{(x, y) \in \mathbb{R}^2 \mid ax + by + c = 0\}$$

för några a, b och c i $\mathbb{R}$, där inte både a och b är lika med noll. Om en punkt p ingår i mängden λ säger vi att punkten *ligger* på linjen. Ett *linjesegment* är

en delmängd av en linje λ som kan skrivas som

$$\{(x, y) \in \lambda \mid d \leq x \leq e\} \text{ eller } \{(x, y) \in \lambda \mid d \leq y \leq e\}$$

där $d < e$.

Definition 2.1.5. Vi definierar *avståndet* mellan punkterna $p = (x, y)$ och $q = (u, v)$ som

$$\text{dist}(p, q) = \sqrt{(x - u)^2 + (y - v)^2}.$$

Utöver dessa använde sig Euklides t.ex. av *arean* av en geometrisk figur. Vi kommer dock inte att behandla area — det är lite för svårt att ge en bra rigorös definition av area, både i det klassiska⁵ och i det icke-euklidiska fallet.

Med hjälp av dessa begrepp kan vi definiera flera olika geometriska figurer. Vi kan tala om en triangel som det begränsade område som fås mellan tre icke-parallella linjer, och en cirkel som mängden av punkter på samma avstånd från en given mittpunkt.

Om vi använder oss av begreppet *båglängd*, som kommer att definieras i nästa kapitel, kan vi tala om längden på ett kurvstycke. Trots att vi ännu inte definierat längden av ett kurvstycke kommer vi nu att använda detta begrepp för att definiera vad en vinkel är för något. Vi hoppas att läsaren har overseende med denna logiska oegentlighet.

Definition 2.1.6. Låt två räta linjesegment λ och μ gå ut från en punkt p , det vill säga, p är ändpunkt till bägge segmenten. Rita cirkeln med mittpunkt p och radie 1. Om segmenten λ och μ förlängs bort från p så att båda har längd minst 1, så kommer båda dessa att skära cirkeln i varsin punkt l respektive m . Vi definierar *vinkeln från λ till μ* som längden av den del av cirkelns omkrets som går moturs från l till m .

Lägg speciellt märke till att definitionen visar att vi alltid kommer att räkna vinklar i *radianer*, inte grader. Vinkeln för ett varv är alltså 2π , vilket ju är omkretsen av en cirkel med radie 1, och inte 360° .

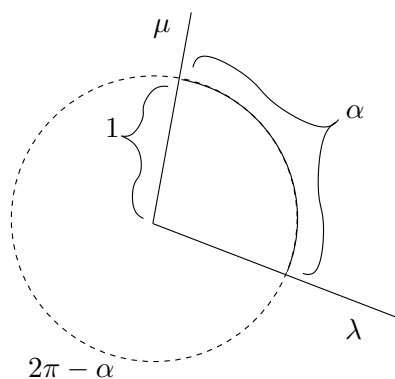
Definition 2.1.7. Vi säger att två vinklar α och β är lika om $\alpha - \beta$ är ett heltal gånger 2π .

Anmärkning 2.1.8. Definitionen 2.1.6 visar att i denna kurs kommer vi att räkna vinklar med tecken. Med andra ord kommer vi att räkna en lika stor vinkel som positiv eller negativ, beroende på om den är mätt medsols eller motsols. I Figur 2.1 så är vinkeln från λ till μ lika med α , medan vinkeln från μ till λ är $2\pi - \alpha$, vilket också är lika med $-\alpha$.

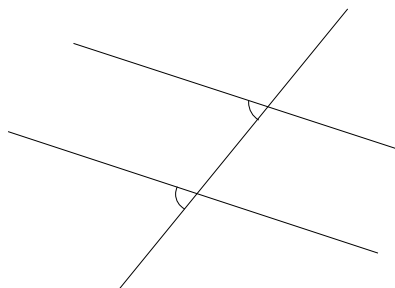
Satserna i euklidisk geometri är de påståenden om geometrin som kan formuleras genom att endast använda dessa nu införda begrepp. Exempel är:

- (i) **Sats:** Om två linjer inte skär varandra i någon punkt, och en tredje linje faller in mot dem, så kommer denna tredje linje att möta bägge två i samma vinkel;

⁵Försöker man lite själv, så märker man snabbt att det inte är helt lätt att ge en allmän definition av vad man ska mena med arean av en figur i planet. Ett mått på komplikationerna som uppstår när man ska definiera area, är att den idag allmänt accepterade definitionen av arean av en delmängd i planet definierades inte förrän i 1900-talets början av Henri Lebesgue!



Figur 2.1: En illustration av definitionen av vinkeln mellan två linjesegment.

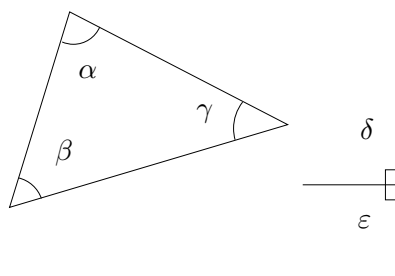


Figur 2.2: De två infallande vinklarna är lika stora.

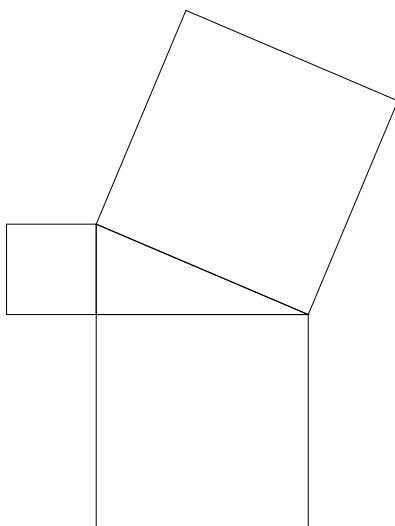
- (ii) **Sats:** I en triangel är summan av alla tre vinklar densamma som summan av två räta vinklar, se Figur 2.3;
- (iii) **Sats:** (Pythagoras) Om man ritar en kvadrat vid varje sida i en rätvinklig triangel, så är arean av kvadraten vid hypotenusan densamma som summan av areorna vid de bägge kateterna, se Figur 2.4.

2.2 Automorfier

Detta avsnitt i vår introduktion till euklidisk geometri kommer att beskriva ett begrepp som dyker upp nästan överallt i matematiken. Med en *automorfi* kommer vi att mena en funktion med egenskapen att den bevarar *all* geometrisk struktur. Mer precist så kräver vi att en automorfi ska vara en *bijektion* (se Definition 1.5.8) med avseende på både punkter och linjer, och att alla storheter



Figur 2.3: Här är $\alpha + \beta + \gamma = \delta + \varepsilon$.



Figur 2.4: Pythagoras sats.

man kan mäta upp (vinklar och linjer) skall vara oförändrade efter automorfin. Så f är en automorfi om:

- Funktionen f är en bijektion från mängden av punkter till sig själv, så $f : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ är en bijektion;
- Funktionen f respekterar *linjer*: så om L är en linje så kommer

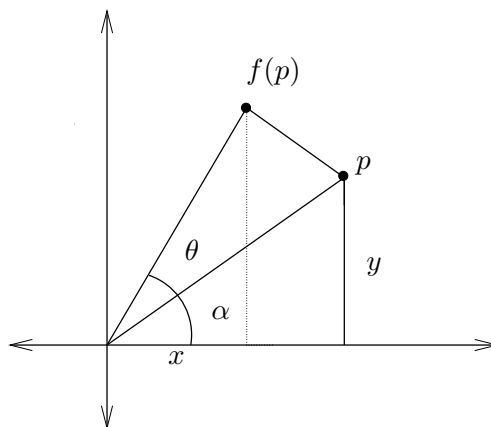
$$f(L) = \{f(x) \mid x \in L\}$$

också att vara en linje, och f ger på detta sätt även en bijektion från mängden av linjer i $\mathbb{R}^2$ till sig själv;

- Funktionen bevarar *avstånd*, i meningen att avståndet mellan x och y alltid är detsamma som avståndet mellan $f(x)$ och $f(y)$;
- Funktionen bevarar *vinklar*, så att vinkeln mellan två linjer som skär varandra är densamma som vinkeln mellan bilden av linjerna.

Helt enkelt betyder ordet ”automorfi” att funktionen f inte får förändra någon enda mätbar egenskap hos det den verkar på (i vårt fall det euklidiska planet) — planet självt ska vara helt oskiljbart från bilden av planet under funktionen f .

Anmärkning 2.2.1. Definitionen av en automorfi som vi har givit är ganska ineffektiv, i meningen att den skulle kunna formuleras mycket kompaktare. Till exempel så kräver vi först att f är en bijektiv funktion, och speciellt kräver vi då att om p och q är skilda från varandra, så kommer $f(p)$ och $f(q)$ vara skilda från varandra. Sedan kräver vi dessutom senare att f ska bevara avstånd mellan punkter. Men det andra villkoret implicerar det första! Om $p \neq q$ är deras avstånd nollskilt, och då måste avståndet mellan $f(p)$ och $f(q)$ vara nollskilt. Vi har använt ett onödigt krångligt sätt att formulera automorfvillkoret för att betona den intuitiva betydelsen av en automorfi, som en funktion som bevarar



Figur 2.5: En rotation.

alla egenskaper som överhuvudtaget kan formuleras utifrån de begrepp som vi har formulerat.

Exempel 2.2.2. En *translation*, eller parallellförflyttning, av planet är en funktion på formen

$$f(x, y) = (x + a, y + b),$$

där a och b är två fixerade reella konstanter. Vi visualiserar det som att funktionen flyttar alla punkter i planet en sträcka a i x -led, och en sträcka b i y -led. Varje translation är en automorfi.

Exempel 2.2.3. En rotation runt en given punkt är alltid en automorfi. Till exempel kan vi rotera planet med en vinkel θ moturs runt origo. Låt oss räkna ut hur denna automorfi kan skrivas i koordinater. Betrakta som i Figur 2.5 en punkt $p = (x, y)$ med avstånd $r = \sqrt{x^2 + y^2}$ till origo. Låt som i figuren α vara vinkeln mellan linjen från origo till p och x -axeln. Punkten $f(p)$ får enligt trigonometri koordinaterna $(r \cos(\alpha + \theta), r \sin(\alpha + \theta))$. Med additionsformler för cosinus och sinus fås att

$$f(p) = (r(\cos \alpha \cos \theta - \sin \alpha \sin \theta), r(\cos \alpha \sin \theta + \sin \alpha \cos \theta)).$$

Använder vi nu att $x = r \cos \alpha$, $y = r \sin \alpha$, fås att

$$f(p) = (x \cos \theta - y \sin \theta, x \sin \theta + y \cos \theta).$$

Vi tänker inte i detalj visa att alla dessa exempel faktiskt är automorfier. Låt oss dock t.ex. visa att varje rotation kring origo avbildar linjer bijektivt. Låt λ vara linjen $ax + by = c$, och f en moturs rotation med vinkel θ runt origo. Vi vill först visa att $f(\lambda)$ är en linje. Från definitionen kommer $(x, y) \in f(\lambda)$ om och endast om (x, y) ligger på linjen som fås av att rotera λ med en vinkel θ moturs, och att detta i sin tur sker om och endast om λ innehåller punkten som fås genom att rotera (x, y) med en vinkel θ medurs. Låt nu f^{-1} vara funktionen som roterar planet med en vinkel θ medurs runt origo. Så f^{-1} fås genom att sätta in vinkeln $-\theta$ i uttrycket för en moturs rotation. Då är

$$f^{-1}(x, y) = (x \cos \theta + y \sin \theta, -x \sin \theta + y \cos \theta),$$

eftersom $\cos x = \cos(-x)$ och $-\sin x = \sin(-x)$. Det vi har sagt i början av denna paragraf är att

$$f(\lambda) = \{f(x, y) \mid (x, y) \in \lambda\} = \{(x, y) \mid f^{-1}(x, y) \in \lambda\}.$$

Men $f^{-1}(x, y) \in \lambda$ gäller precis då

$$a(x \cos \theta + y \sin \theta) + b(-x \sin \theta + y \cos \theta) = c$$

eller

$$(a \cos \theta - b \sin \theta)x + (a \sin \theta + b \cos \theta)y = c.$$

Detta är nu igen ekvationen för en linje. En lustig observation är nu att det enda f gjort var att "rotera" koefficienterna a och b med en vinkel θ ! Om vi jämför med vårt tidigare uttryck för f , ser vi att när linjen λ ges av koefficienterna a, b och c , så kommer $f(\lambda)$ att ges av koefficienterna $f(a, b)$ och c . Speciellt gäller ju att $(a, b) \neq (0, 0)$ om och endast om $f(a, b) \neq (0, 0)$, vilket visar att λ är en linje om och endast om $f(\lambda)$ är det. Utifrån denna sista observation, att f verkar på linjer genom att "rotera" koefficienterna a och b , kan man också visa att f ger en bijektion från mängden av linjer till sig själv.

Vad som visar sig är att *alla* automorfier av det euklidiska planet kan konstrueras genom att sätta ihop dessa två exempel, alltså translationer och rotationer kring origo. Med det menar vi följande:

Sats 2.2.4. *Varje automorfi av det euklidiska planet kan skrivas som en rotation kring origo följt av en translation.*

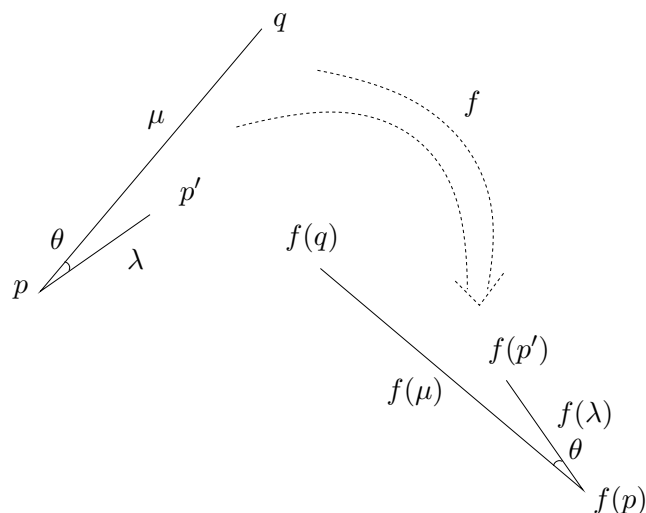
Vi börjar med ett lemma.

Hjälpsats 2.2.5. *Antag att två automorfier f och g har egenskapen att $f(p) = g(p)$, och $f(p') = g(p')$, där p och p' är två distinkta punkter i planet. Då är $f = g$.*

Bevis. Låt $q \in \mathbb{R}^2$ vara en godtycklig punkt i planet: vi skall visa att $f(q) = g(q)$. Om $q = p$ vet vi redan detta, så vi kan utan inskränkning anta att $q \neq p$. Börja med att dra det rätta linjesegmentet μ från q till p . Låt λ vara det rätta linjesegmentet från p till p' . Eftersom både f och g avbildar rätta linjer på rätta linjer, och både f och g avbildar p och p' till samma två punkter, så måste det gälla att $f(\lambda) = g(\lambda)$, det vill säga

$$\{f(x) \mid x \in \lambda\} = \{g(x) \mid x \in \lambda\}.$$

Låt θ beteckna vinkeln mellan λ och μ . Eftersom f och g är automorfier så kommer båda att avbilda μ på ett linjesegment som går ut från punkten $f(p)$, med en vinkel θ från linjen $f(\lambda) = g(\lambda)$, och som har samma längd som μ . Men nu ser vi att det bara finns ett sådant linjesegment, så det gäller även att $f(\mu) = g(\mu)$. Speciellt kommer både f och g att avbilda ändpunkten på μ till samma punkt, så $f(q) = g(q)$. Eftersom detta gäller för varje godtycklig punkt q är vi klara. $\square$



Figur 2.6: En illustration av beviset till 2.2.5.

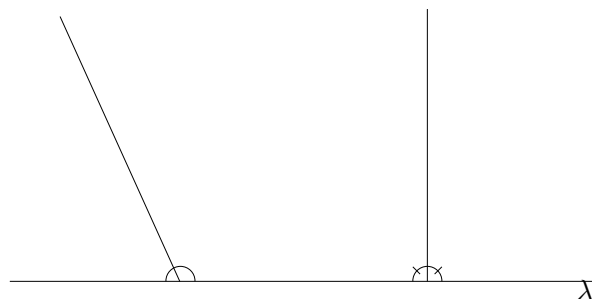
Bevis. (av 2.2.4) Låt f vara en automorfi. Vi ska visa att f kan skrivas som en sammansättning av en rotation kring origo och en translation.

Tag två punkter p och p' . Låt λ vara linjesegmentet från p till p' , och μ linjesegmentet från $f(p)$ till $f(p')$. Rita en markering på de båda linjesegmenten i punkterna p respektive $f(p)$. Om vi roterar planet med en vinkel θ , kommer även vinkeln mellan λ och varje fix linje att ändras med θ . Så det existerar en rotation g runt origo sådan att linjesegmentet $g(\lambda)$ (det vill säga, λ roterad med en vinkel θ) och μ är parallella, och så att den markerade punkten på bägge linjer pekar åt samma håll.

Låt sedan h vara translationen som avbildar $g(p)$ på $f(p)$. Vi hävdar nu att $f = h \circ g$. Enligt konstruktionen vet vi att $f(p) = h(g(p))$, och enligt 2.2.5 behöver vi nu bara visa att $f(p') = h(g(p'))$ för att slutföra beviset. Men enligt definitionen av g kommer linjen från $g(p)$ till $g(p')$ att ha samma längd och riktning som linjen från $f(p)$ till $f(p')$, och eftersom translationen h bevarar både längd och riktningar kommer detta att gälla även linjen från $h(g(p))$ till $h(g(p'))$. Men om $f(p) = h(g(p))$, och linjen från denna punkt till $f(p')$ respektive $h(g(p'))$ är densamma, så måste även $f(p')$ och $h(g(p'))$ vara samma punkt. $\square$

2.3 Axiom

Den presentation av euklidisk geometri vi har gjort i detta kapitel är radikalt annorlunda från Euklides ursprungliga. Euklides använder sig av samma begrepp — punkter, linjer, vinklar — men definierar dessa genom *axiom*. Det vi har gjort är att definiera punkter som par av tal, och definiera geometriska figurer genom vilka ekvationer dessa talpar uppfyller. Och hur skulle Euklides ha kunnat hitta på en sådan definition? Euklides talade aldrig om ekvationer: han använde sig av ett talsystem som inte ens hade en symbol för 0, och det skulle dröja nästan 2000 år innan René Descartes skulle införa bruket av



Figur 2.7: En illustration av fjärde postulatet. Två linjesegment faller in mot linjen λ , och det högra möter λ i rät vinkel.

ekvationer och koordinatsystem i geometrin.⁶

I Euklides *Elementa* följs en annan väg in i geometrin: man definierar aldrig vad en punkt eller en linje faktiskt är för något, endast vilka *egenskaper* dessa skall ha. Vad detta betyder är att Euklides säger att det finns något som heter punkter och linjer och cirklar och vinklar, och att dessa uppfyller en viss uppsättning axiom. I *Elementa* skiljer han på vad han kallar axiom och postulat, även om vi idag skulle säga att axiom och postulat är synonymer. Dessa är Euklides fem postulat:⁷

- (i) Genom två punkter passerar en unik linje;
- (ii) Varje ändligt långt linjesegment kan dras vidare godtyckligt långt i bägge riktningar;
- (iii) Givet en punkt p och en sträcka r , finns det en unik cirkel med p som mittpunkt och r som radie;
- (iv) Vi börjar med en definition: Låt λ vara någon linje, och μ ett linjesegment som möter λ i sin ena ändpunkt. Mellan λ och μ kan vi mäta två olika vinklar, och om dessa två är lika så säger vi att vinklarna är *räta*. Det fjärde postulatet säger att alla räta vinklar är lika stora, se Figur 2.7.
- (v) Givet en linje och en punkt som inte ligger på denna, finns en unik linje genom punkten som inte skär den givna linjen någonstans.

Vi har ersatt Euklides ursprungliga femte postulat med det logiskt ekvivalenta *Playfairs axiom*,⁸ som för många upplevs som geometriskt tydligare. Euklides ursprungliga version ser ut så här: om två linjer skär en tredje i två olika punkter, på ett sådant sätt att de båda inre vinklarna på ena sidan av den tredje linjen summerar till mindre än två räta vinklar, så måste de första två linjerna skära varandra i en punkt på just denna sidan av den tredje linjen.

⁶René Descartes (1596-1650) räknas som den *analytiska geometrins* fader, där ordet analytisk betyder precis att man använder ekvationer för att beskriva geometriska figurer och algebra för att lösa geometriska problem. Från honom kommer namnet "kartesiskt koordinatsystem".

⁷Vi har omformulerat dem något från hur de ser ut i original.

⁸Döpt efter den skotske matematikern John Playfair (1748-1819).

Utöver dessa postulat använder sig Euklides av fem axiomer som man kan föreställa sig att han skilde från postulaten för att han tyckte att de var giltiga för all matematik, inte bara plan geometri:

- (i) Storheter som är lika med samma storhet är lika;
- (ii) Om samma storhet adderas till lika stora storheter, är resultaten lika stora;
- (iii) Om samma storhet subtraheras från lika stora storheter, är resultaten lika stora;
- (iv) Storheter som sammanfaller är lika stora;
- (v) Det hela är större än varje del.

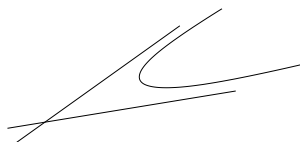
De första två uttrycker två räknelagar i ord: om $a = b$ och $b = c$ är $a = c$, respektive att om $a = b$ är $a + c = b + c$. Det tredje axiomet skulle vi idag se som ett specialfall av det andra eftersom subtraktion endast är addition med ett negativt tal, men Euklides kände inte till begreppet "negativt tal". De sista två är påtagligt mer geometriska och lite luddigare formulerade. Med den första menar han exempelvis att om man genom att flytta och rotera planet kan överföra en triangel till en annan, så kommer dessa att ha samma vinklar, sidlängder, area, och så vidare. Med den sista menar han t.ex. att om en linje delar en vinkel i två delar, så kommer båda dessa delar att vara mindre än den ursprungliga vinkeln.

Idag skulle vi säga att han dessutom använder vissa underförstådda egenskaper hos begreppen han studerar. Till exempel kan han argumentera att ett linjestycke som börjar innanför en cirkel och slutar utanför den, måste någonstans skära cirkeln. Det finns dock inget i hans axiomer som garanterar detta — det går inte ens att visa att det måste existera en in- eller utsida!

Att försöka visa satser om geometri med hjälp av Euklides axiomer, i stället för att använda ekvationer och trigonometri som vi är vana vid, är en utmaning till en början. En kul bok som arbetar med både euklidisk och icke-euklidisk geometri på detta rent axiomatiska sätt är den bitvis ganska tunga [4]. Dock kommer vi inte att göra något sådant i detta kompendium. Vårt fokus kommer vara att arbeta med en konkret modell av geometri, som punkter i ett koordinatplan, och vi kommer försöka konstruera en av de vanligaste modellerna för *hyperbolisk geometri*, den så kallade övre halvplans-modellen.

2.4 Historik över parallellpostulatet

Det femte av Euklides postulat brukar kallas för *parallellpostulatet*, och dess fascinerande historia är idag välkänd bland matematiker — många matematiker kan, i stora drag, återberätta historiken i de följande paragraferna. Materialet i detta avsnitt, och mycket mer detaljer, kan hittas i vilken bok som helst om hyperbolisk geometri eller matematikens historia, såsom [2].



Figur 2.8: Legendres märkliga konfiguration av räta linjer.

Man brukar allmänt definiera två linjer som *parallella* om de inte möts i någon punkt. Parallellpostulatet säger alltså att givet en linje λ och en punkt $p \notin \lambda$, finns en unik linje genom p som är parallell med λ .

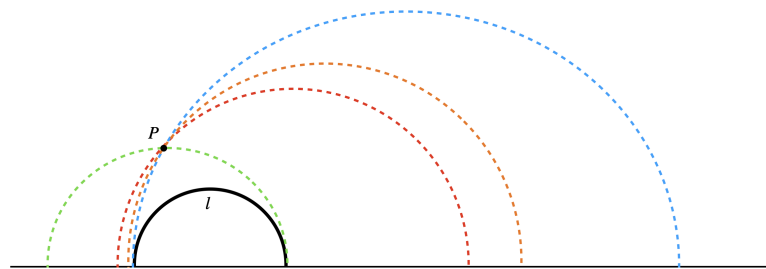
Euklides själv använder inte postulatet i sina första 28 propositioner, något som många idag tolkar som att han själv ogillade det och ansträngde sig för att inte behöva använda det. Det femte postulatet kan tyckas klumpigt och krångligt och mindre självklart än de första fyra, och väldigt länge efter Euklides tid var matematiker övertygade om att det gick att bevisa det femte postulatet från de första fyra. Många felaktiga bevis för det femte postulatet har föreslagits genom åren, men alla har fallit på att de någonstans på ett subtilt sätt använt en egenskap hos planet som är ekvivalent med parallellpostulatet. Till exempel upptäcktes Playfairs axiom av den grekiske matematikern Proclus; det var denna egenskap han implicit använde i sitt eget felaktiga bevis för (Euklides version av) parallellpostulatet.

För en modern läsare är kanske de intressantaste bland dessa felaktiga bevisförsök Girolamo Saccheris (1667 - 1733). Vad han gjorde var att anta att parallellpostulatet är falskt, och försökte hitta en motsägelse. Motsatsen till parallellpostulatet är som följer: det existerar en linje och en punkt ej på linjen, sådana att antingen finns *ingen* parallell linje genom den givna punkten, eller så finns fler än en. Han visade att om det inte finns någon parallell linje finner man en motsägelse, så om parallellpostulatet är falskt så måste det finnas många olika linjer som inte skär den givna. Saccheri hittade flera intressanta resultat innan han upptäckte något han trodde var en motsägelse. (Resultatet han hittade var dock inte egentligen en motsägelse, utan bara ointuitivt.) Han kunde bland annat på vägen visa att om parallellpostulatet är falskt för någon punkt och linje, så är postulatet falskt för alla punkter och alla linjer. Han visade att om det fanns fler än en parallell linje, så existerar oändligt många olika parallella linjer. Vidare så märkte han att om parallellpostulatet är falskt så måste vinkelsumman i varje triangel i planet vara *mindre* än π , inte lika med π som för en vanlig triangel!⁹

Adrien-Marie Legendre (1752-1833) fortsatte Saccheris arbete och visade t.ex. att om parallellpostulatet är falskt så går det att rita två linjer som skär varandra i en godtyckligt liten vinkel, och sedan rita en tredje oändligt lång rät linje som ligger helt innanför det spetsiga området mellan de andra två linjerna, utan att skära någon av dem.

Det är händelseförloppet efter allt detta som ställde hela historien på huvudet. Ungernaren János Bolyai (1802-1860) och ryssen Nikolaj Lobatjevskij (1792-

⁹Euklides själv mätte inte vinklar i vare sig radianer eller grader, utan skulle ha formulerat det som att vinkelsumman är mindre än två rätta vinklar.



Figur 2.9: Illustration av parallellaxiomet i hyperbolisk geometri: givet en linje l och en punkt P som inte ligger på l , finns det oändligt många linjer som går genom P och inte skär l .

1856) insåg båda två, oberoende av varandra, att det *aldrig* kommer att gå att bevisa parallellpostulatet enbart med hjälp av de tidigare fyra postulaten, eftersom det går att konstruera en ny sorts geometri där *alla de fyra första postulaten stämmer, men parallellpostulatet är falskt*. I moderna termer skulle vi säga att de insåg existensen av den hyperboliska geometrin.

I denna geometri betyder alla orden "punkt", "linje", "vinkel", och så vidare, något annat än i den klassiska euklidiska geometrin; men med korrekt ändrade definitioner får man en geometri lika betydelsefull och "sann" som den hittills kända. Med andra ord var alla märkliga resultat som Saccheri och Legendre fann när de antog motsatsen till parallellpostulatet helt enkelt inga motsägelser alls, utan de första kända satserna om hur parallella linjer beter sig i hyperbolisk geometri!

Åtminstone är det så vi ser historien idag. Av deras samtid blev både Bolyai och Lobatjevskij motarbetade och hade svårt att få sina arbeten publicerade. (Så lite genomslag hade deras idéer att det tog Bolyai över tjugo år att få höra talas om Lobatjevskijs arbeten!) För att strö salt i såren skrev den store Carl Friedrich Gauss i ett brev till Bolyai att hans idéer var geniala, men att Gauss själv redan upptäckt dem många år tidigare men inte velat publicera dem!

Inte förrän mot artonhundratalets andra hälft vann den hyperboliska geometrin allmän acceptans, efter flera matematikers arbete (särskilt bör man här nämna Eugenio Beltrami [1835-1900], Felix Klein [1849-1925] och Henri Poincaré [1854-1912]) med att förenkla och förklara dessa teorier. Speciellt kände man vid det här laget till övre halvplansmodellen (som vi kommer visa i detta häfte) och den så kallade diskmodellen för hyperbolisk geometri, d.v.s. man kunde faktiskt rita bilder av hur det hyperboliska planet ser ut. Man kan jämföra med hur svårt det var för komplexa tal att bli accepterade av matematikervärlden innan de kunde åskådliggöras genom det komplexa talplanet, som vi kommer att prata mer om i kapitel 3 i detta häfte.

Övningar

Övning 2.1. (i) Ge exempel på en funktion $f : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ som bevarar vinklar men ej avstånd.

(ii) Går det att tala om en vinkelbevarande funktion som ej respekterar linjer?

(iii) Om (x, y) uppfyller $ax + by = c$ vilken ekvation uppfyller $f((x, y))$ där $f((x, y)) = (x + 1, y + 1)$?

Övning 2.2. Hitta den unika automorfi f som avbildar $(1, 0)$ på $(2, 5)$ respektive $(2, 0)$ på $(2, 6)$. Bestäm även f^{-1} .

Övning 2.3. (i) Visa att translation bevarar vinklar. Antag att ett segment av en cirkel från en punkt A till B har samma längd och riktning efter translation.

(ii) Visa att reflektion $f((x, y)) = ((x, -y))$ ej bevarar vinklar.

Övning 2.4. Låt X vara en mängd. Visa att om en funktion $f : X \rightarrow X$ har en invers, det vill säga det existerar en funktion $f^{-1} : X \rightarrow X$ sådan att $f \circ f^{-1} = f^{-1} \circ f = \text{id}$, så är f en bijektion.

Övning 2.5. En linje som ges av $ax + by = c$ kan också skrivas på så kallad *parameterform*. Detta betyder att vi beskriver kurvan som bilden av en funktion på formen

$$f(t) = (d + rt, e + st)$$

där d, e, r och s är konstanter och inte både r och s är noll. Visa att en möjlig parametrisering av linjen $ax + by = c$ ges av att välja

$$d = \frac{ac}{a^2 + b^2}, \quad e = \frac{bc}{a^2 + b^2}, \quad r = b, \quad s = -a.$$

Övning 2.6. Sats 2.2.4 säger att varje automorfi av det euklidiska planet kan skrivas som en rotation kring origo följt av en translation. Visa att varje automorfi även kan skrivas som en translation följt av en rotation kring origo. (Ledning: om f är en automorfi, är även den inversa avbildningen f^{-1} det.)

Övning 2.7. Antag att ekvationerna $ax + by = c$ och $a'x + b'y = c'$ definierar samma linje i planet. Visa att ekvationerna endast skiljer sig åt med en konstant, det vill säga det finns en nollskild konstant ρ sådan att $a = \rho a'$, $b = \rho b'$ och $c = \rho c'$.

Övning 2.8. Avsluta beviset för att en rotation avbildar linjer bijektivt.

3 Komplexa tal

I det här kapitlet inför vi räknesätten addition, subtraktion, multiplikation och division för punkter i planet. Detta gör att vi kan identifiera punkter i planet med de så kallade *komplexa talen*. Dessa betecknas vanligtvis med $\mathbb{C}$. Vi diskuterar olika sätt att representera komplexa tal och ger en geometrisk tolkning av räknesätten.

Parameterkurvor i planet kan också beskrivas med hjälp av komplexa tal, och automorfierna av $\mathbb{R}^2$ kan på ett elegant sätt identifieras med vissa avbildningar av $\mathbb{C}$. Vi avslutar kapitlet med en kort genomgång av funktioner av komplexa tal och deras egenskaper.

3.1 Operationer med punkter i planet

Alla punkter i planet $\mathbb{R}^2$ kan beskrivas med hjälp av två reella tal, punkternas så kallade koordinater. Vi behärskar sedan tidigare räknesätten för reella tal: vi kan addera, subtrahera, multiplicera och dividera reella tal.

Vi använder nu de kända räknesätten för $\mathbb{R}$ för att införa motsvarande räknesätt för punkter i planet. Vi delar för enkelhetens skull upp definitionerna.

Definition 3.1.1. Låt $p = (a, b)$ och $q = (c, d)$ vara godtyckliga punkter i $\mathbb{R}^2$. *Summan* av p och q ges av

$$p + q = (a + c, b + d)$$

och *differensen* av p och q är

$$p - q = (a - c, b - d).$$

Vi noterar att detta är väldefinierat eftersom $a + c$ och $b + d$ samt $a - c$ och $b - d$ är definierade för reella tal a, b, c, d . Resultatet av additionen och subtraktionen är ett nytt par av reella tal, och vi kan givetvis identifiera dessa par med punkter i planet. Dessutom känns definitionen naturlig: additionen och subtraktionen sker komponentvis. Vi ska senare se hur operationerna kan tolkas geometriskt.

Vi fortsätter med att definiera multiplikation.

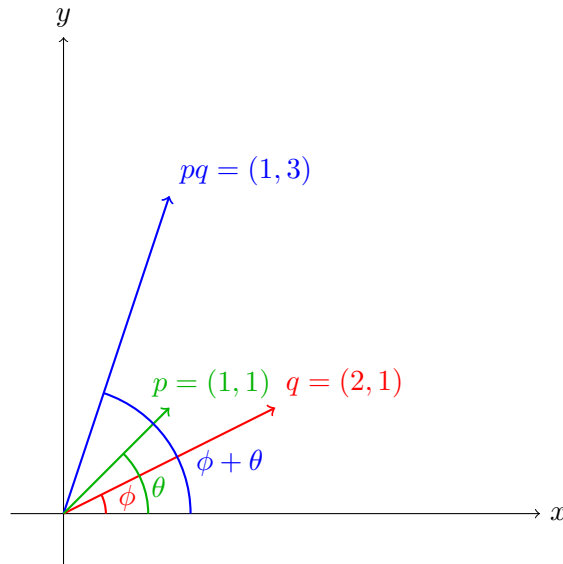
Definition 3.1.2. Låt $p = (a, b)$ och $q = (c, d)$ vara godtyckliga punkter i planet. *Produkten* av p och q definieras som

$$p \cdot q = (ac - bd, ad + bc).$$

Vi ser igen att definitionen är meningsfull eftersom uttrycken ac , bd , och så vidare, samt summor och differensen av dem, är meningsfulla för reella tal a, b, c och d .

Exempel 3.1.3. Sätt $p = (1, 1)$ och $q = (2, 1)$. Då har vi enligt definitionen av addition och subtraktion att

$$p + q = (1, 1) + (2, 1) = (3, 2)$$



Figur 3.1: Vinkeln av produkten pq ges av summan av vinkeln av p och vinkeln av q .

samt

$$p - q = (1, 1) - (2, 1) = (-1, 0).$$

Vi använder definitionen av produkt för att beräkna

$$pq = (1, 1) \cdot (2, 1) = (1 \cdot 2 - 1 \cdot 1, 1 \cdot 1 + 1 \cdot 2) = (1, 3)$$

Definitionen av produkten av två punkter kan se lite godtycklig ut. Det finns dock en fin geometrisk beskrivning av den. Om vinkeln mellan x -axeln och (en linje som går från origo till) p är θ medans vinkeln mellan x -axeln och q är ϕ , visar det sig att vinkeln mellan x -axeln och produkten pq kommer vara $\phi + \theta$. Figur 3.1 illustrerar detta för det föregående exemplet. Dessutom visar det sig att avståndet av pq från origo kommer vara precis produkten av avståndet av p från origo med avståndet av q från origo. Vi tar och bevisar detta i sats 3.4.1.

Vi definierar division av punkter som följande.

Definition 3.1.4. Låt $p = (a, b) \in \mathbb{R}^2$ och $q = (c, d) \neq (0, 0) \in \mathbb{R}^2$. Då definierar vi

$$\frac{p}{q} = \left(\frac{ac + bd}{c^2 + d^2}, \frac{-ad + bc}{c^2 + d^2} \right)$$

Komplexa tal har samma räkneregler som reella tal. Detta beskriver vi i nästa sats.

Sats 3.1.5. För alla punkter $p = (a, b), q = (c, d), r = (e, f) \in \mathbb{R}^2$ gäller det att

- (Associativitet) $(p + q) + r = p + (q + r)$
- (Identitet) $p + (0, 0) = p$

- (Invers) $p + (-p) = (0, 0)$, (där $(-p)$ är definierat som $(-a, -b)$).
- (Kommutativitet) $p + q = q + p$.

att

- (Associativitet) $(pq)r = p(qr)$.
- (Identitet) $p \cdot (1, 0) = p$ för alla $p \in \mathbb{R}^2$
- (Invers) $p \cdot \frac{(1,0)}{p} = (1, 0)$ om $p \neq (0, 0)$.
- (Kommutativitet) $pq = qp$.

samt att

- (Distributivitet) $p(q + r) = pq + pr$.

Bevis. Beviset lämnar vi till en övning. Se övning 3.8. □

3.2 Real- och imaginärdel av komplexa tal

Vi har i föregående avsnitt infört algebraiska operationer på punkter i planet. Detta gör att vi kan uppfatta punkter i planet, tillsammans med dessa operationer, som ett nytt slags tal som vi kan räkna med.

Vi ska nu införa den notation som är bruklig när man arbetar med dessa komplexa tal. Den är framför allt praktisk när man vill betona den algebraiska aspekten av komplexa tal, till skillnad från den geometriska tolkningen som punkter i det vanliga planet.

Vi börjar med två exempel.

Exempel 3.2.1. Betrakta två punkter $p = (a, 0)$ och $q = (c, 0)$ på x -axeln. När vi adderar p och q eller subtraherar dem från varandra får vi nya punkter på x -axeln. Till exempel gäller $q - p = (c - a, 0)$.

Definitionen av multiplikation medför att produkten av p och q är

$$pq = (a \cdot c - 0 \cdot 0, a \cdot 0 + c \cdot 0) = (ac, 0),$$

så pq ligger också på x -axeln.

Exempel 3.2.2. Vi betraktar nu en punkt på x -axeln, $p = (a, 0)$ samt en punkt på y -axeln, $q = (0, b)$. Vi ser genast att summan eller differensen av p och q i allmänhet inte ligger på vare sig x -eller y -axeln. Däremot har vi

$$pq = (a \cdot 0 - 0 \cdot b, a \cdot b + 0 \cdot 0) = (0, ab),$$

så produkten av p och q ligger på y -axeln.

Vi sätter hädanefter

$$i = (0, 1)$$

och kommer lite oegentligt att skriva

$$1 = (1, 0).$$

I det sistnämnda fallet genomför vi i själva verket en identifikation av det speciella paret $(1, 0)$ med det reella talet 1.

Mer allmänt kommer vi att identifiera det *komplexa* talet $z = (a, 0)$ med det reella talet a . Från räkningarna i Exempelen 3.2.1 och 3.2.2 följer att varje punkt $p = (a, b)$ kan skrivas som

$$p = (a, b) = (a, 0)(1, 0) + (0, 1)(b, 0) = a1 + ib = a + ib.$$

Vi använder denna notation när vi vill betona de algebraiska aspekterna av operationer med punkter i planet.

Definition 3.2.3. Ett *komplext tal* är ett tal på formen

$$z = a + ib, \quad \text{där } a, b \in \mathbb{R}.$$

Varje komplext tal svarar mot en punkt $p \in \mathbb{R}^2$ med $p = (a, b)$. En punkt $p = (a, b) \in \mathbb{R}^2$ kan identifieras med det komplexa talet $z = a + ib$.

Det reella talet a kallas för z 's *realdel*, och vi skriver $a = \operatorname{Re}(z)$. Talet b är z 's *imaginärdel* och betecknas $b = \operatorname{Im}(z)$.

Mängden av alla komplexa tal, det vill säga,

$$\{z = a + ib : a, b \in \mathbb{R}\}$$

brukar betecknas med $\mathbb{C}$.

Om $a = 0$ i $z = a + ib$ säger vi att z är rent imaginärt och om $b = 0$ att z är reellt. De reella talen $\mathbb{R}$ kan identifieras med en delmängd av de komplexa talen, nämligen med mängden

$$\{z \in \mathbb{C} : \operatorname{Im}(z) = 0\},$$

vilken sammanfaller med x -axeln i planet.

Vi avslutar detta avsnitt med en intressant observation: vi har

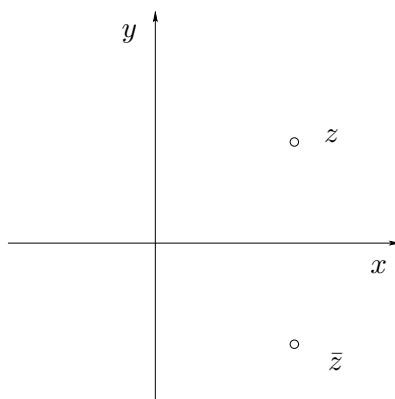
$$i^2 = i \cdot i = (0, 1)(0, 1) = (-1, 0).$$

Vi har ju tidigare identifierat paret $(-1, 0)$ med talet -1 . Vår uträkning visar alltså att i har egenskapen

$$i^2 = -1,$$

det vill säga, i löser ekvationen $x^2 + 1 = 0$.

Historiskt sett var det så studiet av komplexa tal inleddes—man ville hitta rötter till vissa polynomekvationer. Man införde därför en symbol, i , bildade uttryck på formen $a + ib$ och definierade räkneregler för dessa. Därefter visade man att räknereglerna var meningsfulla och reducerades till de vanliga reglerna för reella tal när $b = 0$. Eftersom vi främst är intresserade av geometri har vi börjat med att betrakta komplexa tal som punkter i planet, utrustade med räkneregler, men i själva verket upptäcktes denna tolkning av de komplexa tal efter deras algebraiska egenskaper.



Figur 3.2: Punkten z och dess konjugat $\bar{z}$.

3.3 Konjugat, längd och vinkel.

Att komplex multiplikation motsvarar rotationer gör komplexa tal väldigt användbara. Vad som dock inte kan beskrivas med rotationer är *reflektioner*. Vi inför därför det så kallade konjugatet av ett komplext tal.

Definition 3.3.1. *Konjugatet* av $z = a + bi \in \mathbb{C}$ är

$$a - bi \in \mathbb{C}.$$

Vi skriver detta som $\bar{z}$.

Geometriskt innebär konjugering att vi speglar punkten i x -axeln.

Vi definierar även *längden* eller *absolutbeloppet* av ett komplext tal $z = a + bi \in \mathbb{C}$ som

$$|z| = \sqrt{a^2 + b^2}$$

vilket då motsvarar avståndet från 0 till z .

Om vi testar att beräkna produkten av ett komplext tal med sitt konjugat får vi att

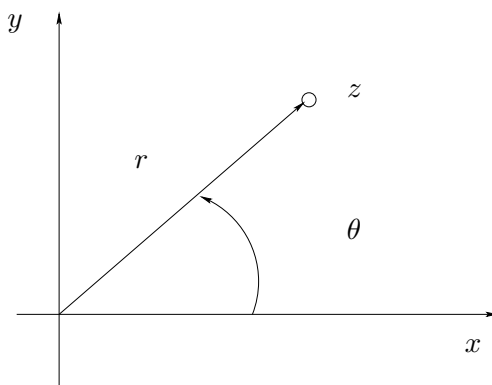
$$z\bar{z} = (a + bi)(a - bi) = a^2 - (bi)^2 = a^2 + b^2 = |z|^2$$

så konjugat-funktionen låter oss även smidigt arbeta med avstånd. Konjugatet ger oss även ett enklare uttryck för kvoten av två komplexa tal, nämligen

$$\frac{z}{w} = \frac{z\bar{w}}{|w|^2}.$$

3.4 Komplexa tal på polär form

Vi kan beskriva punkter i planet genom dess x - och y -koordinater. Ett annat sätt att beskriva en punkt i planet är genom att först säga vilken riktning punkten ligger i, och sedan hur långt vi måste gå i den riktningen för att nå punkten. Detta kalas för polära koordinater (se figur 3.3).



Figur 3.3: Polära koordinater.

För att göra det mer konkret, givet en vinkel θ mot x -axeln och ett avstånd r att gå från origo för att komma till punkten, kan vi beskriva den resulterande punkten i x - och y -koordinater som $(r \cos \theta, r \sin \theta)$, eller som ett komplext tal $r(\cos \theta + i \sin \theta)$.

Vinkeln θ kallas för *argumentet* av z , och vi skriver $\theta = \arg(z)$. Om $z = 0$ är $\arg(z)$ odefinierat. Vi kan därmed skriva

$$z = |z|(\cos \arg(z) + i \sin \arg(z)).$$

Vi ska nu bevisa vad vi nämnde tidigare, att multiplikation av två komplexa tal motsvarar att vi adderar argumenten och multiplicerar absolutvärdena.

Sats 3.4.1. Låt $z, w \in \mathbb{C}$ vara två komplexa tal. Låt $r = |z|$, $s = |w|$, $\theta = \arg z$ och $\phi = \arg w$. Då kommer

$$zw = rs(\cos(\theta + \phi) + i \sin(\theta + \phi)).$$

Bevis. Vi multiplicerar de komplexa talen z och w och får

$$\begin{aligned} z \cdot w &= r(\cos \theta + i \sin \theta) \cdot s(\cos \phi + i \sin \phi) \\ &= rs(\cos \theta \cos \phi - \sin \theta \sin \phi + i(\cos \theta \sin \phi + \sin \theta \cos \phi)) \end{aligned}$$

och genom att använda additionsformlerna för sinus och cosinus får vi

$$z \cdot w = rs(\cos(\theta + \phi) + i \sin(\theta + \phi)) \quad (3.1)$$

vilket vi ville visa. □

Vi kan även sammanfatta detta med formlerna

$$|z \cdot w| = |z| \cdot |w| \quad \text{och} \quad \arg(zw) = \arg(z) + \arg(w).$$

Liknande räkningar kan genomföras för division. I polära koordinater blir kvoten av talen z och w , $w \neq 0$, det komplexa talet som i polära koordinater kan skrivas

$$\frac{z}{w} = \frac{r}{s}[\cos(\theta - \phi) + i \sin(\theta - \phi)],$$

det vill säga, vi dividerar talens absolutbelopp och subtraherar de tillhörande vinklarna:

$$\left| \frac{z}{w} \right| = \frac{|z|}{|w|} \quad \text{och} \quad \arg(z/w) = \arg(z) - \arg(w).$$

Eftersom uttrycket $\cos \theta + i \sin \theta$ används så ofta när man arbetar med komplexa tal är det smidigt att hitta ett förkortat skrivsätt för det. Man kan därför införa den *komplexa exponentialfunktionen*

$$e^{i\theta} = \cos \theta + i \sin \theta,$$

som är en funktion som till varje reellt θ tillordnar ett komplext tal $e^{i\theta}$. Vi kan vidare notera att

$$\left| e^{i\theta} \right| = \sqrt{\cos^2 \theta + \sin^2 \theta} = 1.$$

Vi kan nu skriva det komplexa talet $z = a + ib$ på ett kompakt sätt:

$$z = r e^{i\theta},$$

där $r = |z|$ och $\theta = \arg(z)$. Genom att använda formeln (3.1), med $r = s = 1$, erhåller vi

$$e^{i\theta} e^{i\phi} = e^{i(\theta+\phi)}.$$

Vi har vidare

$$i e^{i\theta} = i(\cos \theta + i \sin \theta) = i \cos \theta + i^2 \sin \theta = -\sin \theta + i \cos \theta$$

eftersom $i^2 = -1$. Detta visar att derivatan av den komplexa exponentialfunktionen med avseende på θ är

$$(e^{i\theta})' = i e^{i\theta}.$$

Funktionen $e^{i\theta}$ beter sig alltså verkligen som en exponentialfunktion! Vi ska i nästa avsnitt förklara exakt vad vi menar med att derivera en funktion som tar komplexa värden.

3.5 Komplexa funktioner

Precis som vi kan ha funktioner av reella tal kan vi även ha funktioner av komplexa tal. I den här kursen kommer vi främst vara intresserade över funktioner $f : I \rightarrow \mathbb{C}$ och funktioner $f : D \rightarrow \mathbb{C}$, där $I \subset \mathbb{R}$ är något intervall och $D \subset \mathbb{C}$ är någon delmängd av de komplexa talen.

En funktion $f : I \rightarrow \mathbb{C}$ kan vi alltid skriva på formen

$$f(t) = f_1(t) + i f_2(t)$$

där $f_1, f_2 : I \rightarrow \mathbb{R}$ båda är reell-värda funktioner. Ifall f_1 och f_2 båda är deriverbara definierar vi derivatan av f att vara

$$f'(t) := f_1'(t) + i f_2'(t)$$

för alla $t \in I$.

I nästa kapitlet kommer vi diskutera dessa funktioner mer samt se hur de kan betraktas som parametriserar av kurvor i planet.

Vi ska nu ägna lite tid åt komplexa funktioner av den andra typen, det vill säga funktioner $f : D \rightarrow \mathbb{C}$ vars argument är komplext, så $D \subset \mathbb{C}$ är någon delmängd av planet. En sådan funktion kan betraktas som en transformation av punkterna i D där alla punkter skickas till en ny punkt i planet.

Den enklaste typen av dessa funktioner är linjära funktioner (eller mer korrekt, affina funktioner). Dessa är funktioner på formen

$$f(z) = \alpha z + \beta$$

där $\alpha, \beta \in \mathbb{C}$ är komplexa tal. Om $|\alpha| = 1$ visar nästa sats att vi redan är bekanta med dessa funktioner.

Sats 3.5.1. *Automorfierna av planet kan identifieras med mängden av funktioner $f: \mathbb{C} \rightarrow \mathbb{C}$ som ges på formen*

$$f(z) = \alpha z + \beta,$$

där $\alpha \in \mathbb{C}$ uppfyller $|\alpha| = 1$ och β är ett godtyckligt komplext tal.

Bevis. Vi börjar med att omtolka translationerna av planet som funktioner av komplexa tal. Låt alltså $T(x, y) = (x + a, y + b)$ vara en translation. Sätt $\beta = a + ib$ och $z = x + iy$. Vi ser då att

$$f(z) = z + \beta = (x + iy) + (a + ib) = (x + a) + i(y + b)$$

har precis samma verkan på punkterna i planet som T .

Betrakta nu en rotation kring origo $R(x, y) = (x \cos \theta - y \sin \theta, x \sin \theta + y \cos \theta)$ och skriv, liksom tidigare, $z = x + iy$. Vi inför $\alpha = \cos \theta + i \sin \theta$. Vi har redan sett att de α som kan skrivas på denna form är exakt de som uppfyller $|\alpha| = 1$. Vi beräknar nu produkten αz med hjälp av definitionen av komplex multiplikation:

$$\alpha z = (\cos \theta + i \sin \theta)(x + iy) = (x \cos \theta - y \sin \theta) + i(x \sin \theta + y \cos \theta).$$

Vi ser alltså att multiplikation med det komplexa talet α roterar punkten som hör till talet $z = x + iy$ med en vinkel θ . Vi ser till slut att $f(z) = \alpha z + \beta$ är en rotation följt av en translation. $\square$

En viktig egenskap hos automorfier är att sammansättningen av två automorfier är en ny automorfi. Vi överlämnar det som en övning (se övning 3.7) åt läsaren att kontrollera att sammansättningen av två funktioner på formen $\alpha z + \beta$ där $\alpha \in \mathbb{C}$ uppfyller $|\alpha| = 1$ är en ny funktion på samma form.

En mer allmän linjär (affin) funktion kan vi skriva på formen

$$f(z) = |\alpha| \frac{\alpha}{|\alpha|} z + \beta$$

givet att $\alpha \neq 0$ (annars är den konstant). Eftersom $|\alpha| \in \mathbb{R}$ och $\left| \frac{\alpha}{|\alpha|} \right| = 1$ kan vi tolka f som en rotation, följt av en skalning, följt av en translation.

3.6 Komplexa polynom

De fyra räknesätten för komplexa tal gör det möjligt för oss att införa mer generella funktioner $f: \mathbb{C} \rightarrow \mathbb{C}$.

Exempel 3.6.1. Ett *komplext polynom av grad n* är en funktion $f: \mathbb{C} \rightarrow \mathbb{C}$ på formen

$$f(z) = a_n z^n + a_{n-1} z^{n-1} + \cdots + a_1 z + c_0, \quad (3.2)$$

där $a_0, a_1, \dots, a_n$ är fixa komplexa tal, polynoms *koefficienter*, och $a_n \neq 0$.

I exemplet ovan väljer man alltså koefficienterna a_i och varierar z : för varje $z \in \mathbb{C}$ ger formeln (3.2) oss ett nytt komplext tal $f(z)$.

Exempel 3.6.2. Vi kommer senare att stöta på funktioner på formen

$$f(z) = \frac{az + b}{cz + d}.$$

I detta fall måste vi vara försiktiga när vi delar med 0, det vill säga, när $z = -d/c$.

3.7 Mer allmänna funktioner av komplexa tal

Man kan också definiera exponentialfunktionen e^z för komplexa tal — kom ihåg att vi hittills definierat $e^{i\theta}$ för θ reellt. Man kan helt enkelt sätta

$$e^z = e^a \cdot e^{ib} \quad \text{för} \quad z = a + ib,$$

där e^a ska tolkas som den vanliga exponentialfunktionen och $e^{ib} = \cos b + i \sin b$. Man kan även införa trigonometriska funktioner av komplexa tal, som till exempel $\sin z$ och $\cos z$, på ett sådant sätt att vi får tillbaka de funktioner vi känner sedan tidigare om vi väljer z reellt. Det uppstår emellertid vissa svårigheter när man studerar komplexvärda funktioner av komplexa tal. När man arbetar med funktioner av en reell variabel brukar man undersöka funktionernas grafer för att förstå dem bättre. Det är som synes svårt att rita en graf över en komplex funktion eftersom vi skulle behöva rita i fyra dimensioner.

3.8 Komplex derivata

Vi avslutar det här kapitlet med att diskutera derivering av komplexa funktioner. Vi kommer dessvärre inte kunna gå in så mycket i detalj här då det är ett väldigt rikt område. Området där komplex derivata studeras kallas för *komplex analys*.

Vi påminner att en reell funktion f definierad på något intervall $I \subset \mathbb{R}$, $f: I \rightarrow \mathbb{R}$ är deriverbar i en punkt x om den kan approximeras med en linjär funktion i den punkten, det vill säga

$$f(t) \approx at + b$$

där $a, b \in \mathbb{R}$, för alla t i ett litet intervall kring x . På samma vis sägs en komplex funktion definierad i något område¹⁰ $D \subset \mathbb{C}$, $f : D \rightarrow \mathbb{C}$ vara *komplext deriverbar*, *holomorf* eller *analytisk* i en punkt z , om

$$f(w) \approx \alpha w + \beta$$

där $\alpha, \beta \in \mathbb{C}$, för alla w nära z .

Talet α säger vi är derivatan av funktionen i punkten z och skriver $f'(z) = \alpha$.

Definitionen ovan är inte så rigorös, men vi kan ändå ge några viktiga konsekvenser.

Vi ser att i små områden så ska f bete sig som en rotation, följt av en skalning, följt av en translation. En konsekvens av detta är att f bevarar vinklar i små områden, givet att $\alpha \neq 0$. Vi kommer förklara mer vad detta innebär i kapitel 6. Vi ser också att om $|\alpha| = 1$ så beter sig f likt en automorfi i det området. Låt oss nu betrakta några exempel.

Exempel 3.8.1. • En linjär funktion $f(z) = \alpha z + \beta$ är deriverbar, med konstant derivata α överallt.

- Alla polynom är också deriverbara, exempelvis är $f(z) = z^2 + 2z$ deriverbart och har derivata $f'(z) = 2z + 2$.
- Funktionen $f(z) = \frac{1}{z}$ är deriverbar i alla punkter förutom $z = 0$, med derivata $f'(z) = -\frac{1}{z^2}$.
- Funktionen $f(z) = \operatorname{Re}(z)$ är inte deriverbar någonstans, då den inte beter sig som en rotation i något litet område. Alternativt ser vi också att den inte bevarar några vinklar
- Funktionen $f(z) = \bar{z}$ är inte heller deriverbar någonstans, vi såg i övning 2.3 att den inte bevarar vinklar, detta gäller även i små områden och den kan därmed inte approximeras med linjära funktioner.
- Funktionen $f(z) = e^z$ är deriverbar överallt och har derivata $f'(z) = e^z$.

Anledningen till att derivatorna ser ut som vi är vana vid är för att produktregeln, kvotregeln samt kedjeregeln är densamma för komplext deriverbara funktioner. Mer precist har vi följande sats.

Sats 3.8.2. Om f och g är analytiska funktioner i ett område $D \subset \mathbb{C}$, gäller det att

- för alla konstanter $\alpha, \beta \in \mathbb{C}$ är $\alpha f + \beta g$ en analytisk funktion i D med derivata $(\alpha f + \beta g)'(z) = \alpha f'(z) + \beta g'(z)$.
- fg är en analytisk funktion med derivata $(fg)'(z) = f'(z)g(z) + g'(z)f(z)$.

¹⁰Ett område D av $\mathbb{C}$ är en delmängd av $\mathbb{C}$ med följande egenskap: om $z \in D$ så ska det finnas något avstånd $r > 0$, så att alla punkter $w \in \mathbb{C}$ inom avstånd r , $|z - w| < r$, också ingår i D , $w \in D$. Detta betyder att alla punkter nära z ingår också i D .

- Där $g \neq 0$ är $\frac{f}{g}$ är en analytisk funktion, med derivata $\left(\frac{f}{g}\right)'(z) = \frac{f'(z)g(z) - g'(z)f(z)}{g(z)^2}$.

- Antag att h är en analytisk funktion i ett område $D' \subset \mathbb{C}$. Om kompositionen $f \circ g$ är väldefinierad är det en analytisk funktion med derivata

$$(f \circ g)'(z) = f'(g(z))g'(z).$$

- Om $f : D \rightarrow D'$ är en bijektion är f^{-1} en analytisk funktion med derivata $f^{-1}'(z) = \frac{1}{f'(f^{-1}(z))}$. Dessutom gäller det att $f'(z) \neq 0$ i D .

Beviset av satsen ligger dock utanför kursens ramar men kan finnas i böcker om komplex analys.

Vi tänkte avsluta kapitlet med att ändå ge en rigorös definition för komplex derivata. Innan det ska vi dock nämna en till version av kedjeregeln som vi kommer ha användning för.

Sats 3.8.3. Om f är en analytisk funktion och $g : I \rightarrow \mathbb{C}$ är en deriverbar funktion från något intervall $I \subset \mathbb{R}$ till C så är kompositionen $f \circ g$ en deriverbar funktion från I till $\mathbb{C}$ med derivata

$$(f \circ g)'(t) = f'(g(t))g'(t)$$

för $t \in \mathbb{R}$.

Definition 3.8.4 (Överkurs). En funktion $f : D \rightarrow \mathbb{C}$ för ett område $D \subset \mathbb{C}$ sägs vara *komplext deriverbar*, *analytisk* eller *holomorf* i en punkt $z \in D$ om det komplexa gränsvärdet

$$\lim_{h \rightarrow 0} \frac{f(z+h) - f(z)}{h}$$

existerar. Vad detta innebär är att för alla reella tal $\varepsilon > 0$ ska det finnas ett reellt tal $\delta > 0$ så att för alla $h \in \mathbb{C}$ där $|h| < \delta$ är

$$\left| \frac{f(z+h) - f(z)}{h} \right| < \varepsilon.$$

Definitionen ovan är överkurs och är inte viktig för att förstå resten av kursen.

Övningar

Övning 3.1. Beräkna $z + w$, $z - w$, zw och z/w för

(i) $z = 3 + i$ och $w = -1 + 4i$

(ii) $z = -\sqrt{2}i$ och $w = 2 - 5i$.

Övning 3.2. Visa att $\operatorname{Re}(z) = (z + \bar{z})/2$ och $\operatorname{Im}(z) = (z - \bar{z})/2i$.

Övning 3.3. Visa att $|zw| = |z||w|$ för $z, w \in \mathbb{C}$.

Övning 3.4. Låt $z, w \in \mathbb{C}$, visa följande regler för komplext konjugat.

(i) $\overline{\bar{z}} = z$

(ii) $\overline{z + w} = \bar{z} + \bar{w}$

(iii) $\overline{z\bar{w}} = \bar{z}w$

(iv) $\overline{\left(\frac{1}{z}\right)} = \frac{1}{\bar{z}}$

(v) $\overline{e^z} = e^{\bar{z}}$.

Övning 3.5. Lös andragradsekvationen $z^2 + iz + 2 = 0$ genom att använda kvadratkomplettering.

Övning 3.6. Visa att derivatan av den komplexa exponentialfunktionen $f(\theta) = e^{i\theta}$ är $ie^{i\theta}$.

Övning 3.7. Verifiera att sammansättningen av funktionerna $f(z) = \alpha z + \beta$ och $g(z) = \tilde{\alpha}z + \tilde{\beta}$, där α och $\tilde{\alpha}$ är komplexa tal med absolutbelopp 1 och β samt $\tilde{\beta}$ är godtyckliga, är en ny funktion på formen $h(z) = Az + B$ där $A, B \in \mathbb{C}$ och $|A| = 1$.

Övning 3.8. Bevisa sats 3.1.5.

Övning 3.9. Beräkna den komplexa derivatan av följande funktioner genom att använda sats 3.8.2 och att $(\alpha z + \beta)' = \alpha$. Beskriv även eventuella punkter där de inte är analytiska.

(i) $f(z) = z^2$

(ii) $f(z) = \frac{1}{z^3}$

(iii) $f(z) = \frac{z^2}{z^3 - 1}$

(iv) $f(z) = (1 + z^3)^3$.

4 Båglängd

I det här kapitlet ska vi studera kurvor i komplexa talplanet, samt deras längder.

Kurvorna kommer kunna vara böjda, börja och sluta i samma punkt och så vidare. Dessa geometriska objekt beskrivas med hjälp av ekvationer. Ofta vill man kunna följa en kurva i en viss riktning, och vi ska se att man kan ge en matematisk beskrivning av detta med hjälp av så kallad parameterframställning.

Med hjälp av parameterkurvor kan vi införa ett längdbegrepp för kurvor, den så kallade *båglängden* som ges av en viss typ av integral. Avståndet mellan två punkter i komplexa talplanet kan formuleras som den minsta möjliga båglängden för en kurva som förbinder de två punkterna, och vi ska visa att det är de rätta linjerna som minimerar detta avstånd.

4.1 Kurvor i komplexa talplanet

Geometriska objekt i komplexa talplanet kan beskrivas genom att ange villkor på de komplexa tal som hör till objektet. Ett sätt att göra detta är att säga att de ska uppfylla vissa ekvationer eller olikheter.

Exempel 4.1.1. En linje Λ i $\mathbb{C}$ är en mängd

$$\Lambda = \{z \in \mathbb{C} : a\operatorname{Re} z + b\operatorname{Im} z + c = 0\} \quad (4.1)$$

för fixa $a, b, c \in \mathbb{R}$.

Exempel 4.1.2. *Enhetscirkeln* $\mathbb{T}$ i komplexa talplanet beskrivas som mängden

$$\mathbb{T} = \{z \in \mathbb{C} : |z| = 1\}.$$

En cirkel är ett exempel på en sluten kurva, det vill säga, en kurva som inte har några ändpunkter.

Exempel 4.1.3. En annan delmängd av planet som kommer vara av intresse för oss är *det övre halvplanet*

$$\mathbb{H} = \{z \in \mathbb{C} : \operatorname{Im} z > 0\}.$$

Notera att punkterna i det här exemplet ska uppfylla en olikhet istället för en ekvation. Det leder till en mycket "större" mängd än i det föregående exemplet.

Eftersom en punkt $z = x+iy$ i komplexa talplanet motsvaras av ett talpar (x, y) i planet, kan motsvarande kurvor ovan beskrivas i vanliga planet. Exempelvis ges en linje Λ i $\mathbb{R}^2$ av en mängd $\Lambda = \{(x, y) \in \mathbb{R}^2 : ax + by + c = 0\}$ för fixa $a, b, c \in \mathbb{R}$.

Ibland är det av intresse att inte bara ange vilka punkter i planet vi vill betrakta, utan också ange en inbördes ordning för punkterna. Vi vill exempelvis kunna ange var man ska börja rita på en viss figur, eller i vilken riktning man

ska genomlöpa en kurva. Ett sätt att uppnå detta är att använda en funktion, eller rättare sagt, ett par av funktioner, för att beskriva vårt objekt.

En funktion $f : [a, b] \rightarrow \mathbb{C}$ tilldelar ett komplext tal $f(t)$ till varje tal i intervallet $[a, b]$. Det betyder att två funktioner $x, y : [a, b] \rightarrow \mathbb{R}$ tillsammans ger oss ett komplext tal $x(t) + iy(t)$ för varje $t \in [a, b]$. Vi kan tolka paret $(x(t), y(t))$ som koordinater för en punkt i planet. När t löper igenom intervallet $[a, b]$ från a till b rör sig punkten $x(t) + iy(t)$ i komplexa talplanet, och genom att rita ut $x(t) + iy(t)$ för alla värden på t får vi ett geometriskt objekt. Genom att välja funktionerna x och y på lämpligt sätt kan vi beskriva en stor klass av geometriska figurer, så kallade kurvor. Man kan tolka vårt upplägg som att funktionen $f(t) = x(t) + iy(t)$ beskriver en bana som vi ska färdas. Variablen t får då tolkningen tid—den säger oss var vi befinner oss under vår färd.

Vi är framför allt intresserade av ”snälla” kurvor, det vill säga, vi vill inte att våra kurvor ska vara alltför hackiga. Ett sätt att precisera detta är att ställa vissa krav på funktionerna som ger oss koordinaterna för punkter på kurvan.

Vi börjar dock med att ge några exempel på kurvor.

Exempel 4.1.4. Låt $p = x_1 + iy_1$ och $q = x_2 + iy_2$ vara punkter i komplexa talplanet med $p \neq q$. Betrakta funktionen $\gamma : [0, 1] \rightarrow \mathbb{C}$ som ges av

$$\gamma(t) = (1 - t)p + tq = (1 - t)x_1 + tx_2 + i((1 - t)y_1 + ty_2). \quad (4.2)$$

Då beskriver $\gamma(t)$ punkterna på den räta linjen som förbinder p och q och som genomlöps från p till q .

Man kan nämligen kontrollera (se Övning 4.1) att $\gamma(t) = x(t) + iy(t)$ uppfyller

$$(x_2 - x_1)y(t) - (y_2 - y_1)x(t) - y_1(x_2 - x_1) + x_1(y_2 - y_1) = 0$$

för varje t , vilket betyder att mängden $\{\gamma(t) : t \in [0, 1]\}$ uppfyller (4.1) med $a = -(y_2 - y_1)$, $b = (x_2 - x_1)$ och $c = x_1(y_2 - y_1) - y_1(x_2 - x_1)$.

I ändpunkten $t = 0$ har vi $\gamma(0) = x_1 + iy_1$ och om vi istället sätter in $t = 1$ i (4.2) får vi $\gamma(1) = x_2 + iy_2$. Det visar i sin tur att linjen verkligen passerar genom de två punkterna p och q .

Exempel 4.1.5. Vi betraktar för $t \in [0, \pi/2]$ funktionen

$$\gamma(t) = 2e^{it} = 2\cos t + 2i\sin t.$$

Vi ser att

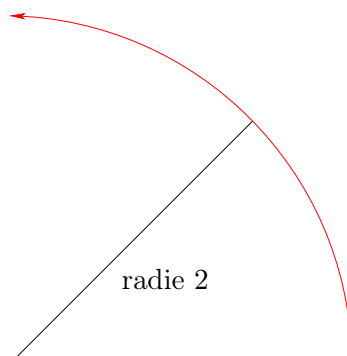
$$\gamma(0) = 2e^{i0} = 2$$

och att

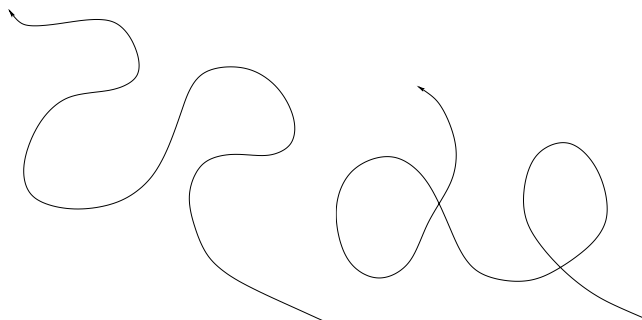
$$\gamma(\pi/2) = 2e^{i\pi/2} = 2i.$$

Vidare är avståndet från punkten 0 till en punkt $\gamma(t)$ på kurvan konstant lika med $|2e^{it}| = 2$ (alternativt kan man beräkna $\sqrt{4\cos^2 t + 4\sin^2 t} = 2$).

Vi inser nu att vi har fått en beskrivning av en kvartscirkel i komplexa talplanet med radie 2, och vi rör oss längs med kvartscirkeln moturs, från 2 till $2i$.



Figur 4.1: Parameterkurvan genomlöper en kvartscirkel.



Figur 4.2: Icke-skärande och skärande kurvor.

Vi är nu redo att ge en formell definition av vad vi ska mena med kurvor i komplexa talplanet.

Definition 4.1.6. En *parameterfunktion* är en funktion

$$\gamma : [a, b] \rightarrow \mathbb{C}, \quad t \mapsto x(t) + iy(t);$$

som är injektiv på $[a, b)$ och är sådan att funktionerna $x(t)$ och $y(t)$ är kontinuerliga och har kontinuerliga derivator $x'(t)$ och $y'(t)$.

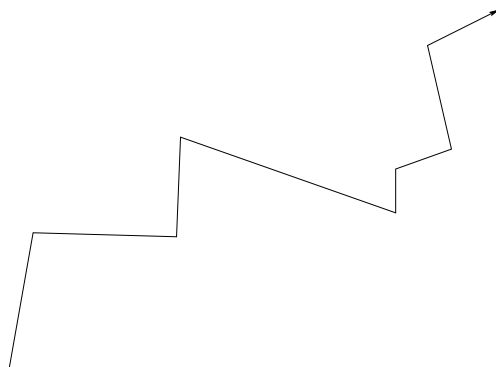
Kurvan Γ är mängden

$$\Gamma = \{\gamma(t) : t \in [a, b]\}$$

det vill säga, bildmängden till parameterfunktionen γ .

Låt oss studera denna definition lite närmare och försöka bygga upp lite intuition. Vi kräver att funktionen $\gamma(t)$ är injektiv, det vill säga, att $\gamma(t_1) = \gamma(t_2)$ medför att $t_1 = t_2$. Detta krav ser till att våra kurvor inte skär sig själva: vi är aldrig i samma punkt vid olika tider (see Figur 4.2)—utom möjligtvis när $t = b$. Om $\gamma(a) = \gamma(b)$ säger vi att kurvan Γ är *sluten*.

Kontinuiteten hos koordinaterna $x(t)$ och $y(t)$ gör att våra kurvor hänger ihop. Grafen till en kontinuerlig funktion har inga “språng”. I vår situation är både x -koordinaten och y -koordinaten kontinuerliga, så punkterna $\gamma(t_1)$ och $\gamma(t_2)$ ligger godtycklig nära varandra när tiderna t_1 och t_2 är tillräckligt nära varandra. Slutligen ger existensen av en derivata att våra objekt har tangenter i varje punkt. Det finns naturliga geometriska objekt som inte har tangenter,



Figur 4.3: Ett polygontåg.

men som ändå är relativt trevliga. Ett exempel är så kallade polygontåg (se Figur 4.3). Vi kommer att anse att ett polygontåg är uppbyggt av flera på varandra följande kurvor—nämligen linjesegment. Hela polygontåget kommer då ha tangenter i alla punkter utom i länkningspunkterna, men denna svårighet kommer vi inte att fästa så stor vikt vid.

Vi observerar att en kurva, som är ett geometriskt objekt, kan ges som bilden av flera olika parameterfunktioner. Vi betraktade i ett av exemplen linjen som förbinder två punkter p och q och angav då parameterfunktionen γ given av (4.2). Vi hade även kunnat betrakta parameterfunktionen

$$\tilde{\gamma}(t) = (1 - t)p + tq \quad t \in [0, 1].$$

Dess bild är samma geometriska objekt, nämligen linjen, men eftersom $\tilde{\gamma}(0) = q$ anger denna parameterfunktion att vi ska färdas åt andra hållet på linjen när t genomlöper intervallet $[0, 1]$.

4.2 Linjeintegraler

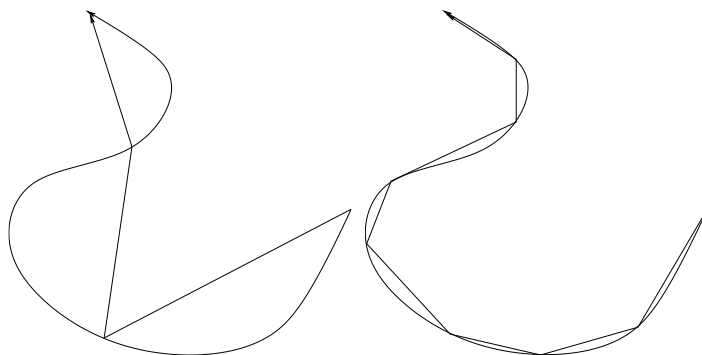
Det är naturligtvis av intresse att kunna mäta längden av en kurva i komplexa planet. Vi måste emellertid börja med att ange vad vi menar med längden av en kurva. Vissa enkla kurvors längder känner vi redan: vi vet hur man mäter längden på ett linjesegment.

Vi ska nu se att vi kan använda integraler för att ta fram ett längdbegrepp för kurvor på parameterform. Låt $q = a_1 + ia_2$ och $p = b_1 + ib_2$ vara två punkter i komplexa planet, och låt Λ vara det rätta linjesegmentet mellan dem. Från Pythagoras sats får vi att längden av Λ ges av

$$l(\Lambda) = |q - p| = \sqrt{(b_1 - a_1)^2 + (b_2 - a_2)^2}.$$

Betrakta nu en parameterfunktion $\gamma: [0, 1] \rightarrow \mathbb{C}$, $\gamma(t) = x(t) + iy(t)$, med bildmängd Γ . Vi låter $n \geq 1$ vara ett heltal och betraktar de $n + 1$ stycken punkterna $t_j \in [0, 1]$ givna av

$$t_j = \frac{j}{n}, \quad j = 0, 1, \dots, n.$$



Figur 4.4: Approximationer av en kurva med polygontåg med $n = 3$ (vänster) respektive $n = 8$ (höger).

Vi ser speciellt att $t_0 = 0$ och $t_n = 1$ för alla n . Vi förbinder nu varje par $\gamma(t_j)$ och $\gamma(t_{j+1})$ av punkter på kurvan med en rät linje (se Figur 4.4).

Detta ger oss ett polygontåg K_n som består av linjesegment Λ_j och som ligger nära den ursprungliga kurvan, i alla fall om vi väljer n , antalet punkter, stort. Längden av varje linjesegment Λ_j ges av

$$l(\Lambda_j) = |\gamma(t_{j+1}) - \gamma(t_j)| = \sqrt{(x(t_{j+1}) - x(t_j))^2 + (y(t_{j+1}) - y(t_j))^2}.$$

Den totala längden av polygontåget K_n ges av summan av längderna av linjesegmenten Λ_j :

$$l(K_n) = \sum_{j=0}^{n-1} |\gamma(t_{j+1}) - \gamma(t_j)| = \sum_{j=0}^{n-1} \sqrt{(x(t_{j+1}) - x(t_j))^2 + (y(t_{j+1}) - y(t_j))^2}.$$

Vi har ju antagit att funktionerna $x(t)$ och $y(t)$ är kontinuerliga och har kontinuerliga derivator $x'(t)$ och $y'(t)$, så vi kan tillämpa medelvärdessatsen A.1.5 från appendix. Vi får då för varje $j = 0, 1, \dots, n$ att

$$x(t_{j+1}) - x(t_j) = (t_{j+1} - t_j)x'(\tau_j)$$

för något $\tau_j \in [t_j, t_{j+1}]$, och

$$y(t_{j+1}) - y(t_j) = (t_{j+1} - t_j)y'(\sigma_j)$$

för något $\sigma_j \in [t_j, t_{j+1}]$. Vi noterar även att $t_{j+1} - t_j = 1/n$.

Detta ger oss att

$$l(K_n) = \frac{1}{n} \sum_{j=0}^{n-1} \sqrt{(x'(\tau_j))^2 + (y'(\sigma_j))^2}. \quad (4.3)$$

Vi undersöker nu vad som händer med detta uttryck när vi väljer n stort.

Eftersom $x'(t)$ och $y'(t)$ är kontinuerliga är även γ' kontinuerlig. Vi ser nu att (4.3) påminner rätt mycket om

$$\frac{1}{n} \sum_{j=0}^n \sqrt{(x'(\tau))^2 + (y'(\tau))^2} = \frac{1}{n} \sum_{j=0}^n |\gamma'(\tau_j)|,$$

som är den typ av Riemannsumma man brukar bilda när man ska beräkna integraler. Vi hänvisar läsaren till häftets appendix för en diskussion. Tyvärr är τ_j och σ_j inte nödvändigtvis lika, men när avståndet mellan t_{j+1} och t_j blir litet, vilket är fallet när $n \rightarrow \infty$, måste ju också τ_j och σ_j ligga nära varandra.

Det här intuitiva argumentet för att vi kan sätta $\tau_j = \sigma_j$ räcker tyvärr inte för att vi formellt ska kunna dra slutsatsen att (4.3) konvergerar mot $\int_0^1 |\gamma'(t)| dt$ när $n \rightarrow \infty$. Man kan ge ett ordentligt bevis för att så är fallet, men vi avstår från det här, utan nöjer oss med att ha gjort troligt att (4.3) går mot

$$\int_0^1 |\gamma'(t)| dt.$$

när $n \rightarrow \infty$.

Man tänker ju ofta på integralen av en funktion som arean (med tecken) under funktionsgrafen, men i det här fallet kan vi tolka integralen som något som har med *längden* av en kurva i planet att göra. Funktionerna $x(t)$ och $y(t)$ anger en position på kurvan som funktion av "tiden" t , vilket betyder att vi ska tolka $x'(t)$ och $y'(t)$ som hastigheterna i x -led och y -led i samma tid. När vi integrerar en hastighet över en tid får vi mycket riktigt en sträcka!

Våra överläggningar motiverar följande definition av längden av en kurva.

Definition 4.2.1. Låt $\gamma : [a, b] \rightarrow \mathbb{C}$ med $\gamma(t) = x(t) + iy(t)$ vara en parameterfunktion som ger upphov till en kurva Γ i komplexa planet. Då ges längden av Γ av

$$l(\Gamma) = \int_a^b |\gamma'(t)| dt = \int_a^b \sqrt{(x'(t))^2 + (y'(t))^2} dt$$

Eftersom vi kan betrakta ett komplexa tal $x + iy$ där x och y är reella tal som en punkt i planet (x, y) kan vi nu också definiera båglängden av en kurva i $\mathbb{R}^2$.

Definition 4.2.2. Låt $\gamma : [a, b] \rightarrow \mathbb{R}^2$ med $\gamma(t) = (x(t), y(t))$ vara en parameterfunktion som ger upphov till en kurva Γ i $\mathbb{R}^2$. Då ges längden av Γ av

$$l(\Gamma) = \int_a^b \sqrt{(x'(t))^2 + (y'(t))^2} dt. \quad (4.4)$$

Vi noterar här att vårt antagande att $x'(t)$ och $y'(t)$ är kontinuerliga medför att integralen i högerledet i (4.4) existerar, se Sats A.2.1.

Exempel 4.2.3. Vi beräknar längden av kvartscirkelbågen i Exempel 4.1.5, som vi parametriserade med

$$\gamma(t) = 2e^{it}, \quad t \in [0, \frac{\pi}{2}].$$

Vi har

$$|\gamma'(t)| = |2ie^{it}| = 2,$$

så båglängden blir

$$\int_0^{\frac{\pi}{2}} 2 dt = 2 \cdot \frac{\pi}{2} = \pi.$$

Exempel 4.2.4. Vi kan också beräkna båglanden av kvartscirkelbågen i $\mathbb{R}^2$, resultatet blir densamma men räkningarna något längre. Motsvarande kvartscirkelbåge i planet har parametrisering

$$\gamma(t) = (2 \cos t, 2 \sin t), \quad t \in [0, \frac{\pi}{2}].$$

Så

$$\gamma'(t) = (-2 \sin t, 2 \cos t)$$

och längden blir

$$\int_0^{\pi/2} \sqrt{(-2 \sin t)^2 + (2 \cos t)^2} dt = \int_0^{\pi/2} 2 dt = \pi.$$

Vårt resultat överensstämmer med vad vi hade fått om vi hade utnyttjat faktumet att omkretsen av en cirkel med radie r är $2\pi r$. I vårt fall är $r = 2$, och vi ska dela resultatet med 4.

Om man tänker efter inser man att det till synes finns några problem med vår definition av längd. En kurva hör ju inte till en entydig parameterfunktion—samma geometriska objekt kan beskrivas av olika parameterfunktioner. Vi kan ju genomlöpa halvcirkelbågen i våra exempel från vänster till höger eller från höger till vänster. Sedan kan vi ju translatera cirkelbågen också, eller rotera den, och då ändras koordinaterna för punkterna på cirkeln.

Det känns naturligt att kräva att längdbegreppet vi har infört inte ska bero av exakt vilken parameterfunktion vi har valt för att beskriva vår kurva, eller var i planet kurvan befinner sig. I nästa avsnitt ska vi diskutera dessa problem. Vi ska se att längden av en kurva inte beror av parametervälet och att längden dessutom är invariant under (oberoende av) automorfier av planet.

4.3 Variabelbyten och automorfier

Vi börjar med det första problemet med längdbegreppet vi tog upp ovan, nämligen att en och samma geometriska objekt Γ kan vara bildmängden till olika funktioner $\gamma(t)$.

Definition 4.3.1. Låt $\gamma : [\alpha, \beta] \rightarrow \mathbb{C}$ vara en parameterfunktion. Låt $h : [a, b] \rightarrow [\alpha, \beta]$ vara en kontinuerligt deriverbar funktion sådan att $h([a, b]) = [\alpha, \beta]$ och $h'(t)$ är nollskild och inte byter tecken. Då säger vi att den sammansätta funktionen $\gamma \circ h : [a, b] \rightarrow \mathbb{C}$ är en omparametrisering av γ .

Kravet att derivatan av h inte byter tecken medför att vi inte "stannar upp" när vi genomlöper intervallet $[a, b]$ enligt h —vi tolkar här $|h'(t)|$ som farten med vilken vi genomlöper intervallet $[a, b]$. Däremot är situationen $h(a) = \beta$ och $h(b) = \alpha$ tillåten. I detta fall kommer vi genomlöpa kurvan Γ åt motsatt håll jämfört med tidigare när vi följer parameterframställningen given av $\gamma \circ h$.

Exempel 4.3.2. : I Exempel 4.2.3 parametriserade vi en kvartscirkel med funktionen $\gamma : [0, \pi/2] \rightarrow \mathbb{C}$ given av

$$\gamma(t) = 2e^{it}.$$

En annan parametrisering av kvartscirkeln är $\gamma_2 : [0, \pi/4] \rightarrow \mathbb{C}$ given av

$$\gamma_2(t) = 2e^{i(\pi/2-2t)}.$$

Notera att $|\gamma_2(t)| = 2$ och att kurvan startar i $2e^{i(\pi/2-2\cdot 0)} = 2i$ och slutar i $2e^{i(\pi/2-2\cdot \pi/4)} = 2$. Vi färdas längs med samma kvartscirkel men rör oss åt andra hållet och med dubbla farten (vi hinner lika långt på halva tiden). Med kedjeregeln får vi

$$\gamma'_2(t) = (-2) \cdot 2e^{i(\pi/2-2t)}$$

och därmed

$$|\gamma'_2(t)| = 4.$$

Vi får längden $\int_0^{\pi/4} 4dt = \pi$, dvs samma som i exempel 4.2.3.

Detta gäller i allmänhet: det spelar inte någon roll precis vilken parameterframställning vi väljer när vi ska räkna ut längden av en given kurva. Ett nyckelresultat som vi använder i beviset är satsen om variabelbyten, Sats A.2.2, som tillåter oss att göra oss av med omparametriseringen vid integration.

Sats 4.3.3. *Låt Γ vara en kurva som hör till parameterfunktionen γ . Då är längden av kurvan, $l(\Gamma)$, oberoende av omparametriseringar.*

Bevis. Låt $\gamma : [\alpha, \beta] \rightarrow \mathbb{C}$ vara en parameterfunktion och $h : [a, b] \rightarrow [\alpha, \beta]$ en omparametrisering. Då är $\gamma \circ h : [a, b] \rightarrow \mathbb{C}$ också en parameterfunktion. Vi vill visa att längden av Γ med parametrisering γ och $\gamma \circ h$ är lika. Vi kan för enkelhetens skull anta att $h'(t) > 0$. Enligt kedjeregeln för derivering gäller att

$$(\gamma \circ h)'(t) = h'(t)(\gamma'(h(t))).$$

Längden av Γ med parametriseringen $\gamma \circ h$ ges av

$$\int_a^b |h'(t)| |\gamma'(h(t))| dt.$$

Detta uttryck påminner mycket om det som definierar $l(\Gamma)$, förutom att vi har en extra faktor $|h'(t)|$ i integranden och att vi evaluerar funktionerna γ i punkten $h(t)$ istället för t .

Vi genomför därför ett variabelbyte genom att sätta $s = h(t)$. Detta ger oss de nya gränserna $s_0 = h(a) = \alpha$ och $s_1 = h(b) = \beta$, samt $ds = h'(t)dt$. Eftersom $h'(t) > 0$ gäller $|h'(t)| = h'(t)$ och vi får från satsen om variabelbyten, A.2.2, att

$$\int_a^b |h'(t)| |\gamma'(h(t))| dt = \int_\alpha^\beta |\gamma'(s)| ds.$$

Integralen i högerledet är precis $l(\Gamma)$.

Beviset i fallet $h'(t) < 0$ är snarlikt och utelämnas därför. $\square$

Vår intuition säger oss att två kurvor som ser likadana ut har samma längd, oberoende om de befinner sig på olika ställen i planet eller är roterade.

Exempel 4.3.4. Betrakta kvartscirkeln Γ i exempel 4.2.3. Låt oss rotera den ett kvarts varv moturs följt av en translation 2 enheter i negativ imaginär-led. Låt oss kontrollera att denna kvartscirkel Γ_2 har samma båglängd som Γ . Vi kan parametrisera Γ_2 genom funktionen

$$\gamma_2 : [0, \pi/2] \rightarrow \mathbb{C}, \quad \gamma_2(t) = 2e^{(t+\pi/2)i} - 2i.$$

Så $\gamma_2'(t) = 2e^{(t+\pi/2)i}$ och $|\gamma_2'(t)| = 2$, därav

$$l(\Gamma_2) = \int_0^{\pi/2} 2dt = \frac{\pi}{2} \cdot 2 = \pi.$$

Detta gäller i allmänhet: längden av en kurva är invariant under automorfier av planet.

Sats 4.3.5. Låt f vara en automorfi av komplexa talplanet. Då gäller $l(\Gamma) = l(f(\Gamma))$ för varje kurva Γ .

Bevis. Vi påminner oss om att alla automorfier av planet är sammansatta av en rotation och en translation, se Sats 2.2.4. Det räcker därför att visa att längden av en kurva är invariant under dessa avbildningar.

Vi börjar med translationerna. Låt alltså betrakta automorfin f som ges av

$$f(z) = z + z_0, z_0 \in \mathbb{C}.$$

Om $\gamma : [0, 1] \rightarrow \mathbb{C}$ är en parameterframställning av Γ gäller att

$$(f \circ \gamma)(t) = \gamma(t) + z_0.$$

Då är

$$(f \circ \gamma)'(t) = \gamma'(t)$$

och vi ser då direkt att

$$\int_0^1 |(f \circ \gamma)'(t)| dt = \int_0^1 |\gamma'(t)| dt.$$

Alltså gäller $l(\Gamma) = l(f(\Gamma))$.

Vi betraktar nu en rotation f av planet moturs med en vinkel θ (se Exempel 2.2.3); det vill säga,

$$f(z) = e^{i\theta}z, \theta \in \mathbb{R}$$

eller

$$f(x + iy) = x \cos \theta - y \sin \theta + i(x \sin \theta + y \cos \theta), \quad \theta \in \mathbb{R}.$$

Vi får nu

$$(f \circ \gamma)(t) = e^{i\theta}\gamma(t),$$

så

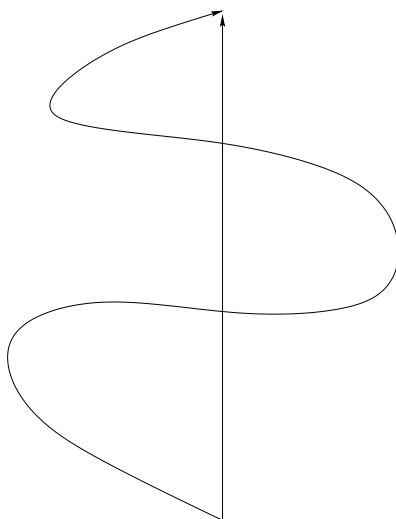
$$(f \circ \gamma)'(t) = e^{i\theta}\gamma'(t).$$

Därmed gäller

$$|(f \circ \gamma)'(t)| = |\gamma'(t)|$$

och precis som ovan fås $l(f(\Gamma)) = l(\Gamma)$.

□



Figur 4.5: En rät linje minimerar avståndet mellan två punkter.

Skriver vi parametriseringen i koordinatform, $\gamma(t) = x(t) + iy(t)$, blir uttrycken längre. Vi får

$$\begin{aligned} (f \circ \gamma)'(t) &= (x(t) \cos \theta - y(t) \sin \theta)' + i(x(t) \sin \theta + y(t) \cos \theta)' \\ &= x'(t) \cos \theta - y'(t) \sin \theta + i(x'(t) \sin \theta + y'(t) \cos \theta). \end{aligned}$$

och en räkning ger $|(f \circ \gamma)'(t)| = |\gamma'(t)|$.

4.4 Avstånd och geodeter

Vi avslutar det här kapitlet med att införa ett avståndsbegrepp i planet med hjälp av längder av kurvor. Betrakta för illustration punkterna 2 och $2i$. Vi har beräknat att längden av kvartscirkeln mellan dem är π . En linje mellan dem kan parametriseras med $\gamma(t) = (1-t)2 + 2it$ för $t \in [0, 1]$, så båglängden är

$$\int_0^1 \sqrt{(-2)^2 + 2^2} dt = 2\sqrt{2} < \pi,$$

dvs mindre än kvartscirkeln. Det finns många andra (oändligt många) kurvor som går genom dessa punkter, men ingen av dessa kommer ha kortare längd än den räta linjen mellan dem. Detta gäller allmänt: låt $p_1 = x_1 + iy_1$ och $p_2 = x_2 + iy_2$ vara två olika punkter i komplexa planet. Vi kan alltid förbinda p_1 och p_2 med hjälp av parameterfunktioner. Mer precist kan vi hitta en kurva Γ som ges av $\gamma : [a, b] \rightarrow \mathbb{C}$ med $\gamma(a) = p_1$ och $\gamma(b) = p_2$. Vi kan till exempel välja en rät linje som förbinder dem och parametrisera den, se (4.2). Det finns många fler möjligheter. De har ändå alla en sak gemensamt—deras båglängd kan inte vara mindre än den räta linjens.

Sats 4.4.1. *Låt p_1 och p_2 vara två punkter i komplexa planet, och låt Λ vara det räta linjesegmentet som förbinder dem. Då gäller*

$$l(\Lambda) \leq l(\Gamma)$$

för varje annan kurva Γ som förbinder p_1 och p_2 .

Bevis. Genom att rotera och translatera punkterna p_1 och p_2 får vi två nya punkter $\tilde{p}_1$ och $\tilde{p}_2$ med

$$\tilde{p}_1 = b_1 i; \quad \tilde{p}_2 = b_2 i.$$

Då båglängd är invariant under automorfier räcker det att visa att $l(\Lambda) \leq l(\Gamma)$ för varje parameterkurva som förbinder $\tilde{p}_1$ och $\tilde{p}_2$.

Eftersom längden av en kurva är invariant under parametrisering, kan vi anta att Γ är bilden av parameterfunktionen $\gamma(t) = x(t) + iy(t)$ definierad på $[0, 1]$ med $\gamma(0) = b_1 i$ och $\gamma(1) = b_2 i$. En parameterfunktion vars bild är linjesegmentet som förbinder $\tilde{p}_1$ och $\tilde{p}_2$ är funktionen

$$\gamma(t) = iy(t), \quad t \in [0, 1].$$

Vi noterar att

$$\sqrt{(x'(t))^2 + (y'(t))^2} \geq |y'(t)|, \quad t \in [0, 1].$$

Därmed gäller

$$l(\Gamma) = \int_0^1 \sqrt{(x'(t))^2 + (y'(t))^2} dt \geq \int_0^1 |y'(t)| dt = b_2 - b_1. \quad (4.5)$$

I sista steget har vi använt att båglängden av den räta linjen är oberoende av valet av parameterfunktion för Λ . Vi kan alltså speciellt välja den vanliga parameterfunktionen $\gamma(t) = ((1-t)b_1 + tb_2)i$ när vi ska beräkna $l(\Lambda)$ i sista steget.

Satsen är nu bevisad. □

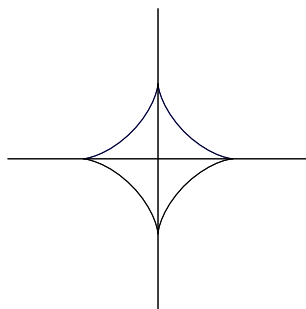
Anmärkning 4.4.2. Man kan visa att likhet inträffar i (4.5) precis när $x'(t) = 0$ för alla t —det vill säga, precis när Γ är en rät linje (för punkter på imaginära axeln blir den vertikal).

En rät linje mellan två punkter i komplexa talplanet har alltså kortare längd än varje annan kurva som förbinder dem. Längden av en rät linje mellan två punkter $p = p_1 + ip_2$ och $q = q_1 + iq_2$ ges ju av $|q - p| = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2}$, och detta är samma sak som avståndet mellan p och q . Vi skulle alltså kunna definiera avståndet mellan p_1 och p_2 som

$$\text{dist}(p_1, p_2) = l(\Lambda),$$

där Λ är den räta linjen som förbinder p_1 och p_2 . Precis vilken parameterfunktion vi väljer spelar ingen roll—längden av Λ är ju oberoende av detta val.

Den här tankegången ter sig kanske onödigt krånglig. Vi ska emellertid senare se att vi kan använda detta synsätt för att införa avståndsbegrepp i andra typer av geometrier: först hittar man en typ av kurva mellan två punkter som alltid har minimal längd och sedan definierar man avståndet mellan punkterna som längden av den speciella kurvan. Avståndsminimerande kurvor har ett eget namn; de brukar kallas *geodeter*.



Figur 4.6: Asteroidkurvan.

Övningar

Övning 4.1. Verifiera att punkterna på kurvan som hör till parameterfunktionen $\gamma: [0, 1] \rightarrow \mathbb{C}$ som ges av

$$\gamma(t) = x(t) + iy(t) = (1-t)x_1 + tx_2 + i((1-t)y_1 + ty_2))$$

uppfyller

$$(x_2 - x_1)y(t) - (y_2 - y_1)x(t) - y_1(x_2 - x_1) + x_1(y_2 - y_1) = 0, \quad (4.6)$$

det vill säga, att Γ är ett linjesegment.

Övning 4.2. Beräkna båglängden av kurvan Γ som hör till parameterfunktionen $\gamma: [0, 1] \rightarrow \mathbb{R}^2$ som ges av

$$\gamma(t) = (t^2, 1 + t^2).$$

Försök även att rita upp kurvan.

Övning 4.3. Bestäm a så att längden av sinuskurvan $y = \sin x$ då $0 \leq x \leq a$ är precis 1. Använd dator för att beräkna en integral.

Övning 4.4. Visa att omkretsen av en ellips med x -radie a och y -radie b ges av uttrycket $4a \int_0^{\pi/2} \sqrt{1 - (1 - \frac{b^2}{a^2}) \sin^2(\theta)} d\theta$. En sådan ellips kan parametriseras som

$$\begin{aligned} x &= a \cos \theta \\ y &= b \sin \theta \end{aligned}$$

Övning 4.5. Beräkna båglängden av den så kallade *asteroidkurvan* vilken ges av parameterfunktionen $\gamma: [0, 2\pi] \rightarrow \mathbb{R}^2$, där

$$\gamma(t) = (\cos^3 t, \sin^3 t), \quad t \in [0, 2\pi].$$

Ledning: Använd symmetrin hos kurvan för att förenkla beräkningarna. Försök sedan att hitta en trigonometrisk etta i integranden.

Övning 4.6. Beräkna båglängden av kurvan Γ som ges av parameterfunktionen $\gamma: [0, 1] \rightarrow \mathbb{C}$ med

$$\gamma(t) = \frac{e^{t(1+i)}}{\sqrt{2}}, \quad t \in [0, 1].$$

Övning 4.7. Antag att längden av en kurva Γ med parametrisering $\gamma : [0, 1] \rightarrow \mathbb{C}$ istället skulle ges av

$$l(\Gamma) = \int_0^1 \frac{|\gamma'(t)|}{\operatorname{Im} \gamma(t)} dt.$$

Beräkna längden av linjesegmentet mellan i and ie . Ges kortaste vägen mellan två punkter fortfarande av en rät linje?

5 Generella automorfier och Riemannsfären

5.1 Automorfier för sfären

I kapitel 2 definierade vi en automorfi som en bijektion som bevarade linjer, avstånd och vinklar. Vi såg också att automorfierna för planet $\mathbb{C}$ bestod av funktioner $f : \mathbb{C} \rightarrow \mathbb{C}$ på formen

$$f(z) = \alpha z + \beta$$

för komplexa tal α, β där $|\alpha| = 1$. Låt oss nu föreställa oss hur en automorfi för en sfär

$$S = \{(a, b, c) \in \mathbb{R}^3 \mid a^2 + b^2 + c^2 = 1\}$$

skulle kunna se ut. Som vi anmärkte i slutet av föregående kapitel kan vi definiera en linje på en sfär som en avståndsminimerande kurva, en geodet, som då är en kurva på S .

Därefter kan vi betrakta avståndet mellan två punkter på en sfär som längden av den kortaste vägen på sfären mellan dem, och med en vinkel menar vi vinkeln mellan två sådana linjer.

En funktion $f : S \rightarrow S$ som bevarar alla dessa egenskaper skulle då bevara alla mätbara egenskaper på sfären, och någon som levde på sfären skulle inte kunna märka någon skillnad före eller efter f appliceras (givet att de inte kunde se upp, bortom sfären). Automorfierna visar sig därför vara rotationer av sfären.

Definition 5.1.1. Automorfierna för sfären $\text{Aut}(S)$ består av alla rotationer av sfären. En funktion $f \in \text{Aut}(S)$ kan skrivas på formen

$$f(x, y, z) = (R_z(\phi) \circ R_y(\theta) \circ R_z(\psi))(x, y, z)$$

för några vinklar ϕ, θ, ψ (kallade Eulervinklar), där funktionerna R_y och R_z är givna av

$$\begin{aligned} R_z(\psi)(x, y, z) &= (x \cos \psi - y \sin \psi, x \sin \psi + y \cos \psi, z) \\ R_y(\theta)(x, y, z) &= (x \cos \theta + z \sin \theta, y, -x \sin \theta + z \cos \theta) \end{aligned}$$

för några vinklar ψ, θ . Alla funktioner på den formen ingår även i $\text{Aut}(S)$.

Vi kommer dock inte gå in på mer detalj varför alla rotationer kan skrivas på precis det här sättet.

5.2 Kartor

Vi har nu en beskrivning både av automorfierna för planet och för sfären, men varför är vi egentligen intresserade av dem? Än så länge har vi kommit fram till automorfierna genom vår kännedom om ytan, det vill säga, planet eller sfären.

Kännedom om $\mathbb{C}$ eller $S \rightarrow$ kännedom om $\text{Aut}(\mathbb{C})$ eller $\text{Aut}(S)$.

Målet i den här kursen är att definiera det hyperboliska planet $\mathcal{H}$. Till skillnad från planet eller sfären går det dock inte att beskriva $\mathcal{H}$ som en delmängd av $\mathbb{R}^3$. Detta innebär att vi måste beskriva det på något annat sätt. Nämligen att gå åt andra hållet, det vill säga

$$\text{kännedom om } \text{Aut}(\mathcal{H}) \rightarrow \text{kännedom om } \mathcal{H}.$$

Anledningen är att om vi känner till automorfierna för ett objekt, så går det att lista ut hur objektet är format. I kapitel 8 kommer vi se precis hur det här går till. Ett problem förekommer dock redan nu, en automorfi $f \in \text{Aut}(\mathcal{H})$ skulle vara definierad som en funktion

$$f : \mathcal{H} \rightarrow \mathcal{H}.$$

Men hur ska vi kunna definiera sådana funktioner om vi inte vet vad $\mathcal{H}$ är i första taget? Ett sätt runt detta är att använda oss av *kartor*.

Låt oss återgå till automorfier av sfären som är funktioner från $S \rightarrow S$. Om vi hade en karta över sfären, det vill säga någon bijektion

$$K : D \rightarrow S$$

för någon delmängd $D \subset \mathbb{C}$, hade vi kunnat se automorfierna som funktioner från mängden D till D istället för mängden S till S . Det vill säga, vi skulle kunna definiera en ekvivalent mängd av automorfier $\text{Aut}(D)$ genom att sätta

$$\text{Aut}(D) = \{K^{-1} \circ f \circ K \mid f \in \text{Aut}(S)\}.$$

Vad ska vi nu välja för kartfunktion K ? Vi vill att det ska vara en bijektion. En annan viktig egenskap som den bör ha är att små områden på kartan bör se ut som de gör i verkligheten. Ett annat sätt att säga det här är att den ska bevara vinklar på små avstånd.

Ett exempel på en karta är den som vi hade i kapitel 1, nämligen Mercator-projektionen.

Som vi såg då är Mercatorprojektionen dock inte en bijektion. Däremot är den vinkelbevarande på små avstånd som vi ville. För att få en karta som är en bijektion kommer vi använda en oändligt stor karta som vi kallar för *Riemannsfären*.

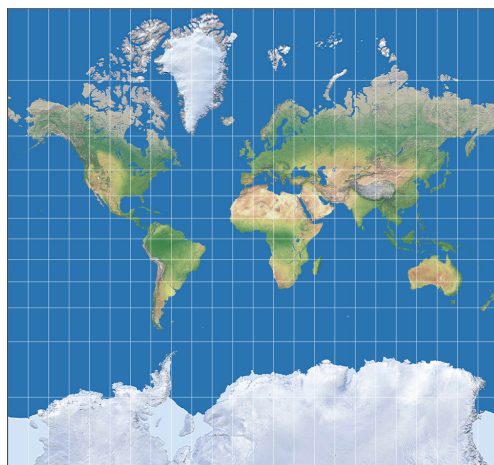
5.3 Riemannsfären

Definition 5.3.1. Riemannsfären betecknas $\hat{\mathbb{C}}$ och definieras som

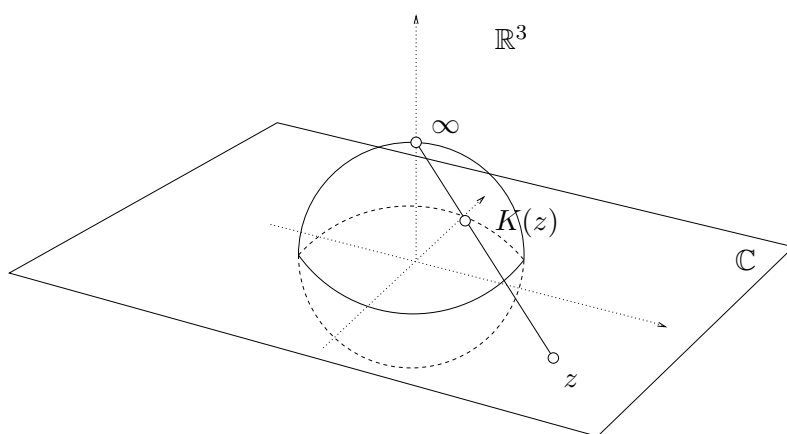
$$\hat{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$$

för något element vi kallar $\infty \notin \mathbb{C}$ som inte befinner sig i de komplexa talen.

Riemannsfären är alltså bara de komplexa talen tillsammans med en till punkt som vi tänker ligger oändligt långt bort från origo. Vi ska nu se hur Riemannsfären kan betraktas som en faktisk sfär.



Figur 5.1: Samma figur som figur 1.10

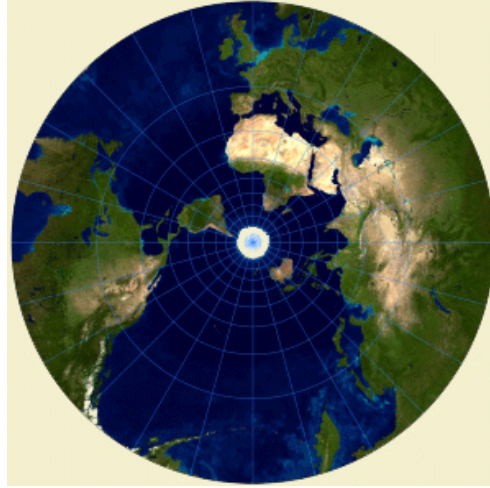


Figur 5.2: Stereografisk projektion avbildar $\mathbb{C}$ till sfären minus nordpolen.

Vi börjar med att identifiera det komplexa planet $\mathbb{C}$ med xy -planet i tredimensionella rummet $\mathbb{R}^3$. Betrakta nu en sfär S i $\mathbb{R}^3$ med radie 1 och centrum i origo. Vi definierar nu en bijektion $K : \widehat{\mathbb{C}} \rightarrow S$ som följer: för varje $z = x + iy \in \mathbb{C}$, drag den unika linjen i $\mathbb{R}^3$ genom $(0, 0, 1)$ och $z = (x, y, 0)$ i $\mathbb{R}^3$. Linjen kommer då skära sfären S i precis en punkt, vi låter denna punkt vara $K(z)$ (Se figur 5.2). Punkterna i $\mathbb{C}$ kommer då avbildas på alla punkter förutom $(0, 0, 1) \in S$. Vi låter därför $K(\infty) = (0, 0, 1)$ för att få en bijektion. Funktionen K som vi definierade här brukar kallas för en *stereografisk projektion*.

Vad är det vi har gjort här? Det vi säger är att om man börjar med en sfär och tar bort en punkt, så får man något som (om man plattar till det, och tänjer ut det) ser ut som ett oändligt stort plan. Vi identifierar sedan detta plan med det komplexa talplanet; punkterna på sfären kan sedan tänkas på som de vanliga komplexa talen men med en extra punkt ”i oändligheten”.

Tittar man på Riemannsfären ser inte punkten i oändligheten annorlunda ut än någon annan punkt. Från bilden kan man se att för en punkt $z \in \mathbb{C}$ gäller att ju större $|z|$ är, desto närmare ligger punkten till ∞ . Med andra ord: om



Figur 5.3: En del av en dtereografisk projektion av jorden.
 (<https://creativecommons.org/licenses/by-nc-nd/4.0/>), via
https://www.atractor.pt/mat/loxodromica/projeccao_estereografica-_en.html

en punkt rör sig *bort* från origo, oavsett i vilken riktning, så kommer den röra sig mot oändligheten. Det här kan först kännas konstigt om man använder sin intuition från vanliga komplexa planet $\mathbb{C}$. Det vi säger är ju att om en punkt rör sig uppåt längs imaginära axeln, och en annan punkt rör sig bort i rakt motsatt riktning, så kommer dessa punkter till slut att istället röra sig mot varandra igen! Det kan vara instruktivt att jämföra detta med situationen för punkten 0, som ligger rakt mittemot ∞ på $\hat{\mathbb{C}}$. För 0 gäller ju att ju *mindre* $|z|$ är, desto närmare ligger z och 0, alltså precis motsatsen mot vad som gäller för ∞ .

Ett annat sätt att se Riemannsfären på är som en oändligt stor karta över en sfär tillsammans med en extra punkt som beskriver beteendet i nordpolen. Figur 5.3. Det går även att definiera K lite mer explicit. Nämligen genom att sätta

$$K(z) = \left(\operatorname{Re} \left(\frac{2z}{|z|^2 + 1} \right), \operatorname{Im} \left(\frac{2z}{|z|^2 + 1} \right), \frac{|z|^2 - 1}{|z|^2 + 1} \right) \in \mathbb{R}^3$$

för $z \in \mathbb{C}$, och $K(\infty) = (0, 0, 1)$. Vi noterar att

$$\begin{aligned} & \operatorname{Re} \left(\frac{2z}{|z|^2 + 1} \right)^2 + \operatorname{Im} \left(\frac{2z}{|z|^2 + 1} \right)^2 + \left(\frac{|z|^2 - 1}{|z|^2 + 1} \right)^2 \\ &= \frac{4|z|^2}{(|z|^2 + 1)^2} + \frac{(|z|^2 - 1)^2}{(|z|^2 + 1)^2} = 1 \end{aligned}$$

så $K(z) \in S$ som vi ville. Vi lämnar det till en övning (se övning 5.6) att visa att linjen som går genom z och $(0, 0, 1)$ skär punkten

$$\left(\operatorname{Re} \left(\frac{2z}{|z|^2 + 1} \right), \operatorname{Im} \left(\frac{2z}{|z|^2 + 1} \right), \frac{|z|^2 - 1}{|z|^2 + 1} \right).$$

Eftersom Riemannsfären identifierar sfären med de komplexa talen och oändligheten så kan vi utföra algebraiska operationer av punkter på Riemannsfären.

För två punkter $z, w \in \widehat{\mathbb{C}} \setminus \{\infty\}$ definierar vi addition, subtraktion, multiplikation och division precis som i kapitel 3. Annars gör vi följande definition.

Definition 5.3.2. Om $z \in \widehat{\mathbb{C}} \setminus \{\infty\}$ så definierar vi

$$\begin{aligned}\infty + z &= \infty, \\ \frac{z}{\infty} &= 0,\end{aligned}$$

om dessutom $z \neq 0$, definierar vi

$$\begin{aligned}\frac{z}{0} &= \infty, \\ \infty \cdot z &= z \cdot \infty = \infty.\end{aligned}$$

Vi tillåter alltså division med 0 om vi är i Riemannsfären! Vi lämnar dock vissa uttryck odefinierade, så som $\infty + \infty$, $\infty - \infty$, $\frac{0}{0}$, $\frac{\infty}{\infty}$.

Vi kan nu definiera mängden automorfier av Riemannsfären som

$$\text{Aut}(\widehat{\mathbb{C}}) = \{K^{-1} \circ f \circ K \mid f \in \text{Aut}(S)\}.$$

Genom att beräkna sammansättningen $K^{-1} \circ f \circ K$ är det möjligt att se att automorfierna av Riemannsfären kan skrivas på en enkel form.

Sats 5.3.3. En funktion $f : \widehat{\mathbb{C}} \rightarrow \widehat{\mathbb{C}}$ är en automorfi $f \in \text{Aut}(\widehat{\mathbb{C}})$ om och endast om den kan skrivas på formen

$$f(z) = \frac{\alpha z - \bar{\beta}}{\beta z + \bar{\alpha}}$$

för komplexa tal $\alpha, \beta \in \mathbb{C}$ som uppfyller $|\alpha|^2 + |\beta|^2 = 1$, där vi definierar

$$f(\infty) = \frac{\alpha}{\beta}$$

Beräkningen är dock lång och vi lämnar det till en övning för den intresserade.

5.4 Automorfier

Vi har än så länge definierat automorfierna för planet $\text{Aut}(\mathbb{C})$ och för sfären $\text{Aut}(\widehat{\mathbb{C}})$. Automorfierna för planet, $\text{Aut}(\mathbb{C})$, såg vi i kapitel 2 att de bevarade alla mätbara egenskaper så som vinklar och avstånd. Detta stämmer även för automorfierna av $\text{Aut}(\widehat{\mathbb{C}})$ men bara om vi tolkar $\widehat{\mathbb{C}}$ som en sfär.

Vi ska nu definiera vad en allmän mängd av automorfier är så att det ska vara möjligt att tolka den underliggande mängden som en yta.

Definition 5.4.1. Låt $D \subset \widehat{\mathbb{C}}$ vara ett område i $\widehat{\mathbb{C}}$. Vi säger att en mängd M av bijektioner från D till D är en mängd *automorfier* om varje bijektion $f \in M$ uppfyller att

- (i) f är analytisk i $D \setminus \{\infty\}$,

dessutom vill vi att

- (ii) för varje par av punkter $z, w \in \widehat{\mathbb{C}}$ finns det en funktion $f \in M$ så att $f(z) = w$,
- (iii) för varje punkt $z \in D \setminus \{\infty\}$ och faktor $\alpha \in \mathbb{C}$, finns det exakt en funktion $f \in M$ så att $f(z) = z$ och $f'(z) = \alpha$ om $|\alpha| = 1$, och ingen sådan funktion om $|\alpha| \neq 1$.

Slutligen vill vi även att M ska vara en *grupp*. Detta innebär att om $f, g \in M$ så är

- (iv) $f \circ g \in M$,
- (v) $f^{-1} \in M$,
- (vi) identitetsfunktionen $\text{id}(z) = z$ ingår också i M .

Anmärkning 5.4.2. Anledningen till att vi utesluter punkten ∞ från (i) och (iii) är för vi inte har definierat vad det innebär för en funktion att vara analytisk där. Det går att definiera vad det innebär, men det visar sig att om vi lägger till ett sådant villkor så kommer vi aldrig utesluta några extra funktioner, så vi kan strunta i det med gott samvete.

Låt oss nu motivera alla ovanstående punkter. Villkor (i) säger att alla automorfier ska vara analytiska¹¹. Tanken är att vi vill att D ska vara en bra representation för vår ytan i ”små områden”. Om vi exempelvis tänker oss en karta över världen vill vi att om vi zoomar in tillräckligt mycket så ska världen där se ut precis som den gör på kartan (förutom möjligtvis roterad, skalad och translaterad). När vi sedan applicerar en automorfi bör denna egenskap fortfarande vara kvar. Vi vill att en automorfi endast ska se ut som en rotation, följt av en skalning, följt av en translation i små områden. Detta är vad det innebär att den ska vara analytisk.

Villkor (ii) säger att automorfierna låter oss relatera vilken punkt till vilken annan punkt som helst. Detta är viktigt för att vi ska ha tillräckligt många automorfier för att kunna beskriva geometrin av hela ytan.

En punkt så att $f(z) = z$ kallas för en fixpunkt. Villkor (iii) säger att för en godtycklig punkt bör det finnas en automorfi som ser ut som en rotation kring punkten. Dock bör det inte finnas flera olika automorfier som alla ser ut som samma rotation, då detta skulle ge motstridig information. Villkor (iii) säger också att det inte kan finnas någon automorfi som ser ut som en skalning kring en punkt, då detta också skulle vara motsägelse.

Om vi kommer ihåg att en automorfi bör bevara alla mätbara egenskaper på den underliggande ytan så är det förståeligt att sammansättningen av två automorfier också bör vara en automorfi. Även inversen av en automorfi bör vara en automorfi och funktionen som inte gör någonting är en automorfi. Detta är vad (iv), (v) och (vi) säger.

¹¹En analytisk bijektion kallas även för en *konform avbildning*. I kapitel 6 ska vi diskutera dem i större detalj.

Anmärkning 5.4.3. En mängd bijektioner som uppfyller egenskaperna (iv), (v) och (vi) kallas för en *grupp*. Vi kommer därför använda begreppet *automorfigrupp* härnäst istället för en mängd av automorfier. Den som vill lära sig mer om grupper kan ta en titt på kompendiet för årgången 2013/2014 som heter "Grupper, mönster och symmetrier", eller för årgången 2006/2007, som heter "Gruppteori".

Sats 5.4.4. *Mängderna $\text{Aut}(\mathbb{C})$ och $\text{Aut}(\widehat{\mathbb{C}})$ är automorfigrupper enligt definition 5.4.1.*

Innan vi tar och bevisar detta ska vi definiera så kallade *Möbiusavbildningar*.

5.5 Möbiusavbildningar

Från Sats 5.3.3 hade vi att alla funktioner i $\text{Aut}(\widehat{\mathbb{C}})$ var på formen

$$f(z) = \frac{\alpha z - \bar{\beta}}{\beta z + \bar{\alpha}}$$

för tal $\alpha, \beta \in \mathbb{C}$, där $|\alpha|^2 + |\beta|^2 = 1$ och där vi även hade konventionen att

$$f(\infty) = \frac{\alpha}{\beta}.$$

Funktionerna i $\text{Aut}(\mathbb{C})$ var däremot på formen

$$f(z) = \alpha z + \beta$$

där $|\alpha| = 1$. Vi ska nu studera en klass av funktioner som generaliserar båda dessa typer av funktioner.

Definition 5.5.1. En *Möbiusavbildning* är en funktion $f : \widehat{\mathbb{C}} \rightarrow \widehat{\mathbb{C}}$ på formen

$$f(z) = \frac{az + b}{cz + d}$$

för komplexa tal $a, b, c, d \in \mathbb{C}$ som uppfyller $ad - bc = 1$, där vi definierar

$$f(\infty) = \frac{a}{c}.$$

Mängden av alla Möbiusavbildningar betecknar vi $\text{Möb}(\widehat{\mathbb{C}})$.

Anmärkning 5.5.2. Vi noterar att funktionerna i $\text{Aut}(\mathbb{C})$ och $\text{Aut}(\widehat{\mathbb{C}})$ är Möbiusavbildningar. En funktion i $\text{Aut}(\mathbb{C})$ kan skrivas som

$$f(z) = e^{i\theta}z + \beta = \frac{e^{i\frac{\theta}{2}}z + e^{-i\frac{\theta}{2}}\beta}{0 \cdot z + e^{-i\frac{\theta}{2}}}$$

medan en funktion i $\text{Aut}(\widehat{\mathbb{C}})$ ses direkt vara en Möbiusavbildning då den redan är på rätt form.

Anmärkning 5.5.3. Bokstäverna a, b, c och d kommer alltid att användas för dessa koefficienter när vi diskuterar Möbiusavbildningar. Ibland kommer vi att slarva och bara tala om koefficienterna a, b, c och d för en Möbiusavbildning utan att varje gång specificera att det är dessa som åsyftas.

Anmärkning 5.5.4. Villkoret att $ad - bc = 1$ kan se konstigt ut. Vi påpekar dock först att om $ad - bc = 0$ så hade

$$\frac{az + b}{cz + d} = \frac{\frac{a}{c}(cz + d) - \frac{a}{c}d + b}{cz + d} = \frac{a}{c} + \frac{b - \frac{a}{c}d}{cz + d} = \frac{a}{c}.$$

Därmed utesluter villkoret $ad - bc \neq 0$ konstantfunktionerna. Det visar sig sedan att vi kan anta att $ad - bc = 1$ och få samma mängd av funktioner, detta visar vi i nästa sats.

Hjälpssats 5.5.5. Låt f vara en avbildning på formen

$$f(z) = \frac{az + b}{cz + d},$$

där $ad - bc \neq 0$. Då existerar ett komplext tal α så att

$$f(z) = \frac{\alpha \cdot (az + b)}{\alpha \cdot (cz + d)}$$

är en Möbiusavbildning, d.v.s. $\alpha^2(ad - bc) = 1$.

Bevis. Låt $ad - bc = \beta$. Om vi kunde hitta ett tal α sådant att $\alpha^2 = 1/\beta$, så skulle $\alpha^2(ad - bc) = 1$. (Notera att vi måste dela med β , så vi använder verkligen att β är nollskilt.) Men detta går alltid i komplexa planet. Om vi skriver $1/\beta$ på polär form,

$$1/\beta = re^{i\theta},$$

kan vi låta $\alpha = \sqrt{r}e^{i\theta/2}$. Eftersom r är reellt och positivt har det en reell och positiv kvadratrots. För detta val av α är

$$\alpha^2 = (\sqrt{r}e^{i\theta/2})^2 = (\sqrt{r})^2(e^{i\theta/2})^2 = re^{i\theta} = 1/\beta.$$

Beviset är klart. □

Anmärkning 5.5.6. Det vi egentligen visat är att alla tal i $\mathbb{C}$ har en kvadratrots.¹² Men om vi till exempel arbetade med reella värden på a, b, c och d gäller det inte längre att vi får samma mängd av funktioner om vi kräver $ad - bc = 1$ respektive $ad - bc \neq 0$: om vi till exempel har $ad - bc = -1$, så skulle vi vilja förlänga bråket med $\sqrt{-1} = i$ för att få samma funktion fast med $ad - bc = 1$. Men det finns ingen reell konstant vi kan förlänga med för att få $ad - bc$ till 1.

Sats 5.5.7. Mängden av funktioner $\text{Möb}(\mathbb{C})$ bildar en grupp, det vill säga, har följande tre egenskaper:

- (i) Identitetsavbildningen id som avbildar varje punkt på sig själv ligger i $\text{Möb}(\mathbb{C})$;

¹²Egentligen två stycken, eftersom om $\alpha^2 = \beta$ är även $(-\alpha)^2 = \beta$.

- (ii) Om f och g ligger i $\text{Möb}(\mathbb{C})$ gör även $f \circ g$ det, vilket alltså säger att $f(g(z))$ är en Möbiusavbildning;
- (iii) För varje $f \in \text{Möb}(\mathbb{C})$ finns en Möbiusfunktion f^{-1} , sådan att $f \circ f^{-1} = f^{-1} \circ f = \text{id}$, det vill säga $f(f^{-1}(z)) = f^{-1}(f(z)) = z$.

Bevis. (i) Identitetsavbildningen får vi t.ex. genom att välja $a = d = 1$, $b = c = 0$. För detta val av a, b, c och d är självklart också $ad - bc = 1$.

- (ii) Låt oss välja två Möbiustransformationer

$$f(z) = \frac{az + b}{cz + d} \quad \text{och} \quad g(z) = \frac{Az + B}{Cz + D}.$$

Då är

$$\begin{aligned} f(g(z)) &= \frac{a \cdot \frac{Az+B}{Cz+D} + b}{c \cdot \frac{Az+B}{Cz+D} + d} = \frac{a(Az+B) + b(Cz+D)}{c(Az+B) + d(Cz+D)} \\ &= \frac{(aA + bC)z + (aB + bD)}{(cA + dC)z + (cB + dD)} \end{aligned}$$

vilket liknar en Möbiustransformation. Det enda som återstår är att kontrollera villkoret $(aA + bC)(cB + dD) - (aB + bD)(cA + dC) = 1$. Men nu händer något förbluffande:

$$(aA + bC)(cB + dD) - (aB + bD)(cA + dC) = (ad - bc)(AD - BC),$$

som läsaren gärna får kontrollera med papper och penna. Båda faktorerna till höger är enligt antagande lika med ett.

- (iii) Låt

$$f(z) = \frac{az + b}{cz + d}.$$

Vi hävdar nu att vi kan låta

$$f^{-1}(z) = \frac{dz - b}{-cz + a},$$

och med detta val kommer $f(f^{-1}(z)) = z$. Detta är samma räkning som i (ii):

$$f(f^{-1}(z)) = \frac{a \cdot \frac{dz-b}{-cz+a} + b}{c \cdot \frac{dz-b}{-cz+a} + d} = \frac{(ad - bc)z + (-ab + ba)}{(cd - dc)z + (-cb + ad)} = \frac{z}{1} = z.$$

Räkningen för $f^{-1} \circ f$ är likadan. Villkoret på $ad - bc$ är också uppenbart uppfyllt för f^{-1} , vilket avslutar beviset.

□

Vi kommer härnäst referera till mängden av Möbiusavbildningar $\text{Möb}(\mathbb{C})$ som *Möbiusgruppen*.

Anmärkning 5.5.8. En konsekvens av Sats 5.5.7 är att alla Möbiusavbildningar är inverterbara, eller med andra ord bijektioner.

Vi avslutar kapitlet med att bevisa att $\text{Aut}(\widehat{\mathbb{C}})$ är en automorfigrupp.

(Del av) bevis av Sats 5.4.4. Vi lämnar beviset att $\text{Aut}(\mathbb{C})$ är en automorfigrupp till en övning. Vi betraktar nu $\text{Aut}(\widehat{\mathbb{C}})$, vi börjar med att visa att funktionerna bildar en grupp.

(vi) Låt $f(z) = \frac{\alpha_1 z - \bar{\beta}_1}{\beta_1 z + \bar{\alpha}_1}$, och $g(z) = \frac{\alpha_2 z - \bar{\beta}_2}{\beta_2 z + \bar{\alpha}_2}$ där

$$|\alpha_1|^2 + |\beta_1|^2 = |\alpha_2|^2 + |\beta_2|^2 = 1.$$

Samma beräkning som i beviset av Sats 5.5.7 ger oss att

$$f(g(z)) = \frac{(\alpha_1 \alpha_2 - \bar{\beta}_1 \bar{\beta}_2)z - \alpha_1 \bar{\beta}_2 - \bar{\beta}_1 \bar{\alpha}_2}{(\beta_1 \alpha_2 + \bar{\alpha}_1 \bar{\beta}_2)z - \beta_1 \bar{\beta}_2 + \bar{\alpha}_1 \bar{\alpha}_2}$$

vilket är på rätt form.

(v) Beräkningen i Sats 5.5.7 ger oss att om $f(z) = \frac{\alpha z - \bar{\beta}}{\beta z + \bar{\alpha}}$ där $|\alpha|^2 + |\beta|^2 = 1$ så är

$$f^{-1}(z) = \frac{\bar{\alpha}z + \bar{\beta}}{-\beta z + \alpha}$$

vilket också är på rätt form.

(vi) Identitetsfunktionen ges genom att sätta $\alpha = 1$ och $\beta = 0$.

Vi går nu vidare till de andra egenskaperna.

- (i) Att alla funktioner $f \in \widehat{\mathbb{C}}$ är analytiska följer från Sats 3.8.2.
- (ii) Vi börjar med att visa att för alla punkter $w \in \widehat{\mathbb{C}}$ finns det någon automorfi $f_w \in \text{Aut}(\widehat{\mathbb{C}})$ så att

$$f_w(0) = w.$$

Skriver vi f_w på formen $\frac{\alpha z - \bar{\beta}}{\beta z + \bar{\alpha}}$ säger detta att

$$-\frac{\bar{\beta}}{\bar{\alpha}} = w.$$

Om $w = \infty$ väljer vi $\alpha = 0$ och $\beta = 1$. Annars väljer vi $\alpha = \sqrt{\frac{1}{1+|w|^2}}$ och $\beta = -\bar{w}\alpha$ vilket ger

$$|\alpha|^2 + |\beta|^2 = |\alpha|^2 + |-\bar{w}\alpha|^2 = |\alpha|^2 (1 + |w|^2) = 1$$

medans $-\frac{\bar{\beta}}{\bar{\alpha}} = w$ som vi önskade. Givet två punkter $z, w \in \widehat{\mathbb{C}}$ betraktar vi nu funktionen

$$f_w \circ f_z^{-1}$$

som ingår i $\text{Aut}(\widehat{\mathbb{C}})$ eftersom vi visade att det var en grupp. Vi har att

$$(f_w \circ f_z^{-1})(z) = f_w(f_z^{-1}(z)) = f_w(0) = w$$

som vi önskade.

(iii) Låt $z \in \mathbb{C}$ och $\alpha \in \mathbb{C}$ vara givna. Vi börjar med att betrakta mängderna

$$S_z = \{f \in \text{Aut}(\widehat{\mathbb{C}}) \mid f(z) = z, f'(z) = \alpha\}$$

och

$$S_0 = \{g \in \text{Aut}(\widehat{\mathbb{C}}) \mid g(0) = 0, g'(0) = \alpha\}.$$

Vi vill visa att S_z har ett element om $|\alpha| = 1$ men inget element om $|\alpha| \neq 1$. Vi börjar med att visa att S_z har lika många element som S_0 , och kan därefter anta att $z = 0$. För att visa detta visar vi att det finns en bijektion mellan mängderna. Definiera funktionen $F : S_z \rightarrow \text{Aut}(\widehat{\mathbb{C}})$ genom att sätta

$$F(f) = f_z^{-1} \circ f \circ f_z$$

där f_z är som vi definierade i (ii). Eftersom $\text{Aut}(\widehat{\mathbb{C}})$ är en grupp, så måste $f_z^{-1} \circ f \circ f_z \in \text{Aut}(\widehat{\mathbb{C}})$. Vi visar först att $F(f) \in S_0$. Vi har att

$$f_z^{-1}(f(f_z(0))) = f_z^{-1}(f(z)) = f_z^{-1}(z) = 0$$

som vi ville. Genom deriveringsreglerna från Sats 3.8.2 räknar vi ut derivatan till att vara

$$\begin{aligned} (F(f))'(0) &= (f_z^{-1} \circ f \circ f_z)'(0) = (f_z^{-1})'(f(f_z(0))) \cdot f'(f_z(0)) \cdot f_z'(0) \\ &= \frac{1}{f_z'(f_z^{-1}(f(f_z(0))))} f'(z) f_z'(0) = \frac{1}{f_z'(0)} f'(z) f_z'(0) \\ &= f'(z) = \alpha \end{aligned}$$

så $F : S_z \rightarrow S_0$. På samma sätt kan vi definiera $G : S_0 \rightarrow S_z$ genom att sätta

$$G(g) = f_z \circ g \circ f_z^{-1}.$$

Vi har nu att

$$G(F(f)) = f_z \circ (f_z^{-1} \circ f \circ f_z) \circ f_z^{-1} = f$$

medans

$$F(G(g)) = f_z^{-1} \circ (f_z \circ g \circ f_z^{-1}) \circ f_z = g$$

så $G = F^{-1}$ och F måste vara en bijektion. Vi kan därmed anta att $z = 0$. Låt $f \in S_0$. Vi skriver $f(z) = \frac{az - \bar{b}}{bz + \bar{a}}$, men då $f(0) = 0$ har vi att

$$\frac{-\bar{b}}{\bar{a}} = 0$$

så $b = 0$ och $|a| = 1$. Funktionen f är därmed på formen

$$f(z) = \frac{a}{\bar{a}} z,$$

och $f'(0) = \frac{a}{\bar{a}}$. Vi ser nu att $|f'(0)| = |\alpha| = 1$ vilket betyder att om $|\alpha| \neq 1$ så finns det ingen funktion i S_0 . Om nu $|\alpha| = 1$ ser vi att det kan finnas som mest en funktion $f \in S_0$, nämligen funktionen $f(z) = \alpha z$. Det återstår att visa att funktionen $f(z) = \alpha z$ ingår i S_0 om $|\alpha| = 1$. Detta kan vi se genom att skriva $\alpha = e^{i\theta}$ och välja $a = e^{\frac{1}{2}i\theta}$ eftersom då kommer

$$\frac{a}{\bar{a}} z = \frac{e^{\frac{1}{2}i\theta}}{e^{-\frac{1}{2}i\theta}} z = e^{i\theta} z = f(z). \quad \square$$

Övningar

Övning 5.1. Antag att en Möbiusavbildning f har egenskapen att $f(0) = 1$, $f(1) = \infty$ och $f(\infty) = 0$. Hur kan f se ut? Skriv ditt svar som

$$f(z) = \frac{az + b}{cz + d},$$

med $ad - bc = 1$.

Övning 5.2. Visa direkt att mängden av funktioner $\mathbb{C} \rightarrow \mathbb{C}$ på formen $f(z) = \alpha z + \beta$, där α och β är komplexa konstanter och $|\alpha| = 1$, bildar en grupp. (Kom ihåg att det är exakt dessa funktioner som är det euklidiska planets automorfier, men du bör inte använda detta i ditt bevis.)

Övning 5.3. Visa att $\text{Aut}(\mathbb{C})$ är en automorfigrupp, det vill säga, uppfyller egenskaperna givna i definition 5.4.1.

Övning 5.4. Visa att varje Möbiusavbildning f alltid har minst en fixpunkt på Riemannsfären, det vill säga en punkt $z \in \hat{\mathbb{C}}$ sådan att $f(z) = z$. (Du kommer nog att behöva använda *algebrans fundamentalsats*, som säger att varje icke-konstant polynom med komplexa koefficienter har minst en rot i $\mathbb{C}$.)

Övning 5.5. Ge ett exempel på en Möbiusavbildning med *exakt* en fixpunkt.

Övning 5.6. Låt z vara en punkt i komplexa planet. Visa att kartfunktionen K avbildar punkten z till

$$\left(\text{Re} \left(\frac{2z}{|z|^2 + 1} \right), \text{Im} \left(\frac{2z}{|z|^2 + 1} \right), \frac{|z|^2 - 1}{|z|^2 + 1} \right)$$

på sfären, det vill säga, att linjen som går genom z och $(0, 0, 1)$ skär sfären just i den punkten.

Övning 5.7. Visa att varje automorfi av sfären $f \in \text{Aut}(\hat{\mathbb{C}})$ har exakt två fixpunkter, det vill säga, punkter $z \in \hat{\mathbb{C}}$ sådana att $f(z) = z$.

6 Vidare egenskaper hos Möb($\mathbb{C}$).

Än så länge har vi undersökt automorfgrupperna $\text{Aut}(\mathbb{C})$ och $\text{Aut}(\widehat{\mathbb{C}})$ samt definierat vad en allmän automorfgrupp är. Vi såg i förra kapitlet att automorfierna i båda dessa fall är Möbiusavbildningar. Automorfgruppen av hyperboliska planet, $\text{Aut}(\mathcal{H})$, som vi ännu inte har definierat kommer också bestå av Möbiusavbildningar. I det här kapitlet kommer vi därför diskutera Möbiusavbildningar i mer detalj.

Vi kommer börja detta kapitel med en egenskap som bevaras av en allmän Möbius avbildning, nämligen mängden av cirklar och linjer.

Nästa resultat vi visar i detta kapitel har en lite annorlunda karaktär. Vi kommer att visa att för varje val av tre punkter på $\widehat{\mathbb{C}}$, så finns exakt en Möbiusavbildning som flyttar dessa tre punkter till vilka tre andra givna punkter som helst. Denna sats kommer att vara till stor hjälp senare när vi ska räkna med Möbiusavbildningar.

Slutligen ska vi gå in i detalj om egenskapen av en funktion att vara *vinkelbevarande*. Vi ska även visa att alla analytiska funktioner med nollskild derivata är vinkelbevarande.

6.1 Cirklar och linjer

I detta delavsnitt kommer vi vissa följande påstående: om kurvan Γ är en cirkel eller en rät linje, så kommer bilden av Γ under varje Möbiusavbildning återigen att vara en cirkel eller en rät linje. Låt Γ vara en cirkel, som exempel. Vi vet då att Γ i komplexa planet kan skrivas som

$$\{z \in \mathbb{C} \mid |z - p| = r\},$$

där $p \in \mathbb{C}$ är cirkelns centrum och $r > 0$ är cirkelns radie. Efter avbildningen

$$f(z) = \frac{az + b}{cz + d}$$

får vi mängden

$$f(\Gamma) = \{z \in \mathbb{C} \mid |f^{-1}(z) - p| = r\},$$

och sätter vi in vårt uttryck för Möbiusavbildningen $f^{-1}(z)$ är det inte direkt uppenbart att detta kommer beskriva en cirkel eller en linje för alla p och r . Vad vi behöver göra är att hitta en mer flexibel form för cirkelns ekvation än $|z - p| = r$. Vi kommer att skriva om cirkelns ekvation på ett sätt som kanske först bara ser mer krångligt ut. Börja med att kvadrera bägge sidor av $|z - p| = r$, och använd $|w|^2 = w\bar{w}$, för att få

$$(z - p)\overline{(z - p)} = z\bar{z} - p\bar{z} - \bar{p}z + p\bar{p} = r^2.$$

Låt $b = -\bar{p}$, och $c = p\bar{p} - r^2$. Då fås i stället ekvationen

$$z\bar{z} + bz + \bar{b}z + c = 0.$$

Vi hävdar att detta är en ny gemensam normalform för cirklar och linjer:

Sats 6.1.1. *Varje ekvation på formen*

$$az\bar{z} + bz + \overline{bz} + c = 0,$$

där a och c är reella, b är komplext, och $b\bar{b} > ac$, beskriver en cirkel i $\mathbb{C}$ eller en linje i $\mathbb{C}$. Varje cirkel och linje kan skrivas unikt på detta sätt, upp till multiplikation med en reell konstant.

Bevis. Vi börjar med andra halvan, att alla cirklar och linjer kan skrivas på denna form. För cirklar har vi nästan redan visat detta i paragrafen ovan. I vårt fall fick vi ut en ekvation precis som den ovan, fast med $a = 1$. Vidare är c reellt, eftersom c var lika med $|p|^2 - r^2$. Slutligen är också $b\bar{b} - ac = |p|^2 - (|p|^2 - r^2) = r^2 > 0$.

Vi vill också visa att varje linje kan skrivas på denna form. Vi kan t.o.m. säga något starkare, att linjer är exakt de kurvor för vilka $a = 0$. Ty en linje i komplexa planet kan alltid skrivas som

$$Ax + By + C = 0,$$

där A, B, C är reella, inte både A och B är noll, och $z = x + iy \in \mathbb{C}$. Men nu är $x = \frac{1}{2}(z + \bar{z})$, och $y = \frac{1}{2i}(z - \bar{z})$ (se Övning 3.2). Alltså kan ekvationen ovan skrivas som

$$\frac{A - Bi}{2}z + \frac{A + Bi}{2}\bar{z} + C = 0,$$

vilket är på samma form som tidigare om vi sätter $a = 0$, $b = (A - Bi)/2$, $c = C$. Vidare är $b \neq 0$ eftersom inte både A och B fick vara noll, så $b\bar{b} - ac = |b|^2 > 0$.

Nu återstår att visa att varje ekvation på denna form beskriver en cirkel om $a \neq 0$, eller en linje om $a = 0$. Om $a \neq 0$, dela hela ekvationen med a , och låt $p = -\bar{b}/a$. Nu har vi ekvationen

$$z\bar{z} - \bar{p}z - p\bar{z} + \frac{c}{a} = 0.$$

Här vet vi att $p\bar{p} = \frac{b\bar{b}}{a^2} > \frac{ac}{a^2} = c/a$. Alltså är $p\bar{p} - c/a > 0$, så vi kan hitta ett reellt tal r sådant att $p\bar{p} - c/a = r^2$. Nu kan vi skriva om ekvationen som

$$z\bar{z} - \bar{p}z - p\bar{z} + p\bar{p} - r^2 = 0,$$

och flyttar vi över r^2 till andra sidan och faktorerisarar fås

$$(z - p)(\overline{z - p}) = r^2,$$

vilket beskriver en cirkel. Fallet $a = 0$ lämnas åt läsaren som övning. $\square$

Hjälpsats 6.1.2. *Låt $h(z) = 1/z$, och Γ en cirkel eller linje i $\mathbb{C}$. Då är även $h(\Gamma)$ en cirkel eller linje.*

Bevis. För att hålla notationen tydlig inför vi variabeln $w = h(z) = 1/z$. Alltså är z koordinat i definitionsmängden och w koordinat i bildmängden. Låt Γ ges av en ekvation

$$az\bar{z} + bz + \overline{bz} + c = 0.$$

Då blir $h(\Gamma)$ en kurva i w -planet (i stället för z -planet), som ges av ekvationen

$$\frac{a}{w\bar{w}} + \frac{b}{w} + \frac{\bar{b}}{\bar{w}} + c = 0.$$

Multipluera med $w\bar{w}$: vi får

$$a + b\bar{w} + \bar{b}w + cw\bar{w} = 0,$$

vilket är en ekvation på samma form som tidigare fast med trippeln a, b, c utbytt mot talen $c, \bar{b}, a$. Villkoret $b\bar{b} - ac > 0$ från 6.1.1 är uppenbarligen fortfarande uppfyllt. $\square$

Sats 6.1.3. *Låt Γ vara en cirkel eller linje, och f en Möbiusavbildning. Då är $f(\Gamma)$ en cirkel eller linje.*

Bevis. Låt

$$f(z) = \frac{az + b}{cz + d}.$$

Om $c = 0$ är $f(z) = \frac{a}{d}z + \frac{b}{d}$. Denna avbildning är endast multiplikation med ett nollskilt komplext tal (vilket kan tolkas som en rotation och en skalning) följt av en translation, och med denna beskrivning är det uppenbart att $f(\Gamma)$ kommer att vara en cirkel eller linje så fort Γ är det.

Så antag nu att $c \neq 0$. Inför nya funktioner

$$g(z) = \left(b - \frac{ad}{c}\right)z + \frac{a}{c}, \quad j(z) = cz + d,$$

och låt som i föregående lemma $h(z) = 1/z$. Anledningen att införa dessa är att $f = g \circ h \circ j$: vi märker att

$$\begin{aligned} g(h(j(z))) &= g(h(cz + d)) = g\left(\frac{1}{cz + d}\right) = \left(b - \frac{ad}{c}\right) \cdot \frac{1}{cz + d} + \frac{a}{c} \\ &= \frac{a(cz + d) + bc - ad}{c(cz + d)} = \frac{acz + bc}{c(cz + d)} = \frac{az + b}{cz + d}. \end{aligned}$$

Men både g och j är Möbiusavbildningar med $c = 0$, så vi vet redan att dessa har egenskapen att de bevarar cirklar och linjer. Att h har denna egenskap visades i 6.1.2. Alltså har även deras sammansättning denna egenskap, så $f(\Gamma)$ måste vara en cirkel. $\square$

6.2 3-transitivitet

Följande egenskap hos gruppen av Möbiusavbildningar brukar kallas för att Möbiusgruppen är en 3-transitiv grupp av funktioner på Riemannsfären.

Sats 6.2.1. *Låt p, q, r vara tre godtyckliga distinkta punkter på $\widehat{\mathbb{C}}$, och p', q' och r' tre andra godtyckliga distinkta punkter. Då finns en unik Möbiusavbildning som tar p till p' , q till q' och r till r' .*

Vi delar upp beviset i flera lemmen. En vanlig bevismetod när man ska visa att det finns en unik sak med en viss egenskap, vilket är precis vad vi ska göra nu, är att dela upp beviset i två delar: först visar man att det finns *minst* en sak med denna egenskap, och sedan visar man att om två saker har denna egenskap så är de i själva verket likadana. Vi kommer nu att göra detta: Lemma 6.2.2 visar att det existerar minst en sådan Möbiusavbildning, och resterande lemmen kommer att användas för att visa entydighet

Hjälpsats 6.2.2. *Låt p, q, r vara tre godtyckliga distinkta punkter på $\widehat{\mathbb{C}}$. Då finns minst en Möbiusavbildning som avbildar p till 0, q till 1 och r till ∞ .*

Bevis. Låt oss först anta att ingen av p, q eller r är lika med ∞ . I så fall kan vi betrakta avbildningen

$$f(z) = \frac{q-r}{q-p} \cdot \frac{z-p}{z-r}.$$

Man kan nu verifiera att $f(p) = 0$, $f(r) = \infty$ och $f(q) = 1$. Dock kommer inte nödvändigtvis villkoret $ad - bc = 1$ att gälla för denna avbildning. Men vi ser att

$$ad - bc = (q-r)(q-p)r - (q-p)(q-r)p = (q-r)(q-p)(r-p).$$

Eftersom de tre punkterna var distinkta är detta nollskilt, så enligt 5.5.5 är vi klara.

Antag nu att någon av p, q och r är lika med ∞ . Välj först en Möbiusavbildning som avbildar 0, 1 eller ∞ på ∞ : vi kan välja $z \mapsto -1/z$, $z \mapsto 1/(1-z)$ eller $z \mapsto z$, beroende på om p, q eller r är ∞ . Vi sätter sedan samman inversen av denna Möbiusavbildning (dvs $z \mapsto -1/z$, $z \mapsto \frac{z-1}{z}$ eller $z \mapsto z$) med en annan som avbildar resterande två på 0 och 1: det är lätt att se att för varje par av punkter p, q i $\mathbb{C}$ så kommer

$$f(x) = \frac{1}{q-p}(x-p)$$

att avbilda p och q på 0 respektive 1. Det är också klart att denna f är en Möbiusavbildning, och att den fixerar punkten i oändligheten, så sammansättningen har den sökta egenskapen. $\square$

Hjälpsats 6.2.3. *Låt f vara en Möbiusavbildning. Antag att $f(0) = 0$, $f(1) = 1$ och $f(\infty) = \infty$. Då är f lika med identitetsavbildningen, d.v.s. $f(z) = z$ för alla $z \in \widehat{\mathbb{C}}$.*

Bevis. Låt $f(z) = \frac{az+b}{cz+d}$. Eftersom

$$f(0) = \frac{b}{d} = 0,$$

måste $b = 0$. Och eftersom

$$f(\infty) = \frac{a}{c} = \infty,$$

måste även $c = 0$. Alltså är

$$f(z) = \frac{a}{d} \cdot z,$$

och vi vet att $f(1) = 1$, så $a/d = 1$. $\square$

Med hjälp av Sats 5.5.7 kan föregående lemma skärpas till. Det kan vara värt att titta på den satsen igen innan man läser nästa lemma.

Hjälpsats 6.2.4. *Låt f vara en Möbiusavbildning. Antag att $f(p) = p$, $f(q) = q$ och $f(r) = r$ gäller för tre olika punkter p, q, r på $\widehat{\mathbb{C}}$. Även då kommer f vara identitetsavbildningen.*

Bevis. Antag att f fixerar punkterna p, q , och r . Låt g vara någon Möbiusavbildning som tar p till 0, q till 1 och r till ∞ (vi vet att g existerar enligt 6.2.2). Betrakta den sammansatta avbildningen

$$g \circ f \circ g^{-1}.$$

(Vi använder Sats 5.5.7 för att konstruera g^{-1}). Vi hävdar att denna måste vara identitetsavbildningen: ty

$$(g \circ f \circ g^{-1})(0) = g(f(g^{-1}(0))) = g(f(p)) = g(p) = 0,$$

och samma argument visar att $g \circ f \circ g^{-1}$ fixerar 1 och ∞ . Här använder vi att då $g(p) = 0$, är $g^{-1}(0) = p$. Alltså är

$$g \circ f \circ g^{-1} = \text{id},$$

enligt föregående lemma. Sätt nu samman funktionerna på höger och vänster led med g^{-1} från vänster, och g från höger. Vi finner att

$$g^{-1} \circ g \circ f \circ g^{-1} \circ g = g \circ \text{id} \circ g^{-1}.$$

Men eftersom $g \circ g^{-1} = \text{id}$, och $h \circ \text{id} = \text{id} \circ h = h$ gäller för alla funktioner h , kan ekvationen ovan förenklas till

$$f = \text{id},$$

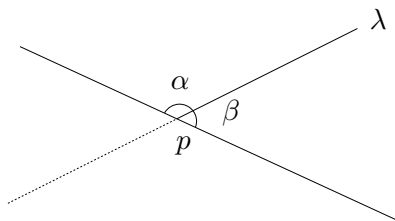
vilket uttryckt i ord säger att f är identitetsavbildningen. $\square$

Vi kan nu bevisa 3-transitivitet.

Bevis av Sats 6.2.1. Låt punkterna p, q, r och p', q', r' vara givna, som i 6.2.1. Låt först f vara en Möbiusavbildning som avbildar p, q, r på 0, 1 respektive ∞ , och g en Möbiusavbildning som avbildar p', q', r' på 0, 1 respektive ∞ . Att f och g existerar följer av 6.2.2. Betrakta nu funktionen $g^{-1} \circ f$. Nu är, på samma sätt som i beviset av föregående lemma,

$$(g^{-1} \circ f)(p) = g^{-1}(f(p)) = g^{-1}(0) = p',$$

och på samma sätt ser vi att $g^{-1} \circ f$ avbildar q på q' och r på r' . Det återstår att visa att $g^{-1} \circ f$ är unik med denna egenskap. Så antag att även h har denna egenskap. Då kommer med samma argument som nyss $h^{-1} \circ g^{-1} \circ f$ att avbilda p på sig själv, q på sig själv och r på sig själv. Vi är nu klara med hjälp av 6.2.4. $\square$



Figur 6.1: Två linjer som möts i en punkt. Här är α vinkeln från λ till μ , och β vinkeln från μ till λ .

Anmärkning 6.2.5. Egenskapen 6.2.1 hos Möbiusgruppen kan till exempel ge oss ett kort bevis för följande observation: *genom tre punkter p, q och r på $\widehat{\mathbb{C}}$ går en unik cirkel eller linje*. Så låt p, q och r vara tre godtyckliga punkter på $\widehat{\mathbb{C}}$. Om $p, q, r = 0, 1, 2$ är påståendet uppenbart — det finns en unik linje (den reella linjen $\mathbb{R}$) genom 0, 1 och 2, och ingen cirkel. Om p, q och r inte är lika med dessa tre, tag en Möbiusavbildning f sådan att $f(p) = 0$, $f(q) = 1$ och $f(r) = 2$. Då kommer $f^{-1}(\mathbb{R})$ att vara en cirkel eller linje enligt 6.1.3, och denna kommer att passera genom p, q och r . Om Γ skulle vara en annan cirkel/linje med denna egenskap, så skulle $f(\Gamma)$ vara en cirkel/linje genom 0, 1 och 2, vilket är omöjligt om inte $f(\Gamma) = \mathbb{R}$.

Vi ser alltså att eftersom Möbiusgruppen avbildar cirklar och linjer på cirklar och linjer, så gäller att om en Möbiusavbildning tar p, q och r till p', q' och r' , kommer den dessutom att avbilda den unika cirkeln/linjen genom p, q och r på den unika cirkeln/linjen genom p', q' och r' .

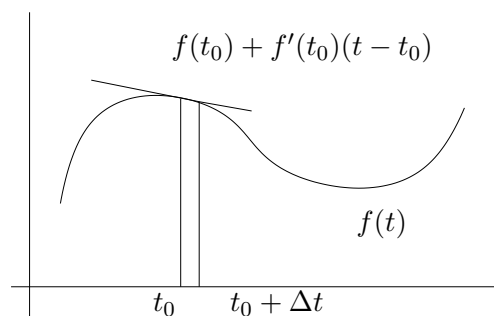
6.3 Konforma avbildningar

Vi vill nu visa påståendet vi utlovat flera gånger tidigare: att analytiska funktioner bevarar vinklar. Låt oss först definiera vad detta betyder. En första komplikation är att vi måste definiera vinkeln mellan två linjer. I Definition 2.1.6 definierade vi vinkeln mellan två *linjesegment* som möts i sina ändpunkter, men när vi har två *linjer* som möts i en punkt kan vi mäta upp två olika vinklar i skärningspunkten mellan linjerna. Vi behöver alltså en definition som säger vilken av dessa två vinklar man skall välja, för att få en väldefinierad vinkel mellan två linjer.

Definition 6.3.1. Låt två räta linjer λ och μ skära varandra i en punkt p . Vi vill definiera vinkeln mellan linjerna λ och μ . Betrakta den ena halvoändliga linjen som går ut från p i λ 's riktning.

Från denna linje kan vi mäta upp två olika vinklar (α och β i Figur 6.1) mot linjen μ . Med vår konvention att vinklar räknas positivt moturs och negativt medurs kommer endast en av dessa två vinklar att ligga i intervallet mellan 0 och π ; vi definierar denna som *vinkeln från λ till μ* . Från figuren ses att vinkeln ej beror på vilken av de två halvoändliga linjerna ut från p i λ 's riktning vi valde.

Definition 6.3.2. Låt två kurvor Γ och Δ mötas i en punkt $p \in \mathbb{C}$. Vi definierar *vinkeln från Γ till Δ* som vinkeln från tangenten till Γ till tangenten



Figur 6.2: Tangentapproximationen för funktioner $f : \mathbb{R} \rightarrow \mathbb{R}$.

till Δ . Detta är väldefinierat eftersom vi vet vad det betyder att mäta vinkeln mellan två räta linjer.

Anmärkning 6.3.3. Det är viktigt att specificera att man mäter vinklar från Γ till Δ och inte tvärtom. Eftersom vi räknar vinklar med tecken så säger vi att vinkeln från Γ till Δ är *minus* vinkeln från Δ till Γ .

Definition 6.3.4. Låt $f : \mathbb{C} \rightarrow \mathbb{C}$ vara en funktion. Skriv $f(x + iy) = u(x, y) + iv(x, y)$, och antag att både u och v är deriverbara funktioner av x för varje fixt y och vice versa. Dessa hypoteser garanterar att om $\Gamma \subset \mathbb{C}$ är en kurva, är även $f(\Gamma)$ en kurva. Om f uppfyller dessa antaganden, säger vi att f är *konform* om följande egenskap gäller: så fort två kurvor Γ och Δ möts i en punkt $p \in \mathbb{C}$, så är vinkeln från Γ till Δ i p densamma som vinkeln från $f(\Gamma)$ till $f(\Delta)$ i $f(p)$.

Vi har redan i kapitel 4 talat om tangentlinjer till kurvor i planet, men vi har faktiskt aldrig definierat vad en tangentlinje till en kurva är för något.

Definition 6.3.5. Om Γ är en kurva i $\mathbb{C}$ med parametrisering γ , och $\gamma'(t_0) \neq 0$, så definierar vi *tangentlinjen* genom $\gamma(t_0)$ som den parametriserade linjen

$$\gamma(t_0) + \gamma'(t_0) \cdot (t - t_0),$$

där t varierar över $\mathbb{R}$.

Anmärkning 6.3.6. Det är värt att vara tydlig med att $\gamma(t_0)$ och $\gamma'(t_0)$ båda är *komplexa* tal, medan t och t_0 överallt är reella tal. Multiplikationen i formeln skall alltså tolkas som multiplikation av ett reellt och ett komplext tal. Vi kan tänka på $\gamma'(t_0)$ som ett tal som anger *riktningen* för tangenten bort från punkten $\gamma(t_0)$. Detta eftersom att linjen som beskrivs av ekvationen inte beror på beloppet av $\gamma'(t_0)$, utan endast av argumentet. Argumentet till $\gamma'(t_0)$ anger exakt vinkeln som tangenten gör mot x -axeln.

Anmärkning 6.3.7. Ett heuristiskt argument för att Definition 6.3.5 är vettig är följande: det är en välbekant approximation att för en deriverbar funktion $f : \mathbb{R} \rightarrow \mathbb{R}$, så gäller för små Δt att $f(t_0 + \Delta t) \cong f(t_0) + f'(t_0) \cdot \Delta t$, eller ekvivalent

$$f(t) \cong f(t_0) + f'(t_0) \cdot (t - t_0).$$

Denna approximation är precis vad man får om man ersätter grafen till f med dess tangentlinje genom $f(t_0)$, som i Figur 6.2. Speciellt gäller att ekvationen för tangenten fås genom att ersätta " $\cong$ " med " $=$ ". Ekvationen i 6.3.5 säger bara att vi kan göra exakt samma sak även när f är en funktion från $\mathbb{R}$ in i det *komplexa* planet. En annan observation som får definitionen att verka rimlig är att notera att tangenten har följande egenskaper: (i) Tangenten är en linje; (ii) Tangenten passerar genom $\gamma(t_0)$ (sätt in $t = t_0$ i ekvationen); (iii) Tangenten har samma derivata i punkten t_0 som kurvan Γ , eftersom tangentkurvans derivata är konstant lika med $\gamma'(t_0)$.

Anmärkning 6.3.8. En märklig sak vid första anblicken av 6.3.5 är att tangentlinjen tycks bero på vilken parametrisering av kurvan Γ vi väljer. Speciellt kan man undra vad som händer om $\gamma'(t_0)$ skulle vara noll — existerar inte tangentlinjen då? Detta beroende på parametriseringen är dock illusoriskt: om vi ändrar parametriseringen av Γ kan vi ändra beloppet på derivatan ("farten") i varje punkt, men inte argumentet ("riktningen"). Och linjen som parametriseras av 6.3.5 blir densamma, oavsett om vi ändrar beloppet på $\gamma'(t_0)$. Man kan visa att varje kurva Γ kan parametriseras så att $\gamma'(t) \neq 0$ för alla t , så det går att definiera tangenten genom varje punkt på Γ .

Hjälpsats 6.3.9. Låt γ och δ vara parametriseringar av kurvorna Γ och Δ som möts i punkten $\gamma(t_0) = \delta(t_0)$. Låt oss anta att γ och δ har nollskilda derivator i denna skärning. I så fall ges vinkeln från Γ till Δ i denna punkt av

$$\arg(\delta'(t_0)) - \arg(\gamma'(t_0)).$$

Bevis. Vi har definierat vinkeln mellan Γ och Δ att vara lika med vinkeln mellan deras respektive tangentlinjer. Från ekvationen i 6.3.5 ser vi att $\gamma'(t_0)$ respektive $\delta'(t_0)$ ger tangentriktningarna för de båda linjerna. Nu är $\arg(\gamma'(t_0))$ vinkeln mellan denna tangentlinje och x -axeln, och samma sak gäller för $\arg(\delta'(t_0))$. Alltså är differensen $\arg(\delta'(t_0)) - \arg(\gamma'(t_0))$ lika med skillnaden i vinkel mätt moturs från x -axeln från Γ till Δ . $\square$

Sats 6.3.10. Låt f vara en analytisk funktion med nollskild derivata. Då är f konform.

Bevis. Låt z_0 vara en punkt i komplexa planet, och antag att kurvorna Γ och Δ möts i denna. Låt kurvorna parametriseras av γ respektive δ . Vi kan utan inskränkning välja parametriseringarna så att $\gamma(t_0) = \delta(t_0) = z_0$, med nollskilda derivator i denna punkt. Vi vet redan att vinkeln från Γ till Δ i z_0 är

$$\arg(\delta'(t_0)) - \arg(\gamma'(t_0)),$$

och vi vill nu hitta vinkeln från bilden av Γ till bilden av Δ under transformationen f . Bilden av kurvan Γ ges av den sammansatta funktionen

$$(f \circ \gamma)(t) = f(\gamma(t)),$$

som vi nu deriverar med kedjeregeln. Vi finner att

$$(f \circ \gamma)'(t) = f'(\gamma(t)) \cdot \gamma'(t).$$

Här är både $f'(\gamma(t))$ och $\gamma'(t)$ komplexa tal. Alltså ges vinkeln mellan $f(\Gamma)$ och $f(\Delta)$ i punkten $f(z_0)$ av

$$\begin{aligned} & \arg(f'(\delta(t_0)) \cdot \delta'(t_0)) - \arg(f'(\gamma(t_0)) \cdot \gamma'(t_0)) \\ &= \arg(f'(z_0) \cdot \delta'(t_0)) - \arg(f'(z_0) \cdot \gamma'(t_0)) \\ &= \arg(f'(z_0)) + \arg(\delta'(t_0)) - \arg(f'(z_0)) - \arg(\gamma'(t_0)) \\ &= \arg(\delta'(t_0)) - \arg(\gamma'(t_0)). \end{aligned}$$

En sak kvarstår dock: formeln $\arg(zw) = \arg(z) + \arg(w)$, som vi använt i beviset ovan, gäller bara för produkten av två *nollskilda* komplexa tal. (Om $z = 0$ är $\arg(z)$ odefinierat.) Men som vi antog är f' nollskilt, därmed är beviset klart! $\square$

Anmärkning 6.3.11. Det motsatta påståendet visar sig också vara sant, en funktion som är konform är också analytisk med nollskild derivata. Vi tar dock inte och bevisar detta, då vi undviker att jobba med den komplexa derivatans definition i den här kursen.

Övningar

Övning 6.1. Hitta någon Möbiusavbildning som avbildar imaginära axeln på enhetscirkeln.

Övning 6.2. (i) Finns det en Möbiusavbildning som avbildar punkterna $2 + 3i, -1 + 4i, -4 - 5i$ och $4 - i$ på $1 + i, -1, -4 - 3i$ och 0 , respektive?

(ii) Visa att $f(z) = 2z$ är en Möbiusavbildning genom att skriva om funktionen på formen $f(z) = \frac{az+b}{cz+d}$ där $a, b, c, d \in \mathbb{R}$ uppfyller $ad - bc = 1$.

(iii) Beskriv alla Möbiusavbildningar som har 0 och ∞ som fixpunkter. Hitta en familj av Möbiusavbildningar som har 1 och ∞ som fixpunkter.

Övning 6.3. Visa direkt att det genom tre godtyckliga punkter p, q och r i komplexa planet passerar exakt en cirkel eller linje (påstående 6.2.5). Ett sätt är så här: dela först upp i två olika fall, antingen ligger punkterna på en linje eller så gör de det inte. Om punkterna ligger på en linje är det enkelt. Om de inte ligger på en linje, betrakta dels mängden av alla punkter som är på samma avstånd till både p och q , och dels mängden av punkter som är på samma avstånd till q och r . Visa att dessa två mängder möts i exakt en punkt, och använd detta för att konstruera den sökta cirkeln.

Övning 6.4. Definiera *dubbelförhållandet* av fyra komplexa tal z_1, z_2, z_3, z_4 som

$$[z_1, z_2; z_3, z_4] = \frac{(z_1 - z_3)(z_2 - z_4)}{(z_1 - z_4)(z_2 - z_3)}.$$

Visa att Möbiusavbildningar bevarar dubbelförhållanden, dvs

$$[z_1, z_2; z_3, z_4] = [f(z_1), f(z_2); f(z_3), f(z_4)].$$

Övning 6.5. I beviset till 6.3.10 visades att om f är en deriverbar funktion $\mathbb{C} \rightarrow \mathbb{C}$, så är f konform i alla punkter där f har nollskild derivata. Visa att $f(z) = z^2$ inte är konform i punkten $z = 0$.

Övning 6.6. Använd resultatet från Övning 5.6 för att visa att skärningen mellan ett plan i $\mathbb{R}^3$ och $\hat{\mathbb{C}}$ alltid är en cirkel eller linje i komplexa planet (förutsatt att skärningen består av mer än en punkt), och att varje cirkel eller linje i komplexa planet kan skrivas som skärningen med ett plan. Detta resultat visar igen att på Riemannsfären bör man tänka på en cirkel och en linje som samma sak.

Övning 6.7. Använd resultatet från föregående uppgift för att ge ett till bevis för att det genom tre punkter på $\hat{\mathbb{C}}$ passerar en unik cirkel eller linje.

Övning 6.8. Hitta en Möbiusavbildning som avbildar enhetscirkeln på linjen $\operatorname{Im} z = \operatorname{Re} z - 1$.

7 Delgruppen $\text{Aut}(\mathbb{H})$.

I det föregående kapitlet nämnde vi att automorfigruppen för det hyperboliska planet också kommer bestå av Möbiusavbildningar, men vi har ännu inte talat om vilka.

I detta kapitel kommer vi först att definiera vad mängden av *punkter* i det hyperboliska planet är. Vi kommer sedan att räkna ut vilka Möbiusavbildningar som avbildar denna mängd av punkter på sig själv; det är exakt denna delmängd av Möbiusavbildningar som sedan blir automorfigruppen till hyperboliska planet.

Efter att vi har gjort detta kommer vi att definiera en mängd av kurvor som kommer att spela rollen som de *räta linjerna* i det hyperboliska planet. I detta kapitel kommer vi helt sonika att definiera en hyperbolisk linje som en av dessa kurvor, och när vi i nästa kapitel definierar hur man räknar ut den hyperboliska längden på en kurva kommer det att visa sig att dessa kurvor är exakt de som minimerar avståndet mellan två punkter. (Så vår definition blir retroaktivt vettig!) Men, som sagt, i detta kapitel kommer vi endast att definiera vilka kurvor som kommer att bli räta linjer i hyperboliska planet, och sedan kommer vi att visa vissa egenskaper hos dessa kurvor.

7.1 Möbiusavbildningar som bevarar övre halvplanet

Den första skillnaden mellan det hyperboliska planet och det euklidiska planet och sfären är att mängden av punkter inte är densamma. I det euklidiska planet består punkterna av elementen i $\mathbb{R}^2$ (vi kan också, som bekant, tänka på $\mathbb{C}$). Sfären kunde vi beskriva genom att lägga till en extra punkt, ∞ , för att bilda mängden $\widehat{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$ som beskrev sfären. För att definiera det hyperboliska planet kommer vi istället använda en *mindre* mängd punkter, nämligen det *övre halvplanet*:

$$\mathbb{H} = \{z \in \mathbb{C} \mid \text{Im}(z) > 0\}.$$

Anmärkning 7.1.1. Varför ska vi beskriva det hyperboliska planet med just det övre halvplanet? Varför inte med det undre halvplanet, enhetsskivan eller någon annan delmängd av $\mathbb{C}$? Det visar sig att det finns två saker som kan hända.

- För många mängder $D \subset \mathbb{C}$ så finns det en analytisk bijektion från $\mathbb{H} \rightarrow D$. Detta innebär att vi skulle kunna skapa två helt ekvivalenta automorfigrupper, en för D och en för $\mathbb{H}$, som båda beskriver det hyperboliska planet. Matematiken blir dock något enklare för just det övre halvplanet vilket är varför vi väljer det som vår modell.
- För andra mängder (ofta lite konstigare mängder som kanske inte är sammanhängande eller har hål), så finns det inte tillräckligt många analytiska bijektioner från mängden till sig själv för att det ska kunna bli en automorfigrupp enligt definition 5.4.1. Speciellt är det (iii) som kräver att vi har tillräckligt många funktioner.

Vi påminner om att en automorfi alltid måste vara en bijektion. En sådan funktion $f : \mathbb{H} \rightarrow \mathbb{H}$ säger vi *bevarar* övre halvplanet, eller att f avbildar övre halvplanet till sig självt, om detta gäller. Det kommer att visa sig att automorfigruppen till det hyperboliska planet är exakt de Möbiusavbildningar som bevarar övre halvplanet. Vi kommer nu att undersöka vilka Möbiusavbildningar som har denna egenskap.

Hjälpsats 7.1.2. *Låt $f \in \text{Möb}(\mathbb{C})$. Följande två påståenden är ekvivalenta:*

- (i) *Möbiusavbildningen f bevarar linjen $\mathbb{R} \cup \{\infty\}$.*
- (ii) *Funktionen f kan skrivas som*

$$f(z) = \frac{az + b}{cz + d},$$

där a, b, c och d alla är reella, men $ad - bc$ endast antas nollskilt och inte nödvändigtvis är lika med 1.

Bevis. Antag först att f bevarar $\mathbb{R} \cup \{\infty\}$. Låt p, q och r avbildas till 0, 1 respektive ∞ av f , och antag för enkelhetens skull att ingen av p, q och r är lika med ∞ . I så fall vet vi att

$$f(z) = \frac{q - r}{q - p} \cdot \frac{z - p}{z - r},$$

med samma argument som i beviset av 6.2.2. Men eftersom f bevarar reella linjen måste p, q och r vara reella tal, och vi ser att vi har skrivit ner ett uttryck för Möbiusavbildningen med reella koefficienter. (Men, som sagt, så vet vi inte att $ad - bc = 1$.)

Antag nu att

$$f(z) = \frac{az + b}{cz + d},$$

där både a, b, c och d är reella. Då kommer uppenbarligen gälla att om z är reellt eller ∞ , så är $f(z)$ reellt eller ∞ . Vi vill dessutom visa att om $f(z)$ är reellt eller ∞ , är z det. Men om

$$\frac{az + b}{cz + d} = w \in \mathbb{R},$$

är

$$z = \frac{wd - b}{a - wc}$$

som är reellt (eller ∞) om alla av a, b, c, d och w är reella. (Om $w = \infty$ blir $z = -d/c$, som ligger i $\mathbb{R} \cup \{\infty\}$.) □

Anmärkning 7.1.3. Formuleringen av föregående lemma kan verka förvirrande. Först antas att f är en Möbiusavbildning, men sedan sägs att $ad - bc$ inte nödvändigtvis är lika med ett! Nyckeln är i formuleringen att f kan skrivas som

$$f(z) = \frac{az + b}{cz + d},$$

där a, b, c och d alla är reella. Att f är en Möbiusavbildning säger på samma sätt att f kan skrivas som

$$f(z) = \frac{az + b}{cz + d},$$

där $ad - bc = 1$. Men dessa två möjliga val av a, b, c och d behöver inte nödvändigtvis vara samma! Jämför diskussionen i 5.5.4.

Varför är egenskapen att bevara $\mathbb{R} \cup \{\infty\}$ viktig? Tänk på hur $\mathbb{H}$ ser ut på Riemannsfären, $\hat{\mathbb{C}}$. Om vi tänker på ∞ som nordpol och 0 som sydpol, så kommer linjen $\mathbb{R} \cup \{\infty\}$ att gå longitudinellt (vertikalt) som en storcirkel runt sfären. Det övre halvplanet kommer att vara den ena halvsfären och det undre halvplanet den andra. Med andra ord, $\mathbb{R} \cup \{\infty\}$ är *randen* till $\mathbb{H}$ om vi betraktar situationen på $\hat{\mathbb{C}}$. Om en Möbiusavbildning f avbildar $\mathbb{H}$ på $\mathbb{H}$ måste den även avbilda randen till $\mathbb{H}$ till sig själv, så ett nödvändigt villkor på en Möbiusavbildning för att kunna bevara $\mathbb{H}$ är att den kan skrivas med reella koefficienter. Dock är det inte nödvändigtvis så att en funktion som avbildar $\mathbb{R} \cup \{\infty\}$ på sig självt bevarar övre halvplanet — funderar man lite så inser man att det finns två möjligheter: antingen så kommer funktionen att avbilda de båda halvsfärerna till sig själva, eller så kommer de att byta plats på de båda halvsfärerna. Vi undersöker hur vi kan se skillnad på de olika fallen:

Hjälpsats 7.1.4. *Låt f vara en funktion på formen*

$$f(z) = \frac{az + b}{cz + d},$$

där a, b, c och d är reella och $ad - bc \neq 0$ (men inte nödvändigtvis lika med 1.) Låt $z \in \mathbb{H}$. Då gäller att om $ad - bc > 0$, så kommer $\text{Im}(f(z)) > 0$ och alltså $f(z) \in \mathbb{H}$. Om däremot $ad - bc < 0$ så kommer $\text{Im}(f(z)) < 0$, så $f(z) \notin \mathbb{H}$.

Bevis. Tag $z = x + iy \in \mathbb{H}$. Då är

$$f(z) = \frac{az + b}{cz + d} = \frac{(az + b)\overline{(cz + d)}}{(cz + d)\overline{(cz + d)}} = \frac{1}{|cz + d|^2}(ax + b + iay)(cx + d - icy).$$

Den reella konstanten $\frac{1}{|cz + d|^2}$ kommer inte att påverka om $f(z)$ ligger i övre halvplanet eller inte. (Kontrollera detta!) Imaginärdelen av $(ax + b + iay)(cx + d - icy)$ är exakt $(cx + d) \cdot ay - (ax + b) \cdot cy = (ad - bc) \cdot y$. Men nu vet vi att z ligger i övre halvplanet precis då $y > 0$, så om $ad - bc > 0$ kommer $f(z)$ att ligga i övre halvplanet, och när $ad - bc < 0$ kommer $f(z)$ ligga in nedre halvplanet. $\square$

Definition 7.1.5. Delmängden $\text{Aut}(\mathbb{H}) \subset \text{Möb}(\mathbb{C})$ består av alla Möbiusavbildningar som kan skrivas som

$$f(z) = \frac{az + b}{cz + d}$$

där a, b, c och d är reella konstanter och $ad - bc = 1$. Mängden $\text{Aut}(\mathbb{H})$ kan också betecknas som $\text{Möb}(\mathbb{H})$.

Sats 7.1.6. Låt $f \in \text{Möb}(\mathbb{C})$. Då gäller att $f \in \text{Aut}(\mathbb{H})$ om och endast om f avbildar övre halvplanet $\mathbb{H}$ på sig självt.

Bevis. Vi är nästan klara med hjälp av våra två tidigare lemman. Antag att f avbildar övre halvplanet på sig självt. Då kommer den även att avbilda randen till övre halvplanet, $\mathbb{R} \cup \{\infty\}$, på sig självt. Alltså kan vi enligt Lemma 7.1.2 skriva

$$f(z) = \frac{az + b}{cz + d}$$

där a, b, c och d är reella, och enligt Lemma 7.1.4 är nu $ad - bc > 0$. Alltså har $ad - bc$ en reell kvadratrots, och genom att förlänga bråket med $1/\sqrt{ad - bc}$ uppnår vi villkoret $ad - bc = 1$.

Om vi i stället har $ad - bc = 1$ och a, b, c och d reella så kommer randen $\mathbb{R} \cup \{\infty\}$ att avbildas till sig själv (enligt 7.1.2), och som vi tidigare observerat finns nu två möjligheter: antingen kommer f att avbilda bägge ”halvkloten” till sig själva, eller så kommer f att byta plats på dem. Men nu visar föregående lemma att alla $z \in \mathbb{H}$ kommer att avbildas in i övre halvplanet av f , så vi är i den första situationen. $\square$

Sats 7.1.7. Delmängden $\text{Aut}(\mathbb{H})$ uppfyller, precis som $\text{Möb}(\mathbb{C})$, alla tre egenkaperna från Sats 5.5.7. (Man säger ibland att $\text{Aut}(\mathbb{H})$ är en delgrupp till Möbiusgruppen.)

Bevis. Beviset är nästan identiskt med beviset för Sats 5.5.7. Det enda vi behöver kontrollera är att koefficienterna till $f \circ g$ och f^{-1} är reella, förutsatt att de för f och g är reella. Men detta är uppenbart från de explicita uttrycken för $f \circ g$ och f^{-1} som ges i beviset. (Om a, b, c, d och A, B, C, D alla är reella så kommer även $aA + bC$ vara det, etc.) $\square$

Sats 7.1.8. $\text{Aut}(\mathbb{H})$ är en automorfigrupp enligt definition 5.4.1.

Bevis. (i) Att funktionerna i $\text{Aut}(\mathbb{H})$ är analytiska i $\mathbb{H}$ följer direkt från Sats 3.8.2.

(ii) Precis som i beviset av Sats 5.4.4 behöver vi bara visa att för alla punkter $w = x + iy \in \mathbb{H}$ finns det någon automorfi $f \in \text{Aut}(\mathbb{H})$ så att

$$f(i) = w.$$

Vi skriver f på formen

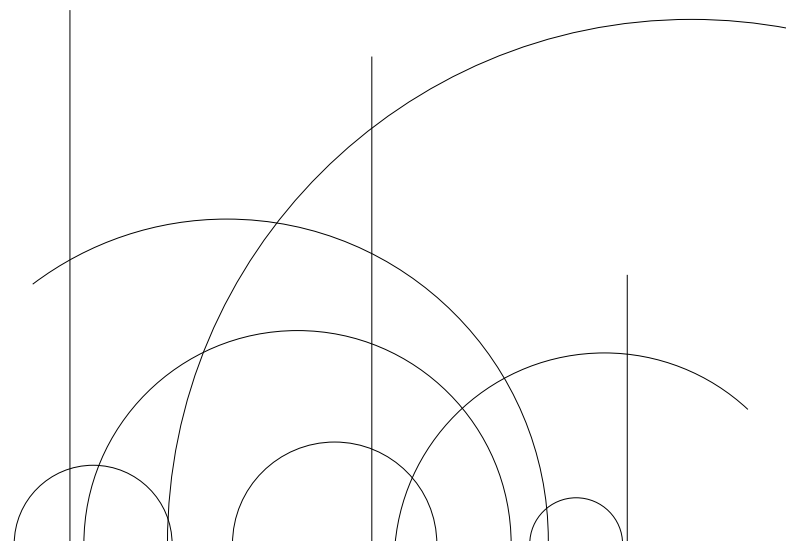
$$f(z) = \frac{az + b}{cz + d}$$

där $ad - bc = 1$. Ekvationen $f(i) = w$ säger att

$$\frac{ai + b}{ci + d} = w.$$

Låt nu $\alpha = ai + b$ och $\beta = ci + d$. Vi har i så fall att

$$\alpha = w\beta$$



Figur 7.1: Reella axeln $\mathbb{R}$, och en samling hyperboliska linjer som möter den i en rät vinkel.

medan villkoret $ad - bc = 1$ kan skrivas om som

$$\operatorname{Im}(\alpha\bar{\beta}) = 1.$$

Substituerar vi in vad α måste vara får vi att

$$1 = \operatorname{Im}(w\beta\bar{\beta}) = |\beta|^2 \operatorname{Im}(w)$$

Vi kan nu välja exempelvis

$$\beta = \frac{1}{\sqrt{\operatorname{Im}(w)}}$$

(detta är möjligt då $w \in \mathbb{H}$ så $\operatorname{Im}(w) > 0$), och sedan $\alpha = w\beta = \frac{w}{\sqrt{\operatorname{Im}(w)}}$.

(iii) Se övning 7.7

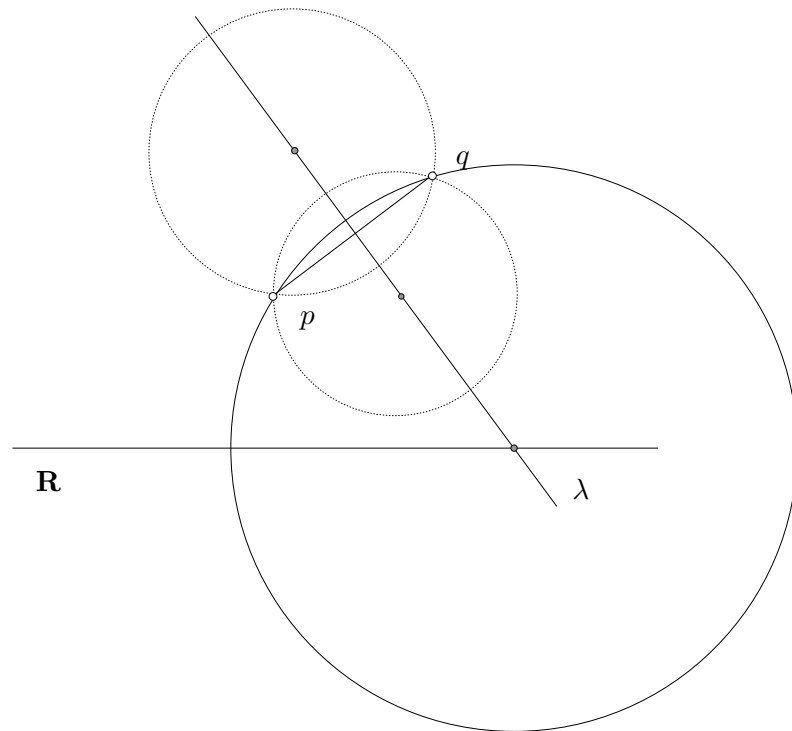
□

7.2 Hyperboliska linjer

Vi kommer nu att definiera vilka kurvor i $\mathbb{H}$ som är de räta linjerna, det vill säga vilka kurvor som vi i nästa kapitel kommer att visa minimerar det hyperboliska avståndet (som kommer att definieras då).

Definition 7.2.1. En delmängd $\lambda \subset \mathbb{H}$ sägs vara en *hyperbolisk linje* om det existerar någon linje eller cirkel Γ i det komplexa planet som skär reella axeln i en rät vinkel, sådan att $\lambda = \mathbb{H} \cap \Gamma$.

Sats 7.2.2. Givet två distinkta punkter $p, q \in \mathbb{H} \cup \mathbb{R} \cup \infty$ finns det en unik hyperbolisk linje som skär dessa punkter.



Figur 7.2: En illustration av beviset till 7.2.2.

Bevis. Antag först att $p, q \neq \infty$. Betrakta sedan alla cirklar som passerar genom p och q . Att en cirkel skär p och q säger exakt samma sak som att cirkelns mittpunkt ligger på samma avstånd från p som q , och att cirkelns radie är exakt detta avstånd. Låt λ vara den rätta linje som består av alla punkter på samma avstånd från p och q , se Figur 7.2. Linjen λ är alltså normalen till det rätta linjesegmentet från p till q som skär denna i segmentets mittpunkt. För varje punkt c på λ kan vi rita en unik cirkel med c som centrum och som skär p och q .

Vi noterar också att en cirkel kommer skära reella axeln i en rät vinkel om och endast om cirkelns mittpunkt ligger *på* reella linjen. Ty antag att en cirkel skär reella axeln i en punkt, och att tangenten till cirkeln i denna punkt är vinkelrät mot reella axeln. Eftersom tangenten till cirkeln dessutom är vinkelrät mot dess radie, ligger nu radien nödvändigtvis längs med reella axeln, så att centrum ligger någonstans på $\mathbb{R}$.

Vi måste nu dela upp i två fall:

- (i) Punkterna p och q ligger inte ovanför varandra. I detta fall kommer λ *inte* att kunna vara parallell med reella linjen, så λ skär reella axeln i en unik punkt. Om vi ritat cirkeln med denna punkt som centrum och som passerar genom p och q , kommer denna att vara vår unika hyperbolisk linje.
- (ii) Punkterna p och q ligger rakt ovanför varandra. I detta fall kommer λ inte att skära reella axeln i någon punkt, så det finns ingen cirkel genom

p och q som skär reella axeln i en rät linje. Dock kan vi i detta fallet (och endast i detta fallet) i stället helt sonika dra den räta linjen genom p och q — denna kommer vara vår unika hyperboliska linje.

Slutligen ifall en av punkterna, säg p , är ∞ så ser vi att endast den vertikala linjen som skär q kommer skära både p och q . $\square$

Sats 7.2.3. *Alla $f \in \text{Aut}(\mathbb{H})$ avbildar hyperboliska linjer på hyperboliska linjer.*

Bevis. Låt oss begrunda definitionen av en hyperbolisk linje. Vi har definierat en hyperbolisk linje som en linje eller cirkel, som skär reella axeln i rät vinkel. Eftersom f är en Möbiusavbildning kommer varje hyperbolisk linje att avbildas på en cirkel eller linje (Sats 6.1.3), så det räcker att visa att även denna skär reella axeln i en rät vinkel. Men detta är lätt att visa med hjälp av vad vi vet om Möbiusavbildningar: vi vet redan att den reella linjen (med oändligheten inräknad) kommer att avbildas till sig själv av f (Lemma 7.1.2) och vi vet dessutom att varje Möbiusavbildning är konform (Sats 6.3.10) och därmed bevarar vinkeln i vilken kurvor skär varandra. Beviset är klart. $\square$

Sats 7.2.4. *Låt λ och μ vara två hyperboliska linjer. Då existerar $f \in \text{Aut}(\mathbb{H})$ som avbildar λ på μ .*

Bevis. Vi visar först att för varje hyperbolisk linje λ så finns en $f \in \text{Aut}(\mathbb{H})$ som avbildar λ på imaginära axeln. Om λ är en rät linje, alltså vertikal, så är detta lätt att visa — använd bara en translation i x -led på formen

$$f(z) = z + b$$

där b är reellt; alla dessa translationer ligger i $\text{Aut}(\mathbb{H})$. Så låt λ vara en cirkel som skär reella axeln i två punkter p och q . Vi kan anta att $p < q$. Betrakta nu avbildningen

$$f(z) = \frac{z - q}{z - p}.$$

För denna gäller att $ad - bc = q - p > 0$, så genom att multiplicera alla koefficienter med $1/\sqrt{q - p}$ (som är reellt) ser vi att denna funktion faktiskt är en Möbiusavbildning. Vidare kommer q avbildas på 0 och p på ∞ . Alltså kommer bilden av λ att vara en hyperbolisk linje som passerar genom 0 och ∞ . Men en hyperbolisk linje som passerar genom oändligheten måste vara en vertikal linje (kontrollera detta!), och om den dessutom passerar origo måste den vara den imaginära axeln. Det följer att f kommer att avbilda λ på imaginära axeln.

Betrakta nu det allmänna fallet med två hyperboliska linjer λ och μ . Tag $f, g \in \text{Aut}(\mathbb{H})$ sådana att f avbildar λ på imaginära axeln och g avbildar μ på imaginära axeln. Då kommer

$$g^{-1} \circ f$$

vara en avbildning i $\text{Aut}(\mathbb{H})$ som avbildar λ på μ . $\square$

Anmärkning 7.2.5. Till skillnad från i förra kapitlet, där vi visade 3-transitivitet, så säger vi inte här att det finns en *unik* avbildning som tar en hyperbolisk linje till en annan. Faktiskt finns det oändligt många olika $f \in \text{Aut}(\mathbb{H})$ som

avbildar en given linje på en annan. Vi kan jämföra detta med situationen i det vanliga Euklidiska fallet: för varje par av vanliga Euklidiska linjer i planet finns det oändligt många olika automorfier av planet som avbildar den ena på den andra. Det är lätt att se att vi kan konstruera *någon* automorfi som gör detta — vi kan ta en translation eller en rotation, beroende på om linjerna är parallella eller inte — men sedan kan vi dessutom sätta samman denna automorfi med en godtycklig translation längs med linjen, och på detta sättet få oändligt många fler automorfier med samma egenskap. Ett motsvarande påstående gäller i det hyperboliska fallet, men vi har inte definierat vad det betyder att translatera övre halvplanet längs med en given hyperbolisk linje.

Övningar

Övning 7.1. Visa att två olika hyperboliska linjer inte kan mötas i mer än en punkt.

Övning 7.2. Hitta ekvationen för den hyperboliska linjen genom $p = -1 + 4i$ och $q = 6 + 3i$.

Övning 7.3. Antag att $f \in \text{Aut}(\mathbb{H})$ har egenskapen att f avbildar imaginära axeln till sig själv. Hur kan f se ut?

Övning 7.4. Betrakta punkten $P = 2i$ och hyperboliska linjen l (som är en halvcirkel) från -1 till 1 . Hitta två andra hyperboliska linjer l_1 och l_2 sådana att l_1 och l_2 går genom P men ej skär l . (Detta visar att parallellpostulatet (v) i 2.3 inte gäller för hyperboliska linjer.)

Övning 7.5. I euklidiska planet känner vi till att givet en linje λ och en punkt $p \in \lambda$, finns en unik linje genom p som möter λ i en rät vinkel. Visa att detta även gäller för hyperboliska linjer i $\mathbb{H}$.

Övning 7.6. Beskriv alla hyperboliska linjer som går genom i .

Övning 7.7. Slutför beviset av sats 7.1.8. Det vill säga, visa att $\text{Aut}(\mathbb{H})$ uppfyller egenskap (iii) av definition 5.4.1.

8 Hyperbolisk längd och hyperboliskt avstånd

Vi ska i det här kapitlet använda det övre halvplanet och dess automorfier för att skapa en typ av geometri som kallas hyperbolisk.

I förra kapitlet såg vi att gruppen av Möbiusavbildningar på formen

$$f(z) = \frac{az + b}{cz + d} \quad (8.1)$$

med

$$a, b, c, d \in \mathbb{R} \quad \text{och} \quad ad - bc = 1$$

bevarar övre halvplanet. En räkning visar att avståndsbegrepp som vi definierade med hjälp av båglängd i Kapitel 2 inte passar ihop med dessa automorfier—detta avstånd förändras om vi tillämpar en automorfi $f \in \text{Aut}(\mathbb{H})$. Vi börjar därför med att modifiera vårt längdbegrepp för att ta hänsyn till egenskaperna hos övre halvplanets automorfier. Vi ska bevisa att detta nya längdbegrepp verkligen är invariant under $\text{Aut}(\mathbb{H})$, och att vi kan använda det för att införa ett hyperboliskt avståndsbegrepp. Vi ska se att de hyperboliska geodeterna, de avståndsminimerande kurvorna, blir de hyperboliska linjerna vi stötte på i föregående kapitel. Vi avslutar med att ge några exempel på hur beräkningar i hyperbolisk geometri kan se ut.

8.1 Hyperbolisk längd

Låt Γ vara bilden av parameterfunktionen $\gamma: [a, b] \rightarrow \mathbb{H}$ med

$$\gamma(t) = x(t) + iy(t), \quad t \in [a, b]$$

Vi har tidigare infört den euklidiska längden av kurvan Γ genom integralen

$$l(\Gamma) = \int_a^b |\gamma'(t)| dt.$$

Följande exempel visar att om $f \in \text{Möb}(\mathbb{H})$ så gäller tyvärr i allmänhet

$$l(\Gamma) \neq l(f(\Gamma)).$$

Exempel 8.1.1. Låt K beteckna det euklidiska linjesegmentet mellan punkterna

$$p_1 = i \quad \text{och} \quad p_2 = 2i.$$

Vi parametriserar denna kurva genom

$$\gamma(t) = it, \quad t \in [1, 2],$$

och noterar att längden av linjesegmentet är $l(K) = 1$.

Betrakta nu avbildningen

$$f(z) = \frac{z-1}{z}$$

som är en automorfi på formen (8.1) med $a = 1$, $b = -1$, $c = 1$ och $d = 0$. Vi tillämpar denna på vår kurva och beräknar båglängden av bilden av K under automorfin:

$$l(f(K)) = \int_1^2 |[f(\gamma(t))]'| dt = \int_1^2 |\gamma'(t)| |f'(\gamma(t))| dt.$$

Vi har $\gamma'(t) = i$, och en räkning visar att

$$f'(z) = \frac{1}{z^2},$$

vilket ger att

$$l(f(K)) = \int_1^2 \left| \frac{1}{(it)^2} \right| dt = \int_1^2 \frac{1}{t^2} dt = \left[-\frac{1}{t} \right]_1^2 = \frac{1}{2}.$$

Vi har alltså $l(K) = 1 \neq 1/2 = l(f(K))$.

Automorfierna i $\text{Möb}(\mathbb{H})$ spelar samma roll i det övre halvplanet som translationer och rotationer gör i det euklidiska planet. Ovanstående resultat visar att båglängdsbegreppet l inte är lämpligt för att mäta kurvors längd i den geometri vi vill bygga upp. Vi såg tidigare att båglängden av en kurva i det euklidiska planet inte förändrades om kurvan roterades eller translaterades, och det är detta vi vill uppnå i $\mathbb{H}$, fast med rotationer och translationer utbytta mot Möbiusavbildningarna på formen (8.1).

Vi är nu redo att definiera begreppet hyperbolisk längd.

Definition 8.1.2. Låt $\gamma: [a, b] \rightarrow \mathbb{H}$ vara en parameterfunktion med bild Γ . Den *hyperboliska längden* av kurvan Γ i övre halvplanet ges av

$$l_{\text{hyp}}(\Gamma) = \int_a^b \frac{1}{\text{Im}|\gamma(t)|} |\gamma'(t)| dt.$$

Idén är alltså att skala om dt i varje punkt $\gamma(t_0)$ på vår kurva med en faktor $1/(\text{Im}(\gamma(t_0)))$ för att kompensera för den extra faktorn vi får ut när vi deriverar en hyperbolisk automorfi f . Ju närmare reella linjen vi är, desto större blir omskalningsfaktorn, och ju längre bort ifrån $\mathbb{R}$ vi rör oss i $\mathbb{H}$, desto mindre blir faktorn. Vi ser också att omskalningen är densamma i alla punkter som ligger på en horisontell linje i $\mathbb{H}$, det vill säga, har samma imaginärdel.

Vi ger ett exempel på hur räkningar med hyperbolisk längd kan se ut.

Exempel 8.1.3. Betrakta det vertikala linjesegment K som förbinder punkterna $p = ia$ och $q = ib$ för $a, b \in \mathbb{R}$ med $a < b$.

En parameterframställning av detta linjesegment är

$$\gamma(t) = it, \quad t \in [a, b].$$

Vi deriverar γ och får att $\gamma'(t) = i$, så $|\gamma'(t)| = 1$. Vi erhåller nu den hyperboliska längden av linjesegmentet genom att beräkna integralen

$$l_{\text{hyp}}(K) = \int_a^b \frac{|\gamma'(t)|}{\text{Im}(\gamma(t))} dt = \int_a^b \frac{1}{t} dt = [\ln t]_a^b.$$

Vi ser alltså att

$$l_{\text{hyp}}(K) = \ln \left(\frac{b}{a} \right),$$

för ett allmänt vertikalt linjesegment, och om vi väljer $b = 2$ och $a = 1$ får vi $l_{\text{hyp}}(K) = \ln 2$.

Vi ser alltså att den hyperboliska längden av ett vertikalt linjesegment skiljer sig från den euklidiska båglängden, som ju är $l(K) = b - a$.

Precis som i definitionen av den euklidiska längden av en kurva verkar det som att valet av parameterframställning skulle kunna påverka integralen som ska beräknas.

Man kan, på ett liknande sätt som tidigare, visa att så inte är fallet — den hyperboliska längden bevaras under omparametrisering.

8.2 Invariants under Möb($\mathbb{H}$)

Vi valde att ändra på vårt tidigare båglängdbegrepp för att skapa en båglängd i övre halvplanet som inte påverkas av automorfier.

Nästa sats visar att vi verkligen har uppnått vårt mål.

Sats 8.2.1. *Låt $f \in \text{Aut}(\mathbb{H})$. Då gäller $l_{\text{hyp}}(\Gamma) = l_{\text{hyp}}(f(\Gamma))$ för varje parameterkurva Γ .*

Bevis. Vi inleder beviset med några räkningar. Låt $f \in \text{Möb}(\mathbb{H})$, det vill säga,

$$f(z) = \frac{az + b}{cz + d},$$

för reella a, b, c, d med $ad - bc = 1$. Vi deriverar f och får

$$f'(z) = \frac{a(cz + d) - c(az + b)}{(cz + d)^2} = \frac{ad - bc}{(cz + d)^2} = \frac{1}{(cz + d)^2}. \quad (8.2)$$

Vi noterar att

$$f(z) = \frac{az + b}{cz + d} = \frac{(az + b)(c\bar{z} + d)}{(cz + d)(c\bar{z} + d)} = \frac{ac|z|^2 + bd + adz + bc\bar{z}}{|cz + d|^2},$$

vilket ger oss att

$$\text{Im}(f(z)) = \text{Im} \left(\frac{ac|z|^2 + bd + adz + bc\bar{z}}{|cz + d|^2} \right).$$

Sätt $z = x + iy$. Eftersom $ac|z|^2 + bd$ är reellt får vi

$$\text{Im}(f(z)) = \frac{1}{|cz + d|^2} \text{Im}(adz + bc\bar{z}) = \frac{(ad - bc)y}{|cz + d|^2}.$$

Vi har ju $ad - bc = 1$, och vi ser att

$$\text{Im}(f(z)) = \frac{y}{|cz + d|^2}.$$

Vi jämför med vårt tidigare uttryck (8.2) och får att

$$|f'(z)| = \frac{\operatorname{Im}(f(z))}{\operatorname{Im}(z)}. \quad (8.3)$$

Vi kommer nu till huvuddelen av beviset. Låt Γ vara en kurva som parametriseras av $\gamma : [a, b] \rightarrow \mathbb{C}$. Den hyperboliska längden av Γ ges av

$$l_{\text{hyp}}(\Gamma) = \int_a^b \frac{|\gamma'(t)|}{\operatorname{Im}(\gamma(t))} dt.$$

Vi tillämpar sedan vår automorfi f på kurvan. Den resulterande kurvan $f(\Gamma)$ har då längden

$$l_{\text{hyp}}(f(\Gamma)) = \int_a^b \frac{|[f(\gamma(t))]'|}{\operatorname{Im}[f(\gamma(t))]} dt$$

Kedjeregeln ger oss att

$$[f(\gamma(t))]' = \gamma'(t) \cdot f'(\gamma(t)),$$

och vi ser att integralen vi ska beräkna blir

$$\int_a^b |\gamma'(t)| \frac{|f'(\gamma(t))|}{\operatorname{Im}[f(\gamma(t))]} dt. \quad (8.4)$$

Från (8.3) följer det att

$$\frac{|f'(\gamma(t))|}{\operatorname{Im}[f(\gamma(t))]} = \frac{1}{\operatorname{Im}(\gamma(t))}.$$

Vi sätter in detta i integralen (8.4) och får att

$$l_{\text{hyp}}(f(\Gamma)) = \int_a^b \frac{|\gamma'(t)|}{\operatorname{Im}(\gamma(t))} dt.$$

Alltså gäller $l_{\text{hyp}}(\Gamma) = l_{\text{hyp}}(f(\Gamma))$ för varje $f \in \operatorname{Möb}(\mathbb{H})$, och detta avslutar beviset. $\square$

8.3 Hyperboliskt avstånd och hyperboliska geodeter

Vi inför nu ett avståndsbegrepp i vår hyperboliska geometri. Vi vill säga att avståndet mellan p och q i $\mathbb{H}$ ges av längden på den parameterkurva som förbinder p och q och har kortast längd, i hyperbolisk mening.

Liksom i fallet med den euklidiska båglängden visar det sig att linjer har kortare båglängd än alla andra kurvor.

Sats 8.3.1. *Låt p och q vara två punkter i det övre halvplanet $\mathbb{H}$ och låt K beteckna det unika hyperboliska linjesegment som förbinder p och q . Då gäller*

$$l_{\text{hyp}}(K) \leq l_{\text{hyp}}(\Gamma)$$

för varje parameterkurva Γ som förbinder p och q .

Bevis. Vi delar upp beviset i två fall: dels $\operatorname{Re}(p) = \operatorname{Re}(q)$ och dels $\operatorname{Re}(p) \neq \operatorname{Re}(q)$. Vi börjar med att behandla det första fallet $\operatorname{Re}(p) = \operatorname{Re}(q)$. Automorfin $f(z) = z - \operatorname{Re}(p)$ avbildar punkterna på den hyperboliska linjen som förbinder p och q , som i det här fallet är en rät linje, på imaginära axeln. Den bevarar vidare den hyperboliska båglängden av alla kurvor som förbinder punkterna p och q . Vi kan alltså förenkla situationen genom att anta att $p = ia$ och $q = ib$ och $a < b$. Låt alltså parameterfunktionen

$$\gamma(t) = x(t) + iy(t), \quad t \in [\alpha, \beta],$$

beskriva en kurva som förbinder $p = \gamma(\alpha)$ och $q = \gamma(\beta)$. Vi observerar att

$$\tilde{\gamma}(t) = iy(t), \quad t \in [\alpha, \beta],$$

ger en parameterfunktion vars bild är den bit av imaginära axeln som förbinder p och q . Eftersom $(x'(t))^2$ och $(y'(t))^2$ är positiva för varje t gäller

$$|\gamma'(t)| = \sqrt{(x'(t))^2 + (y'(t))^2} \geq \sqrt{(y'(t))^2} = |y'(t)|.$$

Denna olikhet medför tillsammans med det faktum att $\operatorname{Im}(\gamma(t)) = \operatorname{Im}(\tilde{\gamma}(t))$ att

$$l_{\text{hyp}}(\Gamma) = \int_{\alpha}^{\beta} \frac{|\gamma'(t)|}{\operatorname{Im}(\gamma(t))} dt \geq \int_{\alpha}^{\beta} \frac{|y'(t)|}{\operatorname{Im}(\tilde{\gamma}(t))} dt = l_{\text{hyp}}(K), \quad (8.5)$$

där K är den räta linjen som förbinder p och q .

Antag nu att $\operatorname{Re}(p) \neq \operatorname{Re}(q)$. Vi kan för enkelhetens skull anta att $\operatorname{Re}(p) > \operatorname{Re}(q)$. Vi använder Sats 7.2.4 för att hitta en automorfi f som avbildar den hyperboliska linjen mellan p och q på imaginära axeln. Vi har då $f(p) = ia$ och $f(q) = ib$ för några reella tal a och b .

Låt Γ vara en kurva som förbinder p och q . Då kommer $f(\Gamma)$ vara en ny kurva som förbinder punkterna ia och ib . Eftersom $l_{\text{hyp}}(\Gamma) = l_{\text{hyp}}(f(\Gamma))$ enligt Sats 8.2.1 kommer den avbildade kurvan ha samma hyperboliska längd som den ursprungliga kurvan Γ . Vårt tidigare resonemang i (8.5) medför att

$$l_{\text{hyp}}(f(\Gamma)) \geq l_{\text{hyp}}(K),$$

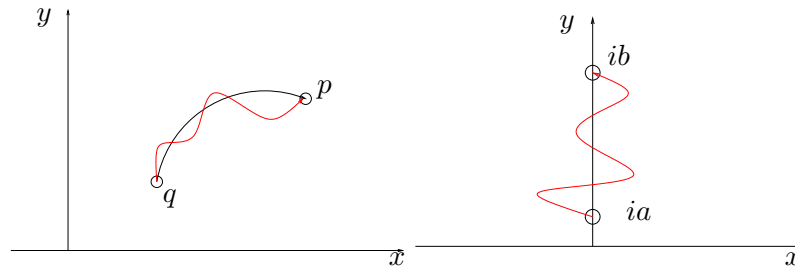
där K liksom tidigare är linjesegmentet som förbinder ia och ib , det vill säga, den unika hyperboliska linjen mellan ia och ib .

Vi utnyttjar nu att f^{-1} också är en Möbiusavbildning och således bevarar hyperbolisk längd. Vi avbildar tillbaka och får

$$l_{\text{hyp}}(\Gamma) \geq l_{\text{hyp}}(f^{-1}(K)).$$

Sats 7.2.3 säger emellertid att automorfierna av $\mathbb{H}$ bevarar de hyperboliska linjerna, och således måste $\Gamma' = f^{-1}(K)$ vara en hyperbolisk linje. Eftersom vi antagit att $\operatorname{Re}(p) \neq \operatorname{Re}(q)$ är Γ' en cirkel som möter $\mathbb{R}$ i en rät vinkel. Detta avslutar beviset. $\square$

Vi använder denna sats för att definiera det hyperboliska avståndet. Om p och q tillhör $\mathbb{H}$ och $\operatorname{Re}(p) = \operatorname{Re}(q)$ säger vi att det hyperboliska avståndet mellan p



Figur 8.1: Vänster: Kurvan Γ och den hyperboliska linjen mellan p och q . Höger: Bilden av Γ under f .

och q är den hyperboliska längden av det vertikala linjesegmentet som förbinder dem, det vill säga

$$\text{dist}_{\text{hyp}}(p, q) = \left| \ln \left(\frac{\text{Im}(p)}{\text{Im}(q)} \right) \right|$$

vilket vi såg i Exempel 8.1.3. Vi har lagt till absolutbeloppet så att vi ska slippa göra antagandet $\text{Im}(q) < \text{Im}(p)$.

Om p och q inte har samma realdel är det inte längre en rät linje mellan p och q som har minst båglängd. Vi låter istället det hyperboliska avståndet mellan p och q ges av den hyperboliska längden av den cirkelbåge som skär reella axeln och innehåller p och q . Lite räkningar, som återfinns i [5, Kapitel 5.4], ger oss formeln

$$\text{dist}_{\text{hyp}}(p, q) = \ln \left(\frac{|p - \bar{q}| + |p - q|}{|p - \bar{q}| - |p - q|} \right).$$

Vi sammanfattar denna diskussion i en definition.

Definition 8.3.2. Låt p och q vara två distinkta punkter i övre halvplanet $\mathbb{H}$. Då ges *hyperboliska avståndet* mellan p och q av

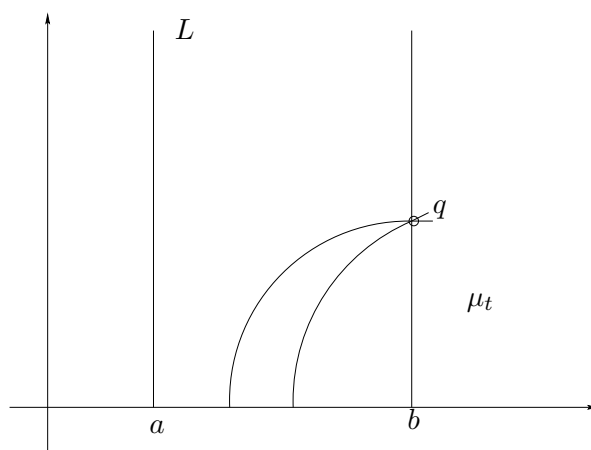
$$\text{dist}_{\text{hyp}}(p, q) = l_{\text{hyp}}(\Lambda),$$

där Λ är den unika hyperboliska linje som förbinder p och q .

För att sammanfatta har vi definierat vad en *punkt* i hyperboliska planet är: det är ett element av $\mathbb{H}$. Vi har definierat den rätta linjen mellan två punkter, och vi har definierat ett sätt att mäta längden på kurvor som gör att dessa rätta linjer är exakt de avståndsminimerande kurvorna. För att kunna arbeta med alla de begrepp från euklidisk geometri vi känner sedan tidigare även i det hyperboliska planet, återstår nu endast att definiera *vinkeln* mellan två kurvor i hyperboliska planet. Detta är nu enkelt:

Definition 8.3.3. Låt två linjesegment λ och μ mötas i en punkt $p \in \mathbb{H}$. Vi definierar vinkeln från λ till μ som den vanliga euklidiska vinkeln, där vi tänker på λ och μ som kurvor i det vanliga euklidiska planet $\mathbb{R}^2$.

För att denna definition ska kunna vara vettig, måste vi veta att en automorfi $f \in \text{Aut}(\mathbb{H})$ inte förändrar vinkeln mellan två kurvor. Men detta har vi redan visat för två kapitel sedan! Sats 6.3.10 visar att varje Möbiusavbildning bevarar den vanliga euklidiska vinkeln mellan kurvor. Vinkelbegreppet från Definition 8.3.3 är alltså invariant under den automorfigrupp vi har sagt att det hyperboliska planet skall ha.



Figur 8.2: Flera parallella linjer.

8.4 Hyperbolisk geometri

Vi avslutar detta kapitel med några geometriska betraktelser. Vi börjar med att definiera vad vi ska mena med att två hyperboliska linjer är parallella.

Definition 8.4.1. Två hyperboliska linjer säges vara *parallella* om de är disjunkta, det vill säga, om de inte har några punkter gemensamma.

Denna definition är helt analog med den i avsnitt 2.4: vi sade att två linjer i planet var parallella om de inte möttes.

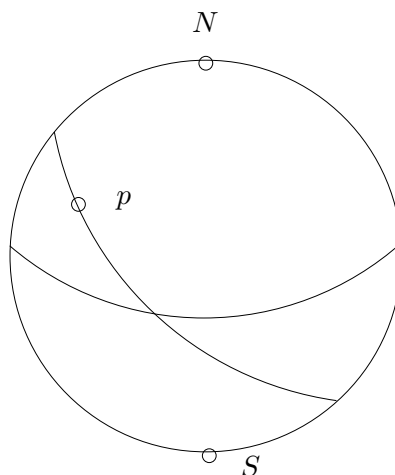
Vi kommer nu till en av höjdpunkterna i detta kompendium: vi ska visa att det för varje punkt q finns *oändligt många* hyperboliska linjer i $\mathbb{H}$ genom punkten q som är parallella med en given linje λ !

Sats 8.4.2. Låt λ vara en hyperbolisk linje, och låt q vara en punkt som inte tillhör λ . Då existerar oändligt många hyperboliska linjer som innehåller q och är parallella med λ .

Bevis. Vi kan utan inskränkning anta att λ är en rät linje som skär reella axeln vinkelrätt och har $\operatorname{Re}(p) = a > 0$ för alla $p \in \lambda$. Om så inte är fallet kan vi tillämpa en Möbiusavbildning för att reducera situationen till detta fall, se Sats 7.2.4.

Låt $q \in \mathbb{H}$ ha $\operatorname{Re}(q) = b$ och låt oss antaga att $b > a$. För varje $t \in (a, b]$ finns det enligt Sats 7.2.2 en unik hyperbolisk linje μ_t som passerar genom t och även q . När $t = b$ är μ_t en rät linje som är vinkelrät mot $\mathbb{R}$. För varje annat $t \in (a, b]$ är μ_t en cirkelbåge med centrum på reella linjen, vilket vi såg i beviset för Sats 7.2.2.

Eftersom vi inte tillåtit $t \neq a$ ser vi att λ och μ_t inte har några punkter gemensamma, och därmed är varje hyperbolisk linje μ_t parallell med λ . $\square$



Figur 8.3: Varje linje genom p skär ekvatorn.

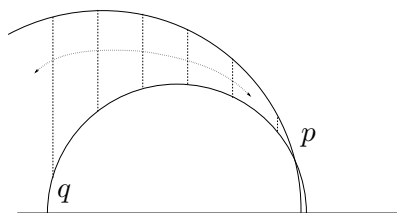
8.5 Sfärisk geometri

Vi tillåter oss en liten utblick mot sfärisk geometri. På en sfär är *storcirkarna* de kurvor som minimerar den motsvarande sfäriska längden mellan två punkter. Vi avstår från att definiera exakt vad vi menar med sfärisk längd. Vi säger att två sfäriska linjer är parallella om de inte har några punkter gemensamt. Vi ser nu att följande gäller på en sfär: givet en sfärisk linje, det vill säga en storcirkel, och en punkt p på sfären existerar det *ingen* sfärisk linje genom p som är parallell med den givna linjen! Vi måste dock observera en sak här: sfärisk geometri med storcirklar som linjer bryter inte bara mot Euklides femte postulat, utan även mot det första! Det finns exempelvis oändligt många linjer som innehåller de två punkterna N och S , nordpolen och sydpolen—alla longitudlinjerna. Det är alltså inte fullt så enkelt som att titta på geometri på en sfär att hitta ett motexempel mot parallellpostulatet.

8.6 Areor, vinklar och Gauss-Bonnets sats

Ett annat intressant exempel på en skillnad mellan euklidisk och icke-euklidisk geometri är vinkelsumman i en triangel. I det euklidiska planet är det välkänt att varje triangel har vinkelsumman π ; så är inte fallet i det hyperboliska planet. Det visar sig i stället att *varje* triangel kommer att ha en vinkelsumma som är strikt mindre än π . Inte nog med detta: det kommer även vara så att olika stora trianglar kommer att ha olika stor vinkelsumma. Vi kommer inte att visa detta, men påståendet illustreras i Figur 8.4.

Vi tänker på bilden som en familj av trianglar, där en vertikal sida (ritad med streckade linjer i figuren) tillåts röra sig fritt, och de andra två är fixerade. När den vertikala sidan är nästan vid skärningspunkten p ser triangeln nästan exakt likadan ut som en vanlig triangel, och vinkelsumman bör därför vara nära π . När den vertikala sidan rör sig åt vänster är det däremot mycket tydligt att vi har en hyperbolisk triangel. När linjen närmar sig punkten q kommer vinkeln mellan den vertikala och nedre sidan att gå mot noll, medan vinkeln mellan



Figur 8.4: En familj av hyperboliska trianglar.

den vertikala och den övre sidan håller sig begränsad. (I vårt exempel blir den inte mycket större än $\pi/2$.) Det som syns i bilden är exempel på:

Sats 8.6.1. *Areal av varje hyperbolisk triangel T är direkt proportionell mot differensen mellan π och T :s vinkelsumma.*

Med andra ord: ju större triangeln är, desto mindre blir vinkelsumman. Speciellt kan vi inte längre prata om kongruenta trianglar i det hyperboliska planet. I det euklidiska fallet kan vi säga att två trianglar är likformiga om alla deras vinklar är desamma, och i så fall kommer de två triangelarna endast skilja sig åt upp till skalning. Men i det hyperboliska fallet kommer två likformiga trianglar att ha samma area och är verkligen likadana! Men den mest förbluffande konsekvens av denna sats är ändå att *det finns en övre gräns för hur stor area en triangel kan ha*. För arean av varje triangel är proportionell mot hur mycket dess vinkelsumma misslyckas med att nå upp till π , och eftersom vinkelsumman definitivt är större än noll är varje triangelns area strikt begränsad av proportionalitetskonstanten gånger π .

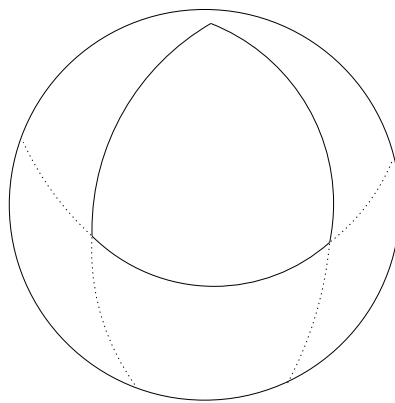
Det är viktigt att här vara tydlig med att vi använder *hyperbolisk area*, som vi inte definierat i den här kursen. Ett sätt att tänka på hyperbolisk area är att det är ett sätt att mäta storleken på delmängder av $\mathbb{H}$, och det är det enda sätt att göra detta så att arean inte förändras av en automorfi i $\text{Aut}(\mathbb{H})$. På samma sätt som med hyperbolisk längd, kan man konstruera den hyperboliska arean genom att modifiera euklidisk area med en förstoringsfaktor som beror på hur nära reella axeln området vars area ska mätas är. Hade vi använt vanlig euklidisk area i föregående sats hade Sats 8.6.1 inte stämt.

Även här kan vi göra en utblick mot den sfäriska geometrin. Även på en sfär kan vi tala om en triangel som en yta som begränsas av tre storcirklar. Funderar man på hur vinklarna på en sfär borde bete sig, anar man snart att det omvända kommer att hända i denna situation. Vinkelsumman blir alltid större än π . På samma sätt som i det hyperboliska planet är skillnaden mellan en triangel på en sfär och en vanlig triangel inte särskilt stor för små trianglar (jämför detta med situationen på jordklotet — det krävs att man ritar en ganska stor triangel på marken för att jordkrökningen ska märkas). Samma sats gäller även här:

Sats 8.6.2. *Areal av varje triangel T på ytan av en sfär är direkt proportionell mot differensen mellan T :s vinkelsumma och π .*

Även för en sfär får vi samma konsekvenser, att likformiga trianglar har samma area och att det existerar en övre gräns för arean på en triangel.

Dessa satser kan formuleras i en enda gemensam formel:



Figur 8.5: En triangel på en sfär, där alla tre vinklarna är räta.

Sats 8.6.3. (*Gauss-Bonnets sats*) För varje geometri — euklidisk, hyperbolisk, sfärisk — existerar en konstant K med egenskapen att för alla trianglar T gäller att

$$K \cdot \text{area}(T) = \alpha + \beta + \gamma - \pi,$$

där α, β och γ är vinklarna i triangeln T . Om $K < 0$ är geometrin hyperbolisk, om $K = 0$ är geometrin euklidisk, och om $K > 0$ är geometrin sfärisk.

Detta vackra resultat är en passande avslutning på detta kompendium.

Övningar

Övning 8.1. Vilka är geodeterna mellan i och $3i$ samt mellan $1 + 5i$ och $-2 + 4i$?

Övning 8.2. Låt $\gamma: [a, b] \rightarrow \mathbb{H}$ vara en parameterfunktion med bild Γ och låt $h: [\alpha, \beta] \rightarrow [a, b]$ vara en omparametrisering, se Definition 4.3.1.

Visa att bilden Γ' av $\gamma \circ h: [\alpha, \beta] \rightarrow \mathbb{H}$ har samma hyperboliska båglängd som Γ .

Övning 8.3. Visa att varje hyperbolisk linje har oändlig hyperbolisk längd.

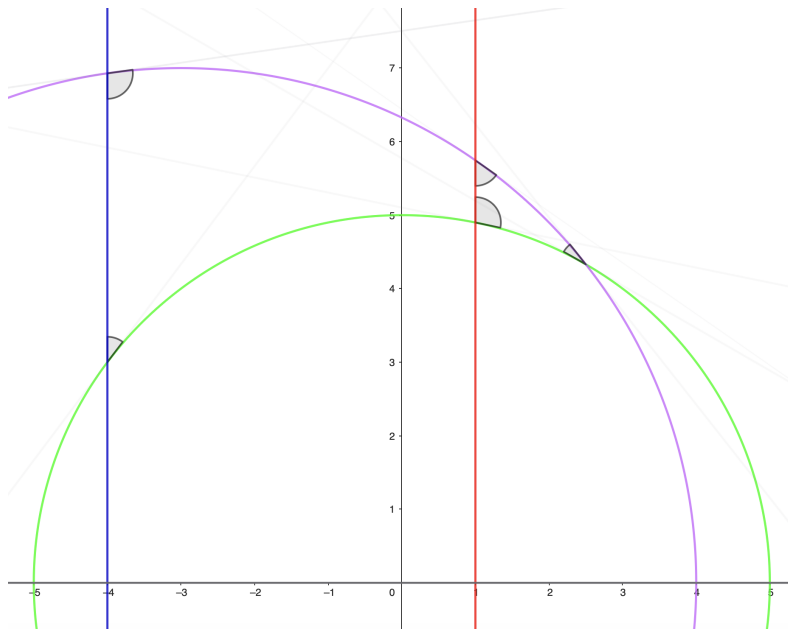
Övning 8.4. Beräkna den hyperboliska båglängden av kurvan $\Gamma \subset \mathbb{H}$ som bestäms av parameterfunktionen

$$\gamma(t) = t + i(1 + t), \quad t \in [0, 1].$$

Övning 8.5. Ge exempel på en funktion som bevarar hyperboliska avstånd men ej euklidiska.

Övning 8.6. Rita fyrhörningen i hyperboliska planet som begränsas av de hyperboliska linjerna vars ekvation i komplexa planet ges av $\text{Re}(z) = 1$, $\text{Re}(z) = -1$, $|z| = \sqrt{2}$ och $|z| = 2$. Räkna ut dess vinkelsumma och jämför med vinkelsumman av en fyrhörning i euklidiska planet.

Övning 8.7. Hitta linjesegment som har samma hyperboliska och euklidiska längd. Går det att hitta vertikala sådana?



Figur 8.6

Övning 8.8. Rita upp cirkelarna $|z| = 5$ och $|z + 3| = 7$ i övre halvplanet. Dra linjerna $\operatorname{Re} z = -4$ och $\operatorname{Re} z = 1$. Detta definierar två hyperboliska trianglar, en mindre som ligger i en större, se Figur 8.6. Beräkna trianglarnas vinkelsummor och jämför med areorna.

Lösningar till udda övningsuppgifter

Kapitel 1

Övning 1.1. (i) 0, 1, 2, 3, 4.

(ii) 1, 2, {2, 3}.

(iii) -3, -2, -1, 0, 1, 2, 3.

(iv) 1, 2.

(v) -3, -2, -1, 1, 2, {2, 3}.

Övning 1.3. (i) B är en äkta delmängd av A .

(ii) A och B är lika.

(iii) Mängderna är disjunkta eftersom $B = \emptyset$ och alla mängder är disjunkta med den tomma mängden. B är en äkta delmängd till A .

(iv) Mängderna är varken disjunkta, lika eller äkta delmängder av varandra.

(v) Mängderna är disjunkta eftersom $A = \emptyset$.

Övning 1.5. Observera att dessa svar är förslag. Uppgifterna har flera korrekta svar.

(i) $\{n \in \mathbb{Z} \mid n = 2k \text{ för något heltal } k \geq 1\}$.

(ii) $\{p/2 \mid p \text{ är ett heltal}\}$.

(iii) $\{r \in \mathbb{R} \mid r \notin \mathbb{Q} \text{ och } |r| < 1\}$.

Övning 1.7. (i) Definitionsmängd: $\{1, 2, 3, \dots\}$. Målmängd: $\mathbb{N}$.

(ii) Definitionsmängd: $\{T \mid T \text{ är en triangel}\}$. Målmängd: $\mathbb{R}$.

(iii) Definitionsmängd: $\{p(x) \mid p(x) \text{ är ett andragradspolynom}\}$. Målmängd: $\{p(x) \mid p(x) \text{ är ett förstegradspolynom}\}$.

Övning 1.9. (i) Detta är en funktion. Den är definierad för alla värden i definitionsmängden, den ger alltid samma värde för ett givet argument, och alla funktionsvärden ligger i målmängden.

(ii) Detta är inte en funktion, då funktionens värden inte ligger i målmängden ($f(2) = \sqrt{2} \notin \mathbb{Q}$).

(iii) Detta är inte en funktion, eftersom dess värden är slumpmässiga.

(iv) Detta är en funktion. Eftersom ordet Balkong börjar på B, så har funktionen ett definierat värde som ligger i målmängden.

Övning 1.11. (i) Funktionerna är lika. De har samma definitionsmängd, målmängd och $\sqrt{x^2} = |x|$ för alla x .

(ii) Funktionerna är inte lika, då deras definitionsmängder inte är samma.

(iii) Funktionerna är inte lika, då deras målmängder är olika.

Övning 1.13. I exempel 1.5.3 är värdemängden $\{3, 4\}$. Funktionen är inte injektiv och inte surjektiv. I exempel 1.5.4 är värdemängden $\{0, 1\}$ och funktionen är bijektiv, alltså surjektiv och injektiv.

Övning 1.15. Vi gör ett direkt bevis. Om $A \cap B = \emptyset$ betyder det att inga element i A tas bort i $A \setminus B$. Alltså har vi $A \cap B = \emptyset \implies A \setminus B = A$. På samma sätt har vi att om $A \setminus B = A$ betyder det att det inte finns några gemensamma element i A och B som tas bort i $A \setminus B$, alltså har vi $A \setminus B = A \implies A \cap B = \emptyset$.

Övning 1.17. Vi gör ett direkt bevis. Mängden $A \setminus B$ är mängden av element som ligger i A och inte i B . Alltså är det mängden vi får när vi 'tar bort elementen som ligger i B från A '. Alltså tar vi egentligen bara bort elementen som ligger i $A \cap B$ från A . De som inte låg i A från första början struntar vi alltså i.

Övning 1.19. Vi gör ett direkt bevis. Hur många par av element A respektive B finns det? Jo, exakt lika många som det finns sätt att välja ett element ur den ena och sedan den andra vilket är $(|A|) \cdot (|B|)$.

Övning 1.21. Om n är jämnt så finns det per definition ett heltal k så att $n = 2k$. Då gäller att $n + 1 = 2k + 1$, det vill säga $n + 1$ är udda.

Övning 1.23. Antag motsatsen, det vill säga att det finns ett rationellt tal p/q och ett irrationellt tal r vars summa är rationell. Skriv summan som kvoten s/t . Då gäller

$$\frac{p}{q} + r = \frac{s}{t} \implies r = \frac{s}{t} - \frac{p}{q} = \frac{sq - pt}{tq}$$

genom att skriva vänsterledet på gemensamt bråkstreck. Men detta bevisar att r är rationellt, vilket är en motsägelse.

Övning 1.25. Antag motsatsen till satsen. Den kan delas in i två fall. I det ena fallet gäller att $ab = c$ och $a < \sqrt{c}$ och $b < \sqrt{c}$. Då gäller att

$$ab < a\sqrt{c} < \sqrt{c}\sqrt{c} = \sqrt{c}^2 = c.$$

Alltså gäller $ab < c$, vilket är en motsägelse.

I det andra fallet gäller att $ab = c$ och $a > \sqrt{c}$ och $b > \sqrt{c}$. Då gäller att

$$ab > a\sqrt{c} > \sqrt{c}\sqrt{c} = \sqrt{c}^2 = c.$$

Alltså gäller $ab > c$, vilket är en motsägelse.

Övning 1.27. Låt $r = p/q$ vara ett rationellt tal. Genom att förlänga med $\sqrt{2}$ kan vi skriva

$$r = \sqrt{2} \frac{r}{\sqrt{2}}.$$

Vi vet att $\sqrt{2}$ är irrationellt. Om vi kunde visa att $r/\sqrt{2}$ är irrationellt så vore vi klara.

Antag motsatsen, det vill säga att $r/\sqrt{2}$ är en kvot av två heltal a och b . Genom algebraiska manipulationer får vi att

$$\frac{r}{\sqrt{2}} = \frac{a}{b} \implies \frac{1}{\sqrt{2}} = \frac{qa}{pb} \implies \sqrt{2} = \frac{pb}{qa}.$$

Men detta innebär att $\sqrt{2}$ är rationellt, vilket är en motsägelse. Alltså är $r/\sqrt{2}$ irrationellt, vilket avslutar vårt bevis.

Kapitel 2

Övning 2.1. (i) Ta $f((x, y)) = 2(x, y)$. Om L är icke-vertikal linje ges den av mängden $L = \{(x, ax+b), x \in \mathbb{R}\}$ för $a, b \in \mathbb{R}$. Då är $f(L) = \{2(x, ax+b), x \in \mathbb{R}\} = \{(x', ax' + 2b), x' \in \mathbb{R}\}$. En linje $x = a$ avbildas på $x = 2a$. I båda fallen translateras bara linjerna, därmed bevaras vinklar. Dock gäller $d((0, 0), (0, 1)) = 1 \neq 2 = d((0, 0), (0, 2))$, så avstånd bevaras ej.

(ii) Nej, om vinkeln två linjer skär varandra i är densamma före och efter funktionen måste speciellt linjerna i fråga avbildas till räta linjer. Vi har bara definierat vinklar mellan linjer.

(iii) Om $ax + by = c$ gäller $a(x+1) + b(y+1) = c + a + b$ så $f((x, y))$ verifierar ekvationen $ax + by = c + a + b$.

Övning 2.3. (i) Betrakta linjesegment L och M som utgår från en punkt $P = (x_p, y_p)$. Förläng L och M så att de har längd minst 1. Kalla skärningspunkterna med cirkeln av radie 1 och medelpunkt P för $P_M = (x_M, y_M)$ och $P_L = (x_L, y_L)$, respektive. Kalla längden av cirkeln från P_L till P_M för l . Då är vinkeln mellan L och M är l . Under $f((x, y)) = (x + a, y + b)$ transformeras punkterna P, P_M och P_L till $P' = (x_p + a, y_p + b)$, $P'_M = (x_M + a, y_M + b)$ och $P'_L = (x_L + a, y_L + b)$, respektive. Eftersom längder av segmenten är $|PP_M| = |PP_L| = 1$ gäller

$$\begin{aligned} |P'P'_M| &= \sqrt{((x_M + a) - (x_p + a))^2 + ((y_M + b) - (y_p + b))^2} \\ &= \sqrt{(x_M - x_p)^2 + (y_M - y_p)^2} \\ &= |PP_M| \\ &= 1 \end{aligned}$$

och analogt gäller $|P'P'_L| = 1$. Om vi antar att cirkelsegmentet bevarar längd under translation har vi att cirkeln med radie 1 med medelpunkt i P' från P'_L till P'_M har längd l . Men detta är precis vinkeln mellan L' och M' och vi är klara.

(ii) Betrakta $P = (0, 0)$, $L = \frac{1}{\sqrt{2}}(-1, 1)$, $M = \frac{1}{\sqrt{2}}(1, 1)$. Då är vinkeln mellan linjesegmenten PM och PL längden av den del av enhetscirkeln som går moturs från M till L . Under reflektion sänds punkterna till $P' = (0, 0)$, $L' = \frac{1}{\sqrt{2}}(-1, -1)$ och $M' = \frac{1}{\sqrt{2}}(1, -1)$, respektive. Segmenten PL och PM sänds till $P'L'$ och $P'M'$, respektive. Vinkeln mellan $P'M'$ och

$P'L'$ är längden av den del av enhetscirkeln som går moturs från M' till L' . Men för att gå från LM' till L' moturs passerar vi M och L moturs så längden av den del av enhetscirkeln som går moturs från M' till L' är längre än den del av enhetscirkeln som går moturs från M till L . Alltså är vinkeln mellan $P'M'$ och $P'L'$ större än vinkeln mellan PM och PL och inte lika.

Övning 2.5. Låt λ vara linjen $ax + by = c$, och $f(t)$ (som i uppgiften) funktionen

$$f(t) = \left(\frac{ac}{a^2 + b^2} + bt, \frac{bc}{a^2 + b^2} - at \right).$$

Vi vill visa att bilden av f är exakt λ . Vi visar först att bilden av f är en delmängd till λ . Så tag en punkt $f(t)$ och sätt in dess koordinater i ekvationen för λ . Vi finner att

$$a \cdot \left(\frac{ac}{a^2 + b^2} + bt \right) + b \cdot \left(\frac{bc}{a^2 + b^2} - at \right) = \frac{a^2c}{a^2 + b^2} + \frac{b^2c}{a^2 + b^2} + abt - bat = c.$$

Alltså satisfierar $f(t)$ linjens ekvation för alla t . Låt nu $(x_0, y_0) \in \lambda$. Vi vill visa att det finns ett t_0 sådant att $f(t_0) = (x_0, y_0)$. Vi har två ekvationer att lösa:

$$\frac{ac}{a^2 + b^2} + bt_0 = x_0,$$

och

$$\frac{bc}{a^2 + b^2} - at_0 = y_0.$$

Eftersom inte både a och b tilläts vara noll kommer minst en av dessa ekvationer att ha en unik lösning. Låt oss t.ex. anta att $b \neq 0$, så att det finns en unik lösning t_0 till den övre ekvationen. Vi vill visa att $f(t_0) = (x_0, y_0)$, så att t_0 även löser den nedre ekvationen. Men vi vet redan att $f(t_0)$ uppfyller ekvationen $ax + by = c$, och eftersom b är nollskilt kan y lösas ut ur linjens ekvation. Alltså vet vi att eftersom x -koordinaten i $f(t_0)$ är lika med x_0 , måste även y -koordinaten vara lika med y_0 .

Övning 2.7. Låt oss börja med en observation. Låt λ vara linjen $ax + by = c$, och antag att $(x_0, y_0) \in \lambda$, $(x_1, y_1) \in \lambda$. I så fall kommer punkten $(x_0 - x_1, y_0 - y_1)$ att uppfylla ekvationen

$$ax + by = 0.$$

Eftersom minst en av a och b är noll, kan vi skriva antingen

$$\frac{a}{b} = -\frac{y_0 - y_1}{x_0 - x_1} \quad \text{eller} \quad \frac{b}{a} = -\frac{x_0 - x_1}{y_0 - y_1}.$$

Geometriskt säger detta att om vi har två punkter kan vi räkna ut lutningen på linjen som en differenskvot, och att lutningen på linjen är lika med $-\frac{b}{a}$ om $a \neq 0$.

Det följer att om $ax + by = c$ definierar samma linje som $a'x + b'y = c'$ så är $a = 0$ precis om $a' = 0$, och $b = 0$ precis om $b' = 0$. Antag utan inskränkning att $a \neq 0$, $a' \neq 0$. Då är

$$\frac{b}{a} = \frac{b'}{a'}.$$

Att dessa bråk är lika betyder att det finns en nollskild konstant ρ sådan att $a = \rho a'$, $b = \rho b'$. Vi är nu nästan klara: vi måste bara visa att det även gäller att $c = \rho c'$. Men detta är nu lätt: vi vet att till exempel (x_0, y_0) uppfyller både ekvationerna $ax + by = c$ och $a'x + b'y = c'$. Alltså kan vi räkna ut att

$$c' = a'x_0 + b'y_0,$$

medan

$$c = ax_0 + by_0 = \rho a'x_0 + \rho b'y_0 = \rho(a'x_0 + b'y_0) = \rho c',$$

vilket avslutar beviset.

Kapitel 3

Övning 3.1.

- (i) Vi beräknar först summan $z + w = 3 + (-1) + (1 + 4)i = 2 + 5i$ och differensen $z - w = 3 - (-1) + (1 - 4)i = 4 - 3i$. Produkten av z och w blir

$$zw = 3 \cdot (-1) - 4 \cdot 1 + i(3 \cdot 4 + 1 \cdot (-1)) = -7 + 11i.$$

Eftersom $|-1 + 4i|^2 = 4^2 + 1^2 = 17$ får vi

$$\frac{z}{w} = \frac{(3 + i)(-1 - 4i)}{(-1 + 4i)(1 - 4i)} = \frac{-3 + 4 + i(-12 - 1)}{17} = \frac{1}{17} - \frac{13}{17}i.$$

- (ii) Vi har $z + w = 0 + 2 + i(-\sqrt{2} - 5) = 2 - (5 + \sqrt{2})i$, och differensen blir $z - w = 0 - 2 + i(-\sqrt{2} - (-5)) = -2 + (5 - \sqrt{2})i$. Vi får vidare

$$zw = 0 \cdot 2 - (-\sqrt{2})(-5) + i(0 \cdot (-5) - 2 \cdot \sqrt{2}) = -5\sqrt{2} - 2\sqrt{2}i,$$

och kvoten av z och w blir

$$\frac{z}{w} = \frac{-\sqrt{2}i \cdot (2 + 5i)}{4 + 25} = \frac{5\sqrt{2}}{29} - \frac{2\sqrt{2}}{29}i.$$

Övning 3.3. Låt $z = a + ib$ och $w = c + id$. Vi har enligt definitionen av komplex multiplikation att $zw = ac - bd + i(ad + bc)$. Absolutbeloppet av produkten zw blir nu

$$\begin{aligned} |zw| &= |(ac - bd) + i(ad + bc)| \\ &= \sqrt{(ac - bd)^2 + (ad + bc)^2} \\ &= \sqrt{(a^2c^2 - 2abcs + b^2d^2) + (a^2d^2 + 2abcd + b^2c^2)} \\ &= \sqrt{(a^2 + b^2)(c^2 + d^2)} = \sqrt{a^2 + b^2} \sqrt{c^2 + d^2}. \end{aligned}$$

Eftersom $|z| = \sqrt{a^2 + b^2}$ och $|w| = \sqrt{c^2 + d^2}$ enligt definitionen av absolutbelopp är det sökta påståendet nu visat.

Övning 3.5. Vi kvadratkompletterar och får:

$$z^2 + iz + 2 = \left(z + \frac{i}{2}\right)^2 + \frac{1}{4} + 2 = \left(z + \frac{i}{2}\right)^2 + \frac{9}{4}.$$

Vi får alltså ekvationen

$$\left(z + \frac{i}{2}\right)^2 = -\frac{9}{4},$$

och därmed har vi

$$z + \frac{i}{2} = \pm \frac{3}{2}i.$$

Detta ger oss de två lösningarna till ekvationen, nämligen

$$z = -2i \quad \text{och} \quad z = i.$$

Övning 3.7. Sammansättningen av f och g ges av

$$(f \circ g)(z) = f((g(z))) = \alpha(\tilde{\alpha}z + \tilde{\beta}) + \beta = \alpha\tilde{\alpha}z + \alpha\tilde{\beta} + \beta.$$

Vi sätter $A = \alpha\tilde{\alpha}$ och $B = \alpha\tilde{\beta} + \beta$ och får att $(f \circ g)(z) = Az + B$.

Det återstår att visa att $|A| = 1$. Eftersom $|\alpha| = |\tilde{\alpha}| = 1$ får vi $|A| = |\alpha||\tilde{\alpha}| = 1$ (se övningen ovan), och vi är klara.

Övning 3.9. (i) $f(z) = z \cdot z$ så enligt produktregeln är $f'(z) = z \cdot 1 + 1 \cdot z = 2z$

(ii) Vi beräknar först derivatan av $z \mapsto z^3$, vi får att $(z^3)' = (z^2 z)' = (z^2)' z + z' (z^2) = 2z \cdot z + z^2 = 3z^2$. Vi har nu att

$$f'(z) = \frac{0 \cdot z^3 - 3z^2 \cdot 1}{z^6} = \frac{1}{z^5}$$

vi noterar att funktionen inte är deriverbar i punkten $z = 0$.

(iii) Derivatan av $z \mapsto z^3 - 1$ är $3z^2$, så vi får att

$$f'(z) = \frac{2z \cdot z^3 - z^2 \cdot 3z^2}{(z^3 - 1)^2} = -\frac{z^4}{(z^3 - 1)^2}.$$

Funktionen är deriverbar i alla punkter där $z^3 \neq 1$. Skriver vi z på formen $z = re^{i\theta}$ säger detta att $re^{3i\theta} = 1$ så vi kräver att $r = 1$ och att $3\theta = 2\pi k$ för något heltal $k \in \mathbb{Z}$. (Detta för att $e^{2\pi k} = 1$ för alla heltal k .) Vi får därmed tre lösningar, $z = 1$, $z = e^{i\frac{2\pi}{3}}$ och $z = e^{i\frac{4\pi}{3}}$.

(iv) Vi använder kedjeregeln för att beräkna derivatan. Vi får att

$$f'(z) = 3(1 + z^3)^2 \cdot (1 + z^3)' = 3(1 + z^3)^2 \cdot 3z^2 = 9(z + z^4)^2$$

Kapitel 4

Övning 4.1. Vi ska sätta in uttrycken som definierar $x(t)$ och $y(t)$ i den givna ekvationen (4.6).

Vi fokuserar först på termerna $(x_2 - x_1)y(t) - (y_2 - y_1)x(t)$. Vi har

$$\begin{aligned} & (x_2 - x_1)y(t) - (y_2 - y_1)x(t) \\ &= (x_2 - x_1)[(1 - t)y_1 + ty_2] - (y_2 - y_1)[(1 - t)x_1 + tx_2] \\ &= (x_2 - x_1)(1 - t)y_1 + t(x_2 - x_1)y_2 + t(y_2 - y_1)x_1 - t(y_2 - y_1)x_2 \\ &= (x_2 - x_1)y_1 - (y_2 - y_1)x_1 \\ & \quad - tx_2y_1 + tx_1y_1 + tx_2y_2 - tx_1y_2 \\ & \quad + tx_1y_2 - tx_1y_1 - tx_2y_2 + tx_2y_1 \\ &= (x_2 - x_1)y_1 - (y_2 - y_1)x_1. \end{aligned}$$

Summan som kvarstår i högerledet uppträder, fast med motsatt tecken, i (4.6). Vi ser nu att $\gamma(t)$ verkligen satisfierar ekvationen för alla t .

Övning 4.3. Sinuskurvan med x -koordinat 0 till a ges av mängden $\{(x(t), y(t)) : t \in [0, a]\} = \{(t, \sin t) : t \in [0, a]\}$. Då är

$$\sqrt{(x'(t))^2 + (y'(t))^2} = \sqrt{1 + \cos^2 t}.$$

Så längden av kurvan är

$$\int_0^a \sqrt{1 + \cos^2 t} dt.$$

Genom att beräkna med olika värden på a i en integralkalkylator ser man att $a = 0.74$ ger värdet 1.002.

Övning 4.5. Låt oss beräkna $x'(t)$ och $y'(t)$. Enligt kedjeregeln får vi

$$x'(t) = (\cos^3 t)' = 3(-\sin t) \cos^2 t = -3 \sin t \cos^2 t$$

samt

$$y'(t) = (\sin^3 t)' = 3 \cos t \sin^2 t.$$

Vi noterar nu att den fyrfaldiga symmetrin hos kurvan medför att $l(\Gamma) = 4l(\Gamma_1)$, där Γ_1 är den delen av kurvan som börjar i punkten $(1, 0)$ och slutar i $(0, 1)$. Denna del av kurvan svarar mot värdena hos parameterfunktionen $\gamma(t)$ för $t \in [0, \pi/2]$. Vi har alltså att beräkna integralen

$$\int_0^{\pi/2} \sqrt{(x'(t))^2 + (y'(t))^2} dt.$$

Vi kan förenkla integranden genom att bryta ut en faktor $\sin^2 t \cos^2 t$ och använda trigonometriska ettan:

$$\begin{aligned} (x'(t))^2 + (y'(t))^2 &= 9(\sin^2 t \cos^4 t + \cos^2 t \sin^4 t) \\ &= 9(\sin^2 t \cos^2 t \cdot \cos^2 t + \sin^2 t \cos^2 t \sin^2 t) \\ &= 9 \sin^2 t \cos^2 t (\cos^2 t + \sin^2 t) = 9 \sin^2 t \cos^2 t. \end{aligned}$$

På intervallet $[0, \pi/2]$ har vi $\cos t \geq 0$ och $\sin t \geq 0$, varför $\sqrt{9 \sin^2 t \cos^2 t} = 3 \sin t \cos t$.

Detta ger oss

$$\begin{aligned} l(\Gamma_1) &= \int_0^{\frac{\pi}{2}} \sqrt{(x'(t))^2 + (y'(t))^2} dt \\ &= 3 \int_0^{\frac{\pi}{2}} \sin t \cos t dt \\ &= \frac{3}{2} \int_0^{\frac{\pi}{2}} \sin(2t) dt \\ &= \frac{3}{2} \left[-\frac{1}{2} \cos(2t) \right]_0^{\frac{\pi}{2}} = \frac{3}{2}. \end{aligned}$$

Båglängden av asteroidkurvan blir alltså $l(\Gamma) = 4l(\Gamma_1) = 4 \cdot (3/2) = 6$.

Övning 4.7. Vi kan parametrisera linjesegmentet mellan i och ie som

$$\gamma(t) = (1 + (e - 1)t)i, \quad t \in [0, 1].$$

Då är $\gamma'(t) = (e - 1)i$ och $|\gamma'(t)| = e - 1$. Så sökta båglängden är

$$\int_0^1 \frac{e - 1}{1 + (e - 1)t} dt = [\ln |1 + (e - 1)t|]_0^1 = 1.$$

Kortaste vägen mellan två punkter ges nu endast av rätta linjer om de har samma x -koordinat. Betrakta $-3 + 4i$ och $3 + 4i$ samt kurvan som är en del av halvcirkeln $|z| = 5$ som passerar dessa punkter. Vi parametriserar kurvan som

$$\gamma(t) = 5e^{it}, \quad t_1 \leq t \leq t_2, \quad \tan t_1 = \tan(\pi - t_2) = 4/3.$$

Då blir längden av halvcirkeln $|z| = 5$ som passerar dessa punkter

$$\int_{t_1}^{\pi - t_1} \frac{|\gamma'(t)|}{\text{Im } \gamma(t)} dt = \int_{t_1}^{\pi - t_1} \frac{1}{\sin t} dt.$$

För att beräkna integralen gör vi variabelbytet $s = \tan \frac{t}{2}$. Notera att

$$\sin t = \sin(2 \arctan s) = 2 \sin(\arctan s) \cos(\arctan s).$$

Genom att rita en rätvinklig triangel med sidor kateter 1 och s ser man att

$$\begin{aligned} \sin(\arctan s) &= \frac{s}{\sqrt{1 + s^2}}, \\ \cos(\arctan s) &= \frac{1}{\sqrt{1 + s^2}}. \end{aligned}$$

Därmed får vi

$$\begin{aligned} \sin t &= \frac{2s}{1 + s^2}, \\ \cos t dt &= \frac{d}{ds} \left(\frac{2s}{1 + s^2} \right) = \frac{2(1 - s^2)}{(1 + s^2)^2} ds, \\ \cos t &= \sqrt{1 - \sin^2 t} = \frac{1 - s^2}{1 + s^2}, \\ dt &= \frac{\frac{2(1 - s^2)}{(1 + s^2)^2} ds}{\frac{1 - s^2}{1 + s^2}} = \frac{2 ds}{1 + s^2}. \end{aligned}$$

Slutligen blir integralen

$$\int \frac{1}{\sin t} dt = \int \frac{1}{\frac{2s}{1+s^2}} \frac{2}{1+s^2} ds = \int 1/s ds = \ln |s| = \ln \left| \tan \frac{t}{2} \right|.$$

Båglängden blir därmed

$$\ln \left| \frac{\tan((\pi - t_1)/2)}{\tan(t_1/2)} \right|.$$

Eftersom $\tan(\pi/2 - x) = \frac{1}{\tan x}$ fås längden

$$\ln\left(\frac{1}{\tan^2(t_1/2)}\right) = -2 \ln(\tan(t_1/2)).$$

Men $\tan(t_1/2) = \frac{\sin t_1}{1+\cos t_1} = \frac{4/5}{1+3/5} = 1/2$. Detta ger längden

$$-2 \ln(1/2) = 2 \ln 2 = \ln 4 \approx 1.386,$$

vilket är mindre än längden av den rätta linjen som är $6/4 = 3/2$.

Kapitel 5

Ekvationen $f(0) = 1$ säger att

$$f(0) = \frac{b}{d} = 1,$$

så $b = d$. Vidare vet vi att $f(1) = \infty$, vilket i sin tur säger att

$$f(\infty) = \frac{a+b}{c+d} = \infty,$$

vilket bara är möjligt om vi har division med noll i detta uttryck. Detta kommer hända precis då $c = -d$. Till sist vet vi att

$$f(\infty) = \frac{a}{c} = 0,$$

så $a = 0$. Det följer att vi kan skriva f som

$$f(z) = \frac{b}{-bz + b},$$

för något val av b . Talet b bestäms av att $ad - bc = 1$, eller $b^2 = 1$. Tar vi t.ex. $b = 1$, fås

$$f(z) = \frac{1}{-z + 1}.$$

Övning 5.3. Från övning 5.2 vet vi att $\text{Aut}(\mathbb{C})$ är en grupp. Det återstår att visa de första tre egenskaperna.

- (i) Att varje funktion i $\text{Aut}(\mathbb{C})$ är analytisk är givet.

- (ii) Genom att argumentera på samma sätt som i beviset av sats 5.4.4 är det nog att visa att för alla $w \in \mathbb{C}$ existerar det en automorfi $f \in \text{Aut}(\mathbb{C})$ så att $f(0) = w$. Vi väljer exempelvis funktionen

$$f(z) = z + w$$

som ingår i $\text{Aut}(\mathbb{C})$

- (iii) Genom att argumentera på samma sätt som i beviset av sats 5.4.4 är det nog att visa att mängden

$$S_0 = \{f \in \text{Aut}(\mathbb{C}) \mid f(0) = 0, f'(0) = \alpha\}$$

har exakt ett element om $|\alpha| = 1$, och annars 0 element. Om $f \in S_0$ skrivs på formen $f(z) = az + b$ ser vi att $b = 0$ för att $f(0) = 0$. Då $f'(0) = a = \alpha$ ser vi att vi har exakt en lösning om $|\alpha| = 1$, men annars ingen lösning då $|\alpha| \neq 1$.

Övning 5.5. Ett exempel ges t.ex. av funktionen $f(z) = z + 1$, vars enda fixpunkt är ∞ . Ett lite klurigare är

$$g(z) = \frac{-1}{z-2},$$

vars enda fixpunkt är 1. För att se att dessa exempel fungerar, notera att fixpunkter i $\mathbb{C}$ ges av lösningar till ekvationen $cz^2 + (d-a)z - b = 0$, enligt lösningen till föregående uppgift. Funktionen f kan skrivas med $a = d = b = 1$ och $c = 0$, vilket ger ekvationen $-1 = 0$. Alltså saknar f fixpunkter i $\mathbb{C}$, så ∞ är den enda fixpunkten. Funktionen g kan skrivas med $a = 0$, $b = -1$, $c = 1$ och $d = 2$. För denna är ∞ inte en fixpunkt, och fixpunkter i $\mathbb{C}$ ges av lösningar till

$$z^2 - 2z + 1 = 0,$$

som faktoriseras till

$$(z-1)^2 = 0.$$

Så den enda lösningen är $z = 1$.

Övning 5.7. Skriv f på formen

$$f(z) = \frac{\alpha z - \bar{\beta}}{\beta z + \bar{\alpha}}$$

där $|\alpha|^2 + |\beta|^2 = 1$. Om $\beta = 0$ är $f(z) = \frac{\alpha}{\bar{\alpha}}z$. Vi ser då direkt att f har två fixpunkter, 0 och ∞ . Vi antar nu att $\beta \neq 0$.

Om z ska vara en fixpunkt ska det gälla att $f(z) = z$. Detta ger oss att

$$\frac{\alpha z - \bar{\beta}}{\beta z + \bar{\alpha}} = z$$

vilket vi kan skriva om som

$$\alpha z - \bar{\beta} = z(\beta z + \bar{\alpha}) = \beta z^2 + z\bar{\alpha}.$$

eller

$$\left(z + \frac{\bar{\alpha} - \alpha}{2\beta}\right)^2 - \left(\frac{\bar{\alpha} - \alpha}{2\beta}\right)^2 + \frac{\bar{\beta}}{\beta} = 0.$$

Låt nu $w = \frac{\bar{\alpha} - \alpha}{2\beta}$ och $A = -\left(\frac{\bar{\alpha} - \alpha}{2\beta}\right)^2 + \frac{\bar{\beta}}{\beta}$. Vi kan då skriva

$$w^2 = A.$$

Om vi sedan tar och skriver $A = re^{i\theta}$ ser vi att w har exakt två lösningar, $w = \sqrt{r}e^{i\frac{\theta}{2}}$ och $w = -\sqrt{r}e^{i\frac{\theta}{2}}$, givet att $r \neq 0$. Eftersom w har lika många lösningar som z återstår det att visa att r , eller lika gärna A , inte är lika med 0. Vi antar motsatsen, det vill säga, att $A = 0$. I så fall har

$$\begin{aligned}\frac{\bar{\beta}}{\beta} &= \left(\frac{\bar{\alpha} - \alpha}{2\beta}\right)^2 \\ \beta\bar{\beta} &= \left(\frac{\bar{\alpha} - \alpha}{2}\right)^2 \\ 1 - |\alpha|^2 &= \frac{\bar{\alpha}^2 - 2|\alpha|^2 + \alpha^2}{2}\end{aligned}$$

där vi har använt att $\beta \neq 0$ och att $|\alpha|^2 + |\beta|^2 = 1$. Men nu kan vi skriva om detta som att

$$2 = \bar{\alpha}^2 + \alpha^2 = 2\operatorname{Re}(\alpha^2).$$

Då $|\alpha|^2 \leq 1$ medför detta att $\alpha = \pm 1$ men i så fall skulle $\beta = 0$ vilket är en motsägelse.

Kapitel 6

Övning 6.1. Eftersom det passerar en unik cirkel eller linje genom tre givna punkter (6.2.5), så räcker det att välja tre punkter på imaginära axeln och tre punkter på enhetscirkeln, samt välja en Möbiusavbildning som tar de första tre till de senare. Så tag $p, q, r = 0, i, \infty$ på imaginära axeln, och $p', q', r' = -1, i, 1$ på enhetscirkeln.

Låt oss nu först hitta en avbildning som tar p, q, r , alltså $0, i, \infty$, på $0, 1$ och ∞ . Vi ser att $f(z) = -iz$ fungerar, och denna är en Möbiusavbildning. Vi vill nu också hitta en Möbiusavbildning som avbildar $p', q', r' = -1, i, 1$ på $0, 1$ respektive ∞ . Här imiterar vi beviset för 6.2.2, och ser att avbildningen

$$g(z) = \frac{i-1}{i+1} \cdot \frac{z+1}{z-1} = \frac{-i(z+1)}{z-1}$$

fungerar. Nu vet vi att $g^{-1} \circ f$ kommer att ha egenskapen att den avbildar p, q, r på p', q' och r' , och enligt beviset av 5.5.7 är

$$g^{-1}(z) = \frac{-z+i}{-z-i}.$$

Alltså är $(g^{-1} \circ f)(z)$ lika med

$$g^{-1}(f(z)) = \frac{iz+i}{iz-i} = \frac{z+1}{z-1},$$

och denna Möbiusavbildning avbildar alltså imaginära axeln på enhetscirkeln. (Självklart finns många andra avbildningar med samma egenskap.) Att denna Möbiusavbildning fungerar kan vi också se direkt genom att titta på uttrycket: om z ligger på imaginära axeln så kommer $z+1$ och $z-1$ att ha samma belopp (rita ut $z+1$ och $z-1$ i komplexa planet!), så för z på imaginära axeln är

$$|f(z)| = \frac{|z+1|}{|z-1|} = 1,$$

där vi använt Övning 3.3. Så $f(z)$ ligger på enhetscirkeln.

Övning 6.3. Vi följer bevisskissen som gavs i uppgiftslydelsen. Om de tre punkterna ligger på en linje så måste vi bara utesluta att de inte *också* ligger på en gemensam cirkel. Men detta är omöjligt eftersom en cirkel och en linje bara kan mötas i högst två punkter.

Om de inte ligger på en cirkel, betrakta först mängden av punkter som är på samma avstånd till både p och q . Denna är lätt att hitta: vi drar det räta linjesegmentet från p till q , och sedan drar vi från denna sträckas mittpunkt en normallinje vinkelrätt ut från segmentet. Punkterna på denna normallinje är exakt de som är på samma avstånd från p och q . Vi gör detsamma med q och r . Den viktiga observationen är att de två linjerna vi får på detta sätt är *aldrig parallella*. Anledningen är att vi antagit att p, q och r inte ligger på samma linje, och därur följer att linjesegmentet mellan p och q inte har samma lutning som linjesegmentet mellan q och r . Därför kommer inte heller dessa segments normaler kunna ha samma lutning.

Alltså möts dessa två linjer i en punkt c , och detta är den unika punkt i komplexa planet som är på samma avstånd R till både p, q och r . Om vi ritar en cirkel med c som centrum och R som radie kommer denna att passera genom p, q och r , och ingen annan cirkel har denna egenskap.

Övning 6.5. Vi måste visa att det finns två kurvor Γ och Δ i komplexa planet som möts i punkten $z = 0$ i en viss vinkel α , sådana att $f(\Gamma)$ och $f(\Delta)$ inte möts i vinkeln α . Låt till exempel Γ vara reella axeln och Δ imaginära axeln. Vinkeln från Γ till Δ i skärningspunkten $z = 0$ är $\pi/2$. Dock kommer $f(\Gamma)$ att vara positiva reella axeln, eftersom om z är reellt är z^2 reellt och positivt. Vidare är $f(\Delta)$ negativa reella axeln, ty om $z = it$ är $z^2 = -t^2$, som är reellt och negativt. Alltså möts $f(\Gamma)$ och $f(\Delta)$ i vinkeln π , så f kan inte ha varit konform i punkten $z = 0$.

Övning 6.7. Låt p, q och r vara tre punkter på $\widehat{\mathbb{C}}$. Vi skall visa att det genom dessa passerar en unik cirkel eller linje. Enligt föregående uppgift är en cirkel eller linje samma sak som en skärning mellan $\widehat{\mathbb{C}}$ och ett plan i $\mathbb{R}^3$, så det räcker alltså att visa att det passerar ett unikt plan genom p, q och r när vi tänker på dem som punkter i $\mathbb{R}^3$.

Men helt allmänt gäller i $\mathbb{R}^3$ att det genom tre punkter passerar ett unikt plan, förutsatt att de tre punkterna inte ligger på en linje. (Om de tre punkterna ligger på en linje, finns oändligt många olika plan.) Vi måste därför endast utesluta att p, q och r ligger på samma linje i $\mathbb{R}^3$. Men vi vet att alla tre ligger på Riemannsfären, som ju är en sfär med radie 1 kring origo. Eftersom en linje

endast kan möta en sfär i högst två punkter, kan det inte vara fallet att p, q och r ligger på en linje i $\mathbb{R}^3$, och vi är klara.

Kapitel 7

Övning 7.1. Antag att två hyperboliska linjer λ och μ möts i två punkter p och q . Vi vet att det passerar en unik hyperbolisk linje genom p och q , och eftersom både λ och μ har denna egenskap kommer dessa nödvändigtvis att vara lika.

Övning 7.3. Vi vet att då f är en Möbiusavbildning, bevarar den cirklar och linjer på $\widehat{\mathbb{C}}$. Vi vet dessutom att då $f \in \text{Aut}(\mathbb{H})$ kommer den att bevara $\mathbb{R} \cup \{\infty\}$. Den linje på $\widehat{\mathbb{C}}$ som svarar mot imaginära axeln möter $\mathbb{R} \cup \{\infty\}$ i de två punkterna 0 och ∞ , och eftersom detta bevaras av f ser vi att det finns två fall: antingen gäller $f(0) = 0$ och $f(\infty) = \infty$, eller $f(0) = \infty$ och $f(\infty) = 0$.

I det första fallet ger $f(0) = 0$ att $b = 0$ och $f(\infty) = \infty$ att $c = 0$. Jämför beviset av 6.2.3. Alltså har f formen $f(z) = \frac{a}{d} \cdot z$, och eftersom $ad - bc = ad = 1$ kommer $a/d > 0$. Så vi finner att f ges av multiplikation med en positiv reell konstant $r = a/d$. Omvänt gäller att om f ges av multiplikation med en positiv reell konstant så ligger f i $\text{Aut}(\mathbb{H})$ och f bevarar imaginära axeln.

I det andra fallet ger i stället $f(0) = \infty$ att $d = 0$, och $f(\infty) = 0$ att $a = 0$. Alltså är

$$f(z) = \frac{b}{cz}.$$

Här är dock $ad - bc = -bc = 1$, så $b/c < 0$. Alltså kan vi i detta fall skriva

$$f(z) = \frac{-r}{z}.$$

Omvänt: för varje $r > 0$ ger detta en avbildning som ligger i $\text{Aut}(\mathbb{H})$ och bevarar imaginära axeln.

Alltså är detta våra två alternativ, f kan skrivas som $f(z) = rz$ eller $f(z) = -rz^{-1}$ där r är reellt och positivt. Geometriskt svarar dessa mot antingen en translation av hyperboliska planet parallellt med imaginära axeln i första fallet, respektive en rotation ett halvt varv runt i följt av en translation parallellt med imaginära axeln.

Övning 7.5. Låt λ vara en hyperbolisk linje och $p \in \lambda$. Drag tangentlinjen T till λ genom punkten p . Vi måste dela upp i två fall:

- (i) Linjen T är parallell med reella axeln. I så fall kommer normallinjen μ till λ genom p att vara *vinkelrät* mot reella axeln, och alltså vara en hyperbolisk linje. Denna möter enligt definition λ i en rät vinkel.
- (ii) Linjen T är inte parallell med reella axeln. I så fall skär T och $\mathbb{R}$ varandra i en unik punkt q . Drag cirkeln μ som har q som mittpunkt och passerar genom p . Denna möter reella axeln i rät vinkel eftersom dess centrum ligger på $\mathbb{R}$, så den är en hyperbolisk linje. Dessutom möter den λ i en rät vinkel: linjen T är vinkelrät mot μ eftersom T passerar cirkeln μ 's mittpunkt, och då T är tangenten till λ följer att λ och μ är vinkelräta.

För att se att dessa är unika, antag att μ möter λ i rät vinkel. Om μ är en rät linje måste tangenten till λ vara parallell med reella axeln i skärningen, eftersom μ då är parallell med imaginära axeln. I detta fall skulle μ ha hittats i fall ett ovan. Om μ är en cirkel måste dess mittpunkt vara reell och den måste skära λ på ett sådant sätt att tangenten till λ är vinkelrät mot cirkeln, och alltså pekar in mot cirkelns centrum.

Kapitel 8

Övning 8.1. Vertikal linje mellan i och $3i$ samt segment av halvcirkel med medelpunkt i 1 och radie 5.

Övning 8.3. Eftersom $\text{Aut}(\mathbb{H})$ bevarar hyperbolisk längd, och varje hyperbolisk linje kan flyttas till en annan hyperbolisk linje av ett element i $\text{Aut}(\mathbb{H})$, räcker det att visa att imaginära axeln är oändligt lång i hyperbolisk längd. Parametrisera imaginära axeln som $\gamma(t) = it$, $0 < t < \infty$. Här är $|\gamma'(t)| = 1$ och $\text{Im}(\gamma(t)) = t$. Längden av imaginära axeln ges av en integral:

$$\int_0^\infty \frac{1}{t} dt = [\ln |t|]_0^\infty = \infty,$$

eftersom $\lim_{t \rightarrow \infty} \ln |t| = \infty$ och $\lim_{t \rightarrow 0} \ln |t| = -\infty$.

Övning 8.5. Funktionen $f(z) = 2z$ betraktad från $\mathbb{H}$ till $\mathbb{H}$ är en Möbiusavbildning, så bevarar hyperboliska avstånd. Betraktad som funktion $\mathbb{R}^2 \rightarrow \mathbb{R}^2$, dubblar den euklidiska avstånd.

Övning 8.7. Horisontala segment mellan två punkter $a + i$ och $b + i$ har hyperboliskt och euklidiskt avstånd $|b - a|$. Vertikala segment mellan ia och ib har hyperbolisk längd $|\ln(b/a)|$ och euklidisk längd $|b - a|$, de kan ej vara lika om $a \neq b$.

A Appendix

Vi formulerar här ett antal definitioner och satser om integraler och derivator som vi använder i häftet. Bevisen för satserna utelämnas här. Vi hänvisar istället till [6] för en mer fullständig framställning av teorin. Vi kommer genomgående att förutsätta att läsaren känner till begreppen derivata och integral sedan tidigare och påminner bara helt kort om några viktiga definitioner.

A.1 Derivator

Definition A.1.1. Låt $\Omega \subset \mathbb{R}$. En funktion $f: \Omega \rightarrow \mathbb{R}$ säges vara *kontinuerlig* i en punkt $x \in \mathbb{R}$ om det för varje $\varepsilon > 0$ existerar ett $\delta > 0$ sådant att

$$|f(x) - f(y)| < \varepsilon \quad \text{om} \quad |x - y| < \delta.$$

Om f är kontinuerlig i varje punkt i definitionsmängden Ω till f säger vi att f är *kontinuerlig*.

En intuitiv, men inte helt korrekt, beskrivning av en kontinuerlig funktion är att dess graf kan ritas utan att man behöver lyfta pennan från pappret.

Exempel A.1.2. Funktionen $f(x) = x$ definierad för alla $x \in \mathbb{R}$ är kontinuerlig. Alla polynom, det vill säga, funktioner på formen $f(x) = a_n x^n + a_{n-1} x^{n-1} + \dots + a_1 x + a_0$, med $a_n, a_{n-1}, \dots, a_0 \in \mathbb{R}$, är kontinuerliga.

Exempel A.1.3. Funktionen $f: \mathbb{R} \rightarrow \mathbb{R}$ som definieras av

$$f(x) = \begin{cases} -1 & x < 0 \\ 1 & x \geq 0 \end{cases}$$

är inte kontinuerlig i punkten $x = 0$, men väl i alla andra punkter.

Kontinuerliga funktioner har många goda egenskaper. Exempelvis är kontinuerliga funktioner på slutna intervall integrerbara, se nedan.

Derivatan av en funktion $f: [a, b] \rightarrow \mathbb{R}$ i en punkt $x \in (a, b)$ definieras som gränsvärdet

$$f'(x) = \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h}$$

om detta existerar, och i så fall är f *deriverbar* i x . Om $f'(x)$ existerar för alla x säger vi att f är *deriverbar*. Tyvärr är det inte sant att alla kontinuerliga funktioner är deriverbara, men de kontinuerliga funktioner vi kommer att stöta på i det här häftet kommer att vara det utom möjligtvis i ett antal speciella punkter. Däremot är deriverbara funktioner automatiskt kontinuerliga.

Vi förutsätter att läsaren känner till räkneregler för derivator samt hur man deriverar produkter och kvoter av deriverbara funktioner. Vi kommer även behöva derivata sammansatta funktioner och påminner om detta i en sats.

Sats A.1.4 (Kedjeregeln). Låt $f: [\alpha, \beta] \rightarrow \mathbb{R}$ och $g: [a, b] \rightarrow [\alpha, \beta]$ vara deriverbara. Då gäller

$$(f \circ g)'(x) = g'(x)f'(g(x)), \quad x \in (a, b).$$

Följande sats, som brukar kallas *medelvärdessatsen*, ger oss ett användbart samband mellan en funktion och dess derivata.

Sats A.1.5. *Låt $f: [a, b] \rightarrow \mathbb{R}$ vara kontinuerlig, och antag att f är deriverbar i det öppna intervallet (a, b) . Då existerar en punkt $\xi \in (a, b)$ sådan att*

$$f(b) - f(a) = (b - a)f'(\xi).$$

Den geometriska tolkningen är att det på funktionsgrafen för f finns minst en punkt $(\xi, f(\xi))$ sådan att tangenten i denna punkt är parallell med den räta linjen som går igenom punkterna $(a, f(a))$ och $(b, f(b))$. Vi uppmanar läsaren att rita en bild för att illustrera detta.

A.2 Integraler

Betrakta en funktion $f: [a, b] \rightarrow \mathbb{R}$ och dess tillhörande funktionsgraf. Man brukar oftast definiera integralen av f , talet $I = \int_a^b f(x)dx$, på följande sätt. Man approximerar f underifrån och ovanifrån med stegfunktioner. Därefter bildar man över- och undersummor genom att lägga ihop areorna av rektanglarna som dessa stegfunktioner bildar tillsammans med x -axeln. Om dessa över- och undersummor konvergerar mot ett tal I när stegen blir godtyckligt små säger vi att f är *integrerbar* och har integralen I .

Ett annat sätt att angripa integraler är genom så kallade *Riemannsummor*. Vi skapar en indelning π av intervallet $[a, b]$ genom att välja n punkter x_j med

$$x_0 = a < x_1 < \cdots < x_{n-1} < x_n = b.$$

Vi låter $\ell(\pi)$ vara längden på det största av intervallen $[x_{j-1}, x_j]$ och använder detta som ett mått på indelningens finhet.

I varje delintervall i indelningen, $[x_{j-1}, x_j]$, väljer vi ut en punkt ξ_j där vi evaluerar funktionen f . Vi bildar därefter Riemannsumman som hör till indelningen:

$$\sum_{j=0}^{n-1} f(\xi_j)(x_j - x_{j-1})$$

Den underliggande idén är att denna summa är godtyckligt nära integralen av f när indelningen blir tillräckligt fin, det vill säga, när $\ell(\pi) \rightarrow 0$.

Följande sats kopplar ihop Riemannsummor, kontinuerliga funktioner och integraler.

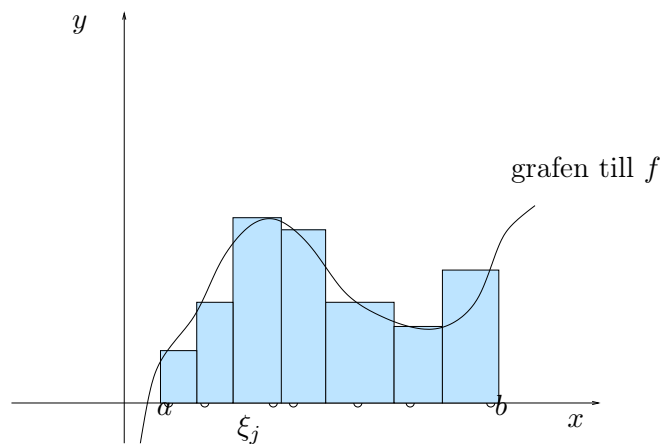
Sats A.2.1. *Låt $f: [a, b] \rightarrow \mathbb{R}$ vara kontinuerlig. Då är f integrerbar över $[a, b]$, och Riemannsummorna*

$$\sum_{j=0}^{n-1} f(\xi_j)(x_{j+1} - x_j)$$

konvergerar mot integralen

$$\int_a^b f(x)dx$$

vid obegränsat förfinad indelning.



Figur A.1: En Riemannsumma.

Nästa sats är av teoretiskt intresse, men kan också underlätta många räkningar med integraler. Den uttrycker hur integralen förändras när vi byter integrationsvariabel. I denna allmänna form anser vi att variabelbytet ges av en funktion g , så att $y = g(x)$ är den nya integrationsvariabeln.

Sats A.2.2. Låt $f: [\alpha, \beta] \rightarrow \mathbb{R}$ vara kontinuerlig, och låt $g: [a, b] \rightarrow [\alpha, \beta]$ vara en kontinuerlig funktion med positiv derivata. Då gäller

$$\int_a^b f(g(x))g'(x)dx = \int_\alpha^\beta f(y)dy$$

med $\alpha = g(a)$ och $\beta = g(b)$.

Vi ger ett exempel för att visa hur satsen kan användas vid praktiska beräkningar.

Exempel A.2.3. Låt oss beräkna integralen

$$\int_0^1 \frac{\cos \sqrt{x+1}}{\sqrt{x+1}} dx.$$

Det är svårt att hitta en primitiv funktion till integranden, men vi ser att funktionen i täljaren är sammansatt av de två funktionerna $\cos x$ och $\sqrt{x+1}$.

Vi sätter $g(x) = \sqrt{x+1}$ och betraktar g som en funktion $g: [0, 1] \rightarrow [1, 2]$. Derivatan av g är $g'(x) = 1/(2\sqrt{x+1})$ och vi har $g(0) = 1$ samt $g(1) = \sqrt{2}$. Satsen om variabelbyten, Sats A.2.2, ger nu att

$$\int_0^1 \frac{\cos \sqrt{x+1}}{\sqrt{x+1}} dx = 2 \int_0^1 \frac{\cos \sqrt{x+1}}{2\sqrt{x+1}} dx = 2 \int_1^{\sqrt{2}} \cos y dy.$$

Den sista integralen är mycket lättare att beräkna:

$$2 \int_1^{\sqrt{2}} \cos y dy = 2 [\sin y]_1^{\sqrt{2}} = 2(\sin \sqrt{2} - \sin 1).$$

Svårigheten vid beräkningen av konkreta integraler är ofta att komma på rätt variabelbyte som gör det möjligt att evaluera integralen. Satsen garanterar i alla fall att en integrals värde är detsamma före och efter variabelbyte, även om vi inte kan beräkna detta värde explicit.

En passande avslutning på detta appendix är en av de mest centrala satserna inom matematiken, *analysens huvudsats*, som kopplar ihop derivator och integraler.

Sats A.2.4. *Låt $f: [a, b] \rightarrow \mathbb{R}$ vara kontinuerlig och bilda funktionen*

$$F(x) = \int_a^x f(t)dt, \quad x \in [a, b].$$

Då är F deriverbar med $F'(x) = f(x)$.

Referenser och förslag till vidare läsning

Referenser

- [1] James W. Anderson: *Hyperbolic geometry, 2nd ed.* Springer Undergraduate Mathematics Series, Springer-Verlag, 2005.
En lärobok om hyperbolisk geometri för tredje årets universitetsstuderande. Kan antagligen läsas efter Persson-Böiers.
- [2] Victor J. Katz: *History of Mathematics: An Introduction*, Addison-Wesley, 1998.
Bred introduktion till matematikens historia.
- [3] Ruel V. Churchill & James Ward Brown: *Complex Variables and Applications, 4th ed.* McGraw-Hill international editions, 1984.
En lärobok i komplex analys för andra årets universitetsstudenter.
- [4] Robin Hartshorne: *Geometry: Euclid and beyond.* Springer-Verlag, 2000.
En mer axiomatisk behandling av euklidisk och icke-euklidisk geometri. Kan vara svår även om det inte finns formella förkunskapskrav.
- [5] Gareth A. Jones & David Singerman: *Complex functions: An algebraic and geometric viewpoint*, Cambridge University Press, 1987.
En mer avancerad bok om gränslandet mellan komplex analys och abstrakt algebra.
- [6] Arne Persson & Lars-Christer Böiers: *Analys i en variabel*, Studentlitteratur, 2001
Standardbok i analys för nybörjare på högskolan.
- [7] Pierre Samuel: *Projective geometry*, Undergraduate Texts in Mathematics Springer-Verlag, 1988
En klassisk bok om ämnet projektiv geometri, som generaliserar vår konstruktion av Riemannsfären.