



KTH Engineering Sciences

KTHs Matematiska Cirkel

HYPERBOLISK GEOMETRI

DAN PETERSEN
ALAN SOLA

INSTITUTIONEN FÖR MATEMATIK, 2009
FINANSIERAT AV MARIANNE OCH MARCUS WALLENBERGS STIFTELSE

Innehåll

0	Mängdlära	1
0.1	Mängder	1
0.2	Funktioner	3
1	Euklidisk geometri	6
1.1	Definitioner	6
1.2	Automorfier	9
1.3	Axiom	13
1.4	Historik över parallellpostulatet	15
1.5	Övningar	17
2	Båglängd	18
2.1	Kurvor i planet	18
2.2	Linjeintegraler	21
2.3	Variabelbyten och automorfier	24
2.4	Avstånd och geodeter	27
2.5	Övningar	28
3	Komplexa tal	30
3.1	Operationer med punkter i planet	30
3.2	Real-och imaginärdel av komplexa tal	32
3.3	Komplexa tal på polär form	34
3.4	Parameterkurvor i komplexa planet	37
3.5	Automorfier av planet	38
3.6	Komplexa polynom	38
3.7	Mer allmänna funktioner av komplexa tal	39
3.8	Komplex derivata	39
4	Möbiusgruppen och Riemannsfären	42
4.1	Möbiusgruppen	42
4.2	Gruppbegreppet	44
4.3	Exempel	45
4.4	Riemannsfären	47

5	Vidare egenskaper hos $\text{Möb}(\mathbf{C})$.	50
5.1	Cirklar och linjer	50
5.2	Konforma avbildningar	53
5.3	3-transitivitet	56
5.4	Övningar	59
6	Delgruppen $\text{Möb}(\mathbf{H})$.	60
6.1	Möbiusavbildningar som bevarar övre halvplanet	60
6.2	Hyperboliska linjer	63
6.3	Övningar	66
7	Hyperbolisk längd och hyperboliskt avstånd	67
7.1	Hyperbolisk längd	67
7.2	Invariants under $\text{Möb}(\mathbf{H})$	69
7.3	Hyperboliskt avstånd och hyperboliska geodeter	71
7.4	Hyperbolisk geometri	73
7.5	Sfärisk geometri	74
7.6	Areor, vinklar och Gauss-Bonnets sats	75
7.7	Övningar	77
A	Appendix	87
A.1	Derivator	87
A.2	Integraler	88
	Förslag till vidare läsning	91

Några ord på vägen

Detta kompendium är skrivet för att användas som litteratur till KTHs MATEMATISKA CIRKEL under läsåret 2009–2010 och består av sju avsnitt samt ett inledande avsnitt om mängdlära. Kompendiet är inte tänkt att läsas enbart på egen hand, utan ska ses som ett skriftligt komplement till undervisningen på de sju träffarna.

Som den mesta matematik 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.

Övningsuppgifterna är fördelade i två kategorier. De med udda nummer har facit, och syftet med dessa är att eleverna ska kunna räkna dem och på egen hand kontrollera att de förstått materialet. De med jämna nummer saknar facit och kan användas som examination. Det rekommenderas dock att man försöker lösa även 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.

Vi bör också nämna att få av uppgifterna är helt enkla. Kika därför inte i facit efter några få minuter, om du inte löst uppgiften, utan prata först med kompisar eller försök litet till. Alla uppgifter ska gå att lösa med hjälp av informationen i detta kompendium.

KTHs Matematiska Cirkel finansieras av Marianne och Marcus Wallenbergs Stiftelse. Vi tackar Dan Laksov, Roy Skjelnes och Kathrin Vorwerk, alla från Institutionen för Matematik vid KTH, samt Johan Wild vid Europaskolan i Strängnäs för deras givande kommentarer om denna skrift.

Några ord om Cirkeln

KTHs Matematiska Cirkel, i dagligt tal benämnd Cirkeln, startade 1999. Dess ambition 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 skall särskilt stimulera elevernas matematikintresse och inspirera dem till fortsatta naturvetenskapliga studier. Lärarna på cirkeln kan vid behov ge eleverna förslag på ämnen till projektarbeten vid gymnasiet.

Till varje kurs skrivs ett kompendium som distribueras gratis till eleverna. Detta material, liksom övriga uppgifter om KTHs Matematiska Cirkel, finns tillgängligt på

<http://www.math.kth.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 och lärare.

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. Detta, och att flera ämnen är på universitetsnivå, gör att lärarna och eleverna kan uppleva programmet som tungt, och alltför långt över gymnasienivån. Meningen är emellertid inte att lärarna och eleverna skall behärska ämnet fullt ut och att lära in det på samma sätt som gymnasiekurserna. Det viktigaste är att eleverna kommer i kontakt med teoretisk matematik och får en inblick i *matematikens väsen*. Vår förhoppning är att lärarna med denna utgångspunkt skall ha lättare att upplysa intresserade elever om KTHs Matematiska Cirkel och övertyga skolledarna om vikten av att låta både elever och lärare delta i programmet.

Några ord om betygssättning

Ett speciellt problem tidigare år har varit betygssättningen. Detta borde emellertid bara vara ett problem om lärarna använder sig av samma standard som de gör när de sätter betyg på ordinarie gymnasiekurser. Om utgångspunkten istället är att eleverna skall få insikt i matematiken genom att gå på föreläsningarna och att eleven gör sitt bästa för att förstå materialet och lösa uppgifterna, blir betygssättningen lättare. Självklart betyder det mycket vad eleverna har lärt av materialet i kursen, men lärarna kan bara förvänta sig att ett fåtal elever behärskar ämnet fullt ut. I det perspektivet blir det lätt att använda de officiella kriterierna:

Godkänd: Eleven har viss insikt i de moment som ingår i kursen och kan på ett godtagbart sätt redovisa valda delar av kursen såväl muntligt som skriftligt. Detta kan ske genom att eleven håller föredrag inför klassen, redovisar eller lämnar en rapport till sin matematiklärare.

Väl godkänd: Eleven har god insikt i flera moment från kursen. Eleven kan redovisa dessa moment både skriftligt och muntligt och dessutom uppvisa lösningar på problem som givits på kursen. Detta kan ske genom att eleven håller föredrag inför klassen, redovisar eller lämnar en rapport till sin matematiklärare.

Mycket väl godkänd: Eleven har mycket god insikt i flera moment av kursen och lämnar skriftliga redovisningar av flera delar av kursen eller lämnar lösningar på problem som givits på kursen. Detta kan ske genom att eleven håller föredrag inför klassen, redovisar eller lämnar en rapport till sin matematiklärare.

Det är också möjligt att skolorna samarbetar, så elever från en skola redovisar eller lämnar rapport för en lärare i en annan skola.

Författarna, september 2009

0 Mängdlära

0.1 Mängder

Låt oss börja med att titta på ett av de mest grundläggande begreppen i matematiken, nämligen mängder. En mängd är en samling matematiska objekt, som till exempel tal, och dessa objekt kallar vi för *element* i mängden. Det enklaste sättet att beskriva en mängd är att räkna upp dess element. Ett sådant exempel är

$$A = \{1, 3, a, 7\}.$$

Detta betyder att A är en mängd som innehåller elementen 1, 3, a och 7. Ett annat sätt att beskriva en mängd är att skriva $\{x \in D \mid \text{villkor på } x\}$. Med detta menar man mängden av alla element i D som uppfyller de givna villkoren. Som exempel tar vi

$$B = \{n \in \{1, 2, 3, \dots\} \mid n \text{ är udda}\}$$

och

$$C = \{y \in \{1, 2, 3, 4\} \mid y > 2\}.$$

Mängden B innehåller alla udda positiva heltal, medan C innehåller alla element från mängden $\{1, 2, 3, 4\}$ som är större än 2. Alltså har vi

$$B = \{1, 3, 5, 7, 9, 11, \dots\} \quad \text{och} \quad C = \{3, 4\}.$$

Vi bryr oss inte om i vilken ordning eller hur många gånger elementen räknas upp och därmed gäller till exempel

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

Om A är en mängd och x är ett element i mängden A så skriver vi $x \in A$ och säger att x *tillhör* A . Exempelvis gäller $17 \in \{n \mid n \text{ är ett udda heltal}\}$ och $b \in \{a, b, 10, 3\}$. Att ett element x inte tillhör mängden A skrivs $x \notin A$. Den *tomma mängden* innehåller ingenting och betecknas $\emptyset$.

Exempel 0.1.1. Låt $A = \{4, 5, 8, 4711, 12, 18\}$ och $B = \{x \in A \mid x > 10\}$. Då är $B = \{12, 18, 4711\}$ medan $\{x \in A \mid x < 3\} = \emptyset$. Vidare har vi att $4 \in A$ men $4 \notin B$. ▲

Definition 0.1.2. Låt A och B vara mängder. Om alla element i mängden A också är element i mängden B så sägs A vara en *delmängd* till B . Detta betecknas $A \subseteq B$.

Exempel 0.1.3. Mängden $\{1, a\}$ är en delmängd till $\{1, 3, a\}$, eftersom alla element i $\{1, a\}$ finns i mängden $\{1, 3, a\}$. Vi skriver $\{1, a\} \subseteq \{1, 3, a\}$. ▲

Definition 0.1.4. Antag att A och B är mängder. *Unionen* av A och B består av de element som ligger i någon av mängderna och betecknas $A \cup B$. *Snittet* av A och B består av de element som ligger i båda mängderna och betecknas $A \cap B$.

Exempel 0.1.5. Låt $A = \{1, 3, 5, 6\}$ och $B = \{5, 8, 3, 4711\}$. Då har vi $A \cup B = \{1, 3, 5, 6, 8, 4711\}$ och $A \cap B = \{3, 5\}$. ▲

Det är dags att titta på några viktiga talmängder. Den mängd vi använder för att räkna föremål är de *naturliga talen* $\{0, 1, 2, 3, \dots\}$. Denna mängd betecknas $\mathbf{N}$. Tar vi med negativa tal får vi heltalen $\mathbf{Z} = \{\dots, -3, -2, -1, 0, 1, 2, 3, \dots\}$. Beteckningen kommer från tyskans *zahl* som betyder tal. Slutligen betecknar vi med $\mathbf{R}$ de *reella talen*, det vill säga alla tal på tallinjen, exempelvis $0, -1, 3/2, -527/3, \sqrt{2}$ och π . Notera att $\mathbf{N} \subseteq \mathbf{Z} \subseteq \mathbf{R}$.

Exempel 0.1.6. Vi har att $\mathbf{N} = \{n \in \mathbf{Z} \mid n \geq 0\}$. ▲

Exempel 0.1.7. Mängden $\{n \in \mathbf{Z} \mid n = 2 \cdot k \text{ för något } k \in \mathbf{Z}\}$ är mängden av alla jämna heltal. Denna mängd kan också skrivas som $\{2 \cdot k \mid k \in \mathbf{Z}\}$, eller som $\{\dots, -4, -2, 0, 2, 4, \dots\}$. ▲

Exempel 0.1.8. Låt oss påpeka att en mängd även kan ha andra mängder bland dess element. Exempelvis kan vi låta

$$A = \{2, 3, \{-1, 1\}, 4\},$$

och vi har att $\{-1, 1\} \in A$, det vill säga mängden $\{-1, 1\}$ är ett element i mängden A . ▲

Låt Ω vara en godtycklig mängd. Vi kommer i följande exempel antaga att alla mängder $A, B, C, \dots$ är delmängder till Ω . Vi definierar två vanligt förekommande mängder:

1. $A \setminus B = \{x \in A \mid x \notin B\}$
2. $A^c = \Omega \setminus A = \{x \in \Omega \mid x \notin A\}$

Mängden A^c kallas vanligtvis för *komplementet till A*.

Låt oss studera några exempel på hur man visar påståenden om allmänna mängder.

Exempel 0.1.9. Två mängder B och C sägs vara *disjunkta* om de inte har några gemensamma element. Visa att B och C är disjunkta om och endast om $B \cap C = \emptyset$.

Lösning. Antag att B och C är disjunkta. Antag att $x \in B \cap C$, det vill säga att $x \in B$ och $x \in C$. Men detta betyder att B och C har x som gemensamt element, vilket motsäger att B och C är disjunkta. Alltså måste $B \cap C = \emptyset$.

Omvänt, antag att $B \cap C = \emptyset$. Det betyder att det inte finns något element som tillhör både B och C . Alltså har B och C inga gemensamma element, det vill säga att B och C är disjunkta.

Nu har vi alltså visat två saker, dels att om B och C är disjunkta så gäller $B \cap C = \emptyset$, dels att om $B \cap C = \emptyset$ så är B och C disjunkta. Tillsammans betyder detta att B och C är disjunkta om och endast om $B \cap C = \emptyset$. ▲

Exempel 0.1.10. Visa att A och A^c är disjunkta.

Lösning. Enligt föregående exempel är det vi ska visa att $A \cap A^c = \emptyset$. Antag att $x \in A \cap A^c$. Det betyder att $x \in A$ och att $x \in A^c$. Det senare betyder per definition att $x \notin A$, vilket är en motsägelse. Alltså måste $A \cap A^c = \emptyset$. ▲

0.2 Funktioner

Innan vi gör en allmän definition av vad en funktion är kan det vara på sin plats att titta på något välbekant, nämligen en formel som $f(x) = x^2 + 1$. Detta är ett exempel på en funktion. Formeln säger att om vi tar ett tal $x \in \mathbf{R}$ så får vi ett nytt tal $f(x) \in \mathbf{R}$ genom att göra beräkningen $x^2 + 1$; till exempel får vi $f(2) = 2^2 + 1 = 5$. Vi säger att f är en funktion från de reella talen till de reella talen, eftersom både det vi stoppar in, x , och det vi får ut, $f(x)$, är reella tal. Vi brukar beteckna detta med $f : \mathbf{R} \rightarrow \mathbf{R}$.

Definition 0.2.1. Låt X och Y vara mängder. En *funktion* $f : X \rightarrow Y$ är ett sätt att till varje element $a \in X$ tilldela ett välbestämt element $b \in Y$. Vi skriver $f(a) = b$. Vi säger att a *avbildas* på b och att b är *bilden* av a .

Anmärkning 0.2.2. Ofta säger man att f är en funktion från X till Y istället för att använda beteckningen $f : X \rightarrow Y$. Ett vanligt alternativ till ordet funktion är *avbildning*.

Exempel 0.2.3. Betrakta mängderna $A = \{1, 2, 3\}$ och $B = \{1, 2, \dots, 100\}$. Ett exempel på funktion $f : A \rightarrow B$ ges av $f(n) = 2n$ för $n \in A$. Vi har alltså att $f(1) = 2$, $f(2) = 4$ och $f(3) = 6$. Per definition måste vi ha $f(x) \in B$ för alla $x \in A$, och detta gäller ju här eftersom

$$f(1) = 2 \in B, \quad f(2) = 4 \in B, \quad \text{och} \quad f(3) = 6 \in B.$$

I detta exempel definieras funktionen f av formeln $f(n) = 2n$, men det är inte alls nödvändigt att det finns en formel som beskriver hur funktionen verkar. Om vi som här har en funktion från en den *ändliga* mängden $A = \{1, 2, 3\}$ kan man till exempel definiera funktionen med hjälp av en tabell:

n	$f(n)$
1	2
2	4
3	6

▲

Exempel 0.2.4. Låt $h(x) = 3/2 \cdot x^2 - x^3$. Detta definierar en funktion h från $\mathbf{R}$ till $\mathbf{R}$. Vi har exempelvis att

$$h(1) = \frac{1}{2}, \quad \text{och} \quad h(-2) = 14.$$

▲

Definition 0.2.5. En funktion $f: X \rightarrow Y$ säges vara *injektiv* om följande är sant: om $f(x) = f(y)$ för $x, y \in X$ så gäller att $x = y$.

Uttryckt i ord säger den här definitionen att funktionen aldrig skickar två olika element i X på samma element i Y .

Definition 0.2.6. En funktion $f: X \rightarrow Y$ säges vara *surjektiv* om följande är sant: för varje $y \in Y$ existerar ett $x \in X$ sådant att $f(x) = y$.

Varje element i Y är alltså bilden av något x under funktionen f om funktionen är surjektiv.

En funktion kan vara surjektiv utan att vara injektiv, och tvärtom.

Exempel 0.2.7. Låt $\mathbf{R}_+$ beteckna de icke-negativa reella talen. Betrakta funktionen $f: \mathbf{R} \rightarrow \mathbf{R}_+$ som definieras av $f(x) = x^2$. Då är f surjektiv, men inte injektiv—till exempel har vi $f(-2) = f(2) = 4$.

Ett exempel på en funktion som är injektiv men inte surjektiv ges av funktionen i 0.2.3. Det finns till exempel inget $n \in \{1, 2, 3\}$ sådant att $f(n) = 3$.

▲

Definition 0.2.8. En funktion $f: X \rightarrow Y$ som är både surjektiv och injektiv säges vara *bijektiv*, eller en *bijektion*.

En bijektion $f: X \rightarrow Y$ har en så kallad *invers*. Detta är en avbildning som brukar betecknas f^{-1} och som låter oss gå tillbaka från bilden av f till den ursprungliga mängden X .

Definition 0.2.9. Låt $f: X \rightarrow Y$ vara en bijektion. *Inversen* till f är avbildningen $f^{-1}: Y \rightarrow X$ som ges av $f^{-1}(y) = x$, där x är det entydiga element i X som uppfyller $f(x) = y$.

Vi ser här att både injektivitet och surjektivitet är viktigt. Om f inte är injektiv kan det finnas många $x \in X$ med $f(x) = y$. Om f inte är surjektiv kan det vara så att det inte finns något x med $f(x) = y$.

Exempel 0.2.10. Betrakta funktionen $f: \mathbf{R} \rightarrow \mathbf{R}$ som ges av $f(x) = x^3$. Denna funktion är injektiv och surjektiv, och därmed en bijektion.

Inversen till f ges av funktionen $f^{-1}: \mathbf{R} \rightarrow \mathbf{R}$ som definieras av $f^{-1}(y) = y^{1/3}$.

▲

Exempel 0.2.11. Både definitionsområdet och bildmängden måste beaktas när vi undersöker om en funktion är en bijektion.

Funktionen $f: \mathbf{R}_+ \rightarrow \mathbf{R}_+$ med $f(x) = x^2$ är en bijektion, med invers $f^{-1}(y) = \sqrt{y}$. Som vi såg tidigare är detta påstående falskt om vi betraktar f definierad på hela $\mathbf{R}$. ▲

Övningar

Övning 0.1. Låt $A = \{1, 2, 3, 4, \dots\}$, $B = \{1, 3, 5, 7, \dots\}$, $C = \{2, 4, 6, 8, \dots\}$ och $D = \{1, 4, 19, 36, 101\}$. Bestäm mängderna

1. $B \cup C$,
2. $B \cap C$,
3. $D \cap C$,
4. $\{x \in D \mid x \in B\}$,
5. $\{x \in A \mid x = y + 1 \text{ för något } y \in D\}$,
6. $\{x + 1 \mid x \in D\}$.

Övning 0.2. Låt $\mathbf{N} = \{0, 1, 2, \dots\}$ och $B_n = \{1, 2, \dots, n\}$ för $n = 1, 2, 3, \dots$. Visa att $\mathbf{N} \setminus \{0\} = B_1 \cup B_2 \cup B_3 \cup \dots$.

Övning 0.3. Låt Ω vara en mängd och $A, B, C \subseteq \Omega$. Visa att

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

Övning 0.4. Låt Ω vara en mängd och $A \subseteq \Omega$. Visa att $\Omega = A \cup A^c$.

Övning 0.5. 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.

1 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.

1.1 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 1.1.1. Mängden $\mathbf{R}^2$ består av alla ordnade par (x, y) av tal, där x och y tillhör de reella talen $\mathbf{R}$. När vi talar om det euklidiska planet, eller helt enkelt planet, så menar vi denna mängd.

Anmärkning 1.1.2. Vi visualiserar $\mathbf{R}^2$ som ett xy -plan, där första talet anger koordinaten i x -led och andra talet koordinaten i y -led.

Definition 1.1.3. En *punkt* p i planet är ett element i mängden $\mathbf{R}^2$.

Definition 1.1.4. Vi säger att L är en *linje* i planet om den är en delmängd som har följande form:

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

för några a, b och c i $\mathbf{R}$, där inte både a och b är lika med noll. Om en punkt p ingår i mängden L säger vi att punkten *ligger* på linjen. Ett *linjesegment* är en delmängd av en linje L som kan skrivas som

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

där $d < e$.

Definition 1.1.5. 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 klassiska¹ 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.

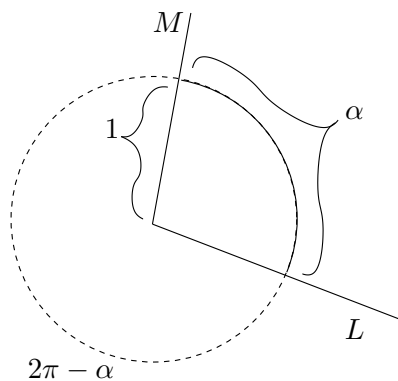
Definition 1.1.6. Låt två räta linjesegment L och M gå ut från en punkt p , det vill säga, p är ändpunkt till bägge segmenten. Rita cirkeln med mittpunkt p och radie 1. Om segmenten L och M förlängs bort från p så att båda har längd minst 1, så kommer båda dessa att skära cirkeln i en varsin punkt l respektive m . Vi definierar *vinkeln från L till M* 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° .

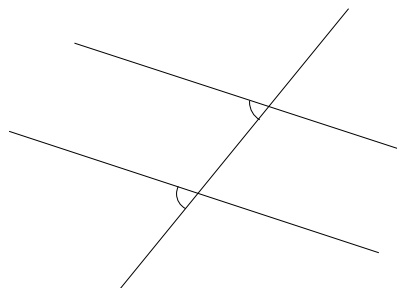
Definition 1.1.7. Vi säger att två vinklar α och β är lika om $\alpha - \beta$ är ett heltal gånger 2π .

Anmärkning 1.1.8. Definitionen 1.1.6 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 1 så är vinkeln från L till M lika med α , medan vinkeln från M till L är $2\pi - \alpha$, vilket också är lika med $-\alpha$.

¹Försöker 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!



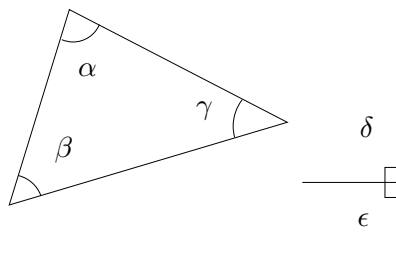
Figur 1: En illustration av definitionen av vinkeln mellan två linjesegment.



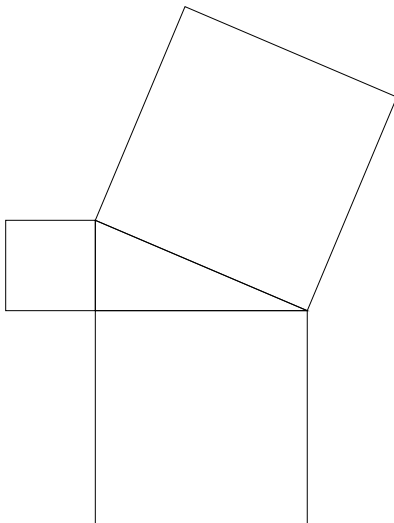
Figur 2: De två infallande vinklarna är lika stora.

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

1. **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;
2. **Sats:** I en triangel är summan av alla tre vinklar densamma som summan av två räta vinklar;
3. **Sats:** (Pythagoras) Om man ritar en kvadrat vid varje sida i en rätvinklig



Figur 3: Här är $\alpha + \beta + \gamma = \delta + \epsilon$.



Figur 4: Pythagoras sats.

triangel, så är arean av kvadraten vid hypotenusan densamma som summan av areorna vid de bägge kateterna.

1.2 Automorfier

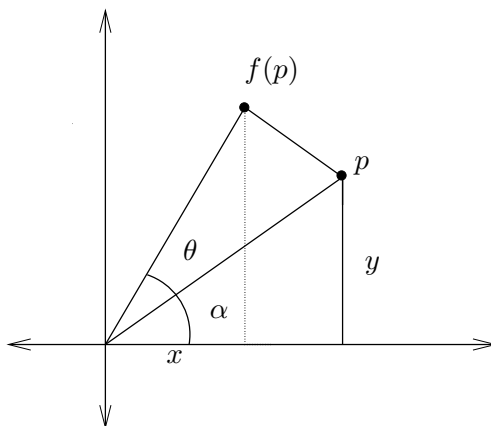
Detta avsnitt i vår introduktion till euklidisk geometri kommer att beskriva ett begrepp som dyker upp nästan överallt i matematiken. Med en *automorfi* kommer vi att mena en funktion med egenskapen att den bevarar *all* geometrisk struktur. Mer precist så kräver vi att en automorfi ska vara en *bijektion* (se Definition 0.2.8) med avseende på både punkter och linjer, och att alla storheter 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 $f : \mathbf{R}^2 \rightarrow \mathbf{R}^2$ 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 $\mathbf{R}^2$ 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 två linjer skär varandra i är densamma före och efter funktionen.



Figur 5: En rotation.

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

Exempel 1.2.2. 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 1.2.3. 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 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 att visa att alla dessa exempel faktiskt är automorfier. Låt oss dock t.ex. visa att varje rotation kring origo avbildar linjer bijektivt. Låt L vara linjen $ax+by=c$, och f en moturs rotation med vinkel θ runt origo. Vi vill först visa att $f(L)$ är en linje. Från definitionen kommer $(x, y) \in f(L)$ om och endast om (x, y) ligger på linjen som fås av att rotera L med en vinkel θ moturs, och att detta i sin tur sker om och endast om L 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

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

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

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

eller

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

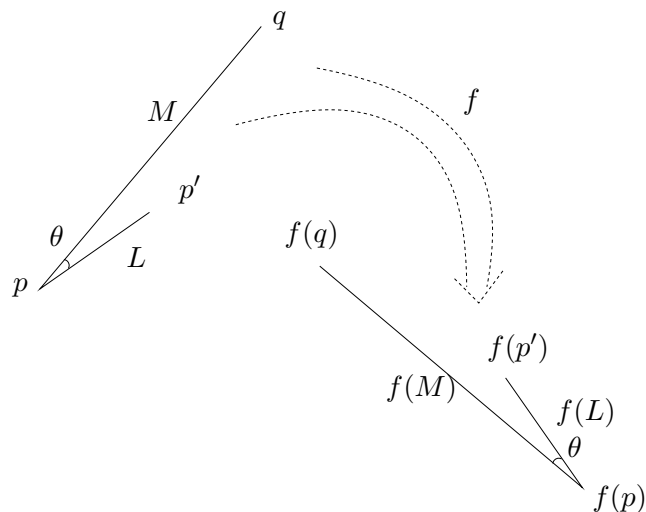
Detta är nu igen ekvationen för en linje. En lustig observation är nu att det enda f gjort var att "rotera" koefficienterna a och b med en vinkel θ ! Om vi jämför med vårt tidigare uttryck för f , ser vi att när linjen L ges av koefficienterna a, b och c , så kommer $f(L)$ att ges av koefficienterna $f(a, b)$ och c . Speciellt gäller ju att $(a, b) \neq (0, 0)$ om och endast om $f(a, b) \neq (0, 0)$, vilket visar att L är en linje om och endast om $f(L)$ 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 1.2.4. *Varja automorfi av det euklidiska planet kan skrivas som en rotation kring origo följt av en translation.*

Vi börjar med ett lemma.

Lemma 1.2.5. *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 $f = g$.*



Figur 6: En illustration av beviset till 1.2.5.

Bevis. Låt $q \in \mathbf{R}^2$ 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 M från q till p . Låt L 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 $f(L) = g(L)$, det vill säga

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

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

Bevis. (av 1.2.4) 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 L vara linjesegmentet från p till p' , och M 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 L och varje fix linje att ändras med θ . Så det existerar en rotation g runt origo sådan att linjesegmentet $g(L)$ (det vill säga, L roterad med en vinkel θ) och M 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 1.2.5 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. $\square$

1.3 Axiom

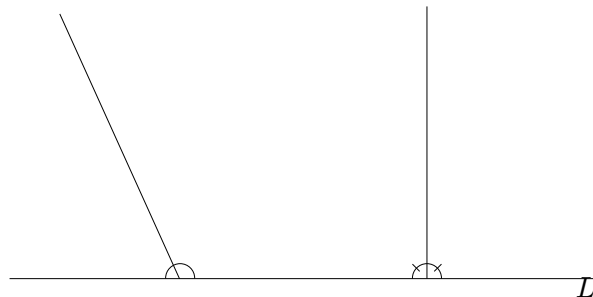
Den presentation av euklidisk geometri vi har gjort i detta kapitel är radikalt annorlunda från Euklides ursprungliga. Euklides använder sig av samma begrepp — punkter, linjer, vinklar — men definierar dessa genom *axiom*. Det vi har gjort är att definiera punkter som par av tal, och definiera geometriska 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 ekvationer och koordinatsystem i geometrin.²

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:³

1. Genom två punkter passerar en unik linje;
2. Varje ändligt långt linjesegment kan dras vidare godtyckligt långt i bägge riktningar;
3. Givet en punkt p och en sträcka r , finns det en unik cirkel med p som mittpunkt och r som radie;
4. Vi börjar med en definition: Låt L vara någon linje, och M ett linjesegment som möter L i sin ena ändpunkt. Mellan L och M 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.
5. Givet en linje och en punkt som inte ligger på denna, finns en unik linje genom punkten som inte skär den givna linjen någonstans.

²René 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”.

³Vi har omformulerat dem något från hur de ser ut i original.



Figur 7: En illustration av fjärde postulatet. Två linjesegment faller in mot linjen L , och det högra möter L i rät vinkel.

Vi har ersatt Euklides ursprungliga femte postulat med det logiskt ekvivalenta *Playfairs axiom*,⁴ 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 rätta vinklar, så måste de första två linjerna skära varandra i en punkt på just denna sidan av den tredje linjen.

Utöver dessa postulat använder sig Euklides av fem axiom 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:

1. Storheter som är lika med samma storhet är lika;
2. Om samma storhet adderas till lika stora storheter, är resultaten lika stora;
3. Om samma storhet subtraheras från lika stora storheter, är resultaten lika stora;
4. Storheter som sammanfaller är lika stora;
5. 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.

⁴Döpt efter den skotske matematikern John Playfair (1748-1819).

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 axiom 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 axiom, 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.

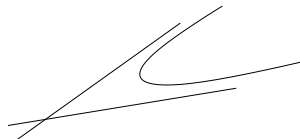
1.4 Historik över parallellpostulatet

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].

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

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 falska bevis för det femte postulatet har föreslagits genom åren, men alla har fallit på att de någonstans på ett subilt 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 falska bevis för (Euklides version av) parallellpostulatet.

För en modern läsare är kanske de intressantaste bland dessa falska 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 hit-



Figur 8: Legendres märkliga konfiguration av räta linjer.

tade 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!⁵

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. Ungraren János Bolyai (1802-1860) och ryssen Nikolaj Lobatjevskij (1792-1856) 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 dittills 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 Poin-

⁵Euklides själv mätte inte vinklar i vare sig radianer eller grader, utan skulle ha formulerat det som att vinkelsumman är mindre än två räta vinklar.

caré [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.

1.5 Övningar

Övning 1.1. Hitta den unika automorfi f som avbildar $(1, 0)$ på $(2, 5)$ respektive $(2, 0)$ på $(2, 6)$. Bestäm även f^{-1} .

Övning 1.2. 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 1.3. Sats 1.2.4 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 1.4. 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 1.5. Avsluta beviset för att en rotation avbildar linjer bijektivt. (Ledning: jämför med Övning 0.5.)

2 Båglängd

I det här kapitlet ska vi studera kurvor i planet, samt deras längder. Vi har redan stött på en speciell sorts kurvor, nämligen räta linjer. En linje kan beskrivas med hjälp av en ekvation som anger vilka punkter i planet som ligger på linjen.

Vi kommer nu att betrakta mer allmänna kurvor som kan vara böjda, börja och sluta i samma punkt och så vidare. Precis som i fallet med räta linjer kan 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. Till skillnad från föregående kapitel, där vi arbetade med begrepp som var kända redan under antiken, kommer vi från och med nu använda mer moderna matematiska begrepp och verktyg som derivator och integraler.

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 planet kan formuleras som den minsta möjliga båglängden för en kurva som förbinder de två punkterna, och vi ska visa att det är de räta linjerna som minimerar detta avstånd.

2.1 Kurvor i planet

En punkt i planet kan beskrivas av två reella tal, (x, y) , som kallas för punktens koordinater. Vi kan beskriva geometriska objekt i planet genom att ange villkor på koordinaterna för de punkter som hör till objektet. Ett sätt att göra detta är att säga att koordinaterna för punkter som hör till vårt geometriska objekt ska uppfylla vissa ekvationer eller olikheter.

Exempel 2.1.1. En linje L i planet är en mängd

$$L = \{(x, y) \in \mathbf{R}^2 : ax + by + c = 0\}$$

för fixa $a, b, c \in \mathbf{R}$. ▲

Exempel 2.1.2. *Enhetscirkeln* $\mathbf{T}$ i planet beskrivas som mängden

$$\mathbf{T} = \{(x, y) \in \mathbf{R}^2 : x^2 + y^2 = 1\}.$$

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

Exempel 2.1.3. En annan delmängd av planet som kommer vara av intresse för oss är *det övre halvplanet*

$$\mathbf{H} = \{(x, y) \in \mathbf{R}^2 : y > 0\}.$$

Notera att koordinaterna för punkter (x, y) i det här exemplet ska satisfiera en olikhet istället för en ekvation. Det leder till en mycket "större" mängd är i det första exemplet. ▲

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 $f : [a, b] \rightarrow \mathbf{R}$ tilldelar ett reellt tal $f(t)$ till varje tal t i intervallet $[a, b]$. Det betyder att två funktioner $x, y : [a, b] \rightarrow \mathbf{R}$ tillsammans ger oss ett par av reella tal $(x(t), y(t))$ för varje $t \in [a, b]$. Vi kan tolka detta par som koordinaterna för en punkt i planet. När t löper igenom intervallet $[a, b]$ från a till b rör sig punkterna $(x(t), y(t))$ i planet, och genom att rita ut $(x(t), y(t))$ 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 $\gamma(t) = ((x(t), y(t)))$ beskriver en bana som vi ska färdas. Variabeln 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 2.1.4. Låt $p = (x_1, y_1)$ och $q = (x_2, y_2)$ vara punkter i planet med $p \neq q$. Betrakta funktionen $\gamma : [0, 1] \rightarrow \mathbf{R}^2$ som ges av

$$\gamma(t) = ((1-t)x_1 + tx_2, (1-t)y_1 + ty_2). \quad (2.1)$$

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

Man kan nämligen kontrollera att $\gamma(t) = (x(t), y(t))$ satisfierar ekvationen

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

för varje t , vilket betyder att mängden $\{\gamma(t) : t \in [0, 1]\}$ uppfyller kravet i Definition 1.1.4 av allmänna linjer i planet.

I ändpunkten $t = 0$ har vi $\gamma(0) = (x_1, y_1)$ och om vi istället sätter in $t = 1$ i (2.1) får vi $\gamma(1) = (x_2, y_2)$. Det visar i sin tur att linjen verkligen passerar genom de två punkterna p och q . ▲

Exempel 2.1.5. Vi betraktar för $t \in [0, \pi/2]$ funktionen

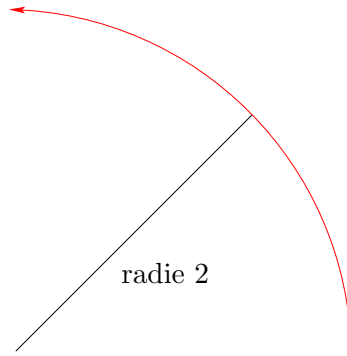
$$\gamma(t) = (x(t), y(t)) = (2 \cos t, 2 \sin t).$$

Vi ser att

$$\gamma(0) = (x(0), y(0)) = (2, 0)$$

och att

$$\gamma(\pi/2) = (x(\pi/2), y(\pi/2)) = (0, 2).$$



Figur 9: Parameterkurvan genomlöper en kvartscirkel.

Vidare är avståndet från punkten $(0, 0)$ till en punkt $\gamma(t)$ på kurvan konstant lika med $\sqrt{4\cos^2 t + 4\sin^2 t} = 2$.

Vi inser nu att vi har fått en beskrivning av en kvartscirkel med radie 2, och vi rör oss längs med kvartscirkeln moturs, från $(2, 0)$ till $(0, 2)$. ▲

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

Definition 2.1.6. En *parameterfunktion* är en funktion

$$\gamma : [a, b] \rightarrow \mathbf{R}^2, \quad t \mapsto (x(t), y(t)); \quad (2.2)$$

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 $y'(t)$.

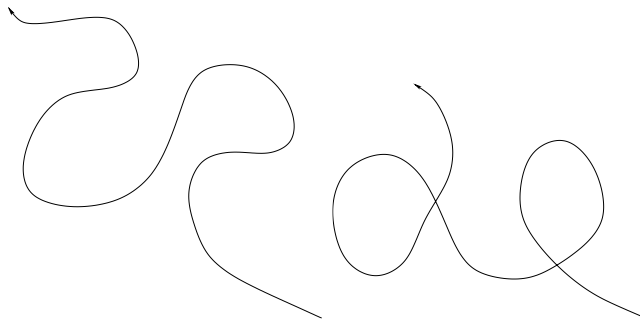
Kurvan Γ är mängden

$$\Gamma = \{\gamma(t) : t \in [a, b]\} \quad (2.3)$$

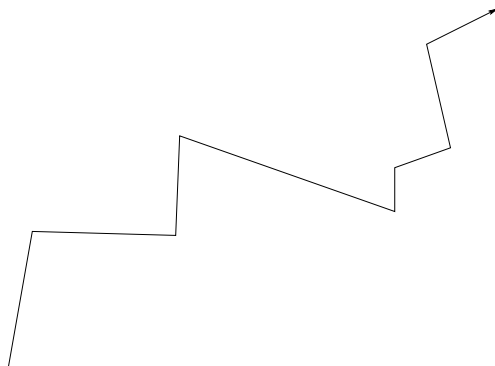
det vill säga, bildmängden till parameterfunktionen γ .

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 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 godtyckligt 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, men som ändå är relativt trevliga. Ett exempel är så kallade polygontåg (se Figur 3). Vi kommer att anse att ett polygontåg är uppbyggt av flera på varandra följande kurvor—nämligen linjesegment. Hela polygontåget kommer



Figur 10: Icke-skärande och skärande kurvor.



Figur 11: Ett polygontåg.

då ha tangenter i alla punkter utom i limningspunkterna, 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 (2.1). Vi hade även kunna betrakta parameterfunktionen

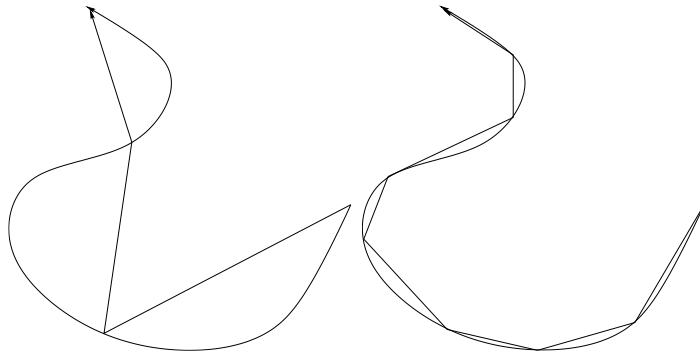
$$\tilde{\gamma}(t) = ((1-t)x_2 + tx_1, (1-t)y_2 + ty_1), \quad t \in [0, 1].$$

Dess bild är samma geometriska objekt, nämligen linjen, men eftersom $\tilde{\gamma}(0) = (x_2, y_2) = q$ anger denna parameterfunktion att vi ska färdas åt andra hållet på linjen när t genomlöper intervallet $[0, 1]$.

2.2 Linjeintegraler

Det är naturligtvis av intresse att kunna mäta längden av en kurva i 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 (a_1, a_2) och (b_1, b_2) vara två punkter i planet,



Figur 12: Approximationer av en kurva med polygontåg med $n = 3$ (vänster) respektive $n = 8$ (höger).

och låt L vara det rätta linjesegmentet mellan dem. Från Pythagoras sats får vi att längden av L ges av

$$l(L) = \sqrt{(b_1 - a_1)^2 + (b_2 - a_2)^2}.$$

Betrakta nu en parameterfunktion $\gamma: [0, 1] \rightarrow \mathbf{R}^2$, 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

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

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).

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

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

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

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

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

$$l(K_n) = \frac{1}{n} \sum_{j=0}^{n-1} \sqrt{(x'(\tau_j))^2 + (y'(\sigma_j))^2}. \quad (2.4)$$

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

Låt oss bilda en funktion $h : \mathbf{R} \rightarrow \mathbf{R}$ genom att sätta

$$h(t) = \sqrt{(x'(t))^2 + (y'(t))^2}$$

Eftersom $x'(t)$ och $y'(t)$ är kontinuerliga är även $h(t)$ kontinuerlig. Vi ser nu att (2.4) påminner rätt mycket om

$$\frac{1}{n} \sum_{j=0}^n h(\tau_j),$$

som är den typ av Riemannsumma man brukar bilda när man ska beräkna integralen av funktionen h . 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 (2.4) konvergerar mot $\int_0^1 h(t)dt$ 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 (2.4) går mot

$$\int_0^1 h(t)dt = \int_0^1 \sqrt{(x'(t))^2 + (y'(t))^2} dt.$$

när $n \rightarrow \infty$.

Man tänker ju ofta på integralen av en funktion som arean (med tecken) under funktionsgraf, 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 2.2.1. Låt $\gamma : [a, b] \rightarrow \mathbf{R}^2$ med $\gamma(t) = (x(t), y(t))$ vara en parameterfunktion som ger upphov till en kurva Γ i planet. Då ges längden av Γ av

$$l(\Gamma) = \int_a^b \sqrt{(x'(t))^2 + (y'(t))^2} dt. \quad (2.5)$$

Vi noterar här att vårt antagande att $x'(t)$ och $y'(t)$ är kontinuerliga medför att integralen i högerleder i (2.5) existerar, se Sats A.2.1.

Exempel 2.2.2. Vi beräknar längden av kvartscirkelbågen i vårt tidigare exempel, där vi alltså hade

$$\gamma(t) = (x(t), y(t)) = (2 \cos t, 2 \sin t).$$

Vi har

$$x'(t) = -2 \sin t, \quad y'(t) = 2 \cos t,$$

och identiteten $\cos^2 t + \sin^2 t = 1$ ger oss

$$\sqrt{(x'(t))^2 + (y'(t))^2} = \sqrt{4 \sin^2 t + 4 \cos^2 t} = \sqrt{4} = 2.$$

Alltså får vi

$$\int_0^{\frac{\pi}{2}} \sqrt{(x'(t))^2 + (y'(t))^2} dt = 2 \int_0^{\frac{\pi}{2}} dt = 2 \cdot \frac{\pi}{2} = \pi$$

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 automorfier av planet.

2.3 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 2.3.1. Låt $\gamma : [\alpha, \beta] \rightarrow \mathbf{R}^2$ 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 sammansatta funktionen $\gamma \circ h : [a, b] \rightarrow \mathbf{R}^2$ är en omparametrisering 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 $\gamma \circ h$.

Nästa sats visar att det faktiskt inte spelar 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 2.3.2. *Låt Γ vara en kurva som hör till parameterfunktionen γ . Då är längden av kurvan, $l(\Gamma)$, oberoende av omparametriseringar.*

Bevis. Vi betraktar parameterfunktionen som ges av sammansättningen av γ med en omparametrisering $h: [a, b] \rightarrow [\alpha, \beta]$:

$$(\gamma \circ h)(t) = (x(h(t)), y(h(t))).$$

Vi kan för enkelhetens skull antaga att $h'(t) > 0$. Enligt kedjeregeln för derivering gäller att

$$(\gamma \circ h)'(t) = h'(t)(x'(h(t)), y'(h(t))).$$

Vi får därför

$$\sqrt{[(x(h(t)))']^2 + [(y(h(t)))']^2} = |h'(t)| \sqrt{(x'(h(t)))^2 + (y'(h(t)))^2}.$$

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

$$\int_a^b |h'(t)| \sqrt{(x'(h(t)))^2 + (y'(h(t)))^2} dt.$$

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 x' och y' 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

$$\int_a^b \sqrt{(x'(h(t)))^2 + (y'(h(t)))^2} |h'(t)| dt = \int_\alpha^\beta \sqrt{(x'(s))^2 + (y'(s))^2} ds.$$

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. Vi ska nu precisera detta genom att visa att längden av en kurva är invariant under automorfier av planet.

Sats 2.3.3. *Låt f vara en automorfi av planet. 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 1.2.4. Det räcker därför att visa att längden av en kurva är invariant under dessa avbildningar.

Vi börjar med translationerna. Låt alltså automorfin f ges av

$$f(x, y) = (x + a, y + b), \quad a, b \in \mathbf{R}.$$

Om γ är en parameterframställning av Γ gäller att

$$(f \circ \gamma)(t) = (x(t) + a, y(t) + b).$$

Eftersom derivatan av en konstant är 0 får vi att

$$(f \circ \gamma)'(t) = ((x(t) + a)', (y(t) + b)') = (x'(t), y'(t)),$$

och vi ser att längden av kurvan $f(\Gamma)$ blir

$$l(f(\Gamma)) = \int_a^b \sqrt{[(x(t) + a)']^2 + [(y(t) + b)']^2} dt = \int_a^b \sqrt{[x'(t)]^2 + [y'(t)]^2} dt,$$

vilket är samma sak som $l(\Gamma)$.

Vi betraktar nu en rotation f av planet moturs med en vinkel θ (se 1.2.3); det vill säga,

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

Vi får nu

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

Vi noterar att

$$(x'(t) \cos \theta - y'(t) \sin \theta)^2 = (x'(t))^2 \cos^2 \theta + (y'(t))^2 \sin^2 \theta - 2 \cos \theta \sin \theta x'(t) y'(t) \quad (2.6)$$

samt att

$$(x'(t) \sin \theta + y'(t) \cos \theta)^2 = (x'(t))^2 \sin^2 \theta + (y'(t))^2 \cos^2 \theta + 2 \cos \theta \sin \theta x'(t) y'(t). \quad (2.7)$$

Vi lägger ihop dessa två uttryck och får

$$(x'(t))^2 (\cos^2 \theta + \sin^2 \theta) + (y'(t))^2 (\cos^2 \theta + \sin^2 \theta) = (x'(t))^2 + (y'(t))^2.$$

Eftersom summan (2.6) och (2.7) är precis det uttryck som dyker upp under rottecknet i integralen som ger oss längden av $f(\Gamma)$ drar vi återigen slutsatsen att $l(f(\Gamma)) = l(\Gamma)$.

Detta avslutar beviset. □

2.4 Avstånd och geodeter

Vi avslutar det här kapitlet med att införa ett avståndsbegrepp i planet med hjälp av längder av kurvor.

Låt $p_1 = (x_1, y_1)$ och $p_2 = (x_2, y_2)$ vara två olika punkter i 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 $\gamma : [a, b] \rightarrow \mathbf{R}^2$ 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 (2.1). Det finns många fler möjligheter. De har ändå alla en sak gemensamt—deras båglängd kan inte vara mindre än den räta linjens.

Sats 2.4.1. *Låt p_1 och p_2 vara två punkter i planet, och låt L vara det räta linjesegmentet som förbinder dem. Då gäller*

$$l(L) \leq l(\Gamma)$$

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

$$\tilde{p}_1 = (0, b_1); \quad \tilde{p}_2 = (0, b_2).$$

Då båglängd är invariant under automorfier räcker det att visa att $l(L) \leq l(\Gamma)$ 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 $\gamma(t) = (x(t), y(t))$ definierad på $[0, 1]$ med $\gamma(0) = (0, b_1)$ och $\gamma(1) = (0, b_2)$. En parameterfunktion vars bild är linjesegmentet som förbinder $\tilde{p}_1$ och $\tilde{p}_2$ är funktionen

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

Vi noterar att

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

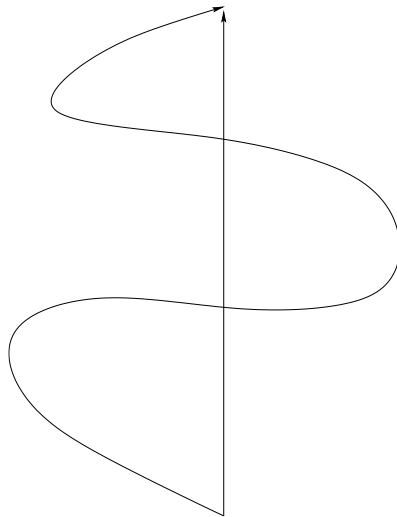
Därmed gäller

$$l(\Gamma) = \int_0^1 \sqrt{(x'(t))^2 + (y'(t))^2} dt \geq \int_0^1 |y'(t)| dt = (b_2 - b_1). \quad (2.8)$$

I sista steget har vi använt att båglängden av den räta linjen är oberoende av valet av parameterfunktion för L . Vi kan alltså speciellt välja den vanliga parameterfunktionen $\gamma(t) = (0, (1-t)b_1 + tb_2)$ när vi ska beräkna $l(L)$ i sista steget.

Satsen är nu bevisad. □

Anmärkning 2.4.2. Man kan visa att likhet inträffar i (2.8) precis när $x'(t) = 0$ för alla t —det vill säga, precis när Γ är en rät linje.



Figur 13: En rät linje minimerar avståndet mellan två punkter.

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

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

där L är den räta linjen som förbinder p_1 och p_2 . Precis vilken parameterfunktion vi väljer spelar ingen roll—längden av L 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*.

2.5 Övningar

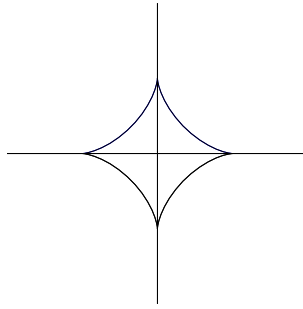
Övning 2.1. Verifiera att punkterna på kurvan som hör till parameterfunktionen $\gamma: [0, 1] \rightarrow \mathbf{R}^2$ som ges av

$$\gamma(t) = (x(t), y(t)) = ((1 - t)x_1 + tx_2, (1 - t)y_1 + ty_2))$$

satisfierar

$$(x_2 - x_1)y(t) - (y_2 - y_1)x(t) - y_1(x_2 - x_1) + x_1(y_2 - y_1) = 0, \quad (2.9)$$

det vill säga, att Γ är ett linjesegment.



Figur 14: Asteroidkurvan.

Övning 2.2. Beräkna båglängden av kurvan Γ som hör till parameterfunktionen $\gamma: [0, 1] \rightarrow \mathbf{R}^2$ som ges av

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

Försök även att rita upp kurvan.

Övning 2.3. Beräkna båglängden av den så kallade *asteroidkurvan* vilken ges av parameterfunktionen $\gamma: [0, 2\pi] \rightarrow \mathbf{R}^2$, 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 2.4. Beräkna båglängden av kurvan Γ som ges av parameterfunktionen $\gamma: [0, 1] \rightarrow \mathbf{R}^2$ med

$$\gamma(t) = \left(\frac{e^t}{\sqrt{2}} \cos t, \frac{e^t}{\sqrt{2}} \sin t \right), \quad t \in [0, 1].$$

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 $\mathbf{C}$. 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 $\mathbf{R}^2$ kan på ett elegant sätt identifieras med vissa avbildningar av $\mathbf{C}$. 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 $\mathbf{R}^2$ 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 $\mathbf{R}$ 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 $\mathbf{R}^2$. *Summan* av p och q ges av

$$p + q = (a + c, b + d) \quad (3.1)$$

och *differensen* av p och q är

$$p - q = (a - c, b - d) \quad (3.2)$$

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.

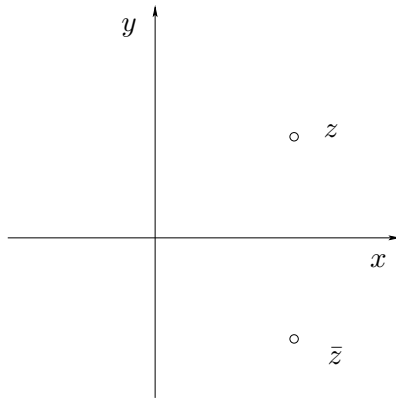
Vi fortsätter med att definiera multiplikation.

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

$$p \cdot q = (ac - bd, ad + bc). \quad (3.3)$$

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 .

Vi kommer för enkelhetens skull oftast att skriva pq för produkten $p \cdot q$. Ibland vill vi påkalla läsaren uppmärksamhet till någon viss faktor i en produkt, och då skriver vi ut multiplikationstecknet.



Figur 15: Punkten z och dess konjugat $\bar{z}$.

Exempel 3.1.3. Sätt $p = (\sqrt{2}, 1)$ och $q = (1/17, -1)$. Då har vi enligt definitionen av addition och subtraktion att

$$p + q = (\sqrt{2}, 1) + (1/17, -1) = \left(\frac{17\sqrt{2} + 1}{17}, 0 \right)$$

samt

$$p - q = p + q = (\sqrt{2}, 1) - (1/17, -1) = \left(\frac{17\sqrt{2} - 1}{17}, 2 \right).$$

Vi använder definitionen av produkt för att beräkna

$$\begin{aligned} pq &= (\sqrt{2}, 1)(1/17, -1) = (\sqrt{2} \cdot 1/17 - 1 \cdot (-1), \sqrt{2} \cdot (-1) + 1 \cdot (1/17)) \\ &= \left(\frac{\sqrt{2} + 17}{17}, \frac{1 - 17\sqrt{2}}{17} \right). \end{aligned}$$

▲

Nu återstår bara att ange vad vi ska mena med division av punkter i planet. Innan vi gör detta ska vi införa lite notation.

Definition 3.1.4. *Konjugatet* av $p \in \mathbf{R}^2$ är

$$\bar{p} = (a, -b). \tag{3.4}$$

Geometriskt innebär konjugering att vi speglar i x -axeln. Vi genomför en liten räkning som visar sig vara användbar:

$$p \cdot \bar{p} = (a, b) \cdot (a, -b) = (aa - (-b)b, a(-b) + ba) = (a^2 + b^2, -ab + ab) = (a^2 + b^2, 0).$$

Vi har här bland annat utnyttjat att $ab = ba$ för reella tal. Vi noterar att andra komponenten av $p \cdot \bar{p}$ alltid är noll. Om x är ett nollskilt reellt tal vet vi

vad som menas med $1/x$, och detta betyder att vi kan ge det komplexa talet $1/p\bar{p}$ tolkningen

$$\frac{1}{p\bar{p}} = \left(\frac{1}{a^2 + b^2}, 0 \right),$$

så länge $p \neq 0 + 0i$.

Vi är nu redo att definiera division för punkter i planet: vi förlänger helt enkelt p/q med $\bar{q}$ i både täljare och nämnare.

Definition 3.1.5. Låt $p = (a, b)$ vara en godtycklig punkt i planet, och låt $q = (c, d)$ vara en punkt i planet skild från origo. Kvoten av p och q definieras som punkten i planet som ges av

$$\frac{p}{q} = \frac{1}{q\bar{q}} \cdot p \cdot \bar{q} \quad (3.5)$$

Exempel 3.1.6. Låt $p = (\sqrt{2}, 1)$ och $q = (1, -1)$. Vi har $q\bar{q} = (2, 0)$ samt $p\bar{q} = (\sqrt{2} + 1, 1 - \sqrt{2})$, vilket ger oss $p/q = ((\sqrt{2} + 1)/2, (1 - \sqrt{2})/2)$. ▲

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

$$i = (0, 1) \quad (3.6)$$

och kommer lite oegentligt att skriva

$$1 = (1, 0). \quad (3.7)$$

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 Exemplet 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

$$z = a + ib, \quad \text{där } a, b \in \mathbf{R}. \quad (3.8)$$

Varje komplext tal svarar mot en punkt $p \in \mathbf{R}^2$ med $p = (a, b)$. En punkt $p = (a, b) \in \mathbf{R}^2$ 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,

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

brukar betecknas med $\mathbf{C}$.

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 $\mathbf{R}$ kan identifieras med en delmängd av de komplexa talen, nämligen med mängden

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

vilken sammanfaller med x -axeln i planet. Våra räkningar i Exempel 3.2.1 visar att denna mängd är sluten under de algebraiska operationerna.

Vi kommer att använda beteckningen $\bar{z} = a - ib$ för konjugatet av det komplexa talet $z = a + ib$. *Absolutbeloppet* av ett komplext tal $z = a + ib$ är den reella storheten

$$|z| = \sqrt{a^2 + b^2}, \quad (3.9)$$

vilket geometriskt sett är inget annat än avståndet från origo till punkten $p = (a, b)$. En liten räkning visar att

$$z \cdot \bar{z} = |z|^2,$$

vilket betyder att vi kan skriva kvoten av två komplexa tal som

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

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.

3.3 Komplexa tal på polär form

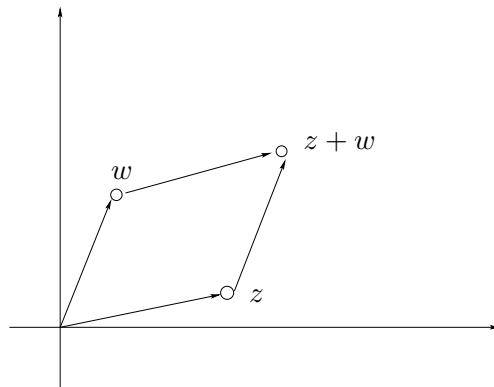
Eftersom de komplexa talen kan identifieras med punkter i planet känns det naturligt att försöka tolka räkneoperationerna på $\mathbf{C}$ geometriskt.

Vi börjar med att betrakta addition och subtraktion. Låt därför $z = a + ib$ och $w = c + id$ vara två komplexa tal. Vi kan rita upp dem som punkter i planet genom att använda a respektive c som x -koordinater och b respektive d som y -koordinater. Summan $z + w$ är enligt definitionen det komplexa talet $a + c + i(b + d)$, och vi ser att motsvarande punkt har x -koordinaten $a + c$ och y -koordinaten $b + d$. Detta betyder att punkten som svara mot summan $z + w$ är hörnpunkten i parallelogrammen som bestäms av z och w (se Figur 7). På ett liknande sätt kan man ge subtraktion av komplexa tal en geometrisk tolkning. Vi överlämnar detta som en övning.

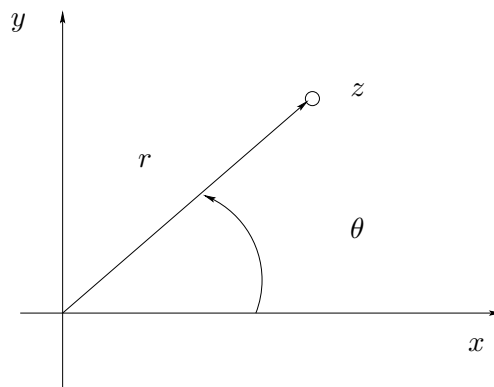
För att kunna ge multiplikation och division en geometrisk tolkning behöver vi införa ett annat sätt att beskriva komplexa tal som punkter i planet. Vi betraktar återigen en figur. Vi kan beskriva en punkt i planet genom att ange dess avstånd från origo, r , samt den vinkel som den räta linjen från origo till punkten bilda med x -axeln. Vi beräknar avståndet med hjälp av Pythagoras sats och får

$$r = \sqrt{a^2 + b^2}, \quad (3.10)$$

vilket är samma sak som absolutbeloppet av det komplexa talet $z = a + ib$. vi noterar att $r = 0$ om och endast om $p = (0, 0)$.



Figur 16: Geometrisk tolkning av $z + w$.



Figur 17: Polära koordinater.

Vinkeln θ kallas för *argumentet* av z , och vi skriver $\theta = \arg(z)$. Om $z = 0$ är $\arg(z)$ odefinierat. Genom att projicera på x -axeln och y -axeln ser vi att x -koordinaten för punkten z kan skrivas med hjälp av θ som $r \cos \theta$, medan y -koordinaten blir $r \sin \theta$. I den sedvanliga $a + ib$ -notationen blir detta

$$z = r(\cos \theta + i \sin \theta) \quad (3.11)$$

eller $z = |z|(\cos \theta + i \sin \theta)$ om man föredrar det.

Låt nu $w = s(\cos \phi + i \sin \phi)$ vara ett annat komplext tal. Vi multiplicerar de komplexa talen z och w och får

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

och genom att använda additionsformlerna för sinus och cosinus får vi

$$z \cdot w = rs[\cos(\theta + \phi) + i \sin(\theta + \phi)]. \quad (3.12)$$

Nu ser vi vad multiplikation av två komplexa tal innebär geometriskt: man multiplicerar talens avstånd från origo med varandra och adderar vinklarna som de rätta linjerna till punkterna bildar med x -axeln. Vi kan sammanfatta detta med formlerna

$$|z \cdot w| = |z| \cdot |w| \quad \text{och} \quad \arg(zw) = \arg(z) + \arg(w). \quad (3.13)$$

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

$$\frac{z}{w} = \frac{r}{s}[\cos(\theta - \phi) + i \sin(\theta - \phi)], \quad (3.14)$$

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

$$\left| \frac{z}{w} \right| = \frac{|z|}{|w|} \quad \text{och} \quad \arg(z/w) = \arg(z) - \arg(w). \quad (3.15)$$

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*

$$e^{i\theta} = \cos \theta + i \sin \theta, \quad (3.16)$$

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:

$$z = re^{i\theta}$$

där alltså $r = |z|$ och $\theta = \arg(z)$. Genom att använda formeln (3.12), med $r = s = 1$, erhåller vi

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

Vi har vidare $ie^{i\theta} = i(\cos \theta + i \sin \theta) = i \cos \theta + i^2 \sin \theta = -\sin \theta + i \cos \theta$ eftersom $i^2 = -1$. Detta visar att derivatan av den komplexa exponentialfunktionen med avseende på θ är

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

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.

3.4 Parameterkurvor i komplexa planet

Vi införde i föregående kapitel ett sätt att beskriva en kurva i planet tillsammans med en riktning i vilken vi vill färdas längs med kurvan. Vi talade om parameterframställningar: funktioner $\gamma : [a, b] \rightarrow \mathbf{R}^2$ som ges av

$$\gamma(t) = (x(t), y(t)).$$

Parameterframställningen anger, för varje värde t_0 , i vilken punkt $(x(t_0), y(t_0))$ i planet vi befinner oss.

Eftersom vi kan identifiera punkter i planet med komplexa tal inser vi att vi även kan skriva parameterkurvor på komplex form. Vi har då

$$\gamma(t) = x(t) + iy(t), \tag{3.17}$$

vilket kan tolkas som en funktion $\gamma : [a, b] \rightarrow \mathbf{C}$. Med denna notation kan paret $(x'(t), y'(t))$ som dyker upp i båglängdsintegralen skrivas

$$\gamma'(t) = x'(t) + iy'(t), \tag{3.18}$$

vilket även det är en funktion som antar komplexa värden. Vi tar högerledet i (3.18) som definition av derivata av en komplexvärd funktion av ett reellt tal: man deriverar real-och imaginärdel för sig.

Definitionen av absolutbelopp visar också att

$$|\gamma'(t)| = \sqrt{(x'(t))^2 + (y'(t))^2}, \tag{3.19}$$

Vilket innebär att båglängden av en parameterkurva Γ i komplex notation ges av

$$l(\Gamma) = \int_a^b |\gamma'(t)| dt. \tag{3.20}$$

3.5 Automorfier av planet

Vi har tidigare sett att alla automorfier av planet är sammansatta av en rotation kring origo följt av en translation. (Sats 1.2.4.) Vi ska nu identifiera rotationer och translationer med vissa speciella funktioner $f: \mathbf{C} \rightarrow \mathbf{C}$, det vill säga, funktioner som tar ett komplext tal z på ett komplext tal $f(z)$.

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

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

där $\alpha \in \mathbf{C}$ 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 åt läsaren att kontrollera att sammansättningen av två funktioner på formen $\alpha z + \beta$ med där $\alpha \in \mathbf{C}$ uppfyller $|\alpha| = 1$ är en ny funktion på samma form.

3.6 Komplexa polynom

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

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

$$f(z) = a_n z^n + a_{n-1} z^{n-1} + \cdots + a_1 z + c_0, \quad (3.21)$$

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

▲

I exemplet ovan väljer man alltså koefficienterna a_i och varierar z : för varje $z \in \mathbf{C}$ ger formeln (3.21) 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

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

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.

Derivering av komplexa funktioner är också möjlig. Man kan till exempel visa att de vanliga komplexa funktionernas derivator uppvisar de egenskaper man väntar sig, som exempelvis

$$(e^z)' = e^z, \quad (\sin z)' = \cos z \tag{3.22}$$

och så vidare. Däremot är det svårt att tolka den komplexa derivatan som lutningen i en punkt på en funktionsgraf. Det finns ändå en naturlig geometrisk tolkning av komplexa derivata men, vi kan tyvärr inte fördjupa oss mer i detta här. Funktioner av komplexa tal studeras ingående i en gren av matematiken som heter *komplex analys*.

3.8 Komplex derivata

Att införa den komplexa derivatan på ett rigoröst sätt med hjälp av gränsvärden ligger dessvärre utanför den här kursens innehåll. Vi ger istället ett slags algebraisk definition av derivata—vi nöjer oss med att derivera polynom, samt produkter och kvoter av dessa.

Definition 3.8.1. Derivatan av ett komplext monom, $p(z) = z^n$ för $n \geq 1$, är det komplexa monomet $p'(z) = nz^{n-1}$. Derivatan av ett konstant monom, $p(z) = c$ för $c \in \mathbf{C}$ är $p'(z) = 0$.

Vi utvidgar definitionen av derivata till allmänna polynom genom att kräva att derivatan ska vara linjär, det vill säga,

$$(a_n z^n + a_{n-1} z^{n-1} + \cdots + a_1 z + a_0)' = n a_n z^{n-1} + (n-1) a_{n-1} z^{n-2} + \cdots + a_1, \quad a_j \in \mathbf{C}.$$

Vi deriverar helt enkelt varje monom-term för sig och flyttar ut konstanterna a_j .

Exempel 3.8.2. Betrakta polynomet $p(z) = (\pi + 2i)z^{17} + 2z^3 + 9z + \sqrt{2}$. Dess derivata är polynomet $p'(z) = 17(\pi + 2i)z^{16} + 6z^2 + 9$. ▲

Man kan, lite förenklat, säga att komplex derivering fungerar på precis samma sätt som vanlig derivering—fastän argumentet z är komplext och inte reellt.

Definition 3.8.3. Produktregeln och kvotregeln gäller vid derivering av komplexa polynom, det vill säga

$$(p(z)q(z))' = p'(z)q(z) + p(z)q'(z) \quad \text{och} \quad \left(\frac{p(z)}{q(z)}\right)' = \frac{p'(z)q(z) - p(z)q'(z)}{(q(z))^2}.$$

Exempel 3.8.4. Vi deriverar funktionen

$$f(z) = \frac{3z^2 + i}{4z - 7}.$$

Vi sätter $p(z) = 3z^2 + i$ och $q(z) = 4z - 7$. Vi beräknar först

$$p'(z) = (3z^2 + i)' = (3z^2)' + i' = 3 \cdot 2z + 0 = 6z$$

och

$$q'(z) = (4z - 7)' = (4z)' - 7' = 4 \cdot 1 + 0 = 4.$$

Sedan använder vi kvotregeln för att få

$$f'(z) = \frac{p'(z)q(z) - p(z)q'(z)}{(q(z))^2} = \frac{6z \cdot (4z - 7) - (3z^2 + i) \cdot 4}{(4z - 7)^2} = \frac{12z^2 - 42z - 4i}{16z^2 - 56z + 49}$$

▲

Definition 3.8.5. Låt $g : \mathbf{C} \rightarrow \mathbf{C}$ vara en kvot av polynom och låt $f : [a, b] \rightarrow \mathbf{C}$ vara en parameterframställning av en kurva Γ . Derivatan av den komplexa funktionen $g \circ f : [a, b] \rightarrow \mathbf{C}$ ges av

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

där f' ska tolkas enligt Ekvation (3.18) och g' är den komplexa derivatan av g evaluerad i punkten $f(t)$.

Övningar

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

1. $z = 3 + i$ och $w = -1 + 4i$

2. $z = -\sqrt{2}i$ 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 $z, w \in \mathbf{C}$.

Övning 3.4. Lös andragradsekvationen $z^2 + iz + 2 = 0$ genom att använda kvadratkomplettering.

Övning 3.5. Visa att derivatan av den komplexa exponentialfunktionen $f(\theta) = e^{i\theta}$ är $ie^{i\theta}$.

Övning 3.6. 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 $A, B \in \mathbf{C}$ och $|A| = 1$.

4 Möbiusgruppen och Riemannsfären

I kapitel ett har vi beskrivit strukturen hos det Euklidiska planet. Speciellt diskuterades *automorfierna* av det Euklidiska planet, som är de funktioner som bevarar all geometrisk struktur — linjer, avstånd, vinklar. Vi fann sedan i kapitel tre att de kan identifieras med komplexa funktioner på formen $f(z) = \alpha z + \beta$, där $|\alpha| = 1$. Vårt mål i den här kursen är att ge en liknande beskrivning av det hyperboliska planet. Det visar sig att vi för att beskriva automorfierna i det hyperboliska fallet behöver använda en större klass av funktioner än de som dyker upp i det euklidiska fallet. Automorfierna i hyperbolisk geometri kommer nämligen att ges av en viss delmängd av de så kallade *Möbiustransformationerna*.

I detta kapitel kommer vi först att definiera mängden av Möbiusavbildningar och visa vissa egenskaper om denna mängd. Till exempel så kommer vi att visa att Möbiusavbildningarna är ett exempel på en *grupp*, vilket är en nödvändighet för att vi ska kunna tänka på mängden av Möbiusavbildningarna som automorfier till något. Dessutom kommer vi att visa en geometrisk konstruktion som kallas för *Riemannsfären*. Genom att tänka på Möbiusavbildningarna som funktioner från denna sfär till sig själv kan vissa lite märkliga egenskaper hos Möbiusavbildningarna göras begripliga.

4.1 Möbiusgruppen

Definition 4.1.1. En *Möbiustransformation* (eller *Möbiusavbildning*) är en funktion $\mathbf{C} \rightarrow \mathbf{C}$ som ges av

$$z \mapsto \frac{az + b}{cz + d}, \quad (4.1)$$

där a, b, c och d är komplexa tal som uppfyller $ad - bc = 1$. Mängden av alla Möbiustransformation brukar kallas för *Möbiusgruppen*, $\text{Möb}(\mathbf{C})$.

Anmärkning 4.1.2. 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 4.1.3. Egentligen är definitionen lite slarvig: om $c \neq 0$ kommer det att finnas en punkt där funktionen är odefinierad, ty om $z = -d/c$ får vi division med noll. Man borde skriva att en Möbiusavbildning är en funktion $\mathbf{C} \rightarrow \mathbf{C}$ om $c = 0$, och en funktion $\mathbf{C} \setminus \{-d/c\} \rightarrow \mathbf{C}$ om $c \neq 0$. Vi kommer senare att se vad det här beror på och hur man ska tänka på att Möbiusavbildningar ibland är odefinierade i en punkt.

Anmärkning 4.1.4. Det är viktigt att hålla isär själva *funktionen* f , och det algebraiska uttrycket som används för att definiera funktionen. Till exempel är funktionerna

$$f(z) = \frac{1}{z+1} \quad \text{och} \quad g(z) = \frac{5}{5z+5}$$

exakt samma funktion — detta betyder endast att $f(z) = g(z)$ för alla z — även om de är skrivna på olika sätt. Eftersom f uppenbarligen ligger i Möbiusgruppen, och $f = g$, ser vi att även g är en Möbiusavbildning. Med andra ord kan vi (om vi vill) multiplicera alla talen a, b, c och d i uttrycket för en Möbiusavbildning med någon nollskild konstant α , och resultatet är fortfarande en Möbiusavbildning. Notera att $ad - bc$ då kommer multipliceras med α^2 , och inte nödvändigtvis längre vara lika med ett. Det här innebär att kan vi ersätta villkoret $ad - bc = 1$ med villkoret $ad - bc \neq 0$, och vi får exakt samma mängd av funktioner. Låt oss visa detta.

Lemma 4.1.5. *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 4.1.6. Det vi egentligen visat är att alla tal i $\mathbf{C}$ har en kvadratrots.⁶ Men om vi till exempel arbetade med reella värden på a, b, c och d gäller 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.

Anmärkning 4.1.7. Vid första åsynen av definitionen av en Möbiusavbildning ser kanske villkoret $ad - bc = 1$ ut att vara helt taget ur luften. En bra sak är förstås att det utesluter att c och d båda är noll, som skulle göra funktionen odefinierad i varje punkt. Men vad händer om vi tar bort det? Enligt 4.1.5 är detta samma sak som att tillåta att $ad - bc = 0$. Med andra ord, vad är $f(z) = (az + b)/(cz + d)$ för slags funktion om $ad - bc = 0$? För att se vad

⁶Egentligen två stycken, eftersom om $\alpha^2 = \beta$ är även $(-\alpha)^2 = \beta$.

som gör att dessa funktioner ska uteslutas ur $\text{Möb}(\mathbf{C})$, låt oss *derivera* f med kvotregeln:

$$f'(z) = \frac{a(cz + d) - c(az + b)}{(cz + d)^2} = \frac{ad - bc}{(cz + d)^2} = 0,$$

så f är bara en konstant funktion när $ad - bc = 0$! Eftersom vi infört Möbiusgruppen för att vi vill beskriva automorfier av hyperboliska planet, och en konstant funktion absolut inte kan vara en automorfi, är det rimligt att utesluta dessa. I argumentet har vi använt två saker vi inte visat: dels att alla vanliga deriveringsregler gäller även för funktioner $\mathbf{C} \rightarrow \mathbf{C}$, och dels att påståendet att en funktion med derivatan noll överallt är en konstant funktion fortfarande är sant för komplexa funktioner. Egentligen bör man ge argument för dessa påståenden, men bevisen skulle ta oss långt från ämnet och är i princip likadana som i det reella fallet.

4.2 Gruppbegreppet

Sats 4.2.1. *Mängden av funktioner $\text{Möb}(\mathbf{C})$ har följande tre egenskaper:*

1. *Identitetsavbildningen id som avbildar varje punkt på sig själv ligger i $\text{Möb}(\mathbf{C})$;*
2. *Om f och g ligger i $\text{Möb}(\mathbf{C})$ gör även $f \circ g$ det, vilket alltså säger att $f(g(z))$ är en Möbiusavbildning;*
3. *För varje $f \in \text{Möb}(\mathbf{C})$ 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. 1. 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 $ad - bc = 1$.

2. Låt oss välja två Möbiustransformationer

$$f(z) = \frac{az + b}{cz + d} \quad \text{och} \quad g(z) = \frac{Az + B}{Cz + D}. \quad (4.2)$$

Då är

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

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.

3. 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 2:

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

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

Anmärkning 4.2.2. En mängd av funktioner med de tre egenskaperna i 4.2.1 brukar i matematiken kallas för en grupp. (Därav namnet Möbiusgruppen!) Man kan tänka på de tre egenskaperna som att de garanterar att det är vetigt att tänka på mängden av funktioner som en mängd av automorfier. Vi har hittills beskrivit mängden av automorfier till det euklidiska planet, och när vi senare i kursen har definierat det hyperboliska planet med dess avstånd och vinklar och linjer, kan vi på samma sätt definiera automorfierna av hyperboliska planet. I många sammanhang i matematiken talar man på ett liknande sätt om automorfier som bijektioner som "bevarar alla egenskaper". Som absolut minimum kan vi förvänta oss att dessa tre egenskaper ska gälla för en mängd av automorfier: Funktionen som gör ingenting kommer att vara en automorfi; om två funktioner är automorfier kommer även deras sammansättning att vara det, och den tredje egenskapen säger att en automorfi alltid kan göras ogjord.

Men detta är ingen kurs i gruppteori och vi kommer inte att diskutera grupper mer än detta. Vill man veta mer om ämnet kan man läsa i kompendiet till 2005 års Cirkel, som heter precis "Gruppteori".

4.3 Exempel

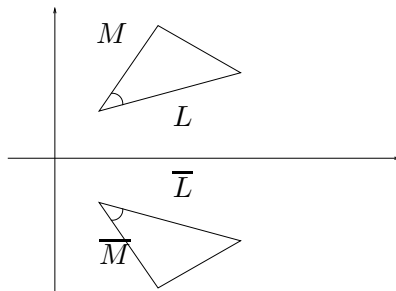
Exempel 4.3.1. Varje avbildning på formen

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

där $\alpha \neq 0$, är en Möbiusavbildning. Dessa avbildningar inkluderar exempelvis de euklidiska automorfierna av planet (som fås när $|\alpha| = 1$ enligt 3.5.1), men även exempelvis avbildningen $f(z) = rz$ där $r > 0$, som förstör hela planet med en faktor r . För att se att dessa ligger i Möbiusgruppen, låt oss välja $\alpha = a$, $\beta = b$, $c = 0$ och $d = 1$. Då är

$$\frac{az + b}{cz + d} = \frac{\alpha z + \beta}{1} = \alpha z + \beta = f(z),$$

som vi ville uppnå. Här är också $ad - bc = \alpha \neq 0$, och enligt 4.1.5 är nu f en Möbiusavbildning. $\blacktriangle$



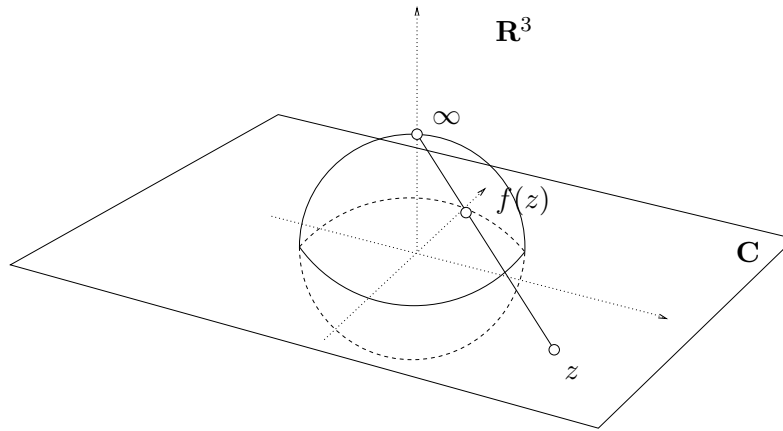
Figur 18: Den nedre triangeln är komplexkonjugatet av den övre. Från bilden ses att vinkeln från triangelnsidan L till M är *minus* vinkeln från $\overline{L}$ till $\overline{M}$, de motsvarande speglade sidorna.

Exempel 4.3.2. Ett icke-exempel: funktionen $f(z) = \overline{z}$, som har som effekt att spela hela planet i reella axeln, är inte en Möbiusavbildning. Det kan kännas intuitivt rimligt (efter några försök) att oavsett vilka värden på a, b, c och d vi väljer, kan vi inte få

$$\frac{az + b}{cz + d} = \overline{z}$$

att gälla för alla $z \neq -d/c$. I stället för att visa detta i detalj nu, kommer vi dock bara hänvisa till ett senare resultat som säger att *varje Möbiusavbildning bevarar vinklar*: det vill säga, om två kurvor möts i en vinkel α , kommer bilden av de två kurvorna under en Möbiusavbildning vara två nya kurvor som igen möts i vinkeln α . Men funktionen som komplexkonjugerar varje tal kommer inte att bevara vinklar utan byta tecken på dem, som i bilden nedan. Avbildningen $f(z) = \overline{z}$ kan alltså inte vara en Möbiusavbildning. Mer allmänt kan man på detta sätt se att en reflektion (spegling) av planet i *någon* linje aldrig kan vara en Möbiusavbildning. ▲

Exempel 4.3.3. Ett sista exempel är om vi tar $a = d = 0$, $b = c = i$. Då får vi exakt avbildningen $f(z) = 1/z$, som alltså är en Möbiustransformation. Vad är detta för funktion? Om vi skriver $z = re^{i\theta}$, kommer $1/z = (1/r)e^{-i\theta}$. (Kontrollera detta!) Så den här avbildningen kommer att göra två saker: den ändrar beloppet på varje nollskilt komplext tal genom avbildningen $r \mapsto 1/r$, som byter plats på insidan och utsidan av enhetscirkeln där $r = 1$. Punkter nära origo kommer att hamna långt ut mot oändligheten och vice versa. Dessutom byter argumentet på varje komplext tal tecken, vilket är samma sak som komplexkonjugering. Alltså kan $1/z$ tänkas på som en sammansättning av två funktioner: dels $1/\overline{z}$, som byter plats på insida och utsida av enhetscirkeln men bevarar argument, och dels $z \mapsto \overline{z}$, som speglar i reella axeln men bevarar beloppet. Notera att komplexkonjugering inte är en Möbiusavbildning enligt föregående exempel (och inte heller $1/\overline{z}$ är en Möbiusavbildning), men deras sammansättning ligger i Möbiusgruppen. ▲



Figur 19: Stereografisk projektion avbildar $\mathbf{C}$ till sfären minus nordpolen.

4.4 Riemannsfären

Låt oss nu återvända till ett problem vi nämnde i början av detta kapitel, och sedan dess har försökt skyla över, nämligen att de flesta Möbiusavbildningar (alla för vilka $c \neq 0$) är odefinierade i en punkt! Anledningen att detta problem uppstår är att man egentligen inte bör tänka på Möbiustransformationer som funktioner $\mathbf{C} \rightarrow \mathbf{C}$, utan som avbildningar från den så kallade *Riemannsfären* till sig själv!

Konstruktionen av Riemannsfären går till som följer. Identifiera först det komplexa planet $\mathbf{C}$ med xy -planet i tredimensionella rummet $\mathbf{R}^3$. Betrakta en sfär i $\mathbf{R}^3$ med radie 1 och centrum i origo. Låt oss beteckna denna sfär med $\hat{\mathbf{C}}$. Vi kallar "nordpolen" $(0, 0, 1)$ för *punkten i oändligheten*, och betecknar den med ∞ . Vi definierar nu en funktion $f : \mathbf{C} \rightarrow \hat{\mathbf{C}}$ som följer: för varje $z = x + iy \in \mathbf{C}$, drag den unika linjen i $\mathbf{R}^3$ genom $\infty = (0, 0, 1)$ och $z = (x, y, 0)$ i $\mathbf{R}^3$. Denna kommer att skära sfären $\hat{\mathbf{C}}$ i en unik punkt w , och vi definierar f genom att kräva att $f(z) = w$.

Funktionen f kommer avbilda varje punkt i $\mathbf{C}$ på en unik punkt i $\hat{\mathbf{C}}$, och alla punkter i $\hat{\mathbf{C}}$ utom ∞ kommer vara bilden av någon punkt i $\mathbf{C}$. Med andra ord: f kommer på ett bijektivt sätt att para ihop elementen av $\mathbf{C}$ med de i $\hat{\mathbf{C}} \setminus \{\infty\}$. Funktionen f brukar kallas för *stereografisk projektion*. Vi illustrerar f i Figur 19

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 $z \in \mathbf{C}$ gäller att ju större $|z|$ är, desto närmare ligger punkten till ∞ . Med andra ord: om

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 $\mathbf{C}$. 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 $\hat{\mathbf{C}}$. För 0 gäller ju att ju *mindre* $|z|$ är, desto närmare ligger z och 0, alltså precis motsatsen mot vad som gäller för ∞ .

Vi kommer nu att glömma konstruktionen av Riemannsfären som en delmängd av $\mathbf{R}^3$, utan i stället göra följande definition:

Definition 4.4.1. Riemannsfären $\hat{\mathbf{C}}$ är mängden $\mathbf{C} \cup \{\infty\}$. Ett element av $\hat{\mathbf{C}}$ skrivs som antingen ett komplext tal $x + iy$ eller ∞ .

Denna definition gör vi för att det är lättare att skriva en punkt på Riemannsfären som $x + iy$ än som ett komplicerat uttryck med tre koordinater. Dock kommer vi fortfarande geometriskt att tänka på $\hat{\mathbf{C}}$ som ytan av ett klot. Det är i detta sammanhang vi egentligen vill förstå Möbiusavbildningar:

Definition 4.4.2. Tag $a, b, c, d \in \mathbf{C}$ med $ad - bc = 1$. Om $c \neq 0$ definierar de en Möbiusavbildning $\hat{\mathbf{C}} \rightarrow \hat{\mathbf{C}}$ genom

$$f(z) = \begin{cases} \frac{az+b}{cz+d} & \text{om } z \neq \infty, z \neq -d/c \\ a/c & \text{om } z = \infty \\ \infty & \text{om } z = -d/c \end{cases}.$$

Om i stället $c = 0$ så får vi en Möbiusavbildning $\hat{\mathbf{C}} \rightarrow \hat{\mathbf{C}}$ genom

$$f(z) = \begin{cases} \frac{az+b}{d} & \text{om } z \neq \infty \\ \infty & \text{om } z = \infty \end{cases}.$$

Lägg märke till att om $c = 0$ kan inte samtidigt $d = 0$, eftersom vi antagit att $ad - bc = 1$.

Anmärkning 4.4.3. Vad som har hänt är att vi har modifierat den gamla definitionen av en Möbiusavbildning genom att kräva att så fort vi får något nollskilt delat med noll, ska vi låta resultatet vara lika med ∞ . För att tolka $f(\infty)$ gör vi i stället så här: vi skriver

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

Att z är "nära" ∞ på Riemannsfären betyder exakt samma sak som att $|z|$ är väldigt stort, så att b/z och d/z är försumbart små. Så det är logiskt att definiera $f(\infty) = \frac{a}{c}$. När $c = 0$ är $a \neq 0$ eftersom $ad - bc = 1$, så vi ska tolka a/c som ∞ , och punkten i oändligheten skickas till sig själv i detta fall.

Anmärkning 4.4.4. Genom att införa punkten i oändligheten behöver man (i princip) inte längre särbehandla de Möbiusavbildningar som har $c \neq 0$ respektive $c = 0$. Vi kan säga att *alla* Möbiusavbildningar är väldefinierade i *alla* punkter på Riemannsfären! Anledningen att vi inte ”märker av” punkten i oändligheten när $c = 0$ är att den avbildas till sig själv. Men så fort $c \neq 0$ kommer någon annan punkt i $\mathbf{C}$ att skickas ut i oändligheten, och oändligheten kommer att flyttas in i komplexa planet. Det är detta vi tyckte såg ut som att Möbiusavbildningen var odefinierad i en punkt.

Övning 4.1. 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 4.2. Visa direkt att mängden av funktioner $\mathbf{C} \rightarrow \mathbf{C}$ 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 4.3. Visa att varje Möbiusavbildning f alltid har minst en fixpunkt på Riemannsfären, det vill säga en punkt $z \in \widehat{\mathbf{C}}$ 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 $\mathbf{C}$.)

Övning 4.4. Ge ett exempel på en Möbiusavbildning med *exakt* en fixpunkt.

Övning 4.5. Låt $x + iy$ vara en punkt i komplexa planet. Till vilken punkt i $\mathbf{R}^3$ kommer z att avbildas under stereografisk projektion? (Varning: svaret är inte särskilt vackert.)

5 Vidare egenskaper hos Möb(C).

De vanliga euklidiska automorfierna bevarar enligt deras definition många egenskaper hos planet. De bevarar avstånd och vinklar; alla geometriska figurer vi kan föreställa oss, såsom cirklar, linjer och polygoner, blir "likadana" efter en automorfi. Det är ganska lätt att se att inte så många av dessa vanliga geometriska egenskaper bevaras av den större klassen av Möbiusavbildningar. Till exempel innehåller Möbiusgruppen funktionen som förstör planet med en faktor r , så avstånd bevaras inte. Möbiusgruppen innehåller även avbildningen $1/z$ som byter plats på in- och utsidan av enhetscirkeln och definitivt inte kan bevara räta linjer.

I detta kapitel kommer vi att diskutera vilka egenskaper som faktiskt bevaras av Möbiusavbildningar. Vi kommer att visa två stora resultat av denna typ: ett första är att mängden av cirklar och linjer *tillsammans* kommer att bevaras av varje Möbiusavbildning. När vi har visat detta kommer vi att definiera vad en konform (vinkelbevarande) avbildning är för något, och sedan visa att varje Möbiusavbildning är konform.

Det sista resultatet vi visar i detta kapitel har en lite annorlunda karaktär. Vi kommer att visa att för varje val av tre punkter på $\widehat{\mathbf{C}}$, 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.

5.1 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

$$\{z \in \mathbf{C} \mid |z - p| = r\},$$

där $p \in \mathbf{C}$ är cirkelns centrum och $r > 0$ är cirkelns radie. Efter avbildningen

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

får vi mängden

$$f(\Gamma) = \{z \in \mathbf{C} \mid |f^{-1}(z) - p| = r\},$$

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 $|w|^2 = w\overline{w}$, för att få

$$(z - p)\overline{(z - p)} = z\overline{z} - p\overline{z} - \overline{p}z + p\overline{p} = r^2.$$

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 5.1.1. *Varje ekvation på formen*

$$az\bar{z} + bz + \bar{b}z + c = 0,$$

där a och c är reella, b är komplext, och $b\bar{b} > ac$, beskriver en cirkel i $\mathbf{C}$ eller en linje i $\mathbf{C}$. 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 $z = x + iy \in \mathbf{C}$. 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 faktoreriserar fås

$$(z - p)(\overline{z - p}) = r^2,$$

vilket beskriver en cirkel. Fallet $a = 0$ lämnas åt läsaren som övning. $\square$

Lemma 5.1.2. Låt $h(z) = 1/z$, och Γ en cirkel eller linje i $\mathbf{C}$. 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

$$az\bar{z} + bz + \bar{b}\bar{z} + c = 0.$$

Då blir $h(\Gamma)$ en kurva i w -planet (i stället för z -planet), som ges av ekvationen

$$\frac{a}{w\bar{w}} + \frac{b}{w} + \frac{\bar{b}}{\bar{w}} + c = 0.$$

Multiplikera 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 5.1.1 är uppenbarligen fortfarande uppfyllt. $\square$

Sats 5.1.3. 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.

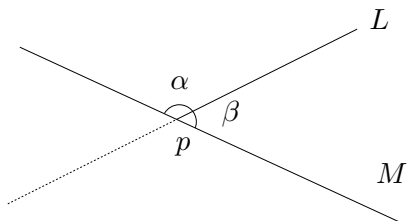
Så antag nu att $c \neq 0$. Inför nya funktioner

$$g(z) = \left(b - \frac{ad}{c}\right)z + \frac{a}{c}, \quad j(z) = cz + d,$$

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 5.1.2. Alltså har även deras sammansättning denna egenskap, så $f(\Gamma)$ måste vara en cirkel. $\square$



Figur 20: Två linjer som möts i en punkt. Här är α vinkeln från L till M , och β vinkeln från M till L .

5.2 Konforma avbildningar

Vi vill nu visa påståendet vi utlovat flera gånger tidigare: att en Möbiusavbildning bevarar vinklar. Låt oss först definiera vad detta betyder. En första komplikation är att vi måste definiera vinkeln mellan två linjer. I Definition 1.1.6 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.

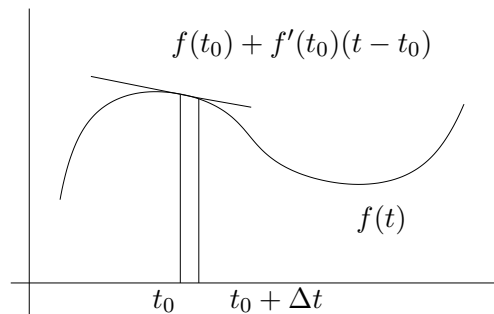
Definition 5.2.1. Låt två rätta linjer L och M skära varandra i en punkt p . Vi vill definiera vinkeln mellan linjerna L och M . Betrakta den ena halvoändliga linjen som går ut från p i L 's riktning. Från denna linje kan vi mäta upp två olika vinklar (α och β i Figur 20) mot linjen M . Med vår konvention 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 L till M* . Från figuren ses att vinkeln ej beror på vilken av de två halvoändliga linjerna ut från p i L 's riktning vi valde.

Definition 5.2.2. Låt två kurvor Γ och Δ mötas i en punkt $p \in \mathbf{C}$. Vi definierar *vinkeln från Γ till Δ* som vinkeln från tangenten till Γ till tangenten till Δ . Detta är väldefinierat eftersom vi vet vad det betyder att mäta vinkeln mellan två rätta linjer.

Anmärkning 5.2.3. 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 5.2.4. Låt $f : \mathbf{C} \rightarrow \mathbf{C}$ 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 $\Gamma \subset \mathbf{C}$ 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 $p \in \mathbf{C}$, så är vinkeln från Γ till Δ i p densamma som vinkeln från $f(\Gamma)$ till $f(\Delta)$ i $f(p)$.

Strikt talat är vi här lite slarviga, eftersom vi bara definierat konformitet för avbildningar från $\mathbf{C}$ till $\mathbf{C}$. Men påståendet vi vill visa är att en Möbiusavbildning



Figur 21: Tangentapproximationen för funktioner $f : \mathbf{R} \rightarrow \mathbf{R}$.

är konform, och en Möbiusavbildning bör egentligen ses som en avbildning $\widehat{\mathbf{C}} \rightarrow \widehat{\mathbf{C}}$. Vi kommer alltså endast att visa att en Möbiustransformation (tänkt på som en funktion $\mathbf{C} \rightarrow \mathbf{C}$ som tillåts vara odefinierad i en punkt) är konform i alla punkter där den är definierad, och inte det precisare påståendet. Det går dock att definiera konforma avbildningar på Riemannsfären med lite extra tekniska komplikationer, och när man har gjort detta visar det sig att en Möbiusavbildning är konform på hela $\widehat{\mathbf{C}}$.

Vi har redan i kapitel två talat om tangentlinjer till kurvor i planet, men vi har faktiskt aldrig definierat vad en tangentlinje till en kurva är för något.

Definition 5.2.5. Om Γ är en kurva i $\mathbf{C}$ 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 $\mathbf{R}$.

Anmärkning 5.2.6. 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 5.2.7. Ett heuristiskt argument för att Definition 5.2.5 är riktig är följande: det är en välbekant approximation att för en deriverbar funktion $f : \mathbf{R} \rightarrow \mathbf{R}$, 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 21. Speciellt gäller att ekvationen för tangenten fås genom att ersätta " $\cong$ " med " $=$ ". Ekvationen i 5.2.5 säger bara att vi kan göra exakt samma sak även när f är en funktion från $\mathbf{R}$ 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 tangentskurvans derivata är konstant lika med $\gamma'(t_0)$.

Anmärkning 5.2.8. En märklig sak vid första anblicken av 5.2.5 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 5.2.5 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 .

Lemma 5.2.9. *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 5.2.5 ser vi att $\gamma'(t_0)$ respektive $\delta'(t_0)$ ger tangenteriktningarna 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 5.2.10. *Låt f vara en Möbiustransformation. 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.) Alltså måste vi visa att $f'(z_0)$ är nollskilt. Så låt

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

och derivera med kvotregeln:

$$f'(z) = \frac{ad - bc}{(cz + d)^2} = \frac{1}{(cz + d)^2}.$$

Men denna funktion *är* nollskild överallt i $\mathbf{C} \setminus \{-d/c\}$. Beviset är klart! $\square$

Anmärkning 5.2.11. Vi har egentligen visat ett starkare påstående än att en Möbiusavbildning är konform – vi har visat att varje komplext deriverbar funktion (en så kallad *holomorf funktion*) är konform överallt där dess derivata är nollskild. Den enda egenskapen hos Möbiusavbildningarna som vi använde i beviset var just att deras derivata är nollskild överallt. Dock har vi inte definierat vad det betyder för en allmän funktion $\mathbf{C} \rightarrow \mathbf{C}$ att vara deriverbar (holomorf), och vi har tyvärr inte plats att göra det i denna text. Den intresserade läsaren hänvisas till [3]. Vi har dessutom inte ens försökt att definiera deriverbarhet för en funktion $\widehat{\mathbf{C}} \rightarrow \widehat{\mathbf{C}}$, även om vi egentligen vill tänka på Möbiustransformationer som funktioner på Riemannsfären. Med en korrekt definition av deriverbarhet på $\widehat{\mathbf{C}}$ kan man visa att en Möbiusavbildning har nollskild derivata överallt, även i oändligheten och i $-d/c$.

5.3 3-transitivitet

Följande egenskap hos gruppen av Möbiusavbildningar brukar kallas för att Möbiusgruppen är en *3-transitiv* grupp av funktioner på Riemannsfären.

Sats 5.3.1. *Låt p, q, r vara tre godtyckliga distinkta punkter på $\widehat{\mathbf{C}}$, 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 lemman. 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 5.3.2 visar att det existerar minst en sådan Möbiusavbildning, och resterande lemmen kommer att användas för att visa entydighet

Lemma 5.3.2. *Låt p, q, r vara tre godtyckliga distinkta punkter på $\widehat{\mathbf{C}}$. 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 om $f(p) = 0$, $f(r) = \infty$ 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 4.1.5 är vi klara.

Antag nu att någon av p, q och r är lika med ∞ . Välj först en Möbiusavbildning som avbildar rätt punkt av dessa 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 denna Möbiusavbildning 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 $\mathbf{C}$ 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. $\square$

Lemma 5.3.3. *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 $z \in \widehat{\mathbf{C}}$.*

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 4.2.1 kan föregående lemma skärpas till. Det kan vara värt att titta på den satsen igen innan man läser nästa lemma.

Lemma 5.3.4. *Låt f vara en Möbiusavbildning. Antag att $f(p) = p$, $f(q) = q$ och $f(r) = r$ gäller för tre olika punkter 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 5.3.2). Betrakta den sammansatta avbildningen

$$g^{-1} \circ f \circ g.$$

(Vi använder Sats 4.2.1 för att konstruera g^{-1}). Vi hävdar att denna måste vara identitetsavbildningen: ty

$$(g^{-1} \circ f \circ g)(0) = g^{-1}(f(g(0))) = g^{-1}(f(p)) = g^{-1}(p) = 0,$$

och samma argument visar att $g^{-1} \circ f \circ g$ fixerar 1 och ∞ . Här använder vi att då $g(p) = 0$, är $g^{-1}(0) = p$. Alltså är

$$g^{-1} \circ f \circ g = \text{id},$$

enligt föregående lemma. Sätt nu samman funktionerna på höger och vänster led med g från vänster, och g^{-1} från höger. Vi finner att

$$g \circ g^{-1} \circ f \circ g \circ g^{-1} = g \circ \text{id} \circ g^{-1}.$$

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. □

Vi kan nu bevisa 3-transitivitet.

Bevis. (av 5.3.1.) Låt punkterna p, q, r och p', q', r' vara givna, som i 5.3.1. 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 5.3.2. 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 avbildar p på sig själv, q på sig själv och r på sig själv. Vi är nu klara med hjälp av 5.3.4. □

Anmärkning 5.3.5. Egenskapen 5.3.1 hos Möbiusgruppen kan till exempel ge oss ett kort bevis för följande observation: *genom tre punkter p, q och r på $\widehat{\mathbf{C}}$ går en unik cirkel eller linje*. Så låt p, q och r vara tre godtyckliga punkter på $\widehat{\mathbf{C}}$. Om $p, q, r = 0, 1, 2$ är påståendet uppenbart — det finns en unik linje (den reella linjen $\mathbf{R}$) 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 $f^{-1}(\mathbf{R})$ att vara en cirkel eller linje enligt 5.1.3, 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 $f(\Gamma) = \mathbf{R}$.

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' .

5.4 Övningar

Övning 5.1. Hitta någon Möbiusavbildning som avbildar imaginära axeln på enhetscirkeln.

Övning 5.2. Visa direkt att det genom tre godtyckliga punkter p, q och r i komplexa planet passerar exakt en cirkel eller linje (påstående 5.3.5). 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 5.3. I beviset till 5.2.10 visades att om f är en deriverbar funktion $\mathbf{C} \rightarrow \mathbf{C}$, 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 5.4. Använd resultatet från 4.5 för att visa att skärningen mellan ett plan i $\mathbf{R}^3$ och $\widehat{\mathbf{C}}$ 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 5.5. Använd resultatet från föregående uppgift för att ge ett till bevis för att det genom tre punkter på $\widehat{\mathbf{C}}$ passerar en unik cirkel eller linje.

6 Delgruppen $\text{Möb}(\mathbf{H})$.

I de två föregående kapitlen har vi lovat visa att automorfgruppen till det hyperboliska planet — som vi fortfarande inte har definierat — kommer att vara en delmängd av $\text{Möb}(\mathbf{C})$, men vi har inte specificerat vilken delmängd av $\text{Möb}(\mathbf{C})$ vi talar om. Under resten av kursen kommer dessa oklarheter förhoppningsvis att redas ut.

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 automorfgruppen 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.

6.1 Möbiusavbildningar som bevarar övre halvplanet

Den första skillnaden mellan det hyperboliska planet och det euklidiska är att mängden av punkter inte är densamma. I det euklidiska planet består punkterna av elementen i $\mathbf{R}^2$ (vi kan också, som bekant, tänka på $\mathbf{C}$); i det hyperboliska planet är punkterna exakt de som ligger i *övre halvplanet*:

$$\mathbf{H} = \{z \in \mathbf{C} \mid \text{Im}(z) > 0\}.$$

Vi har inte ännu definierat avstånd och vinklar i det hyperboliska planet, men så fort vi har gjort detta kan vi definiera vilka funktioner som kommer att vara automorfer till det hyperboliska planet genom att imitera definitionen i 1.2. Vi kommer dock att göra saker i ”fel” ordning. I detta kapitel kommer vi i förväg att avslöja vilka automorfierna till hyperboliska planet är, och i nästa kapitel kan vi efter att ha inför avstånd och vinklar kontrollera att dessa bevaras av vår klass av funktioner. Ett första villkor på en automorfi, innan vi ens kan kräva något om att bevara vinklar eller avstånd, är att automorfin skall vara en bijektion från mängden av punkter i geometrin till sig själv.

Vi vill alltså att $f: \mathbf{H} \rightarrow \mathbf{H}$ ska vara en bijektion. Vi kommer att säga att f *bevarar* övre halvplanet, eller att f avbildar övre halvplanet till sig självt, om detta gäller. Det kommer att visa sig att automorfgruppen till hyperboliska planet är exakt de Möbiusavbildningar som bevarar övre halvplanet. Vi kommer nu att undersöka vilka Möbiusavbildningar som har denna egenskap.

Lemma 6.1.1. Låt $f \in \text{Möb}(\mathbf{C})$. Följande två påståenden är ekvivalenta:

1. Möbiusavbildningen f bevarar linjen $\mathbf{R} \cup \{\infty\}$.
2. 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 $\mathbf{R} \cup \{\infty\}$. 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 5.3.2. 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

$$\frac{az + b}{cz + d} = w \in \mathbf{R},$$

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 $\mathbf{R} \cup \{\infty\}$.) $\square$

Anmärkning 6.1.2. 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 4.1.4.

Varför är egenskapen att bevara $\mathbf{R} \cup \{\infty\}$ viktig? Tänk på hur $\mathbf{H}$ ser ut på Riemannsfären, $\widehat{\mathbf{C}}$. Om vi tänker på ∞ som nordpol och 0 som sydpol, så kommer linjen $\mathbf{R} \cup \{\infty\}$ 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, $\mathbf{R} \cup \{\infty\}$ är *randen* till $\mathbf{H}$ om vi betraktar situationen på $\widehat{\mathbf{C}}$. Om en Möbiusavbildning f avbildar $\mathbf{H}$ på $\mathbf{H}$ måste den även avbilda randen till $\mathbf{H}$ till sig själv, så ett nödvändigt villkor på en Möbiusavbildning för att kunna bevara $\mathbf{H}$ är att den *kan* skrivas med reella koefficienter. Dock är det inte nödvändigtvis så att en funktion som avbildar $\mathbf{R} \cup \{\infty\}$ 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:

Lemma 6.1.3. *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 $z \in \mathbf{H}$. Då gäller att om $ad - bc > 0$, så kommer $\operatorname{Im}(f(z)) > 0$ och alltså $f(z) \in \mathbf{H}$. Om däremot $ad - bc < 0$ så kommer $\operatorname{Im}(f(z)) < 0$, så $f(z) \notin \mathbf{H}$.

Bevis. Tag $z = x + iy \in \mathbf{H}$. 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 i nedre halvplanet. $\square$

Definition 6.1.4. Delmängden $\operatorname{Möb}(\mathbf{H}) \subset \operatorname{Möb}(\mathbf{C})$ 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 $ad - bc = 1$.

Sats 6.1.5. *Låt $f \in \operatorname{Möb}(\mathbf{C})$. Då gäller att $f \in \operatorname{Möb}(\mathbf{H})$ om och endast om f avbildar övre halvplanet $\mathbf{H}$ 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, $\mathbf{R} \cup \{\infty\}$, på sig självt. Alltså kan vi enligt Lemma 6.1.1 skriva

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

där a, b, c och d är reella, och enligt Lemma 6.1.3 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 $\mathbf{R} \cup \{\infty\}$ att avbildas till sig själv (enligt 6.1.1), 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 $z \in \mathbf{H}$ kommer att avbildas in i övre halvplanet av f , så vi är i den första situationen. $\square$

Sats 6.1.6. *Delmängden $\text{Möb}(\mathbf{H})$ uppfyller, precis som $\text{Möb}(\mathbf{C})$, alla tre egenskaperna från Sats 4.2.1. (Man säger ibland att $\text{Möb}(\mathbf{H})$ är en delgrupp till Möbiusgruppen.)*

Bevis. Beviset är nästan identiskt med beviset för Sats 4.2.1. 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.) $\square$

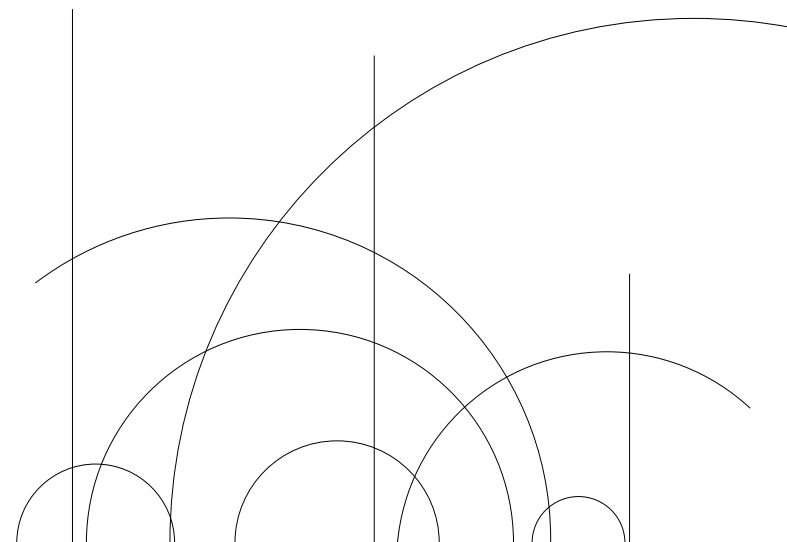
6.2 Hyperboliska linjer

Vi kommer nu att definiera vilka kurvor i $\mathbf{H}$ 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).

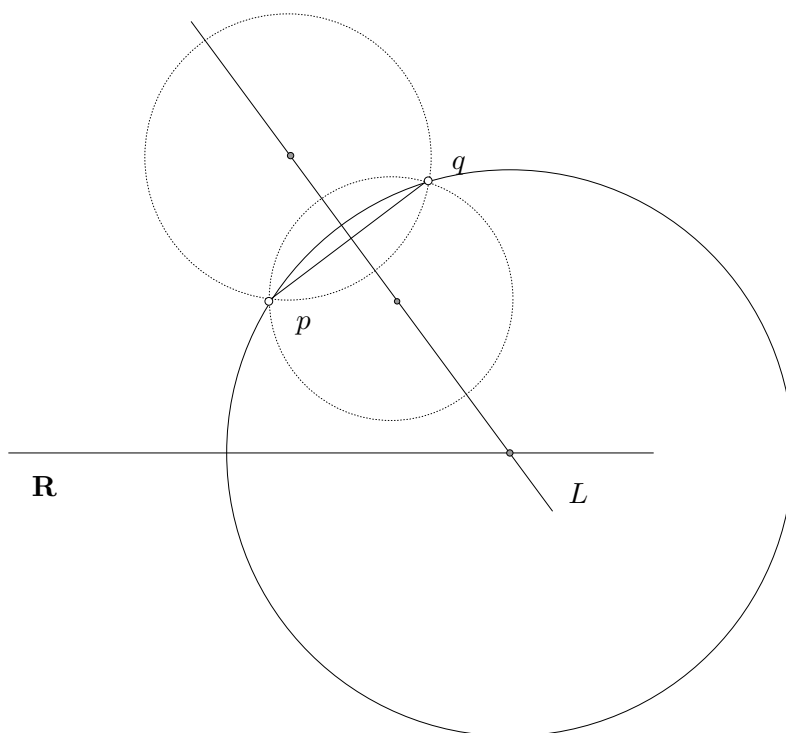
Definition 6.2.1. En delmängd $L \subset \mathbf{H}$ sägs vara en *hyperbolisk linje* om det existerar någon linje eller cirkel Γ i det komplexa planet som skär reella axeln i en rät vinkel, sådan att $L = \mathbf{H} \cap \Gamma$.

Sats 6.2.2. *Genom två distinkta punkter $p, q \in \mathbf{H}$ passerar en unik hyperbolisk linje.*

Bevis. Betrakta först alla cirklar som passerar genom p och q . Att en cirkel passerar genom 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 L vara den räta linje som består av alla punkter på samma avstånd från p och q , se Figur 23. Linjen L är alltså normalen till det räta linjesegmentet från p till q som skär denna i segmentets mittpunkt. För varje punkt c på L kan vi rita en unik cirkel med c som centrum och som passerar genom p och q .



Figur 22: Reella axeln $\mathbf{R}$, och en samling hyperboliska linjer som möter den i en rät vinkel.



Figur 23: En illustration av beviset till 6.2.2.

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 $\mathbf{R}$.

Vi måste nu dela upp i två fall:

1. Punkterna p och q ligger inte ovanför varandra. I detta fall kommer L inte att kunna vara parallell med reella linjen, så L skär reella axeln i en unik punkt. Om vi ritar cirkeln med denna punkt som centrum och som passerar genom p och q , kommer denna att vara vår unika hyperbolisk linje.
2. Punkterna p och q ligger rakt ovanför varandra. I detta fall kommer L 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.

□

Sats 6.2.3. *Alla $f \in \text{Möb}(\mathbf{H})$ 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 5.1.3), 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 6.1.1) och vi vet dessutom att varje Möbiusavbildning är konform (Sats 5.2.10) och därmed bevarar vinkeln i vilken kurvor skär varandra. Beviset är klart. □

Sats 6.2.4. *Låt L och M vara två hyperboliska linjer. Då existerar $f \in \text{Möb}(\mathbf{H})$ som avbildar L på M .*

Bevis. Vi visar först att för varje hyperbolisk linje L så finns en $f \in \text{Möb}(\mathbf{H})$ som avbildar L på imaginära axeln. Om L är en rät linje 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 $\text{Möb}(\mathbf{H})$. Så låt L 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 L att vara en hyperbolisk linje som passerar genom 0 och ∞ . Men en hyperbolisk linje som passerar genom oändligheten måste vara en vanlig rät linje (kontrollera detta!), och om den dessutom passerar origo måste den vara den imaginära axeln. Det följer att f kommer att avbilda L på imaginära axeln.

Betrakta nu det allmänna fallet med två hyperboliska linjer L och M . Tag $f, g \in \text{Möb}(\mathbf{H})$ sådana att f avbildar L på imaginära axeln och g avbildar M på imaginära axeln. Då kommer

$$g^{-1} \circ f$$

vara en avbildning i $\text{Möb}(\mathbf{H})$ som avbildar L på M . □

Anmärkning 6.2.5. 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 $f \in \text{Möb}(\mathbf{H})$ 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.

6.3 Övningar

Övning 6.1. Visa att två olika hyperboliska linjer inte kan mötas i mer än en punkt.

Övning 6.2. Hitta ekvationen för den hyperboliska linjen genom $p = -1 + 4i$ och $q = 6 + 3i$.

Övning 6.3. Antag att $f \in \text{Möb}(\mathbf{H})$ har egenskapen att f avbildar imaginära axeln till sig själv. Hur kan f se ut?

Övning 6.4. I euklidiska planet känner vi till att givet en linje L och en punkt $p \in L$, finns en unik linje genom p som möter L i en rät vinkel. Visa att detta även gäller för hyperboliska linjer i $\mathbf{H}$.

7 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

$$f(z) = \frac{az + b}{cz + d} \quad (7.1)$$

med

$$a, b, c, d \in \mathbf{R} \quad \text{och} \quad ad - bc = 1$$

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 $f \in \text{Möb}(\mathbf{H})$. 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 $\text{Möb}(\mathbf{H})$, och att vi kan använda det för att införa ett hyperboliskt avstånds-begrepp. 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.

7.1 Hyperbolisk längd

Låt Γ vara bilden av parameterfunktionen $\gamma: [a, b] \rightarrow \mathbf{H}$ 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 $f \in \text{Möb}(\mathbf{H})$ så gäller tyvärr i allmänhet

$$l(\Gamma) \neq l(f(\Gamma)).$$

Exempel 7.1.1. 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 (7.1) 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 $\text{Möb}(\mathbf{H})$ 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 $\mathbf{H}$, fast med rotationer och translationer utbytta mot Möbiusavbildningarna på formen (7.1).

Vi är nu redo att definiera begreppet hyperbolisk längd.

Definition 7.1.2. Låt $\gamma: [a, b] \rightarrow \mathbf{H}$ vara en parameterfunktion med bild Γ . Den *hyperboliska längden* av kurvan Γ i övre halvplanet ges av

$$l_{\text{hyp}}(\Gamma) = \int_a^b \frac{1}{\text{Im}|\gamma(t)|} |\gamma'(t)| dt. \quad (7.2)$$

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 $\mathbf{R}$ vi rör oss i $\mathbf{H}$, desto mindre blir faktorn. Vi ser också att omskalningen är densamma i alla punkter som ligger på en horisontell linje i $\mathbf{H}$, det vill säga, har samma imaginärdel.

Vi ger ett exempel på hur räkningar med hyperbolisk längd kan se ut.

Exempel 7.1.3. Betrakta det vertikala linjesegment K som förbinder punkterna $p = ia$ och $q = ib$ för $a, b \in \mathbf{R}$ 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 $l(K) = b - a$. $\blacktriangle$

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.

7.2 Invariants under Möb(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 7.2.1. *Låt $f \in \text{Möb}(\mathbf{H})$. Då gäller $l_{\text{hyp}}(\Gamma) = l_{\text{hyp}}(f(\Gamma))$ för varje parameterkurva Γ .*

Bevis. Vi inleder beviset med några räkningar. Låt $f \in \text{Möb}(\mathbf{H})$, 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

$$f'(z) = \frac{a(cz + d) - c(az + b)}{(cz + d)^2} = \frac{ad - bc}{(cz + d)^2} = \frac{1}{(cz + d)^2}. \quad (7.3)$$

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

$$\text{Im}(f(z)) = \text{Im} \left(\frac{ac|z|^2 + bd + adz + bc\bar{z}}{|cz + d|^2} \right).$$

Eftersom $ac|z|^2 + bd$ är reellt får vi

$$\operatorname{Im}(f(z)) = \frac{1}{|cz + d|^2} \operatorname{Im}(adz + bc\bar{z}) = \frac{(ad - bc)y}{|cz + d|^2}.$$

Vi har ju $ad - bc = 1$, och vi ser att

$$\operatorname{Im}(f(z)) = \frac{y}{|cz + d|^2}.$$

Vi jämför med vårt tidigare uttryck (7.3) och får att

$$|f'(z)| = \frac{\operatorname{Im}(f(z))}{\operatorname{Im}(z)}. \quad (7.4)$$

Vi kommer nu till huvuddelen av beviset. Låt Γ vara en kurva som parametriseras av $\gamma : [a, b] \rightarrow \mathbf{C}$. 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

$$\int_a^b |\gamma'(t)| \frac{|f'(\gamma(t))|}{\operatorname{Im}[f(\gamma(t))]} dt. \quad (7.5)$$

Från (7.4) följer det att

$$\frac{|f'(\gamma(t))|}{\operatorname{Im}[f(\gamma(t))]} = \frac{1}{\operatorname{Im}(\gamma(t))}.$$

Vi sätter in detta i integralen (7.5) 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 $f \in \operatorname{Möb}(\mathbf{H})$, och detta avslutar beviset. $\square$

7.3 Hyperboliskt avstånd och hyperboliska geodeter

Vi inför nu ett avståndsbegrepp i vår hyperboliska geometri. Vi vill säga att avståndet mellan p och q i $\mathbf{H}$ 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 7.3.1. *Låt p och q vara två punkter i det övre halvplanet $\mathbf{H}$ och låt K beteckna det unika hyperboliska linjesegment som förbinder p och q . Då gäller*

$$l_{\text{hyp}}(K) \leq l_{\text{hyp}}(\Gamma) \quad (7.6)$$

för varje parameterkurva Γ som förbinder p och q .

Bevis. Vi delar upp beviset i två fall: dels $\text{Re}(p) = \text{Re}(q)$ och dels $\text{Re}(p) \neq \text{Re}(q)$. Vi börjar med att behandla det första fallet $\text{Re}(p) = \text{Re}(q)$. Automorfin $f(z) = z - \text{Re}(p)$ 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 w . 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 att tillsammans med det faktum att $\text{Im}(\gamma(t)) = \text{Im}(\tilde{\gamma}(t))$ att

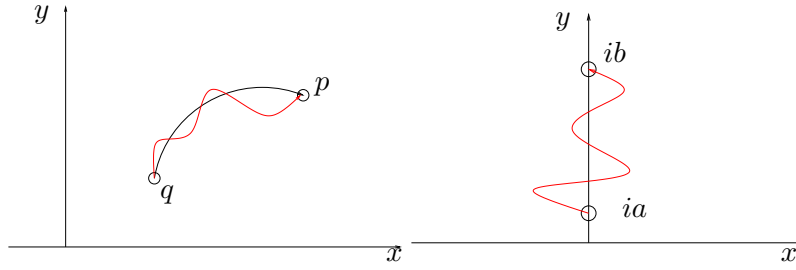
$$l_{\text{hyp}}(\Gamma) = \int_{\alpha}^{\beta} \frac{|\gamma'(t)|}{\text{Im}(\gamma(t))} dt \geq \int_{\alpha}^{\beta} \frac{|y'(t)|}{\text{Im}(\tilde{\gamma}(t))} dt = l_{\text{hyp}}(K), \quad (7.7)$$

där K är den räta linjen som förbinder p och q .

Antag nu att $\text{Re}(p) \neq \text{Re}(q)$. Vi kan för enkelhetens skull anta att $\text{Re}(p) > \text{Re}(q)$. Vi använder Sats 6.2.4 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 7.2.1 kommer den avbildade kurvan ha samma hyperboliska längd som den ursprungliga kurvan Γ . Vårt tidigare resonemang i (7.7) medför att

$$l_{\text{hyp}}(f(\Gamma)) \geq l_{\text{hyp}}(K),$$



Figur 24: Höger: Kurvan Γ och den hyperboliska linjen mellan p och q . Vänster: Bilden av Γ under f .

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 6.2.3 säger emellertid att automorfierna bevarar av $\mathbf{H}$ de hyperboliska linjerna, och således måste $\Gamma' = f^{-1}(K)$ vara en hyperbolisk linje. Eftersom vi antagit att $\text{Re}(p) \neq \text{Re}(q)$ är Γ' en cirkel som möter $\mathbf{R}$ i en rät vinkel. Detta avslutar beviset. $\square$

Vi använder denna sats för att definiera det hyperboliska avståndet. Om p och q tillhör $\mathbf{H}$ och $\text{Re}(p) = \text{Re}(q)$ säger vi att det hyperboliska avståndet mellan p 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 7.1.3. 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 7.3.2. Låt p och q vara två distinkta punkter i övre halvplanet $\mathbf{H}$. Då ges *hyperboliska avståndet* mellan p och q av

$$\text{dist}_{\text{hyp}}(p, q) = l_{\text{hyp}}(L), \quad (7.8)$$

där L 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 $\mathbf{H}$. 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 7.3.3. Låt två linjesegment L och M mötas i en punkt $p \in \mathbf{H}$. Vi definierar vinkeln från L till M som den vanliga euklidiska vinkeln, där vi tänker på L och M som kurvor i det vanliga euklidiska planet $\mathbf{R}^2$.

För att denna definition ska kunna vara vettig, måste vi veta att en automorfi $f \in \text{Möb}(\mathbf{H})$ inte förändrar vinkeln mellan två kurvor. Men detta har vi redan visat för två kapitel sedan! Sats 5.2.10 visar att varje Möbiusavbildning bevarar den vanliga euklidiska vinkeln mellan kurvor. Vinkelbegreppet från Definition 7.3.3 är alltså invariant under den automorfigrupp vi har sagt att det hyperboliska planet skall ha.

7.4 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 7.4.1. Två hyperboliska linjer säges vara *parallella* om de är disjunkta, det vill säga, om de inte har några punkter gemensamma.

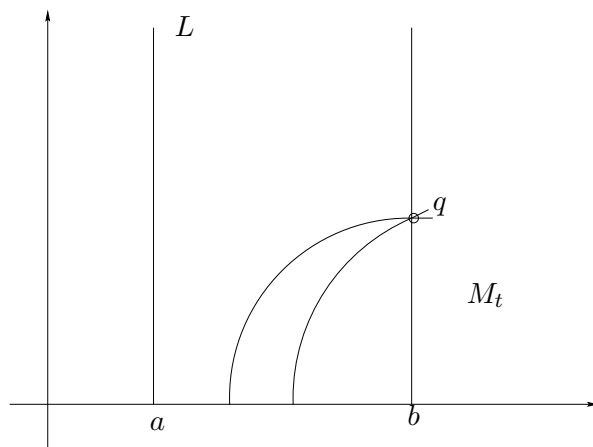
Denna definition är helt analog med den i kapitel 1: vi sade att två linjer i planet var parallella om de inte möttes.

Vi kommer nu till en av höjdpunkterna i detta kompendium: vi ska visa att det för varje punkt finns *oändligt många* hyperboliska linjer i $\mathbf{H}$ som är parallella med en given linje L !

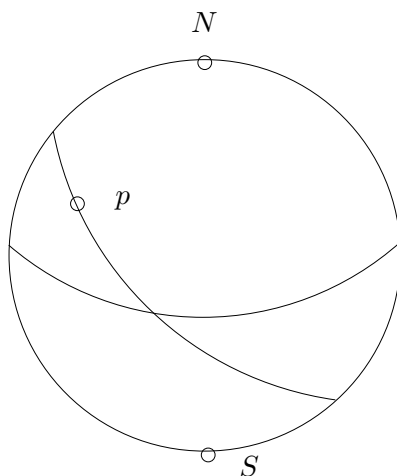
Sats 7.4.2. Låt L vara en hyperbolisk linje, och låt q vara en punkt som inte tillhör L . Då existerar oändligt många hyperboliska linjer som innehåller q och är parallella med L .

Bevis. Vi kan utan inskränkning anta att L är en rät linje som skär reella axeln vinkelrätt och har $\text{Re}(p) = a > 0$ för alla $p \in L$. Om så inte är fallet kan vi tillämpa en Möbiusavbildning för att reducera situationen till detta fall, se Sats 6.2.4.

Låt $q \in \mathbf{H}$ ha $\text{Re}(q) = b$ och låt oss antaga att $b > a$. För varje $t \in (a, b]$ finns enligt Sats 6.2.2 det en unik hyperbolisk linje M_t som passerar genom t och även q . När $t = b$ är M_t en rät linje som är vinkelrät mot $\mathbf{R}$. För varje annat $t \in (a, b]$ är M_t en cirkelbåge med centrum på reella linjen, vilket vi såg i beviset för Sats 6.2.2.



Figur 25: Flera parallella linjer.

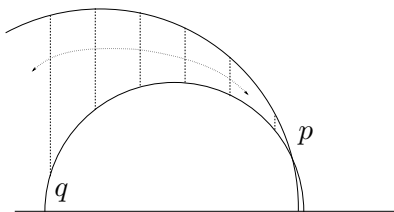


Figur 26: Varje linje genom p skär ekvatorn.

Eftersom vi inte tillåtit $t \neq a$ ser vi att L och M_t inte har några punkter gemensamma, och därmed är varje hyperbolisk linje M_t parallell med L . $\square$

7.5 Sfärisk geometri

Vi tillåter oss en liten utblick mot sfärisk geometri. På en sfär är *storcirklarna* de kurvor som minimerar den motsvarande sfärisk 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—allra



Figur 27: En familj av hyperboliska trianglar.

longitudlinjerna. Det är alltså inte fullt så enkelt som att titta på geometri på en sfär att hitta ett motexempel mot parallellpostulatet.

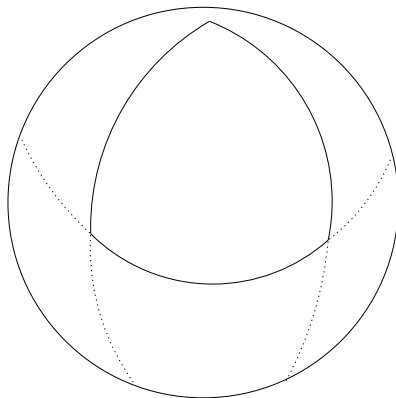
7.6 Areor, vinklar och Gauss-Bonnets sats

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 27.

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 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 ett exempel på:

Sats 7.6.1. *Arean 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 kongruenta om alla deras vinklar är desamma, och i så fall kommer de två trianglarna endast skilja sig åt upp till skalning. Men i det hyperboliska fallet kommer två kongruenta 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 triangels area strikt begränsad av proportionalitetskonstanten gånger π .



Figur 28: En triangel på en sfär, där alla tre vinklarna är räta.

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 $\mathbf{H}$, och det är det enda sätt att göra detta så att arean inte förändras av en automorfi i $\text{Möb}(\mathbf{H})$. 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 7.6.1 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 storcirkel. 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 7.6.2. *Arean 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 kongruenta 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:

Sats 7.6.3. *(Gauss-Bonnets sats) För varje geometri — euklidisk, hyperbolisk, sfärisk — existerar en konstant K med egenskapen att för alla trianglar T gäller att*

$$K \cdot \text{area}(T) = \pi - \alpha - \beta - \gamma,$$

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.

7.7 Övningar

Övning 7.1. Låt $\gamma: [a, b] \rightarrow \mathbf{H}$ vara en parameterfunktion med bild Γ och låt $h: [\alpha, \beta] \rightarrow [a, b]$ vara en omparametrisering, se Definition 2.3.1.

Visa att bilden Γ' av $\gamma \circ h: [\alpha, \beta] \rightarrow \mathbf{H}$ har samma hyperboliska båglängd som Γ .

Övning 7.2. Visa att varje hyperbolisk linje har oändlig hyperbolisk längd.

Övning 7.3. Beräkna den hyperboliska båglängden av kurvan $\Gamma \subset \mathbf{H}$ som bestäms av parameterfunktionen

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

Övning 7.4. Rita fyrhörningen i hyperboliska planet som begränsas av de hyperboliska linjerna vars ekvation i komplexa planet ges av $\operatorname{Re}(z) = 1$, $\operatorname{Re}(z) = -1$, $|z| = \sqrt{2}$ och $|z| = 2$. Räkna ut dess vinkelsumma och jämför med vinkelsumman av en fyrhörning i euklidiska planet.

Lösningar till udda övningsuppgifter

Övning 0.1.

1. $B \cup C = A$.
2. $B \cap C = \emptyset$.
3. $D \cap C = \{4, 36\}$.
4. $\{x \in D \mid x \in B\} = D \cap B = \{1, 19, 101\}$.
5. $\{x \in A \mid x = y + 1 \text{ för något } y \in D\} = \{2, 5, 20, 37, 102\}$.
6. $\{x + 1 \mid x \in D\} = \{2, 5, 20, 37, 102\}$.

Övning 0.3. Tag $x \in ((A \cap C) \cup (B \cap C^c))^c$. Det betyder att $x \in \Omega$ och $x \notin (A \cap C) \cup (B \cap C^c)$. Alltså har vi att $x \notin A \cap C$ och $x \notin B \cap C^c$. Det finns nu två möjligheter: $x \in C$ och $x \notin C$.

I det första fallet, det vill säga $x \in C$, måste $x \notin A$ eftersom om $x \in A$ så skulle $x \in A \cap C$ vilket är falskt. Alltså gäller $x \in A^c$, vilket tillsammans med $x \in C$ ger att $x \in A^c \cap C$ i detta fall. I synnerhet har vi att $x \in (A^c \cap C) \cup (B^c \cap C^c)$.

I det andra fallet gäller $x \in C^c$ och då måste $x \in B^c$ eftersom om $x \in B$ så skulle $x \in B \cap C^c$ vilket är falskt. Alltså gäller $x \in B^c \cap C^c$, och i synnerhet $x \in (A^c \cap C) \cup (B^c \cap C^c)$.

I båda fallen gäller alltså $x \in (A^c \cap C) \cup (B^c \cap C^c)$, och eftersom x var godtycklig så visar detta att $((A \cap C) \cup (B \cap C^c))^c \subseteq (A^c \cap C) \cup (B^c \cap C^c)$.

Omvänt, tag $x \in (A^c \cap C) \cup (B^c \cap C^c)$. Då gäller $x \in A^c \cap C$ eller $x \in B^c \cap C^c$ (eller båda). Vi har alltså dessa två fall.

I det första fallet, det vill säga $x \in A^c \cap C$, har vi att $x \notin A$ och $x \in C$. I synnerhet har vi att $x \notin A \cap C$ (eftersom $x \notin A$) och att $x \notin B \cap C^c$ (eftersom $x \in C$). Alltså tillhör x varken $A \cap C$ eller $B \cap C^c$, vilket betyder att $x \in ((A \cap C) \cup (B \cap C^c))^c$.

I det andra fallet, det vill säga $x \in B^c \cap C^c$ har vi att $x \notin B$ och att $x \notin C$. Det följer att $x \notin A \cap C$ och att $x \notin B \cap C^c$. Alltså gäller $x \notin (A \cap C) \cup (B \cap C^c)$, vilket betyder att $x \in ((A \cap C) \cup (B \cap C^c))^c$.

I båda fallen har det alltså visats att $x \in ((A \cap C) \cup (B \cap C^c))^c$, och eftersom x var godtycklig visar detta att $(A^c \cap C) \cup (B^c \cap C^c) \subseteq ((A \cap C) \cup (B \cap C^c))^c$.

Vi har alltså visat att $((A \cap C) \cup (B \cap C^c))^c \subseteq (A^c \cap C) \cup (B^c \cap C^c)$ och att $(A^c \cap C) \cup (B^c \cap C^c) \subseteq ((A \cap C) \cup (B \cap C^c))^c$ och därmed att $((A \cap C) \cup (B \cap C^c))^c = (A^c \cap C) \cup (B^c \cap C^c)$.

Övning 0.5. Antag att f har en invers. Vi skall visa två saker: att om $f(x) = f(y)$, så är $x = y$, och vi skall visa att för alla $y \in X$ existerar x sådan att $f(x) = y$. Låt oss börja med det första. Antag att

$$f(x) = f(y).$$

Låt funktionen f^{-1} verka på både höger- och vänsterled ovan. Vi finner att

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

Men eftersom $f^{-1} \circ f = \text{id}$, är $f^{-1}(f(x)) = x$ och $f^{-1}(f(y)) = y$. Alltså är $x = y$.

Låt oss nu ta ett $y \in X$; vi vill visa att y är i bilden av f . Låt $x = f^{-1}(y)$. Om vi låter funktionen f verka på höger- och vänsterled på

$$x = f^{-1}(y),$$

finner vi att

$$f(x) = f(f^{-1}(y)),$$

och som tidigare vet vi att då $f \circ f^{-1} = \text{id}$ är $f(f^{-1}(y)) = y$. Alltså är $f(x) = y$, så y är i bilden av f .

Övning 1.1. Vi vet att f kan skrivas som en rotation följt av en translation; låt oss börja med att bestämma rotationsdelen. Vi ser att linjen från $(1, 0)$ till $(2, 0)$ skall avbildas på linjen som pekar från $(2, 5)$ till $(2, 6)$. Ritar man dessa punkter ser man att f kommer att rotera planet en vinkel $\pi/2$. Låt g vara avbildningen som roterar planet $\pi/2$ kring origo: enligt 1.2.3 kan vi skriva

$$g(x, y) = (-y, x).$$

Vi ser att $g(1, 0) = (0, 1)$ och $g(2, 0) = (0, 2)$, så vi behöver dessutom translatera med en term $(2, 4)$ för att få $(1, 0)$ och $(2, 0)$ till rätt punkt. Så om $h(x, y) = (x + 2, y + 4)$ är $f = h \circ g$, vilket ger att

$$f(x, y) = (2 - y, 4 + x).$$

I det här fallet kan vi lätt bestämma f^{-1} . Om $u = 2 - y$ och $v = 4 + x$, skall $f^{-1}(u, v) = (x, y)$. Löser vi ut x och y ser vi att $y = 2 - u$ och $x = v - 4$, så $f^{-1}(u, v) = (v - 4, 2 - u)$.

Övning 1.3. Låt f vara en automorfi. Enligt Sats 1.2.4 kan vi skriva $f^{-1} = g \circ h$ där h är en rotation kring origo och g är en translation. Notera nu att h^{-1} och g^{-1} igen är rotationer kring origo respektive translationer.

Vi ser också att $f = h^{-1} \circ g^{-1}$. Detta säger bara att om f^{-1} ges av att utföra först h , sedan g , så kommer f (inversen till f^{-1}) att ges av att först göra g ogjord, och sedan göra h ogjord.

Men eftersom f var godtycklig, och vi har en uppdelning $f = h^{-1} \circ g^{-1}$ där h^{-1} är en rotation kring origo och g^{-1} är en translation, har vi alltså visat att f kan skrivas som först en translation och sedan en rotation.

Övning 1.5. Vi använder oss av resultatet i Övning 0.5. Låt $\mathcal{L}$ vara mängden av linjer i planet, och $f : \mathbf{R}^2 \rightarrow \mathbf{R}^2$ en rotation. Vi vet att f är en bijektion, med en invers $f^{-1} : \mathbf{R}^2 \rightarrow \mathbf{R}^2$ som roterar *medurs* med samma

vinkel som f . I texten efter Exempel 1.2.3 visas att om L är en linje i planet, är även $f(L)$ en linje i planet. Vi får alltså även en funktion

$$f : \mathcal{L} \rightarrow \mathcal{L},$$

genom att avbilda L på den nya linjen $f(L)$. Vi hävdar att denna funktion har en invers, så att f även är en bijektion när vi tänker på f som en funktion $\mathcal{L} \rightarrow \mathcal{L}$. Inversen är precis den omvända rotationen f^{-1} : om L är en linje så kommer

$$f^{-1}(f(L)) = f(f^{-1}(L)) = L.$$

För att se detta, notera att $f(L)$ enligt definitionen är mängden

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

så att $f^{-1}(f(L))$ är mängden

$$\{f^{-1}(f(x)) \mid x \in L\} = \{x \mid x \in L\} = L,$$

eftersom $f^{-1}(f(x)) = x$. Vi har alltså visat att de två observationerna att f är en bijektion $\mathbf{R}^2 \rightarrow \mathbf{R}^2$, och att f och dess invers avbildar linjer på linjer, räcker för att visa att f även avbildar linjer bijectivt.

Övning 2.1. Vi ska sätta in uttrycken som definierar $x(t)$ och $y(t)$ i den givna ekvationen (2.9).

Vi fokuserar först på termerna $(x_2 - x_1)y(t) - (y_2 - y_1)x(t)$. Vi har

$$\begin{aligned} (x_2 - x_1)y(t) - (y_2 - y_1)x(t) &= (x_2 - x_1)[(1 - t)y_1 + ty_2] - (y_2 - y_1)[(1 - t)x_1 + tx_2] \\ &= (x_2 - x_1)(1 - t)y_1 + t(x_2 - x_1)y_2 + t(x_2 - x_1)y_2 \\ &\quad - (y_2 - y_1)x_1 + t(y_2 - y_1)x_1 - t(y_2 - y_1)x_2 \\ &= (x_2 - x_1)y_1 - (y_2 - y_1)x_1 \\ &\quad - tx_2y_1 + tx_1y_1 + tx_2y_2 - tx_1y_2 \\ &\quad + tx_1y_2 - tx_1y_1 - tx_2y_2 + tx_2y_1 \\ &= (x_2 - x_1)y_1 - (y_2 - y_1)x_1. \end{aligned}$$

Summan som kvarstår i högerledet uppträder, fast med motsatt tecken, i (2.9). Vi ser nu att $\gamma(t)$ verkligen satisfierar ekvationen för alla t .

Övning 2.3. 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

$$\int_0^{\frac{\pi}{2}} \sqrt{(x'(t))^2 + (y'(t))^2} dt.$$

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 3.1.

- 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

$$\frac{z}{w} = \frac{(3 + i)(-1 - 4i)}{(-1 + 4i)(1 - 4i)} = \frac{-3 + 4 + i(-12 - 1)}{17} = \frac{1}{17} - \frac{13}{17}i.$$

- 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. Den komplexa exponentialfunktionen är definierad genom

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

Vi deriverar real-och imaginärdelarna var för sig med avseende på det reella talet θ och får

$$(\cos \theta)' = -\sin \theta \quad \text{och} \quad (\sin \theta)' = \cos \theta.$$

Vi har således

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

Notera att $i^2 = i \cdot i = -1$, så genom att bryta ut i får vi

$$(e^{i\theta})' = i[i \sin \theta + \cos \theta] = ie^{i\theta},$$

vilket skulle visas.

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 4.3. Låt

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

Om $c = 0$ är ∞ en fixpunkt, så vi kan anta att $c \neq 0$. Enligt definitionen kommer $z \in \mathbf{C}$ att vara en fixpunkt precis då

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

vilket vi genom att multiplicera med $cz + d$ kan arrangera om till

$$cz^2 + (d - a)z - b = 0.$$

Eftersom $c \neq 0$ vet vi att detta polynom är icke-konstant, och har därför minst en rot. Denna rot är en fixpunkt till f !

Övning 4.5. Låt oss först dra den rätta linjen mellan $x + iy$ och ∞ . Vi hävdar att denna kan parametriseras som

$$(x(1 - t), y(1 - t), t),$$

där t varierar över $\mathbf{R}$. För att se detta, notera att ekvationen uppenbarligen beskriver en linje, och att den passerar genom både $x + iy$ och ∞ (sätt in $t = 0$ respektive $t = 1$). Vi vill nu hitta den punkt på linjen som också ligger på sfären i $\mathbf{R}^3$ med radie 1 och mittpunkt i origo. Med andra ord, vi vill hitta punkter på linjen vars avstånd från origo är 1. Avståndet mellan en punkt i $\mathbf{R}^3$ och origo ges av Pythagoras sats:

$$\text{dist}(0, p) = \sqrt{a^2 + b^2 + c^2} \quad \text{om } p = (a, b, c).$$

Alltså ska vi lösa ekvationen

$$\sqrt{x^2(1 - t)^2 + y^2(1 - t)^2 + t^2} = 1,$$

eller

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

Högerledet kan faktoriseras som $(1 + t)(1 - t)$. Vi ser alltså att ekvationen har en trivial lösning för $t = 1$, men denna lösning ger bara skärningspunkten $\infty = (0, 0, 1)$. Det är den andra skärningen vi är intresserade av, så vi delar bort faktorn $(1 - t)$. Vi finner att

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

eller

$$t = \frac{x^2 + y^2 - 1}{x^2 + y^2 + 1}.$$

Det följer att skärningen $f(x + iy)$ fås genom att sätta in detta värde på t i linjens ekvation. Man beräknar att

$$\begin{aligned} f(x + iy) &= \left(x \left(1 - \frac{x^2 + y^2 - 1}{x^2 + y^2 + 1} \right), y \left(1 - \frac{x^2 + y^2 - 1}{x^2 + y^2 + 1} \right), \frac{x^2 + y^2 - 1}{x^2 + y^2 + 1} \right) \\ &= (\rho 2x, \rho 2y, \rho(x^2 + y^2 - 1)) \quad \text{där} \quad \rho = (x^2 + y^2 + 1)^{-1}. \end{aligned}$$

Som en kontroll kan man t.ex. se att enhetscirkeln i xy -planet avbildas till sig själv under stereografisk projektion, och att tredje koordinaten blir positiv respektive negativ precis då $x+iy$ ligger utanför respektive innanför enhetscirkeln, vilket vi kan se är rimligt från bilden av stereografisk projektion (Figur 19).

Övning 5.1. Eftersom det passerar en unik cirkel eller linje genom tre givna punkter (5.3.5), 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 5.3.2, 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 4.2.1 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 5.3. 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 5.5. Låt p, q och r vara tre punkter på $\widehat{\mathbf{C}}$. 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 $\widehat{\mathbf{C}}$ och ett plan i $\mathbf{R}^3$, 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 $\mathbf{R}^3$.

Men helt allmänt gäller i $\mathbf{R}^3$ 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 $\mathbf{R}^3$. 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 $\mathbf{R}^3$, och vi är klara.

Övning 6.1. Antag att två hyperboliska linjer L och M möts i två punkter p och q . Vi vet att det passerar en unik hyperbolisk linje genom p och q , och eftersom både L och M har denna egenskap kommer dessa nödvändigtvis att vara lika.

Övning 6.3. Vi vet att då f är en Möbiusavbildning, bevarar den cirklar och linjer på $\widehat{\mathbf{C}}$. Vi vet dessutom att då $f \in \text{Möb}(\mathbf{H})$ kommer den att bevara $\mathbf{R} \cup \{\infty\}$. Den linje på $\widehat{\mathbf{C}}$ som svarar mot imaginära axeln möter $\mathbf{R} \cup \{\infty\}$ 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 5.3.3. 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 $\text{Möb}(\mathbf{H})$ 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 $\text{Möb}(\mathbf{H})$ 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 7.1. Enligt definitionen av hyperbolisk båglängd har vi $l_{\text{hyp}}(\Gamma) = \int_a^b |f'(t)|/\text{Im}[f(t)] dt$.

Låt oss för enkelhetens skull anta att $g'(t) \geq 0$. Det andra fallet leder till helt analoga räkningar. Längden av den omparametriserade kurvan ges av

$$l_{\text{hyp}}(\Gamma') = \int_{\alpha}^{\beta} \frac{|(f \circ g)'(t)|}{\text{Im}[(f \circ g)(t)]} dt. \quad (7.9)$$

Kedjeregeln medför att $(f \circ g)'(t) = g'(t)f'(g(t))$. Insättning av detta i (7.9) ger

$$\int_{\alpha}^{\beta} \frac{f'(g(t))}{\text{Im}[f(g(t))]} |g'(t)| dt.$$

Vi tillämpar satsen om variabelbyten A.2.2 med $s = g(t)$. Då fås $ds = |g'(t)|dt = g'(t)dt$ och de nya gränserna $a = g(\alpha)$ och $b = g(\beta)$, och vi får

$$\int_{\alpha}^{\beta} \frac{f'(g(t))}{\text{Im}[f(g(t))]} |g'(t)| dt = \int_a^b \frac{|f'(s)|}{\text{Im}(f(s))} ds,$$

vilket sammanfaller med hyperboliska båglängden av den ursprungliga kurvan Γ .

Övning 7.3. Vi har att beräkna integralen

$$l_{\text{hyp}}(\Gamma) = \int_0^1 \frac{|f'(t)|}{\text{Im}(f(t))} dt.$$

Vi har $f'(t) = 1 + i$ och därmed $|f'(t)| = \sqrt{2}$. Vidare har vi $\text{Im}(f(t)) = 1 + t$. Detta ger oss

$$l_{\text{hyp}}(\Gamma) = \int_0^1 \frac{\sqrt{2}}{1+t} dt = \sqrt{2} [\ln(1+t)]_0^1 = \sqrt{2} (\ln(1+1) - \ln(1+0)) = \sqrt{2} \ln 2.$$

Den hyperboliska längden av Γ är således $l_{\text{hyp}}(\Gamma) = \sqrt{2} \ln 2$.

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. En funktion $f: \Omega \subset \mathbf{R} \rightarrow \mathbf{R}$ säges vara *kontinuerlig i en punkt* $x \in \mathbf{R}$ om det för varje $\epsilon > 0$ existerar ett $\delta > 0$ sådant att

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

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 $x \in \mathbf{R}$ 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 $a_n, a_{n-1}, \dots, a_0$, är kontinuerliga. ▲

Exempel A.1.3. Funktionen $f: \mathbf{R} \rightarrow \mathbf{R}$ 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 $f: [a, b] \rightarrow \mathbf{R}$ i en punkt $x \in (a, b)$ definieras som gränsvärdet

$$f'(x) = \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h} \tag{A.1}$$

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 derivera sammansatta funktioner och påminner om detta i en sats.

Sats A.1.4. Låt $f: [\alpha, \beta] \rightarrow \mathbf{R}$ 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 $f: [a, b] \rightarrow \mathbf{R}$ 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 funktionsgrafén för f finns minst en punkt $(\xi, f(\xi))$ sådan att tangenten i denna punkt är parallell med den rätta 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 $f: [a, b] \rightarrow \mathbf{R}$ 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 rektangelarna 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

$$x_0 = a < x_1 < \cdots < x_{n-1} < x_n = b.$$

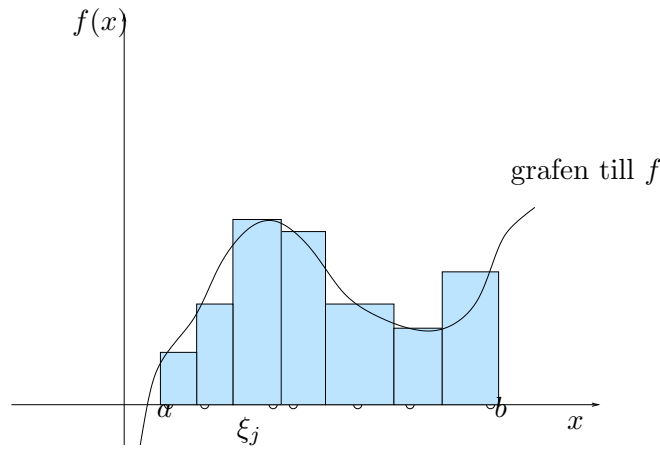
Vi låter $\ell(\pi)$ vara längden på det största av intervallen $[x_{j-1}, x_j]$ och använder detta som ett mått på indelningens finhet.

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.



Figur 29: En Riemannsumma.

Sats A.2.1. Låt $f: [a, b] \rightarrow \mathbf{R}$ vara kontinuerlig. Då är f integrerbar över $[a, b]$, och Riemannsummorna

$$\sum_{k=0}^{n-1} f(\xi_j)(x_{j+1} - x_j)$$

konvergerar mot integralen

$$\int_a^b f(x)dx$$

vid obegränsat förfinad indelning.

Nästa sats är teoretiskt intressant, 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 $f: [\alpha, \beta] \rightarrow \mathbf{R}$ 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:

$$2 \int_1^{\sqrt{2}} \cos y dy = 2 [\sin y]_1^{\sqrt{2}} = 2(\sin 2 - \sin 1).$$

▲

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 $f: [a, b] \rightarrow \mathbf{R}$ vara kontinuerlig och bilda funktionen

$$F(x) = \int_a^x f(t) dt, \quad x \in [a, b].$$

Då är F deriverbar med $F'(x) = f(x)$.

Förslag till vidare läsning

- [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.

Böckerna ovan, och många andra böcker, finns att låna på Matematikbiblioteket, Lindstedtsvägen 25 (bottenvåningen). Biblioteket är öppet för alla.