ANALYSEDRYPP

Et forkurs til MAT2400

av

Tom Lindstrøm

Matematisk insitutt og CMA
Universitetet i Oslo
2011

Forord

Hensikten med dette heftet er å bygge en bro mellom kalkuluskursene og det første kurset i matematiske analyse — ved Universitetet i Oslo vil det si mellom kursene MAT1100 og MAT1110 på den ene siden og MAT2400 på den andre. Egentlig burde det være unødvendig med en slik bro; temamessig er det slett ikke noe gap mellom disse kursene, tvert imot blir en del av teorien fra MAT1100 og MAT1110 repetert i nesten samme form i MAT2400. Den store forskjellen er at dette stoffet nå må tas på alvor av alle, mens det i MAT1100/1110 nok ble oppfattet som "krydder" for de aller ivrigste og flinkeste. Denne forskjellen kommer tydelig frem i eksamensoppgavene: Der man tidligere har kunnet redde seg på regneoppgavene, må man i MAT2400 få til noen av de teoretiske oppgavene for å oppnå en anstendig karakter. Trøsten er at teorien ikke er så forferdelig vanskelig hvis man bare arbeider systematisk med den fra starten av.

Heftet består av fem kapitler. Det første inneholder en del bakgrunnsmateriale om bevis, mengder og funksjoner — temaer som ofte blir skjøvet litt i bakgrunnen i de første matematikkemnene, men som er viktige i det første analysekurset. De tre midterste kapitlene repeterer en del sentrale resultater og teknikker fra kalkuluskursene: ϵ-δ-bevis, kompletthet samt sentrale teoremer som skjæringssetningen, ekstremalverdisetningen, middelverdisetningen og Bolzano-Weierstrass' teorem. Det siste kapitlet gir en forsmak på teorien for metriske rom, kanskje det mest sentrale temaet i MAT2400. Poenget er å vise hvordan man kan generalisere begreper og resultater fra $\mathbb{R}^n$ til ethvert rom der man har et fornuftig avstandsbegrep.

Heftet er ikke ment som en lærebok, til det er det for knapt skrevet og forutsetter at studentene har kjennskap til de sentrale begrepene og teknikkene på forhånd. Knappheten skyldes ikke bare dovenskap, men også en tanke om at det er lettere å få oversikt over strukturen og ideene dersom man holder seg til det vesentligste.

Blindern, 5. januar 2011

Tom Lindstrøm

Kapittel 1

Bevis, mengder og funksjoner

I det første kapitlet skal vi ta en kikk på noen temaer det er nyttig å vite litt om før man begynner på et skikkelig analysekurs. Mengder og funksjoner kommer til å opptre i mer abstrakte former enn tidligere, og du vil bli nødt til å gjennomføre fullstendige bevis for påstandene dine.

1.1 Litt om bevis

Matematisk teori er bygget opp av bevis. Bevisene har minst to formål: I tillegg til å sikre at resultatene våre faktisk *er* sanne, skal de fortelle oss *hvorfor* de er sanne. Det siste aspektet er kanskje det viktigste når du lærer en ny matematisk teori — det er nesten umulig å få oversikt og sammenheng dersom du ikke skjønner hvordan resultatene bygger på hverandre. En matematisk teori består av hundrevis av resultater, men bare noen få ideer og teknikker, og behersker du disse få ideene og teknikkene, får de mange hundre resultatene en sammenheng og en struktur som gjør det mulig å huske og forstå dem. I tillegg er bevisene i lærebøkene eksempler på resonnementstyper du selv må kunne gjennomføre for å løse oppgaver.

Det er ikke noe hokus-pokus med matematiske bevis, de er bare sunn fornuft satt i system. Et bevis er en argumentkjede som gir deg en logisk uangripelig bro fra noe du allerede vet, til det du skal bevise. Formålet med argumentene er litt annerledes enn i diskusjoner og debatter — her gjelder det ikke med alle midler å overbevise andre om at du har rett, her gjelder det å finne ut hva som faktisk er riktig. Er det en svakhet i argumentasjonen din, lurer du først og fremst deg selv.

Selv om matematiske bevis ikke er noe hokus-pokus, men bare vanlig, sunn fornuft, er det noen få ting det er lurt å være klar over. I matematikk er det mange påstander av typen "Hvis A, så B". Det betyr bare at hver gang A inntreffer, så inntreffer også B. Et typisk eksempel er: "Hvis $f : \mathbb{R} \to \mathbb{R}$ er deriverbar, så er f kontinuerlige". Dette betyr bare at hvis du treffer på en funksjon, og den viser seg å være deriverbar, så vil den alltid være kontinuerlig. Med symboler skriver vi $A \Longrightarrow B$ for "Hvis A, så B", og vi sier også "A impliserer B" eller "A medfører B".

Vær oppmerksom på at $A \Longrightarrow B$ og $B \Longrightarrow A$ betyr helt forskjellige ting: I det første tilfellet sier vi at hver gang A inntreffer, så inntreffer også B; i det andre tilfellet sier vi at hver gang B inntreffer, så inntreffer også A. Det er åpenbart forskjell på å si "Hvis $n = 17$, så er n et primtall" og "Hvis n er et primtall, så er $n = 17$" — det første er riktig, det andre er galt siden det finnes andre primtall enn 17.

Hvis vi både har $A \Longrightarrow B$ og $B \Longrightarrow A$, så sier vi at A og B er *ekvivalente* og skriver $A \Longleftrightarrow B$. Dette betyr at de to påstandene A og B er sanne i nøyaktig de samme tilfellene. Med ord sier vi gjerne "A hvis og bare hvis B". Et eksempel er:

"I en trekant er alle vinklene like store hvis og bare hvis alle sidene er like lange."

Dette betyr at desom du tar en titt på alle verdens trekanter, vil de der alle vinklene er like store, være nøyaktig de samme som de der alle sidene er like lange.

Når man skal bevise $A \Longleftrightarrow B$-utsagn, lønner det seg ofte å vise $A \Longrightarrow B$ og $B \Longrightarrow A$ hver for seg.

Hvis du tenker deg om litt, vil du innse at $A \Longrightarrow B$ betyr akkurat det samme som ikke-$B \Longrightarrow$ ikke-A. Det betyr at istedenfor å bevise $A \Longrightarrow B$, kan vi bevise ikke-$B \Longrightarrow$ ikke-A. I noen tilfeller lønner dette seg fordi hypotesen "ikke-B" gir oss mer å arbeide med enn hypotesen A. Når vi beviser $A \Longrightarrow B$ ved isteden å argumentere for ikke-$B \Longrightarrow$ ikke-A, utfører vi et *kontrapositivt bevis*. Her er et enkelt eksempel på et kontrapositivt bevis:

Setning 1.1.1 *Hvis n^2 er et partall, så er n et partall.*

Bevis: Det kontrapositive utsagnet er

"Hvis n er et oddetall, så er n^2 et oddetall."

For å bevise dette, bruker vi at hvis n er et oddetall, så er $n = 2k + 1$ for et helt tall k. Dermed er $n^2 = (2k + 1)^2 = 4k^2 + 4k + 1 = 2(2k^2 + 2k) + 1$ som åpenbart er et oddetall. □

I beviset ovenfor er det naturlig å argumentere kontrapositivt siden "n er et oddetall" er en enklere antagelse å arbeide med enn "n^2 er et partall".

Det er en bevisform til som det kanskje er lurt å peke på, såkalte *motsigelsesbevis*. I et motsigelsesbevis antar du det omvendte av det du skal bevise, og viser at dette leder til noe som er åpenbart galt. Dermed må det være det du skulle bevise, som er riktig. Her er et berømt eksempel på et motsigelsesbevis:

Setning 1.1.2 *$\sqrt{2}$ er et irrasjonalt tall.*

Bevis: Anta for motsigelse at $\sqrt{2}$ er rasjonal. Da kan $\sqrt{2}$ skrives som en brøk

$$\sqrt{2} = \frac{m}{n}$$

der m og n er naturlige tall uten felles faktorer. Kvadrerer vi denne ligningen og ganger med n^2 på begge sider, får vi

$$2n^2 = m^2$$

Dette betyr at m^2 er et partall, og ifølge setning 1.1.1 er da m et partall, dvs. $m = 2k$ for et naturlig tall k. Setter vi dette inn i den siste ligningen ovenfor, ser vi $2n^2 = 4k^2$, dvs. $n^2 = 2k^2$. Etter setning 1.1.1 betyr dette at også n er et partall, men det er umulig siden m og n ikke har felles faktorer. Dermed har vi vist at antagelsen om at $\sqrt{2}$ er rasjonal, fører til en selvmotsigelse, og følgelig må $\sqrt{2}$ være irrasjonal, $\square$

1.2 Mengder

Mengde er det grunnleggende begrepet i matematikken — alle andre begreper kan bygges opp fra mengdebegrepet. Så ambisiøse skal ikke vi være, for oss skal mengder bare være et praktisk arbeidsredskap for å holde styr på ganske enkle fenomener. Vi skal derfor ikke forsøke å gi en dypsindig definisjon av hva en mengde er, bare si at det er en samling matematiske objekter. Mengdene du støter på i MAT2400 vil som regel være samlinger av tall, punkter, vektorer eller funksjoner.

En mengde kan godt være endelig, som mengden $S = \{1, 2, 3, 4, 5\}$ bestående av tallene 1, 2, 3, 4 og 5, eller uendelig som mengden $B = [0, 1]$, bestående av alle reelle tall x slik at $0 \leq x \leq 1$. For å si at et objekt x hører med i mengden A, skriver vi med symboler $x \in A$, og for å si at objektet y *ikke* er med i A, skriver vi $y \notin A$. De objektene som er med i en mengde A, kaller vi *elementene* til A. En litt spesiell mengde er *den tomme mengden* $\emptyset$, mengden uten elementer. Det kan virke som en selvmotsigelse å kalle dette en mengde, men den tomme mengden er så nyttig at det ville være en stor ulempe å ekskludere den på grunn av filologisk finfølelse.

To mengder A og B er like dersom de har nøyaktig de samme elementene, og i så fall skriver vi (ikke overraskende!) $A = B$. Dersom alle elementene i A er med i B (men ikke nødvendigvis omvendt), sier vi at A er en *delmengde* av B og skriver $A \subset B$. Legg merke til at hvis både $A \subset B$ og $B \subset A$, så er

$A = B$. Når vi skal bevise at to tilsynelatende forskjellige mengder faktisk er like, lønner det seg ofte å bevise $A \subset B$ og $B \subset A$ hver for seg. Vi skal se eksempler på dette etter hvert.

Det er to måter å skrive mengder på som ofte er nyttige. Den ene er *listeform*

$$A = \{1, 3, 7, 11, 13\}$$

$$B = \{1, 3, 5, 7, 9, 11, \ldots\}$$

Den andre er ved hjelp av en såkalt *mengdebygger*;

$$A = \{a \in B \,|\, a \text{ tilfredsstiller betingelsen } C\}$$

Her er A definert som alle elementene i B som tilfredsstiller en betingelse C. Vi kan for eksempel definere mengden av positive tall ved

$$A = \{a \in \mathbb{R} \,|\, a > 0\}$$

Husk at en del tallmengder har etablerte navn

$$\mathbb{N} = \{1, 2, 3, 4 \ldots\}, \text{ mengden av naturlige tall}$$

$$\mathbb{Z} = \{\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots\}, \text{ mengden av hele tall}$$

$$\mathbb{Q} = \left\{\frac{m}{n} \,|\, m \in \mathbb{Z}, n \in \mathbb{N}\right\}, \text{ mengden av rasjonale tall}$$

$$\mathbb{R}, \text{ mengden av relle tall}$$

$$\mathbb{C}, \text{ mengden av komplekse tall}$$

De to viktigste operasjonene for mengder er *snitt* $\cap$ og *union* $\cup$. Dersom vi har mengder $A_1, A_2 \ldots, A_n$, er snittet definert ved

$$A_1 \cap A_2 \cap \ldots \cap A_n = \{x \,|\, x \text{ er med i } \textit{alle} \text{ mengdene } A_1, A_2, \ldots, A_n\}$$

og unionen ved

$$A_1 \cup A_2 \cup \ldots \cup A_n = \{x \,|\, x \text{ er med i } \textit{minst én av} \text{ mengdene } A_1, A_2, \ldots, A_n\}$$

Legg merke til at hvis A og B ikke har felles elementer, så er $A \cap B = \emptyset$. I så fall sier vi at A og B er *disjunkte*.

Man kan "gange" mengder inn og ut av paranteser omtrent som man kan med tall, dvs. man har de *distributive lovene*

$$B \cup (A_1 \cap A_2 \cap \ldots \cap A_n) = (B \cup A_1) \cap (B \cup A_2) \cap \ldots \cap (B \cup A_n) \quad (1.2.1)$$

og

$$B \cap (A_1 \cup A_2 \cup \ldots \cup A_n) = (B \cap A_1) \cup (B \cap A_2) \cup \ldots \cup (B \cap A_n) \quad (1.2.2)$$

La oss bevise (1.2.1) som et eksempel på hvordan man beviser likheter mellom mengder. Vi skal først vise at dersom x er med i mengden til venstre, så må den også være med i mengden til høyre, og deretter at hvis x er med i mengden til høyre, så må den også være med i mengden til venstre. Dette vil vise at mengdene er like.

Anta $x \in B \cup (A_1 \cap A_2 \cap \ldots \cap A_n)$, da er x med i enten B eller $A_1 \cap A_2 \cap \ldots \cap A_n$. Dette betyr at x enten er med i B eller i alle $A_1, A_2, \ldots, A_n$. I begge tilfeller er $x \in B \cup A_1, x \in B \cup A_2, \ldots, x \in B \cup A_n$, dvs.

$$x \in (B \cup A_1) \cap (B \cup A_2) \cap \ldots \cap (B \cup A_n)$$

Anta nå omvendt at $x \in (B \cup A_1) \cap (B \cup A_2) \cap \ldots \cap (B \cup A_n)$. Da er x med i $B \cup A_i$ for alle i. Det betyr at enten er $x \in B$, eller så må $x \in A_i$ for alle i, dvs. $x \in A_1 \cap A_2 \cap \ldots \cap A_n$. Uansett hvilket av disse alternativne som gjelder, er $x \in B \cup (A_1 \cap A_2 \cap \ldots \cap A_n)$, og dermed er (1.2.1) bevist.

Det finnes andre mengdeoperasjoner enn snitt og union, f.eks. *differansen*

$$A \setminus B = \{x \in A \,|\, x \notin B\}$$

Den er ofte nyttig når vi skal unnta punkter, slik som i

$$A = \mathbb{R} \setminus \{-1, 3\}$$

der A består av alle reelle tall unntatt -1 og 3.

Som regel er alle mengdene vi er interessert i, delmengder av en større mengde U — et "univers" som de andre mengdene lever innenfor. Jobber vi med tall på tallinjen, er det ofte naturlig å tenke på $\mathbb{R}$ som universet; jobber vi med punkter i planet, er det naturlig å tenke på $\mathbb{R}^2$ som universet; jobber vi med et vektorrom V, er det naturlig å tenke på dette som universet osv. Hvis vi er blitt enig om et univers U, er det nyttig å definere *komplementet* til en delmengde A av U ved

$$A^c = \{x \in U \,|\, x \notin A\}$$

To nyttige formler binder sammen komplementet med snitt og union. De kalles *De Morgans lover*:

$$(A_1 \cap A_2 \cap \ldots \cap A_n)^c = A_1^c \cup A_2^c \cup \ldots \cup A_n^c \qquad (1.2.3)$$

$$(A_1 \cup A_2 \cup \ldots \cup A_n)^c = A_1^c \cap A_2^c \cap \ldots \cap A_n^c \qquad (1.2.4)$$

Den enkleste måten å huske disse reglene på er å si at dersom vi skal ta komplementet til en union eller et snitt, kan vi flytte komplementtegnet inn på hver enkelt mengde, men da må vi gjøre om snittene til unioner og omvendt.

La oss bevise (1.2.3) som enda et eksempel på hvordan vi viser likhet mellom mengder. Vi må vise at det er nøyaktig de samme elementene som med i begge mengder. At $x \in (A_1 \cap A_2 \cap \ldots \cap A_n)^c$, er det samme som at $x \notin A_1 \cap A_2 \cap \ldots \cap A_n$. Det er det samme som at det finnes minst én i slik at $x \notin A_i$, og det er det samme som at det finnes minst én i slik at $x \in A_i^c$, som igjen er det samme som at $x \in A_1^c \cup A_2^c \cup \ldots \cup A_n^c$. Dermed er (1.2.3) bevist.

1.3 Funksjoner

Du er nok mest vant til funksjoner mellom tallmengder, men funksjoner kan vi faktisk ha mellom alle mengder. Hvis A og B er to mengder, er en *funksjon fra A til B* bare en regel som til hvert element i A tilordner et element $f(x)$ i B. Vi skriver $f : A \to B$ for å markere at f er en funksjon fra A til B. Funksjoner mellom generelle mengder er egentlig ikke vanskeligere enn de funksjonene du er mest vant til, men er du blant dem som først og fremst forbinder funksjoner med funksjonsgrafer, må du nok arbeide litt for å få et mer abstrakt bilde — funksjoner mellom generelle mengder har ikke grafer i tradisjonell forstand.

Anta at C er en delmengde av A. *Bildet av C* under f er da mengden

$$f(C) = \{f(x) \,|\, x \in C\}$$

dvs. mengden av alle verdier $f(x)$ når x ligger i C. Bildet $f(A)$ av hele A kalles ofte *bildet til f*.

Hvis vi isteden starter med en mengde $D \subset B$, kan vi danne det *inverse bildet*

$$f^{-1}(D) = \{x \in A \,|\, f(x) \in D\}$$

Det inverse bildet $f^{-1}(D)$ består altså av alle de elementene i A som havner i D når vi bruker f på dem. Bilder og inverse bilder spiller en sentral rolle i all videregående matematikk.

En funksjon $f : A \to B$ kaller *surjektiv* dersom $f(A) = B$, dvs. hvis alle elementene i B treffes av minst ett element i A. Vi kaller f *injektiv* dersom $x_1 \neq x_2$ medfører $f(x_1) \neq f(x_2)$, dvs. dersom f sender ulike elementer i A på ulike elementer i B. En funksjon som både er injektiv og surjektiv, kalles *bijektiv.* For en bijektiv funksjon vil det for hver $y \in B$ finnes nøyaktig én $x \in A$ slik at $y = f(x)$. En bijektiv funksjon har derfor en *omvendt funksjon* $g : B \to A$ gitt ved

$$g(y) = x \text{ der } x \text{ er elementet i } A \text{ slik at } y = f(x)$$

Den omvendte funksjonen til f betegnes ofte med f^{-1}. Legg merke til at dersom f er injektiv, men ikke surjektiv, så kan vi gjøre f surjektiv ved å tenke på den som en funksjon $f : A \to f(A)$. Den får da en invers funksjon $f^{-1} : f(A) \to A$.

La oss avslutte denne seksjonen med et kort blikk på sammensatte funksjoner. Dersom vi har to funksjoner $f : A \to B$, $g : B \to C$, er den sammensatte funksjonen $h : A \to C$ gitt ved $h(x) = g(f(x))$ (du kjenner slike funksjoner fra kjerneregelen). Ofte skriver man $g \circ f$ for den sammensatte funksjonen; altså $(g \circ f)(x) = g(f(x))$.

1.4 Tellbarhet

La oss helt til slutt i dette kapitlet se på et litt mer avansert begrep fra mengdelæren. En mengde A er *tellbar* dersom det går an å lage en liste $a_1, a_2, a_3, \dots$ som inneholder alle elementene i A (samme element må gjerne komme flere ganger). En endelig mengde $A = \{a_1, a_2, \dots, a_n\}$ er åpenbart tellbar siden vi f.eks. kan liste den på denne måten

$$a_1, a_2, \dots, a_n, a_n, a_n \dots$$

Mengden $\mathbb{N}$ av naturlige tall er også tellbar siden den er ferdig listet:

$$1, 2, 3, 4, \dots$$

Vi kan også liste mengden $\mathbb{Z}$ av alle hele tall:

$$0, 1, -1, 2, -2, 3, -3, \dots$$

Det neste resultatet viser at det finnes ganske mange tellbare mengder. Dersom A og B er mengder, skriver vi $A \times B$ for mengden av alle par (a, b) der $a \in A$, $b \in B$.

Setning 1.4.1 *Dersom A og B er tellbare, så er $A \times B$ det også.*

Bevis: Vi vet at vi kan liste alle elementene i A:

$$a_1, a_2, a_3, a_4, \dots$$

og B:

$$b_1, b_2, b_3, b_4, \dots$$

Vi kan nå liste alle parene i $A \times B$ på denne måten

$$(a_1, b_1), (a_2, b_1), (a_1, b_2), (a_3, b_1)(a_2, b_2), (a_1, b_3), (a_4, b_1), \dots$$

Det skulle ikke være så vanskelig å gjennomskue metoden: Først lister vi alle par der summen av indeksene er 2 (det er bare ett slikt par), så lister vi alle par der summen av indeksene er 3, deretter alle der summen av indeksene er 4, osv. Dette gir en opplisting av alle parene i $A \times B$. $\square$

Setningen ovenfor har en overraskende konsekvens:

Teorem 1.4.2 *Mengden $\mathbb{Q}$ av rasjonale tall er tellbar.*

Bevis: Siden både $\mathbb{Z}$ og $\mathbb{N}$ er tellbare, forteller setningen ovenfor oss at $\mathbb{Z}\times\mathbb{N}$ er tellbar. La

$$(a_1, b_1), (a_2, b_2), (a_3, b_3), (a_4, b_4) \ldots$$

være en opplisting av alle elementene i $\mathbb{Z} \times \mathbb{N}$, Da er

$$\frac{a_1}{b_1}, \frac{a_2}{b_2}, \frac{a_3}{b_3}, \frac{a_4}{b_4}, \ldots$$

en opplisting av alle elementene i $\mathbb{Q}$, og følgelig er $\mathbb{Q}$ tellbar (det spiller ikke noen rolle at alle elementene i $\mathbb{Q}$ vil komme mange ganger på listen i forkortet og uforkortet form). □

Man kan jo begynne å lure på om alle mengder er tellbare, men det er ikke tilfellet:

Teorem 1.4.3 *Mengden $\mathbb{R}$ av alle reelle tall er* ikke *tellbar.*

Bevis: (Cantors diagonalargument) Anta for motsigelse at $\mathbb{R}$ er tellbar og kan bli listet: $r_1, r_2, r_3, \ldots$. Vi skriver opp desimalutviklingen til tallene på listen:

$$\begin{array}{rcl}
r_1 & = & w_1.a_{11}a_{12}a_{13}a_{14}\ldots \\
r_2 & = & w_2.a_{21}a_{22}a_{23}a_{24}\ldots \\
r_3 & = & w_3.a_{31}a_{32}a_{33}a_{34}\ldots \\
r_4 & = & w_4.a_{41}a_{42}a_{43}a_{44}\ldots \\
\vdots & \vdots & \quad\vdots
\end{array}$$

(w_i er heltallsdelen til r_i, og $a_{i1}, a_{i2}, a_{i3}, \ldots$ desimalene). For å få motsigelsen vår innfører vi et nytt tall c med desimalutvikling $c = 0.c_1c_2c_3c_4\ldots$ der desimalene er definert ved:

$$c_i = \begin{cases} 1 & \text{if } a_{ii} \neq 1 \\ \\ 2 & \text{if } a_{ii} = 1 \end{cases}$$

Dette tallet må være forskjellig fra det i-te tallet r_i på listen siden de i-te desimalene er forskjellige (siden c har bare 1 og 2 som desimaler, er det ingen problemer med at c kan ja to desimalutviklinger). Dermed er c forskjellig fra alle tallene på listen, og vi har fått en motsigelse siden vi antok at alle reelle tall (inkludert c) var på listen. □

Ofte er det en stor ulempe at $\mathbb{R}$ ikke er tellbar. Teorem 1.4.2 redder imidlertid en del — i mange av de situasjonen der vi skulle ønske oss at de reelle tallene var tellbare, kan vi komme oss ut av problemene ved å bruke at $\mathbb{Q}$ er tellbar. Det skyldes at de rasjonale tallene ligger tett i $\mathbb{R}$, og at alle reelle tall derfor kan tilnærmes med rasjonale tall.

Oppgaver

1. Bevis formel (1.2.2).
2. Bevis formel (1.2.4).
3. Vis at $A \setminus B = A \cap B^c$.
4. Anta at $f : A \to B$ er en funksjon. Vis at hvis $C, D \subset B$, så er $f^{-1}(C \cap D) = f^{-1}(C) \cap f^{-1}(D)$ og $f^{-1}(C \cup D) = f^{-1}(C) \cup f^{-1}(D)$. Vi sier at snitt og union er *bevart under inverse bilder*.
5. Anta at $f : A \to B$ er en funksjon. Vis at hvis $C, D \subset A$, så er $f(C \cup D) = f(C) \cup f(D)$. Vis også at $f(C \cap D) \subset f(C) \cap f(D)$. Finn et eksempel som viser at vi ikke alltid har $f(C \cap D) = f(C) \cap f(D)$. Vi sier at union (men ikke snitt) er *bevart under direkte bilder*.
6. Vis at hvis A, B er tellbare, så er $A \cup B$ tellbar.
7. Anta at mengdene $A_1, A_2, \ldots, A_n, \ldots$ er tellbare. Bevis at $A = \bigcup_{n=1}^{\infty} A_n$ er tellbar. (Her består $\bigcup_{n=1}^{\infty} A_n$ av alle elementer som er med i minst én av mengdene $A_1, A_2, \ldots, A_n, \ldots$. Det kan for øvrig være lurt å kikke litt på beviset for setning 1.4.1.)
8. Vis at mengden av alle delmengder av $\mathbb{N}$ *ikke* er tellbar. (*Hint*: Prøv å tilpasse Cantors diagonalargument).

Kapittel 2

ϵ-δ og alt det der

Mange strever med ϵ-δ-argumenter. Det er flere grunner til dette: Noen har problemer med å forstå den underliggende tankegangen, mens andre sliter med de grunnleggende regneteknikkene. De heldigste faller først av lasset når regnestykkene blir for kompliserte. I dette kapitlet skal jeg prøve å løse opp noen av flokene som folk kanskje har rotet seg inn i.

2.1 Tallverdier og normer

Én av tingene som gjør ϵ-δ-argumenter vanskelige, er at de involverer tallverdier og normer, typisk i uttrykk av typen $|x-a| < \delta$ eller $|\mathbf{F}(\mathbf{x}) - \mathbf{F}(\mathbf{a})| < \epsilon$. Mange synes tallverdier og normer er litt ekle, men i ϵ-δ-sammenheng er de lette å tolke: *Du skal alltid tenke på* $|x-a|$ *som avstanden mellom* x *og* a! At $|x-a| < \epsilon$, betyr altså bare at avstanden mellom x og a er mindre enn ϵ. Er du på tallinjen og tenker på a som et fast tall mens x kan variere, betyr $|x-a| < \epsilon$ dermed at x ligger mellom $a-\epsilon$ og $a+\epsilon$. Akkurat det samme gjelder i $\mathbb{R}^n$: *Du skal alltid tenke på* $|\mathbf{x}-\mathbf{a}|$ *som avstanden mellom punktene* $\mathbf{x}$ *og* $\mathbf{a}$. Tenker du på $\mathbf{a}$ som et fast punkt mens $\mathbf{x}$ kan variere, betyr altså $|\mathbf{x}-\mathbf{a}| < \epsilon$ at $\mathbf{x}$ ligger innenfor kulen om $\mathbf{a}$ med radius ϵ. Legg for øvrig merke til at rekkefølgen av leddene ikke spiller noen rolle: $|\mathbf{x}-\mathbf{a}| = |\mathbf{a}-\mathbf{x}|$.

Når man arbeider med tallverdier og normer, er det viktig å kunne veksle mellom det geometriske bildet og de regnetekniske uttrykkene. I MAT1100 lærte du to ulikheter som hjelper deg med det regnetekniske:

Setning 2.1.1 (Trekantulikheten) *For alle* $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$ *gjelder*

$$|\mathbf{a}+\mathbf{b}| \leq |\mathbf{a}| + |\mathbf{b}|$$

Trekantulikheten er viktig i mange argumenter fordi den tillater oss å bryte kompliserte uttrykk ned i enklere deler — vi kan kontrollere det sammensatte uttrykket $|\mathbf{a} + \mathbf{b}|$ dersom vi kan kontrollere hver del $\mathbf{a}$ og $\mathbf{b}$. Den neste ulikheten gir oss en tilsvarende kontroll over skalarprodukter.

Setning 2.1.2 (Schwartz' ulikhet) *For alle* $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$ *gjelder*

$$|\mathbf{a} \cdot \mathbf{b}| \leq |\mathbf{a}||\mathbf{b}|$$

Når a og b er tall, er Schwartz' ulikhet en likhet: Vi har

$$|ab| = |a||b| \qquad \text{når } a, b \in \mathbb{R}$$

Vi skal senere se hvordan vi kan bruke ulikheter i ϵ-δ-argumenter.

2.2 Konvergens av følger

Før vi går løs på ϵ-δ-argumenter for alvor, skal vi ta en titt på de litt enklere ϵ-N-argumentene som dukker opp når vi studerer konvergens av følger. La oss først gruble litt over hva vi er på jakt etter: Vi ønsker å fange ideen om at $a \in \mathbb{R}$ er grensen til følgen $\{x_n\}$ når n går mot uendelig. Intuitivt betyr det at vi kan få x_n så nær a vi måtte ønske (men ikke nødvendigvis helt lik!) ved å gå tilstrekkelig langt ut i følgen. Sagt på en litt annen måte: Dersom vi spesifiserer hvor nær vi ønsker at elementene i følgen skal være til a, må det finnes en N slik at kravet er oppfylt for alle leddene i følgen fra x_N og utover. Dersom spesifikasjonen vår er at avstanden mellom leddene og a skal være mindre enn ϵ, ender vi opp med denne definisjonen:

Definisjon 2.2.1 *Følgen* $\{x_n\}$ *i* $\mathbb{R}$ *konvergerer mot* a *dersom det til hver* $\epsilon > 0$ *finnes en* $N \in \mathbb{N}$ *slik at* $|x_n - a| < \epsilon$ *for alle* $n \geq N$.

Med litt andre ord krever definisjonen at x_n skal ligge mellom $a - \epsilon$ og $a + \epsilon$ når $n \geq N$.

Vi har akkurat samme definisjon for følger i $\mathbb{R}^n$:

Definisjon 2.2.2 *Følgen* $\{\mathbf{x}_n\}$ *i* $\mathbb{R}^m$ *konvergerer mot* $\mathbf{a}$ *dersom det til hver* $\epsilon > 0$ *finnes en* $N \in \mathbb{N}$ *slik at* $|\mathbf{x}_n - \mathbf{a}| < \epsilon$ *for alle* $n \geq N$.

Den "geometriske" tolkningen i dette tilfellet er at for $n \geq N$, skal $\mathbf{x}_n$ ligge inni kulen med sentrum i $\mathbf{a}$ og radius ϵ.

Legg merke til at definisjonene ovenfor sier at det for *alle* ϵ skal finnes en N som tilfredsstiller kravet. Denne N'en vil normalt avhenge av ϵ — jo mindre ϵ blir, dess større må vi vanligvis velge N (noen bøker skriver $N(\epsilon)$ for å markere at N avhenger av ϵ). Dersom vi ønsker å vise at følgen *ikke* konvergerer, er det nok å vise at det finnes én ϵ slik at uansett hvor stor vi velger N, så finnes det en $n \geq N$ slik at $|\mathbf{x}_n - \mathbf{a}| \geq \epsilon$.

Bemerkning: Noen liker å tenke på definisjonen ovenfor som et spill mellom to spillere, I og II. Spiller I ønsker å vise at følgen *ikke* konvergerer, mens spiller II ønsker å vise at den konvergerer. I hver runde i spillet velger spiller I først en ϵ, og deretter velger spiller II en N. Spiller II vinner runden dersom $|\mathbf{x}_n - \mathbf{a}_n| < \epsilon$ for alle $n \geq N$, hvis ikke vinner spiller I.

Følgen $\{\mathbf{x}_n\}$ konvergerer dersom spiller II har en gevinststrategi i dette spillet — dvs. dersom hun uansett hvilken $\epsilon > 0$ spiller I måtte velge, kan svare med en N slik at $|\mathbf{x}_n - \mathbf{a}| < \epsilon$ for alle $n \geq N$. Følgen konvergerer *ikke* dersom spiller I har en gevinststrategi, dvs. dersom spiller I kan velge en ϵ som ikke kan "pareres" av noen N.

La oss se på et enkelt eksempel på hvordan trekantulikheten kan brukes til å vise konvergens av sammensatte uttrykk.

Setning 2.2.3 *Anta at $\{\mathbf{x}_n\}$ og $\{\mathbf{y}_n\}$ er to følger i $\mathbb{R}^m$ som konvergerer mot henholdsvis $\mathbf{a}$ og $\mathbf{b}$. Da konvergerer følgen $\{\mathbf{x}_n + \mathbf{y}_n\}$ mot $\mathbf{a} + \mathbf{b}$.*

Bevis: Vi må vise at gitt en $\epsilon > 0$, kan vi alltid finne en $N \in \mathbb{N}$ slik at $|(\mathbf{x}_n + \mathbf{y}_n) - (\mathbf{a} - \mathbf{b})| < \epsilon$ for alle $n \geq N$. Vi starter med å samle de leddene som "hører sammen", og bruker deretter trekantulikheten:

$$|(\mathbf{x}_n + \mathbf{y}_n) - (\mathbf{a} - \mathbf{b})| = |(\mathbf{x}_n - \mathbf{a}) + (\mathbf{y}_n - \mathbf{b})| \leq |\mathbf{x}_n - \mathbf{a}| + |\mathbf{y}_n - \mathbf{b}|$$

Siden $\mathbf{x}_n$ konvergerer mot $\mathbf{a}$, vet vi at det finnes en $N_1 \in \mathbb{N}$ slik at $|\mathbf{x}_n - \mathbf{a}| < \frac{\epsilon}{2}$ for alle $n \geq N_1$ (hvis du ikke skjønner dette, se bemerkningen nedenfor). Siden $\mathbf{y}_n$ konvergerer mot $\mathbf{b}$, kan vi tilsvarende finne en $N_2 \in \mathbb{N}$ slik at $|\mathbf{y}_n - \mathbf{b}| < \frac{\epsilon}{2}$ for alle $n \geq N_2$. Setter vi $N = \max\{N_1, N_2\}$, ser vi at når $n \geq N$, så er

$$|(\mathbf{x}_n + \mathbf{y}_n) - (\mathbf{a} - \mathbf{b})| \leq |\mathbf{x}_n - \mathbf{a}| + |\mathbf{y}_n - \mathbf{b}| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

Dermed er oppdraget utført, og setningen er bevist. □

Bemerkning: Mange blir forvirret av at $\frac{\epsilon}{2}$ plutselig dukker opp i beviset ovenfor og overtar rollen til ϵ: Vi finner en N_1 slik at $|\mathbf{x}_n - \mathbf{a}| < \frac{\epsilon}{2}$ for alle $n \geq N_1$. Det er imidlertid ikke noe galt i det; siden $\mathbf{x}_n \to \mathbf{a}$, kan vi takle en hvilken som helst "epsilon" vi blir utfordret med, også halvparten av den opprinnelige epsilon'en.

2.3 Kontinuitet

"Ekte" ϵ-δ-formuleringer dukker først opp i forbindelse med kontinuitet og grenser. La oss holde oss til kontinuitet:

Definisjon 2.3.1 *En funksjon $f : \mathbb{R} \to \mathbb{R}$ er kontinuerlig i punktet a dersom det for hver $\epsilon > 0$, finnes en $\delta > 0$ slik at $|f(x) - f(a)| < \epsilon$ når $|x - a| < \delta$.*

I denne definisjonen er det viktig ikke å gå seg vill i tallverdiene, men holde seg til vår "avstandsfilosofi": Definisjonen sier at uansett hvilken $\epsilon > 0$ vi blir gitt, så finnes det en $\delta > 0$ slik at når avstanden mellom x og a er mindre enn δ, så er avstanden mellom $f(x)$ og $f(a)$ mindre enn ϵ.

Akkurat som før kan vi tenke på definisjonen som et spill mellom to spillere; spiller I som ønsker å vise at funksjonen *ikke* er kontinuerlig i a, og spiller II som ønsker å vise at den *er* kontinuerlig i a. En runde i spillet foregår ved at spiller I velger en $\epsilon > 0$, og at spiller II svarer med en $\delta > 0$. Spiller II vinner runden dersom $|f(x) - f(a)| < \epsilon$ for alle x slik at $|x - a| < \delta$, ellers vinner spiller I. Funksjonen er kontinuerlig dersom spiller II har en gevinststrategi; dvs. dersom hun uansett hvilken $\epsilon > 0$ spiller I velger, kan finne en $\delta > 0$ som vinner runden, dvs. som er slik at $|f(x) - f(a)| < \epsilon$ når $|x - a| < \delta$. Spiller II vinner dersom hun finner en ϵ som ikke kan pareres, dvs. dersom det uansett hvor liten δ spiller II velger, finnes tall x slik at $|x - a| < \delta$ og $|f(x) - f(a)| \geq \epsilon$.

La oss for en gangs skyld se på et eksempel der spiller I vinner, dvs. der funksjonen f *ikke* er kontinuerlig.

Eksempel 1: La

$$f(x) = \begin{cases} 1 & \text{hvis } x \leq 0 \\ \\ 2 & \text{hvis } x > 0 \end{cases}$$

Intuitivt er denne funksjonen diskontinuerlig i 0 fordi den gjør et sprang der. Hvordan fanges dette opp av ϵ-δ-definisjonen? Vi ser at $f(0) = 1$, men at det ligger punkter helt opptil 0 der funksjonsverdien er 2. Velger vi (nå i rollen til spiller I) en $\epsilon < 1$, har spiller

II ingen mulighet til å parere: Uansett hvor liten hun velger $\delta > 0$, vil det finnes punkter x, $0 < x < \delta$ der $f(x) = 2$, og følgelig er $|f(x) - f(0)| = |2 - 1| = 1 > \epsilon$. Altså er f diskontinuerlig i 0.

La oss nå se på et litt mer sammensatt eksempel.

Setning 2.3.2 *Funksjonen $f : \mathbb{R} \to \mathbb{R}$ er kontinuerlig i a hvis og bare hvis $\lim_{n\to\infty} f(x_n) = f(a)$ for alle følger $\{x_n\}$ som konvergerer mot a.*

Bevis: Anta først at f er kontinuerlig i a, og at $\lim_{n\to\infty} x_n = a$. Vi må vise at $f(x_n)$ konvergerer mot $f(a)$. Gitt en $\epsilon > 0$, må vi altså vise at det finnes en $N \in \mathbb{N}$ slik at $|f(x_n) - f(a)| < \epsilon$ når $n \geq N$. Siden f er kontinuerlig i a, finnes det en $\delta > 0$ slik at $|f(x) - f(a)| < \epsilon$ når $|x - a| < \delta$. Men vi vet at x_n konvergerer mot a, og dermed finnes det en $N \in \mathbb{N}$ slik at $|x_n - a| < \delta$ når $n \geq N$ (legg merke til at δ nå spiller den rollen som ϵ pleier å ha, men det er ikke noe problem). Vi ser nå at dersom $n \geq N$, så er $|x_n - a| < \delta$, og dermed er $|f(x_n) - f(a)| < \epsilon$. Dermed har vi vist at $\{f(x_n)\}$ konvergerer mot $f(a)$.

Det gjenstår å vise at dersom f *ikke* er kontinuerlig i a, så finnes det minst én delfølge $\{x_n\}$ som konvergerer mot a uten at $\{f(x_n)\}$ konvergerer mot $f(a)$. Siden f ikke er kontinuerlig i a, finnes det en $\epsilon > 0$ slik at uansett hvor liten vi velger $\delta > 0$, så finnes det et punkt x slik at $|x - a| < \delta$, men $|f(x) - f(a)| \geq \epsilon$. Velger vi $\delta = \frac{1}{n}$, finnes det derfor et et punkt x_n slik at $|x_n - a| < \frac{1}{n}$, men $|f(x_n) - f(a)| \geq \epsilon$. Dermed konvergerer $\{x_n\}$ mot a, mens $\{f(x_n)\}$ *ikke* konvergerer mot $f(a)$ (fordi $f(x_n)$ alltid har avstand minst ϵ til $f(a)$). $\square$

Beviset ovenfor viser hvordan vi kan sette sammen ulike former av avhengighet. Legg spesielt merke til hvordan gamle størrelser dukker opp i nye roller — plutselig hadde δ den rollen som ϵ pleier å ha. Dette er uproblematisk så lenge vi husker på at navnet på en størrelse ikke spiller noen rolle; det vi vanligvis kaller ϵ, kunne likså godt ha hett a, b — eller δ. Grunnen til at vi forsøker å bruke samme navn på størrelser som har samme funksjon i ulike bevis, er at det forenkler tankearbeidet — vi slipper å bruke krefter på å huske hva symbolene står for.

La oss helt til slutt se litt på kontinuitet i $\mathbb{R}^n$. Definisjonen er den samme som før:

Definisjon 2.3.3 *En funksjon $\mathbf{F} : \mathbb{R}^n \to \mathbb{R}^m$ er kontinuerlig i punktet $\mathbf{a}$ dersom det for hver $\epsilon > 0$, finnes en $\delta > 0$ slik at $|\mathbf{F}(\mathbf{x}) - \mathbf{F}(\mathbf{a})| < \epsilon$ når $|\mathbf{x} - \mathbf{a}| < \delta$.*

Lar vi $\mathrm{B}(\mathbf{c}, r)$ betegne (den åpne) kulen med sentrum i c og radius r, har definisjonen denne geometriske tolkningen: Hvis vi velger en kule $\mathrm{B}(\mathbf{F}(\mathbf{a}), \epsilon)$ med sentrum i $\mathbf{F}(\mathbf{a})$, så vil det uansett hvor liten radiusen ϵ er, finnes en kule $\mathrm{B}(\mathbf{a}, \delta)$ med sentrum i $\mathbf{a}$ som avbildes inn i $\mathrm{B}(\mathbf{F}(\mathbf{a}), \epsilon)$, dvs. slik at $\mathbf{x} \in \mathrm{B}(\mathbf{a}, \delta) \Longrightarrow \mathbf{F}(\mathbf{x}) \in \mathrm{B}(\mathbf{F}(\mathbf{a}), \epsilon)$.

2.4 Estimater og overslag

Det er viktig å være klar over at når vi gjennomfører et ϵ-δ-bevis, er vi bare på jakt etter én δ som gjør jobben, og det er ikke noen vits å lete etter den beste/største δ'en som virker. Det betyr at vi ofte kan bruke overslag for å forenkle argumentene — vi sier ting som: "denne faktoren kan aldri bli større enn 10, og dermed holder det å velge δ lik $\frac{\epsilon}{10}$". Overslag er uvant for mange studenter og bidrar nok til ϵ-δ-bevis ofte virker vanskeligere enn de er. Her er et litt mer teoretisk eksempel på hvordan man går frem i et sammensatt ϵ-δ-argument. Jeg har prøvd å forklare hvordan man tenker underveis:

Setning 2.4.1 *Anta at $g : \mathbb{R} \to \mathbb{R}$ er kontinuerlig i punktet a, og at $g(a) \neq 0$. Da er funksjonen $h(x) = \frac{1}{g(x)}$ er kontinuerlig i a.*

Bevis: Gitt en $\epsilon > 0$, må vi vise at det finnes en $\delta > 0$ slik at $|\frac{1}{g(x)} - \frac{1}{g(a)}| < \epsilon$ når $|x - a| < \delta$.
La oss først skrive om uttrykket på en enklere form. Setter vi på felles brøkstrek, ser vi at

$$\left|\frac{1}{g(x)} - \frac{1}{g(a)}\right| = \frac{|g(a) - g(x)|}{|g(x)||g(a)|}$$

Siden $g(x) \to g(a)$, kan vi få telleren i dette uttrykket så liten vi måtte ønske ved å velge x nær nok a. Problemet er bare at dersom nevneren er liten, kan brøken likevel bli stor (husk at små nevnere, gir store brøker — vi er nødt til å tenke litt opp-ned her!) Den ene faktoren i nevneren, $|g(a)|$, har vi rimelig kontroll på siden den er konstant. Hva med den andre faktoren $|g(x)|$? Siden $g(x) \to g(a)$, kan heller ikke denne faktoren bli for liten; det må f.eks. finnes en $\delta_1 > 0$ slik at $|g(x)| > \frac{|g(a)|}{2}$ når $|x - a| < \delta_1$ (tenk nøye hva som foregår her — dette er egentlig et separat lite ϵ-δ-resonnement med $\epsilon = \frac{|g(a)|}{2}$). For alle x'er slik at $|x - a[< \delta_1$, er dermed

$$\left|\frac{1}{g(x)} - \frac{1}{g(a)}\right| = \frac{|g(a) - g(x)|}{|g(x)||g(a)|} < \frac{|g(a) - g(x)|}{\frac{|g(a)|}{2}|g(a)|} = \frac{2}{|g(a)|^2}|g(a) - g(x)|$$

Hvordan kan vi få dette uttrykket mindre enn ϵ? Vi trenger åpenbart å få $|g(a) - g(x)| < \frac{|g(a)|^2}{2}\epsilon$, og siden g er kontinuerlig i a, vet vi at det finnes en δ_2 slik at $|g(a) - g(x)| < \frac{|g(a)|^2}{2}\epsilon$ når $|x - a| < \delta_2$. Velger vi $\delta = \min\{\delta_1, \delta_2\}$, har vi dermed

$$\left|\frac{1}{g(x)} - \frac{1}{g(a)}\right| \leq \frac{2}{|g(a)|^2}|g(a) - g(x)| < \frac{2}{|g(a)|^2}\frac{|g(a)|^2}{2}\epsilon = \epsilon$$

og beviset er fullført. □

Oppgaver

1. Vis at dersom følgen $\{x_n\}$ konvergerer mot a, så konvergerer følgen $\{Mx_n\}$ (der M er en konstant) mot Ma. Bruk definisjonen av konvergens og tenk nøye gjennom hvordan du finner N når ϵ er gitt.
2. Bruk definisjonen av kontinuitet til å vise at dersom funksjonen $f : \mathbb{R} \to \mathbb{R}$ er kontinuerlig i punktet a, så er funksjonen $g(x) = Mf(x)$, der M er en konstant, også kontinuerlig i a.
3. Bruk definisjonen av kontinuitet til å vise at dersom $f, g : \mathbb{R} \to \mathbb{R}$ er kontinuerlige i punktet a, så er $f + g$ det også.
4. Bruk definisjonen av kontinuitet til å vise at dersom $f, g : \mathbb{R} \to \mathbb{R}$ er kontinuerlige i punktet a, så er fg det også. (*Hint:* Skriv $|f(x)g(x) - f(a)g(a)| = |(f(x)g(x) - f(a)g(x)) + (f(a)g(x) - f(a)g(a))|$ og bruk trekantulikheten.)
5. Bruk definisjonen av kontinuitet til å vise at funksjonen $f(x) = \frac{1}{\sqrt{x}}$ er kontinuerlig i alle punkter $a > 0$.
6. Bruk trekantulikheten til å vise at $\big|\,|\mathbf{a}| - |\mathbf{b}|\,\big| \leq |\mathbf{a} - \mathbf{b}|$ for alle $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$.

Kapittel 3

Kompletthet

Kompletthet er et begrep som står sentralt i MAT1100 og MAT1110, og som vil stå enda mer sentralt i MAT2400. I de tidligere kursene fremstår begrepet på litt forskjellig måte — i MAT1100 er det først og fremst knyttet til begrepet minste øvre skranke, mens det i MAT1110 er knyttet til konvergens av Cauchy-følger — og det kan derfor være greit med en gjennomgang som viser sammenhengen mellom de to formuleringene.

3.1 Minste øvre skranke

Husk at en delmengde $A \subset \mathbb{R}$ er *oppad begrenset* dersom det finnes et tall c som er større enn alle elementene i A — et slikt tall kalles en *øvre skranke* for A. En oppad begrenset mengde vil selvfølgelig ha mange øvre skranker (hvis c er en øvre skanke, vil også alle tall større enn c være øvre skranker), men kompletthetsprinsippet fra MAT1100 forteller oss at det alltid finnes en *minste* øvre skranke.

Kompletthetsprinsippet. Enhver ikke-tom, oppad begrenset delmengde av $\mathbb{R}$ har en minste øvre skranke.

Den minste øvre skranken til A kalles også *supremum* til A og betegnes med $\sup A$. En nedad begrenset mengde vil på samme måte ha en *største nedre skranke* (eller et *infimum*) som betegnes med $\inf A$. Dersom mengden ikke er oppad begrenset, setter vi $\sup A = \infty$, og dersom den ikke er nedad begrenset, setter vi $\inf A = -\infty$

Husk at for noen mengder er $\sup A$ ($\inf A$) med i mengden (f.eks. hvis A=[0,1]), mens i andre tilfeller er $\sup A$ ($\inf A$) ikke med i mengden (f.eks. når $A = (0, 1)$).

Tenker man geometrisk, kan kompletthetsprinsippet virke opplagt — det må da være punkter som markerer hvor en begrenset mengde begynner og hvor den slutter? Bytter vi ut de reelle tallene med de rasjonale, ser vi at dette ikke er så opplagt; mengden

$$A = \{q \in \mathbb{Q} \mid q^2 < 2\}$$

er oppad og nedad begrenset i $\mathbb{Q}$, men den har ikke infimum og supremum i $\mathbb{Q}$ — infimum og supremum er henholdsvis $-\sqrt{2}$ og $\sqrt{2}$ som ikke er rasjonale tall. Dette eksemplet viser en viktig forskjell på den reelle tallinjen $\mathbb{R}$ og den rasjonale tallinjen $\mathbb{Q}$ — den rasjonale tallinjen har "hull", mens den reelle tallinjen er "tett". Skjæringssetningen er et slående eksempel på denne forskjellen. Den sier som kjent at enhver kontinuerlig funksjon $f : \mathbb{R} \to \mathbb{R}$ som skifter fortegn i intervallet fra a til b, må ha et nullpunkt i dette intervallet. Dette gjelder ikke for den rasjonale tallinjen: Funksjonen $f : \mathbb{Q} \to \mathbb{Q}$ gitt ved $f(x) = x^2 - 2$ skifter fortegn i intervallet fra 0 til 2, men den har ikke noe (rasjonalt) nullpunkt der — funksjonen sniker seg gjennom x-aksen i "hullet" det $\sqrt{2}$ burde ha vært.

Dette eksemplet viser poenget med kompletthet: Vi ønsker å arbeide med strukturer uten "hull"; med strukturer der det alltid er elementer der vi med rimelighet kan vente at de skal være. Det er tilfellet med $\mathbb{R}$, men ikke med $\mathbb{Q}$, og det skyldes at $\mathbb{R}$ er "komplett" — i en betydning vil bli klarere etter hvert — mens $\mathbb{Q}$ ikke er det.

MAT2400 er på mange måter studiet av kompletthetsbegrepet og dets konsekvenser, men før vi nærmer oss det generelle begrepet, må vi ta en titt på hvordan det fungerer i $\mathbb{R}^m$.

3.2 Cauchy-følger

Det er ikke noen naturlig måte å ordne punktene i $\mathbb{R}^m$ på når $m > 1$, og derfor er det heller ikke naturlig å bruke øvre og nedre skranker til å beskrive kompletthet i $\mathbb{R}^m$. I MAT1110 brukte vi derfor Cauchy-følger til å beskrive kompletthet i disse rummene. Husk definisjonen:

Definisjon 3.2.1 *En følge* $\{\mathbf{x}_n\}$ *i* $\mathbb{R}^m$ *kalles en* Cauchy-følge *dersom det for enhver* $\epsilon > 0$ *finnes en* $N \in \mathbb{N}$ *slik at* $|\mathbf{x}_n - \mathbf{x}_k| < \epsilon$ *når* $n, k \geq N$.

Intuitivt er en Cauchy-følge en følge der leddene skvises tettere og tettere sammen dess lenger ut i følgen vi kommer. Kompletheten av $\mathbb{R}^n$ kan formuleres som et teorem (som ble bevist i MAT1110):

Teorem 3.2.2 (Kompletthet av $\mathbb{R}^m$) *En følge* $\{\mathbf{x}_n\}$ *i* $\mathbb{R}^m$ *konvergerer hvis og bare hvis den er en Cauchy-følge.*

I utgangspunktet er det ikke så lett å se den logiske sammenhengen mellom dette resultatet og kompletthetsprinsippet i seksjon 1, men det er i hvert fall en viss idémessig sammenheng — i et "rom uten hull" bør leddene i en Cauchy-følge skvises sammen mot et grensepunkt.

Vi skal nå se hvordan vi kan bruke kompletthetsprinsippet til å bevise teorem 3.2.2. Beviset er i flere skritt, og underveis skal vi innføre noen begreper du får bruk for i MAT2400. Først ser vi hva som skjer med monotone følger, dvs. følger i $\mathbb{R}$ som er enten voksende eller avtagende.

Setning 3.2.3 *Enhver begrenset, monoton følge i $\mathbb{R}$ konvergerer.*

Bevis: Vi ser på voksende følger, de avtagende kan behandles på samme måte. Siden følgen $\{a_n\}$ er begrenset, er mengden

$$A = \{a_1, a_2, a_3, \ldots, a_n, \ldots\}$$

av alle elementer i følgen også begrenset og har følgelig en minste øvre skranke $a = \sup A$. For å vise at følgen konvergerer mot a, må vi vise at for enhver $\epsilon > 0$, finnes det en $N \in \mathbb{N}$ slik at $|a - a_n| < \epsilon$ når $n > N$.

Det er ikke så vanskelig. Siden a er den minste øvre skranken til A, kan ikke $a - \epsilon$ være en øvre skranke, og dermed må det finnes en a_N slik at $a_N > a - \epsilon$. Siden følgen er voksende er da $a - \epsilon < a_n \leq a$ for alle $n > N$, og følgelig er $|a - a_n| < \epsilon$ for slike n. □

Legg merke til hvordan vi brukte kompletthetsprinsippet i beviset ovenfor. For å bruke setningen på Cauchy-følger, trenger vi å vite at de er begrensede.

Lemma 3.2.4 *En Cauchy-følge er begrenset.*

Bevis: Vi kan bruke definisjonen av Cauchy-følge med en hvilken som helst ϵ, la oss si $\epsilon = 1$. Ifølge definisjonen finnes det da en N slik at $|a_n - a_k| < 1$ når $n, k \geq N$. Spesielt er $|a_n - a_N| < 1$ for alle $n > N$. Dette betyr at

$$K = \max\{a_1, a_2, \ldots, a_{N-1}, a_N + 1\}$$

er en øvre skranke for følgen, mens

$$k = \min\{a_1, a_2, \ldots a_{N-1}, x_N - 1\}$$

er en nedre skranke. □

Vi skal nå innføre to begreper for følger som du sannsynligvis ikke har sett før. Gitt en følge $\{a_n\}$ av reelle tall, innfører vi to nye følger $\{M_n\}$ og $\{m_n\}$ ved

$$M_n = \sup\{a_n, a_{n+1}, a_{n+2}, \ldots\}$$

$$m_n = \inf\{a_n, a_{n+1}, a_{n+2}, \ldots\}$$

Vi ser at følgen $\{M_n\}$ er avtagende (siden vi tar sup over stadig mindre mengder), mens følgen $\{m_n\}$ er voksende (siden vi tar inf over stadig mindre mengder). I tillegg er åpenbart $m_n \leq M_n$. Intuitivt måler $\{M_n\}$ hvordan toppene til $\{a_n\}$ avtar når vi går utover i følgen, mens $\{m_n\}$ viser hvordan bunnene vokser.

Vi definerer nå den øvre og nedre grensen til $\{a_n\}$ ved

$$\overline{\lim}_{n\to\infty} a_n = \lim_{n\to\infty} M_n$$

$$\underline{\lim}_{n\to\infty} a_n = \lim_{n\to\infty} m_n$$

Siden følgene M_n og m_n er monotone, vil disse grensene alltid finnes og

$$\overline{\lim}_{n\to\infty} a_n \geq \underline{\lim}_{n\to\infty} a_n$$

Generelt kan disse grensene være uendelige, men ikke når utgangsfølgen $\{a_n\}$ er begrenset.

Det er en nær sammenheng mellom øvre og nedre grenser og vanlige grenser:

Lemma 3.2.5 *En følge $\{a_n\}$ av reelle tall konvergerer mot $a \in \mathbb{R}$ hvis og bare hvis*

$$\overline{\lim}_{n\to\infty} a_n = \underline{\lim}_{n\to\infty} a_n = a$$

Bevis: Siden $m_n \leq a_n \leq M_n$, ser vi at hvis

$$\lim_{n\to\infty} M_n = \overline{\lim}_{n\to\infty} a_n = a$$

og

$$\lim_{n\to\infty} m_n = \underline{\lim}_{n\to\infty} a_n = a$$

så er $\lim_{n\to\infty} a_n = a$.

Anta omvendt at $\lim_{n\to\infty} a_n = a$. For enhver $\epsilon > 0$ finnes det da en N slik at $|a - a_n| < \epsilon$ for alle $n \geq N$. Men da er $|a - M_n| \leq \epsilon$ og $|a - m_n| \leq \epsilon$ for alle $n \geq N$, og følgelig er

$$\overline{\lim}_{n\to\infty} a_n = \underline{\lim}_{n\to\infty} a_n = a$$

□

Vi kan nå vise at Cauchy-følger i $\mathbb{R}$ konvergerer.

Setning 3.2.6 *Alle Cauchy-følger* $\{a_n\}$ *i* $\mathbb{R}$ *konvergerer.*

Bevis: Siden $\{a_n\}$ er begrenset, er grenseverdiene

$$M = \overline{\lim}_{n\to\infty} a_n \qquad \text{og} \qquad m = \underline{\lim}_{n\to\infty} a_n$$

endelige, og ifølge lemmaet holder det å vise at $M = m$.

Gitt en $\epsilon > 0$, finnes det finnes det en N slik at $|a_n - a_k| < \epsilon$ når $n, k \geq N$. Spesielt er $|a_n - a_N| < \epsilon$ for alle $n \geq N$. Dette betyr at $m_k \geq a_N - \epsilon$ og $M_k \leq a_N + \epsilon$ for $k \geq N$. Følgelig er $M_k - m_k \leq 2\epsilon$ for alle $k \geq N$. Dermed $M - m \leq 2\epsilon$ for enhver $\epsilon > 0$, og det er bare mulig hvis $M = m$. □

Det er lett å utvide dette resultatet til Cauchy-følger i $\mathbb{R}^m$.

Setning 3.2.7 *Alle Cauchy-følger* $\{x_n\}$ *i* $\mathbb{R}^m$ *konvergerer.*

Bevis: Dersom leddene i følgen har komponenter

$$\mathbf{a}_n = (a_n^{(1)}, a_n^{(2)}, \ldots, a_n^{(m)}),$$

vil komponentfølgene $\{a_n^{(k)}\}$ være Cauchy-følger, og følgelig konvergere. Men hvis komponentene konvergerer, konvergerer også den opprinnelige følgen $\mathbf{a}_n$. □

Vi har nå bevist den ene implikasjonen i teorem 3.2.2. Den andre (at alle konvergente følger er Cauchy-følger) er mye enklere og overlates til leserne (se oppgave 5).

Argumentet ovenfor viser hvordan man kan bruke kompletthetsprinsippet til å vise at alle Cauchy-følger konvergerer. Det går an å argumentere motsatt også — starte med å anta at alle Cauchy-følger i $\mathbb{R}$ konvergerer og utlede kompletthetsprinsippet. Kompletthetsprinsippet og teorem 3.2.2 kan derfor sees på som to innfallsvinkler til samme fenomen — de fanger begrepet "kompletthet" på to forskjellige måter. Begge har sine fordeler og ulemper: Kompletthetsprinsippet er enklere å formulere og forstå, men konvergens av Cauchy-følger er enklere å generalisere til nye sammenhenger. I MAT2400 skal vi studere kompletthet i såkalte *metriske rom*, og da er det Cauchy-følgene vi skal ta utgangspunkt i. Du vil få en forsmak på dette i kapittel 5.

Oppgaver

I oppgave 1-5 vil du se at en del sentrale teoremer fra MAT1100/1110 ikke ville ha holdt dersom vi hadde insistert på bare å arbeide med rasjonale tall. Dette skyldes at $\mathbb{Q}$ ikke tilfredsstiller kompletthetsprinsippet som er en viktig ingrediens i beviset for alle disse resultatene. Hvis du ikke husker hva de ulikene teoremene sier, kan du ta en titt i neste kapittel.

1. Vis at funksjonen $f : \mathbb{Q} \to \mathbb{Q}$ definert ved $f(x) = \frac{1}{x^2-2}$ er kontinuerlig for alle $x \in \mathbb{Q}$, men at den er ubegrenset på intervallet $[0, 2]$. Sammenlign med ekstremalverdisetningen fra MAT1100 og MAT1110.
2. Vis at funksjonen $f : \mathbb{Q} \to \mathbb{Q}$ definert ved $f(x) = x^3 - 6x$ er kontinuerlig for alle $x \in \mathbb{Q}$, men at den ikke har et maksimumspunkt på intervallet $[0, 2]$. Sammenlign med ekstremalverdisetningen fra MAT1100 og MAT1110.
3. Vis at funksjonen $f : \mathbb{Q} \to \mathbb{Q}$ definert ved $f(x) = x^3 - 9x$ tilfredsstiller $f(0) = f(3) = 0$, men at det ikke finnes noe rasjonalt punkt i intervallet $[0, 3]$ der den deriverte er null. Sammenlign med middelverdisetningen fra MAT1100.
4. Finn en Cauchy-følge i $\mathbb{Q}$ som ikke konvergerer mot et tall i $\mathbb{Q}$. (*Hint:* Finn en følge av rasjonale tall som konvergerer mot $\sqrt{2}$).
5. Finn en begrenset delfølge i $\mathbb{Q}$ som ikke har noen delfølge som konvergerer i $\mathbb{Q}$.
6. Vis at alle konvergente følger i $\mathbb{R}^m$ er Cauchy-følger. (*Hint:* Anta at $\mathbf{x}_n$ konvergerer not $\mathbf{x}$. Gitt $\epsilon > 0$, finnes det en $N \in \mathbb{N}$ slik at $|\mathbf{x} - \mathbf{x}_n| < \epsilon/2$ når $n \geq N$. Hvordan er det med $|\mathbf{x}_n - \mathbf{x}_k|$ når $n, k \geq N$?)

Kapittel 4

Gjensyn med noen viktige teoremer fra kalkulus

I dette kapitlet skal vi raskt repetere noen viktige resultater fra kalkuluskursene: skjæringssetningen, middelverdisetningen, ekstremalverdisetningen og Bolzano-Weierstrass' teorem. Disse setningene (og deres bevis) er både et grunnlag og en inspirasjonskilde for mye av det som skjer i MAT2400.

4.1 Skjæringssetningen

Skjæringssetningen sier at en kontinuerlig funksjon ikke kan skifte fortegn uten å gå innom 0. Mer presist:

Teorem 4.1.1 (Skjæringssetningen) *Anta at $f : [a, b] \to \mathbb{R}$ er kontinuerlig, og at $f(a)$ og $f(b)$ har motsatte fortegn. Da finnes det et punkt $c \in (a, b)$ slik at $f(c) = 0$.*

Bevis: Vi ser på tilfellet der $f(a) < 0 < f(b)$, det andre tilfellet går på tilsvarende måte. La

$$c = \sup\{x \in [a, b] \mid f(x) < 0\}$$

Vi må vise at $f(c) = 0$. Hvis $\{x_n\}$ er en følge fra $[a, b]$ som konvergerer mot c ovenfra, er $f(x_n) \geq 0$ for alle n, og følgelig er $f(c) = \lim_{n\to\infty} f(x_n) \geq 0$. Per definisjon av c finnes det omvendt en følge $\{z_n\}$, der $f(z_n) < 0$, som konvergerer mot c nedenfra. Dermed er $f(c) = \lim_{n\to\infty} f(z_n) \leq 0$. Til sammen viser dette at $f(c) = 0$. □

Skjæringssetningen gjelder åpenbart ikke for diskontinuerlige funksjoner som kan hoppe fra positive til negative verdier, og det er derfor viktig at du skjønner hvor i beviset vi brukte at f er kontinuerlig.

Skjæringssetningen har mange generaliseringer. Her er en som ofte brukes:

Korollar 4.1.2 *Anta at $f, g : [a, b] \to \mathbb{R}$ er to kontinuerlige funksjoner slik at $f(a) < g(a)$, men $f(b) > g(b)$. Da finnes det et punkt $c \in [a, b]$ slik at $f(c) = g(c)$*

Bevis: Bruk skjæringssetningen på funksjonen $h = f - g$. □

4.2 Bolzano-Weierstrass' teorem

Dette teoremet sier at alle begrensede følger i $\mathbb{R}^m$ har en konvergent delfølge. La oss først minne om at dersom $\{\mathbf{x}_n\}$ er en følge, så får vi en delfølge $\{\mathbf{y}_k\}$ ved å plukke ut noen (men sannsynligvis ikke alle) leddene i $\{\mathbf{x}_n\}$ og sette dem sammen til den nye følgen $\{\mathbf{y}_k\}$. Mer presist: dersom

$$n_1 < n_2 < \ldots < n_k < \ldots$$

er indeksene til leddene vi plukker ut, så vil $\{\mathbf{y}_k\} = \{\mathbf{x}_{n_k}\}$ være delfølgen vår.

Vi beviser først Bolzano-Weierstrass' teorem for tallinjen.

Setning 4.2.1 *Enhver begrenset følge i $\mathbb{R}$ har en konvergent delfølge.*

Bevis: Siden følgen er begrenset, finnes det et intervall $I_0 = [a_0, b_0]$ som inneholder alle elementene x_n. Hvis vi deler dette intervallet i to like lange delintervaller $[a_0, \frac{a_0+b_0}{2}], [\frac{a_0+b_0}{2}, b_0]$, så må minst ett av disse intervallene inneholde uendelig mange elementer fra følgen. Kall dette intervallet I_1 (hvis begge delintervallene inneholder uendelig mange elementer, velger vi bare ett av dem). Nå deler vi på samme måten I_1 i to like store delintervaller, og observerer at minst ett av dem må inneholde uendelig mange elementer fra følgen. Kall dette intervallet I_2. Ved å fortsette på denne måten får vi en uendelig følge $\{I_n\}$ av intervaller som alle inneholder uendelig mange elementer i følgen. Hvert intervall ligger inni det foregående, og lengden til intervallene går mot 0.

Vi er nå klar til å konstruere delfølgen. La y_1 være det første elementet i den opprinnelige følgen $\{x_n\}$ som ligger i I_1. La så y_2 være det første elementet etter y_1 som ligger i I_2, la y_3 være det første elementet etter y_2 som ligger i I_3 osv. Siden alle intervallene inneholder uendelig mange ledd fra følgen, vil denne prosessen aldri stoppe opp, og vi får dermed en delfølge $\{y_k\}$ av den opprinnelige. Siden y_k-ene ligger nøstet inni stadig kortere intervaller, er $\{y_k\}$ en Cauchy-følge, og dermed konvergent. $\Box$

Teorem 4.2.2 (Bolzano-Weierstrass' teorem) *Enhver begrenset følge i $\mathbb{R}^m$ har en konvergent delfølge.*

Bevis: La $\{\mathbf{x}_n\}$ være følgen, og skriv den på komponentform

$$\mathbf{x}_n = (x_n^{(1)}, x_n^{(2)}, \dots, x_n^{(m)})$$

Ifølge setningen ovenfor, finnes det en delfølge $\{\mathbf{x}_{n_k}\}$ der førstekomponentene $\{x_{n_k}^{(1)}\}$ konvergerer. Bruker vi setningen ovenfor en gang til, får vi en delfølge av $\{\mathbf{x}_{n_k}\}$ der annenkomponentene konvergerer (førstekomponentene vil fortsette å konvergere mot samme grense som før). Ved å fortsette på denne måten ender vi til slutt opp med en delfølge der alle komponentene (og dermed delfølgen selv) konvergerer. $\Box$

4.3 Ekstremalverdisetningen

Maksimums- og minimumsproblemer er viktige i mange deler av matematikken. Før man prøver å løse slike problemer, er det ofte lurt å sjekke at det faktisk finnes løsninger. Ekstremalverdisetningen er et verktøy for å gjøre det.

Teorem 4.3.1 (Ekstremalverdisetningen) *Anta at K er en lukket, begrenset delmengde av $\mathbb{R}^m$, og at $f : K \to \mathbb{R}$ er en kontinuerlig funksjon. Da har f maksimums- og minimumspunkter, dvs. det finnes punkter $c, d \in K$ slik at*

$$f(d) \leq f(x) \leq f(c)$$

for alle $x \in K$.

Bevis: Jeg viser at f har et maksimumspunkt; argumentet for at f har et minimumspunkt går på akkurat samme måte.

La

$$M = \sup\{f(x) \mid x \in K\}$$

(siden vi ikke vet ennå om f er begrenset, kan det godt tenkes at $M = \infty$). Velg en følge $\{x_n\}$ i K slik at $f(x_n) \to M$ (en slik følge finnes uansett om M er endelig eller uendelig). Siden K er begrenset, har

$\{x_n\}$ en konvergent delfølge $\{y_k\}$, og siden K er lukket, må grensepunktet $c = \lim_{k\to\infty} y_k$ ligge i K. Da må på den ene siden $f(y_k) \to M$, mens på den andre siden er $\lim f(y_k) = f(c)$ ifølge setning 2.3.2. Dermed er $f(c) = M$, og siden $M = \sup\{f(x) \mid x \in K\}$, er c et maksimumspunkt for f på K. $\square$.

4.4 Middelverdisetningen

Det siste teoremet i dette kapitlet er litt annerledes enn de andre siden det forutsetter at funksjonene er deriverbare og ikke bare kontinuerlige. Husk definisjonen av derivert

$$f'(a) = \lim_{x\to a} \frac{f(x) - f(a)}{x - a}$$

Funksjonen er deriverbar i a dersom grenseverdien til høyre eksisterer.

Vi trenger et par lemmaer:

Lemma 4.4.1 *Anta at $f : [a, b] \to \mathbb{R}$ har et maksimums- eller minimumspunkt i et indre punkt $c \in (a, b)$ der funksjonen er deriverbar. Da er $f'(c) = 0$.*

Bevis: Anta for motsigelse at $f'(c) > 0$. Siden

$$f'(c) = \lim_{x\to c} \frac{f(x) - f(c)}{x - c}$$

må $\frac{f(x)-f(c)}{x-c} > 0$ for alle x tilstrekkelig nær c. Hvis $x > c$, betyr det at $f(x) > f(c)$, og hvis $x < c$, betyr det at $f(x) < f(c)$. Følgelig kan c hverken være et minimums- eller et maksimumspunkt, motsigelse. På akkurat samme måte viser man at antagelsen $f'(c) < 0$ fører til en motsigelse, og dermed må $f'(c) = 0$. $\square$

Lemma 4.4.2 (Rolles teorem) *Anta at $f : [a, b] \to \mathbb{R}$ er kontinuerlig i hele $[a, b]$ og deriverbar i alle indre punkter $x \in (a, b)$. Anta videre at $f(a) = f(b)$. Da finnes det et punkt $c \in (a, b)$ der $f'(c) = 0$,*

Bevis: Hvis f er konstant, er $f'(x) = 0$ i alle indre punkter, og det er ingen ting å bevise. Ifølge ekstremalverdisetningen har funksjonen minimums- og maksimumspunkter, og hvis den ikke er konstant, må minst ett av disse være et indre punkt c. Ifølge foregående lemma er $f'(c) = 0$. $\square$

Da er vi klare til å bevise teoremet. Det sier at for deriverbare funksjoner finnes det i ethvert intervall et punkt der den øyeblikkelige veksten er lik gjennomsnittsveksten over intervallet. Setningen kalles også ofte for *sekantsetningen*

Teorem 4.4.3 (Middelverdisetningen) *Anta at $f : [a, b] \to \mathbb{R}$ er kontinuerlig i hele $[a, b]$ og deriverbar i alle indre punkter $x \in (a, b)$. Da finnes det et punkt $c \in (a, b)$ slik at*

$$f'(c) = \frac{f(b) - f(a)}{b - a}$$

Bevis: La g være funksjonen

$$g(x) = f(x) - \frac{f(b) - f(a)}{b - a}(x - a)$$

Ved innsetting ser vi at både $g(a)$ og $g(b)$ er lik $f(a)$, og ifølge Rolles teorem finnes det da et punkt $c \in (a, b)$ der $g'(c) = 0$. Deriverer vi og setter inn, ser vi at dette er det samme som at

$$f'(c) = \frac{f(b) - f(a)}{b - a}$$

$\square$

Oppgaver

1. Hvor brukes antagelsen om at f er kontinuerlig i beviset for skjæringssetningen?
2. Forklar hvorfor det finnes en følge $\{z_n\}$ i beviset for skjæringssetningen?.
3. Forklar hvorfor følgen $\{y_k\}$ i beviset for Setning 4.2.1 er en Cauchyfølge.
4. Forklar hvorfor det finnes en følge $\{x_n\}$ i beviset for ekstremalverdisetningen?
5. Anta at f og f' er kontinuerlige på intervallet $[a, b]$. Vis at det finnes en konstant M slik at $|f(x) - f(y)| \leq M|x - y|$ for alle $x, y \in [a, b]$.

Kapittel 5

Metriske rom

Vi har tidligere sett at begreper som konvergens og kontinuitet har med avstand å gjøre — at f er kontinuerlig i punktet a, betyr f. eks. at det for enhver $\epsilon > 0$, finnes en $\delta > 0$ slik at avstanden mellom $f(x)$ og $f(a)$ er mindre enn ϵ når avstanden mellom x og a er mindre enn δ. I dette kapitlet skal vi bruke denne observasjonem til å generalisere de grunnleggende begrepene i matematisk analyse til mer generelle rom, såkalte *metriske rom.*

5.1 Definisjon og eksempler

Et metrisk rom er bare en mengde X som er utstyrt med et "avstandsmål" (en *metrikk*) som forteller oss hvor stor avstanden er mellom punkter i X; vi skriver gjerne $d(x, y)$ for avstanden fra x til y, der x og y er to punkter i X. For å få en teori som virker, må vi kreve at metrikken d har egenskaper som ligner på de avstandsmålene vi er vant til fra før. Før vi ser på den formelle definisjonen av metriske rom, kan det være lurt å tenke gjennom hva slags egenskaper dette er.

Avstanden $d(x, y)$ fra x til y bør være et ikke-negativt tall, og den bør være 0 hvis $x = y$ og større enn 0 ellers. Vi bør med andre ord ha $d(x, y) \geq 0$ med likhet hvis og bare hvis $x = y$. Den neste egenskapen vi bør ha, er at avstanden fra x til y er lik avstanden fra y til x, dvs. at $d(x, y) = d(y, x)$. Man kan innvende at dette ikke alltid er tilfellet i virkeligheten; bruker vi f.eks. tiden det tar å gå fra x til y som mål på avstanden fra x til y, er det ikke sikkert at $d(x, y) = d(y, x)$ — er terrenget bratt, kan det hende at det tar lenger tid å gå den ene veien enn den andre. Vi velger likevel å holde oss til avstandsmål som er symmetriske, dvs. der $d(x, y) = d(y, x)$. Det siste kravet vi skal stille, er at hvis x, y, z er tre punkter i X, så er $d(x, y) \leq d(x, z) + d(z, y)$.

Denne formelen sier at avstanden direkte fra x til y er mindre enn eller lik den avstanden vi får ved å gå via et tredje punkt z.

La oss oppsummere disse egenskapene i en formell definisjon:

Definisjon 5.1.1 *Et metrisk rom* (X, d) *består av en mengde* X *og en funksjon* $d : X \times X \to [0, \infty)$ *slik at:*

(i) (positivitet) $d(x, y) = 0$ *hvis og bare hvis* $x = y$.

(ii) (symmetri) For alle $x, y \in X$ *er* $d(x, y) = d(y, x)$.

(iii) (trekantulikhet) For alle $x, y, z \in X$ *er* $d(x, y) \leq d(x, z) + d(z, y)$.

En funksjon d *som oppfyller kravene (i)-(iii), kalles en* metrikk *på* X.

Bemerkning: Når det er klart hvilken metrikk d vi har i tankene, skal vi ofte referere til "det metriske rommet X" istedenfor "det metriske rommet (X, d)".

La oss se på noen eksempler på metriske rom.

Eksempel 1: Lar vi $d(x, y) = |x - y|$, så er $(\mathbb{R}, d)$ et metrisk rom. De to første kravene til en metrikk er åpenbart oppfylt, og det tredje kravet følger fra den vanlige trekantulikheten på tallinjen:

$$d(x, y) = |x-y| = |(x-z)+(z-y)| \leq |x-z|+|z-y| = d(x, z)+d(z, y)$$

Eksempel 2: Lar vi $d(\mathbf{x}, \mathbf{y}) = |\mathbf{x} - \mathbf{y}|$, så er $(\mathbb{R}^n, d)$ et metrisk rom. De to første kravene til en metrikk er åpenbart oppfylt, og det tredje kravet følger fra den vanlige trekantulikheten for vektorer på samme måte som ovenfor:

$$d(\mathbf{x}, \mathbf{y}) = |\mathbf{x}-\mathbf{y}| = |(\mathbf{x}-\mathbf{z})+(\mathbf{z}-\mathbf{y})| \leq |\mathbf{x}-\mathbf{z}|+|\mathbf{z}-\mathbf{y}| = d(\mathbf{x}, \mathbf{z})+d(\mathbf{z}, \mathbf{y})$$

Eksempel 3: Anta at vi skal forflytte oss fra et punkt $\mathbf{x} = (x_1, x_2)$ i planet til et annet $\mathbf{y} = (y_1, y_2)$, men at vi bare kan bevege oss loddrett og vannrett. Vi kan da først flytte oss vannrett fra (x_1, x_2) til (y_1, x_2) og deretter loddrett fra (y_1, x_2) til (y_1, y_2) slik at den tilbakelagte avstanden blir

$$d(\mathbf{x}, \mathbf{y}) = |y_1 - x_1| + |y_2 - x_2|$$

Dette gir oss en metrikk på $\mathbb{R}^2$ som er forskjellig fra den vanlige metrikken i eksempel 2. Den kalles ofte *Manhattan-metrikken* fordi den måler de avstandene vi tilbakelegger i et rektangulært gatenett av den typen som ofte finnes i amerikanske byer.

Også i dette tilfellet er de to første kravene til en metrikk åpenbart oppfylt. For å vise trekantulikheten observerer vi at for et tredje punkt $\mathbf{z} = (z_1, z_2)$, er

$$d(\mathbf{x}, \mathbf{y}) = |y_1 - x_1| + |y_2 - x_1| =$$

$$= |(y_1 - z_1) + (z_1 - x_1)| + |(y_2 - z_2) + (z_2 - x_2)| \leq$$

$$\leq |y_1 - z_1| + |z_1 - x_1| + |y_2 - z_2| + |z_2 - x_2| =$$

$$= |z_1 - x_1| + |z_2 - x_2| + |y_1 - z_1| + |y_2 - z_2| =$$

$$= d(x, z) + d(z, y)$$

der vi har brukt den vanlige trekantulikheten på tallinjen for å komme fra den andre til den tredje linjen.

Eksempel 4: Vi skal nå se på et eksempel av en annen type. Anta at vi skal sende meldinger i et språk med N symboler (bokstaver, tall, skilletegn, mellomrom osv.) Vi antar at alle meldinger har den samme lengden K (hvis de er for korte eller for lange, blir de enten fylt ut med mellomrom eller sendt i flere omganger). Vi lar X være mengden av alle mulige meldinger, altså alle mulige sekvenser av symboler fra språket vårt med lengde K. Hvis $\mathbf{x} = (x_1, x_2, \ldots, x_K)$ og $\mathbf{y} = (y_1, y_2, \ldots, y_K)$ er to meldinger, lar vi

$$d(\mathbf{x}, \mathbf{y}) = \text{antall } n \text{ slik at } x_n \neq y_n$$

Det er ikke så vanskelig å sjekke at d er en metrikk. Den kalles gjerne *Hamming-metrikken* og brukes mye i kodeteori, blant annet som et mål på hvor mye en forvrengt melding avviker fra den opprinnelige meldingen.

Eksempel 5: Det er mange måter å måle avstander mellom funksjoner på, og i dette eksemplet skal vi se på noen. La X være mengden av alle kontinuerlige funksjoner $f : [a, b] \to \mathbb{R}$. Da er

$$d_1(f, g) = \sup\{|f(x) - g(x)| \;:\; x \in [a, b]\}$$

en metrikk på X. Denne metrikken måler avstanden mellom to funksjoner ved å se på de x-verdiene der avstanden mellom funksjonsgrafene er størst. Det betyr at to funksjoner kan ha stor avstand selv om de stort sett ligger nær hverandre. Metrikken

$$d_2(f, g) = \int_a^b |f(x) - g(x)| \, dx$$

legger større vekt på det gjennomsnittlige avviket mellom $f(x)$ og $g(x)$. En tredje metrikk som er mye brukt, er

$$d_3(f,g) = \left(\int_a^b |f(x) - g(x)|^2 \, dx\right)^{\frac{1}{2}}$$

Denne metrikken er en direkte generalisering av den vanlige (*euklidske*) metrikken i $\mathbb{R}^n$

$$d(\mathbf{x},\mathbf{y}) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} = \left(\sum_{i=1}^{n}(x_i - y_i)^2\right)^{\frac{1}{2}}$$

(tenk på integral som en generalisert sum). At vi har flere metrikker på X, betyr ikke at én av dem er "riktig" og de andre "gale", men at de er nyttige for ulike formål.

Eksempel 6: Dette eksemplet kan virke litt tullete, men det er greit å vite om fordi det gir oss en metrikk som er helt forskjellig fra dem vi er vant til i $\mathbb{R}^n$. La X være en hvilken som helst ikke-tom mengde, og definer

$$d(x,y) = \begin{cases} 0 & \text{if } x = y \\ \\ 1 & \text{if } x \neq y \end{cases}$$

Da er d en metrikk på X. Den kalles den *diskrete* metrikken og er hovedsakelig av interesse fordi den kan brukes til å lage moteksempler mot altfor raske generaliseringer.

Eksempel 7: Det er mange metoder for å lage nye metriske rom fra dem vi allerede har. Den enkleste er underromsmetrikken: Dersom (X,d) er et metrisk rom, og A er en delmengde av X, kan vi lage en metrikk d_A på A rett og slett ved å sette $d_A(x,y) = d(x,y)$ for alle $x, y \in A$ — vi restrikterer altså metrikken til A. Det er trivielt å sjekke at d_A er en metrikk på A. I praksis er det sjelden vi tar oss bryet med å skifte navn på metrikken når vi restrikterer til en delmengde: vi kaller d_A også for d, og bare underforstår at den nå virket på en mindre mengde.

Det finnes mange andre typer metriske rom enn de vi har sett på hittil, men eksemplene ovenfor gir forhåpentligvis et visst inntrykk av hvor mangeartet begrepet er. Da du lærte om funksjoner av flere variable, la du kanskje merke til at mange av argumentene var svært like de du tidligere hadde sett i én-variabelteorien, spesielt argumenter som hadde med kontinuitet og konvergens å gjøre. Ofte er argumentene så like at det nesten føles litt tullete å gjøre dem på nytt; den eneste som skiller det ene argumentet fra det andre, er at det foregår i en annen setting og med andre objekter. Noe av hensikten med teorien for metriske rom er å slippe slikt dobbeltarbeid — vi ønsker å lage en teori om gjelder for *alle* metriske rom slik at vi slipper å gå gjennom argumentene på nytt hver gang vi tar i bruk et nytt rom.

5.2 Konvergens

I resten av dette kapitlet skal vi få noen smakebiter på hvordan ideer fra $\mathbb{R}$ og $\mathbb{R}^n$ kan overføres til metriske rom. Vi begynner med konvergens. En følge $\{x_n\}$ i et metrisk rom X er bare en nummerert sekvens $\{x_1, x_2, x_3, \dots, x_n, \dots\}$ av elementer i X.

Definisjon 5.2.1 *La (X, d) være et metrisk rom. En følge $\{x_n\}$ i X* konvergerer mot *et punkt $a \in X$ dersom det til enhver $\epsilon > 0$ finnes en $N \in \mathbb{N}$ slik at $d(x_n, a) < \epsilon$ for alle $n \geq N$. Vi skriver $\lim_{n\to\infty} x_n = a$ eller $x_n \to a$.*

Legg merke til at denne definisjonen er en direkte oversettelse av den vi har brukt for følger i $\mathbb{R}$ og $\mathbb{R}^n$. Her er en alternativ variant av definisjonen.

Lemma 5.2.2 *En følge $\{x_n\}$ i et metrisk rom (X, d) konvergerer mot a hvis og bare hvis $\lim_{n\to\infty} d(x_n, a) = 0$.*

Bevis: Avstandene $\{d(x_n, a)\}$ danner en følge av ikke-negative tall. Denne følgen konvergerer mot 0 hvis og bare hvis det for hver $\epsilon > 0$, finnes en $N \in \mathbb{N}$ slik at $d(x_n, a) < \epsilon$ når $n \geq N$. Men det er nøyaktig hva definisjonen ovenfor sier. □

Kan en følge konvergere mot mer enn ett punkt? Vi vet at den ikke kan det i $\mathbb{R}^n$, men noen av disse nye metriske rommene er så uvante at man ikke kan være sikker uten å føre et bevis.

Setning 5.2.3 *En følge i et metrisk rom kan ikke konvergere mot mer enn ett punkt.*

Bevis: Anta $\lim_{n\to\infty} x_n = a$ og $\lim_{n\to\infty} x_n = b$. Vi må vise at dette bare er mulig dersom $a = b$. Ifølge trekantulikheten er

$$d(a, b) \leq d(a, x_n) + d(x_n, b)$$

Tar vi grenseverdier, har vi dermed

$$d(a,b) \leq \lim_{n\to\infty} d(a,x_n) + \lim_{n\to\infty} d(x_n,b) = 0 + 0 = 0$$

Det betyr at $d(a,b) = 0$, og ifølge punkt (i) i definisjonen av metriske rom, er $a = b$. □

Legg merke til hvordan vi har brukt punktene i definisjonen av metriske rom i dette beviset. Det er ikke så rart; foreløpig oppsummerer disse punktene alt vi vet om metriske rom. Etter hvert som vi beviser resultater, får vi flere og flere verktøy å arbeide med.

Det er lett å generalisere begrepet Cauchy-følge til metriske rom.

Definisjon 5.2.4 *En følge* $\{x_n\}$ *i et metrisk rom* (X,d) *er en* Cauchy-følge *dersom det til enhver* $\epsilon > 0$ *finnes en* $N \in \mathbb{N}$ *slik at* $d(x_n,x_m) < \epsilon$ *når* $n, m \geq N$

Setning 5.2.5 *Enhver konvergent følge er en Cauchy-følge.*

Bevis: Hvis a er grensen til følgen, vil det for en gitt $\epsilon < 0$, finnes en $N \in \mathbb{N}$ slik at $d(x_n,a) < \frac{\epsilon}{2}$ når $n \geq N$. Hvis $n, m \geq N$ har vi da ved trekantulikheten

$$d(x_n,x_m) \leq d(x_n,a) + d(a,x_m) < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

og følgelig er $\{x_n\}$ en Cauchy-følge. □

Den omvendte implikasjonen er ikke alltid oppfylt, så vi innfører følgende begrep:

Definisjon 5.2.6 *Et metrisk rom kalles* komplett *dersom alle Cauchy-følger konvergerer.*

Vi vet at $\mathbb{R}^n$ er komplett, men at $\mathbb{Q}$ ikke er komplett dersom vi bruker den vanlige avstandsmetrikken $d(x,y) = |x-y|$. Kompletthet vil være et av de mest sentrale begrepene i MAT2400. De komplette rommene er på mange måter de "snille" metriske rommene, og vi skal bruke mye tid på å studere egenskapene til slike rom. Vi skal også bruke en del tid på å se hvordan vi kan gjøre ikke-komplette rom komplette ved å legge til flere elementer. Viktige tilfeller av dette finner du i eksempel 5 i forrige seksjon (der X er rommet av alle kontinuerlige funksjoner $f : [a,b] \to \mathbb{R}$); her er X komplett hvis man bruker metrikken d_1, men ikke komplett hvis man bruker metrikkene d_2 og d_3. Ved å innføre et mer generelt integralbegrep (Lebesgue-integralet) kan man imidlertid gjøre metrikkene d_2 og d_3 komplette ved å la dem virke på et rikere rom av integrerbare funksjoner.

5.3 Kontinuitet

Kontinuitet er et annet begrep som er lett å generalisere til metriske rom.

Definisjon 5.3.1 *Anta at* (X, d), (Y, ρ) *er to metriske rom. En funksjon* $f : X \to Y$ *er* kontinuerlig *i punktet* $a \in X$ *dersom det til enhver* $\epsilon > 0$ *finnes en* $\delta > 0$ *slik at* $\rho(f(x), f(a)) < \epsilon$ *når* $d(x, a) < \delta$.

Denne definisjonen sier nøyaktig det samme som våre tidligere definisjoner, nemlig at vi kan få avstanden mellom $f(x)$ og $f(a)$ til å bli mindre enn ϵ ved å velge x slik at avstanden mellom x og a er mindre enn δ. Den eneste forskjellen er at vi nå bruker metrikkene d og ρ til å måle avstandene.

Vi har den samme koblingen mellom kontinuitet og konvergens som vi har hatt tidligere:

Setning 5.3.2 *En funksjon* $f : X \to Y$ *mellom to metriske rom er kontinuerlig i et punkt* a *hvis og bare hvis* $f(x_n) \to f(a)$ *for alle følger* $\{x_n\}$ *som konvergerer mot* a.

Proof: Anta først at f er kontinuerlig i a og at $x_n \to a$. For å vise at $f(x_n) \to f(a)$, må vi vise at for enhver $\epsilon > 0$ finnes det en en $N \in \mathbb{N}$ slik at $\rho(f(x_n), f(a)) < \epsilon$ for alle $n \geq N$. Siden f er kontinuerlig i a, finnes det en $\delta > 0$ slik at hvis $d(x, a) < \delta$, så er $\rho(f(x), f(a)) < \epsilon$. Siden $x_n \to a$, finnes det en $N \in \mathbb{N}$ slik at $d(x_n, a) < \delta$ når $n \geq N$. Men det betyr at når $n \geq N$, så er $\rho(f(x_n), f(a)) < \epsilon$, akkurat som vi skulle vise.

Anta så at f *ikke* er kontinuerlig i a. Da må det finnes en $\epsilon > 0$ som ikke kan pareres, dvs. som er slik at uansett hvor liten vi velger $\delta > 0$, så finnes det et punkt x slik at $d(x, a) < \delta$, men $\rho(f(x), f(a)) \geq \epsilon$. Spesielt vil det for $\delta = \frac{1}{n}$ finnes et punkt x_n slik at $d(x_n, a) < \frac{1}{n}$, men $\rho(f(x_n), f(a)) \geq \epsilon$. Da konvergerer $\{x_n\}$ mot a, mens $\{f(x_n)\}$ ikke konvergerer mot $f(a)$ (siden $\rho(f(x_n), f(a)) \geq \epsilon$, kan ikke $\rho(f(x_n), f(a))$ gå mot 0). $\square$

Vi skal nå se på et viktig teorem som forbinder en viss klasse kontinuerlige funksjoner med kompletthet. Teoremet har flere anvendelser i MAT2400 (og vil bli bevist på nytt der). Først litt terminologi:

Anta at (X, d) er et metrisk rom. En funksjon $f : X \to X$ kalles en *kontraksjon* dersom det finnes et positivt tall $s < 1$ slik at

$$d(f(x), f(y)) \leq s\, d(x, y) \quad \text{for alle } x, y \in X$$

Vi kaller s en *kontraksjonsfaktor* for f. Alle kontraksjoner er kontinuerlige (bevis det!), og ved induksjon er det lett å se at

$$d(f^{\circ n}(x), f^{\circ n}(y)) \leq s^n d(x,y)$$

der $f^{\circ n}(x) = f(f(f(\ldots f(x))))$ er resultatet av å iterere f n ganger. Vi sier at a er et *fikspunkt* for f dersom $f(a) = a$.

Resultatet vi skal vise, er en generalisering av et resultat du kanskje husker fra MAT1110.

Teorem 5.3.3 (Banachs fikspunktteorem) *Anta at (X,d) er et komplett metrisk rom og at $f : X \to X$ er en kontraksjon. Da har f et entydig fikspunkt a, og uansett hvilket startpunkt $x_0 \in X$ vi velger, vil følgen*

$$x_0,\ x_1 = f(x_0),\ x_2 = f^{\circ 2}(x_0), \ldots,\ x_n = f^{\circ n}(x_0), \ldots$$

konvergere mot a.

Bevis: La oss først vise at f ikke kan ha mer enn ett fikspunkt. Hvis a og b er to fikspunkter, og s er en kontraksjonsfaktor for f, har vi

$$d(a,b) = d(f(a), f(b)) \leq s\, d(a,b)$$

Siden $0 < s < 1$ er dette bare mulig når $d(a,b) = 0$, dvs. når $a = b$.

For å vise at f har et fikspunkt, velger vi et startpunkt x_0 i X og ser på følgen

$$x_0,\ x_1 = f(x_0),\ x_2 = f^{\circ 2}(x_0), \ldots,\ x_n = f^{\circ n}(x_0), \ldots$$

Anta at vi kan vise at dette er en Cauchy-følge. Siden (X,d) er komplett, vil følgen konvergere mot et punkt a. Tar vi grenseverdien på begge sider av ligningen $x_{n+1} = f(x_n)$, får vi $a = f(a)$, så a er et fikspunkt til f, og teoremet er bevist. Det gjenstår derfor bare å vise at $\{x_n\}$ virkelig er en Cauchy-følge.

Velg to elementer x_n og x_{n+k} i følgen. Ved gjentatt bruk av trekantulikheten, er

$$d(x_n, x_{n+k}) \leq d(x_n, x_{n+1}) + d(x_{n+1}, x_{n+2}) + \ldots + d(x_{n+k-1}, x_{n+k}) =$$

$$= d(f^{\circ n}(x_0), f^{\circ n}(x_1)) + d(f^{\circ(n+1)}(x_0), f^{\circ(n+1)}(x_1)) + \ldots$$

$$\ldots + d(f^{\circ(n+k-1)}(x_0), f^{\circ(n+k-1)}(x_1)) \leq$$

$$\leq s^n d(x_0, x_1) + s^{n+1} d(x_0, x_1) + \ldots + s^{n+k-1} d(x_0, x_1) =$$

$$= \frac{s^n(1-s^k)}{1-s}\, d(x_0, x_1) \leq \frac{s^n}{1-s}\, d(x_0, x_1)$$

der vi blant annet har summert en geometrisk rekke for å komme fra den nest siste til den siste linjen. Siden $s < 1$, kan vi få det siste uttrykket så lite vi måtte ønske ved å velge n stor nok. Gitt en $\epsilon > 0$, kan vi derfor finne en N slik at $\frac{s^N}{1-s}\, d(x_0, x_1) < \epsilon$. For $n, m = n + k$ større enn eller lik N er dermed

$$d(x_n, x_m) \leq \frac{s^n}{1-s}\, d(x_0, x_1) < \epsilon$$

og følgelig er $\{x_n\}$ en Cauchy-følge. □

5.4 Åpne, lukkede og kompakte mengder

I MAT1110 studerte vi åpne og lukkede mengder i $\mathbb{R}^n$. Siden disse begrepene kan defineres ved hjelp av avstand, er de lette å generalisere til metriske rom. Vi begynner med å definere kuler:

Definisjon 5.4.1 *La a være et punkt i et metrisk rom (X, d), og anta at r er et positivt reelt tall. Da er* kulen om a med radius r *mengden*

$$\mathrm{B}(a; r) = \{x \in X \,:\, d(x, a) < r\}$$

Kulene ovenfor oppfører seg på mange måter som kulene vi er vant til, men geometrisk kan de se ganske annerledes ut. En kule i Manhattan-metrikken (eksempel 3 i seksjon 1) ser ut som et ruteress, mens kulene i den diskrete metrikken (eksempel 6 i seksjon 1) består enten bare av ett punkt eller av hele rommet.

Anta at A er en delmengde av X, og at at x er et punkt i X. Da er det tre muligheter:

(i) Det finnes en kule $\mathrm{B}(x; r)$ om x som er inneholdt i A. I så fall kalles x et *indre punkt* til A.

(ii) Det finnes en kule $\mathrm{B}(x; r)$ om x som er inneholdt i komplementet A^c. I så fall kalles x et *ytre punkt* til A.

(iii) Alle kulene $\mathrm{B}(x; r)$ om x inneholder både punkter som er med i A og punkter som er med i A^c. I så fall kalles x et *randpunkt* til A.

Legg merke til at et indre punkt *alltid* er med i A, et ytre punkt *aldri* er med i A, mens et randpunkt noen ganger vil være med og andre ganger ikke.

Vi kan nå definere åpne og lukkede mengder på samme måte som i $\mathbb{R}^n$:

Definisjon 5.4.2 *En mengde er* åpen *hvis den ikke inneholder noen av sine randpunkter, og den er* lukket *hvis den inneholder alle sine randpunkter.*

De fleste mengder inneholder noen av sine randpunkter, men ikke alle, og er dermed hverken åpne eller lukkede. Vi trenger følgende enkle observasjon:

Lemma 5.4.3 *Anta at F er en lukket delmengde av et metrisk rom, og at $\{x_n\}$ er en konvergent følge av elementer i F. Da ligger grensepunktet $a = \lim_{n\to\infty} x_n$ i F.*

Bevis: Anta (for motsigelse) at a ikke ligger i F. Siden F er lukket og inneholder alle sine randpunkter, må a være et ytre punkt, og følgelig finnes det en kule $\mathrm{B}(a;\epsilon)$ som bare inneholder punkter fra komplementet til F. Det betyr at $\mathrm{B}(a;\epsilon)$ ikke inneholder noen av elementene i følgen, og følgelig er $d(a, x_n) \geq \epsilon$ for alle n, i strid med antagelsen om at $\{x_n\}$ konvergerer mot a. □

Fra MAT1110 og kapittel 4 husker du sikkert Bolzano-Weierstrass' teorem som sier at alle begrensede delfølger i $\mathbb{R}^n$ har en konvergent delfølge. Det er utgangspunktet for følgende definisjon (husk at en delfølge er en ny følge som vi får ved å plukke ut noen av leddene i den opprinnelige følgen).

Definisjon 5.4.4 *En delmengde K av et metrisk rom (X, d) kalles* kompakt *dersom enhver følge i K har en delfølge som konvergerer mot et punkt i K. Rommet (X, d) kalles* kompakt *dersom X er en kompakt mengde.*

Kompakthet er et svært nyttig begrep som kommer til å stå sentralt i MAT2400. Her kan vi bare gi en liten smakebit på hva begrepet kan brukes til. Før vi går igang, trenger vi en definisjon til.

Definisjon 5.4.5 *En delmengde A av et metrisk rom (X, d) er* begrenset *dersom det finnes et punkt $b \in X$ og en konstant $K \in \mathbb{R}$ slik at $d(a, b) \leq K$ for alle $a \in A$ (det spiller ingen rolle hvilket punkt $b \in X$ vi bruker i denne definisjonen).*

Her er vårt første resultat om kompakte mengder.

Setning 5.4.6 *Enhver kompakt mengde K i et metrisk rom (X, d) er lukket og begrenset.*

Bevis: Vi resonnerer kontrapositivt. Først skal vi vise at dersom en mengde K ikke er lukket, så er den heller ikke kompakt. Deretter viser vi at dersom en mengde K ikke er begrenset, så er den ikke kompakt.

Anta at K ikke er lukket, da finnes det et randpunkt a som ikke er med i K. For hver $n \in \mathbb{N}$ finnes det derfor en $x_n \in K$ slik at $d(x_n, a) < \frac{1}{n}$. Følgen $\{x_n\}$ konvergerer mot a, og det samme vil enhver delfølge gjøre. Men a ligger ikke i K, og følgelig kan ingen delfølge av $\{x_n\}$ konvergere mot et element i K. Dermed kan ikke K være kompakt.

Anta så at K ikke er begrenset. For enhver $n \in \mathbb{N}$ finnes det da et element $x_n \in K$ slik at $d(x_n, b) > n$. Hvis $\{y_n\}$ er en delfølge av x_n, er da $\lim_{n\to\infty} d(y_n, b) = \infty$. Det er lett å se at $\{y_n\}$ ikke kan konvergere mot noe element $y \in X$. Ifølge trekantulikheten er nemlig

$$d(y_n, b) \leq d(y_n, y) + d(y, b)$$

og siden $d(y_n, b) \to \infty$, må $d(y_n, y) \to \infty$. Altså har $\{x_n\}$ ingen konvergente delfølger, og følgelig er ikke K kompakt. □

I $\mathbb{R}^n$ har setningen ovenfor en omvending som sier at lukkede, begrensede mengder er kompakte (dette er bare en omformulering av Bolzano-Weierstrass' teorem). Denne omvendingen gjelder ikke generelt i metriske rom, og én av utfordringene i MAT2400 blir å styrke begrensethetsbegrepet (til noe som kalles "totalt begrensethet") slik at vi får en beskrivelse av kompakte mengder som gjelder for alle metriske rom.

Hva er sammenhengen mellom de to viktige begepene kompakthet og kompletthet? Komplette rom behøver ikke å være kompakte ($\mathbb{R}$ er komplett, men ikke kompakt), men omvendt har vi:

Setning 5.4.7 *Ethvert kompakt metrisk rom er komplett.*

Bevisskisse: Anta at $\{x_n\}$ er en Cauchy-følge. Siden X er kompakt, finnes det en delfølge $\{y_k\}$ som konvergerer mot et punkt c i X. Siden den opprinnelige følgen $\{x_n\}$ er Cauchy, vil den også konvergere mot c (bevis det!), og følgelig er X komplett. □

La oss avslutte med et resultat som antyder hvorfor kompakte mengder er viktige.

Teorem 5.4.8 (Ekstremalverdisetningen) *Anta at K er en kompakt delmengde av et metrisk rom (X, d) og at $f : K \to \mathbb{R}$ er en* *kontinuerlig funksjon. Da har f maksimums- og minimumspunkter i K, dvs. det finnes punkter $c, d \in K$ slik at*

$$f(d) \leq f(x) \leq f(c)$$

for alle $x \in K$.

Bevis: Jeg viser at f har et maksimumspunkt; argumentet for at f har et minimumspunkt går på akkurat samme måte.

La

$$M = \sup\{f(x) \mid x \in K\}$$

(siden vi ikke vet om f er begrenset, kan det godt tenkes at $M = \infty$). Velg en følge $\{x_n\}$ i K slik at $f(x_n) \to M$ (en slik følge finnes uansett om M er endelig eller uendelig). Siden K er kompakt, har $\{x_n\}$ en delfølge $\{y_k\}$ som konvergerer mot et punkt $c \in K$. Da må på den ene siden $f(y_k) \to M$, mens på den andre siden er $\lim f(y_k) = f(c)$ ifølge setning 5.3.2. Dermed er $f(c) = M$, og siden $M = \sup\{f(x) \mid x \in K\}$, er c et maksimumspunkt for f på K. $\Box$.

Oppgaver

1. Vis at (X, d) i eksempel 4 er et metrisk rom.
2. Vis at d_1 i eksempel 5 er en metrikk.
3. Vis at d_2 i eksempel 5 er en metrikk.
4. Vis at d i eksempel 6 er en metrikk.
5. Anta at d_1 og d_2 er to metrikker på X. Vis at

   $$d(x, y) = d_1(x, y) + d_2(x, y)$$

   er en metrikk på X.
6. Vis at hvilket punkt b vi bruker i definisjon 5.4.5 ikke spiller noen rolle.
7. Anta at A er en delmengde av et komplett metrisk rom (X, d). Vis at (A, d) er komplett hvis og bare hvis A er lukket.
8. Bruk Bolzano-Weierstass' teorem til å vise at alle lukkede, begrensede delmengder av $\mathbb{R}^n$ er kompakte.
9. Vis at hvis en delfølge av en Cauchy-følge konvergerer til et punkt c, så konvergerer også Cauchy-følgen selv til c.
10. (Utfordrende) Vis at en funksjon $f : X \to Y$ er kontinuerlig hvis og bare hvis $f^{-1}(O)$ er åpen for alle åpne mengder $O \subset Y$.