Kapittel 2

Funksjoner fra $\mathbb{R}^n$ til $\mathbb{R}^m$

I dette kapitlet skal vi studere funksjoner $\mathbf{F}$ fra $\mathbb{R}^n$ til $\mathbb{R}^m$. En slik funksjon er bare en regel som til hvert n-tuppel $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ tilordner et m-tuppel $\mathbf{y} = \mathbf{F}(\mathbf{x})$. Dersom $n > 1$ kaller vi $\mathbf{F}$ en *funksjon av flere variable.* Funksjoner av flere variable har mange likhetstrekk med de funksjonene av én variabel som du kjenner fra før, men de har også en del nye og litt uvante egenskaper. Spesielt blir geometrien mer komplisert når vi får mer enn én variabel å arbeide med.

2.1 Funksjoner av flere variable

Som allerede nevnt er en funksjon ~~funksjon~~ $\mathbf{F}$ fra $\mathbb{R}^n$ til $\mathbb{R}^m$ bare en regel som til hver $\mathbf{x} \in \mathbb{R}^n$ gir oss en $\mathbf{y} = \mathbf{F}(\mathbf{x})$ i $\mathbb{R}^m$. Ofte er disse reglene gitt ved formler, f.eks. kan en funksjon fra $\mathbb{R}^4$ til $\mathbb{R}^3$ være gitt ved

$$\mathbf{F}(x, y, z, u) = (2x^2 z + u, xy^2 zu^3, ye^{x^2 y + u \sin y})$$

Gitt et 4-tuppel $\mathbf{x} = (x, y, z, u)$ forteller denne formelen oss hvordan vi kan regne ut 3-tuplet $\mathbf{F}(\mathbf{x})$. Som du ser, veksler vi mellom skrivemåtene $\mathbf{F}(\mathbf{x})$ og $\mathbf{F}(x_1, x_2, \ldots, x_n)$; den første er ofte mest praktisk når vi snakker om en generell, uspesifisert funksjon, mens den siste som regel er greiest når vi snakker om en bestemt funksjon gitt ved en formel. Husk at $\mathbb{R}^1 = \mathbb{R}$, slik at teorien vår også dekker funksjoner $F : \mathbb{R}^n \to \mathbb{R}$ der $\mathbf{x}$ er en vektor, men $F(\mathbf{x})$ er et tall (i prinsippet dekker teorien også funksjoner av typen $\mathbf{F} : \mathbb{R} \to \mathbb{R}^m$, men for øyeblikket er vi ikke så interessert i dem).

Vi husker at vanlige funksjoner $f(x)$ ikke alltid er definert for alle reelle tall x, og på samme måte vil heller ikke disse nye funksjonene $\mathbf{F}(\mathbf{x})$ nødvendigvis være definert for alle vektorer $\mathbf{x} \in \mathbb{R}^n$. Den mengden $A \subset \mathbb{R}^n$ av vektorer $\mathbf{x}$ som $\mathbf{F}(\mathbf{x})$ er definert for, kaller vi *definisjonsmengden* eller *definisjonsområdet* til $\mathbf{F}$. Vi skal av og til bruke symbolet $D_{\mathbf{F}}$ for definisjonsområdet til $\mathbf{F}$. Når vi skriver

$$\mathbf{F} : A \to \mathbb{R}^m$$

$\mathbb{R}^n$ $\mathbb{R}^m$

mener vi at $\mathbf{F}$ er en funksjon definert på mengden A med verdier i $\mathbb{R}^m$. Hvis en funksjon er gitt ved formler, og definisjonsområdet ikke er spesifisert, regner vi med at funksjonen er definert der alle formlene gir mening.

En funksjon definert på en delmengde av $\mathbb{R}^n$ kaller vi en funksjon av n-variable; f.eks. er

$$\mathbf{G}(x, y) = (x^y, \ln(1 - xy), \sin(x^2 y))$$

en funksjon av to variable med verdier i $\mathbb{R}^3$, mens

$$g(x, y, z) = x^2 + ye^{z+y^2}$$

er en funksjon av tre variable med verdier i $\mathbb{R}$ (når funksjonene tar verdier i $\mathbb{R}$ skriver vi dem gjerne uten fete typer og ofte med små bokstaver).

En funksjon som tar verdier i $\mathbb{R}$ kalles gjerne et *skalarfelt* når vi vil understreke at verdiene er tall og ikke vektorer. Legg merke til at en funksjon $\mathbf{F} : \mathbb{R}^n \to \mathbb{R}^m$ kan skrives

$$\mathbf{F}(x_1, x_2, \ldots, x_n) = (F_1(x_1, x_2, \ldots, x_n), \ldots, F_m(x_1, x_2, \ldots, x_n))$$

der $F_1, \ldots, F_m$ er skalarfelt. Vi kaller $F_1, \ldots, F_m$ for *komponentene* til $\mathbf{F}$. Funksjonen $\mathbf{G}$ ovenfor har altså komponentene

$$G_1(x, y) = x^y, \quad G_2(x, y) = \ln(1 - xy), \quad G_3(x, y) = \sin(x^2 y)$$

Vi skal ofte gjøre bruk av at funksjoner med verdier i $\mathbb{R}^m$ er bygget opp av skalarfelt på denne måten, for eksempel ved at vi først innfører begreper og beviser resultater for skalarfelt og så utvider til funksjoner med verdier i $\mathbb{R}^m$.

Eksempler

Det er naturlig å spørre seg selv om *hvorfor* man må studere funksjoner av flere variable — greier det seg ikke med de funksjonene av bare én variabel som man kjenner fra før? Det viser seg imidlertid at funksjoner av flere variable dukker naturlig opp i svært mange sammenhenger. Her er et lite utvalg:

- BMI (body mass index) brukes ofte som en indikator på overvekt og undervekt. For å finne din BMI, tar du vekten din v (målt i kilo) og deler på kvadratet av høyden din h (målt i meter). Vi kan tenke på BMI som en funksjon f av to variable med verdier i $\mathbb{R}$:

$$f(v, h) = \frac{v}{h^2}$$

- Når en gjenstand varmes opp, vil temperaturen avhenge av når og hvor vi måler. Det er naturlig å angi temperaturen i punktet (x, y, z) ved tiden t som $T(x, y, z, t)$. Dette er en funksjon av fire variable med verdier i $\mathbb{R}$.

- Gravitasjonskraften fra jorden på en gjenstand i verdensrommet avhenger av posisjonen til gjenstanden. Tenker vi på kraften som en vektor $\mathbf{F}(\mathbf{x})$ (vi er interessert i både størrelse og retning) som er avhengig av posisjonen $\mathbf{x} = (x, y, z)$, får vi en funksjon av tre variable med verdier i $\mathbb{R}^3$. Dersom vi plasserer jorden i origo, kan $\mathbf{F}$ i dette tilfellet skrives

$$\mathbf{F}(\mathbf{x}) = -\gamma \frac{Mm}{|\mathbf{x}|^3}\,\mathbf{x}$$

$-\frac{GMm}{|x|^2} \frac{\vec{x}}{|x|}$

der γ er en naturkonstant (gravitasjonskonstanten), M er massen til jorden og m massen til gjenstanden. I dette eksemplet er $\mathbf{F}$ en funksjon fra $\mathbb{R}^3$ til $\mathbb{R}^3$

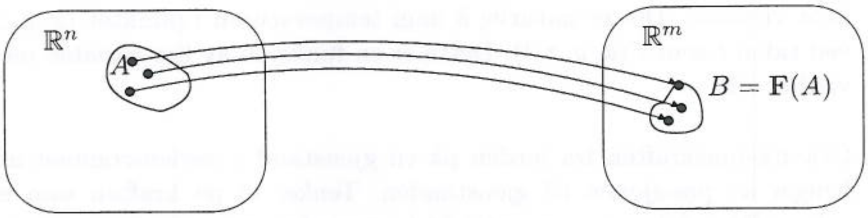
- Meteorologer arbeider med vind i atmosfæren. Vinden i et punkt med koordinater (x, y, z) ved tiden t vil være en vektor $\mathbf{v}(x, y, z, t)$ (vi er interessert i både vindstyrken og retningen). Vi kan tenke på dette som en funksjon av fire variable med verdier i $\mathbb{R}^3$.

- I eksempel 3 i seksjon 1.5 så vi på fire produsenter som leverte epler av tre ulike kvaliteter til en fruktpresse. Dersom produsentene leverte hhv. x, y, z og u tonn epler, kunne vi regne ut hvor mange tonn vi ville få av hver kvalitet. Dette gir oss en funksjon av fire variable med verdier i $\mathbb{R}^3$. Vi kan skrive funksjonen som $\mathbf{F}(\mathbf{x}) = A\mathbf{x}$, eller med koordinater

$$\mathbf{F}(x, y, z, u) = \begin{pmatrix} 0.5x + 0.3y + 0.25z + 0.2u \\ 0.3x + 0.4y + 0.4z + 0.6u \\ 0.2x + 0.3y + 0.35z + 0.2u \end{pmatrix}$$

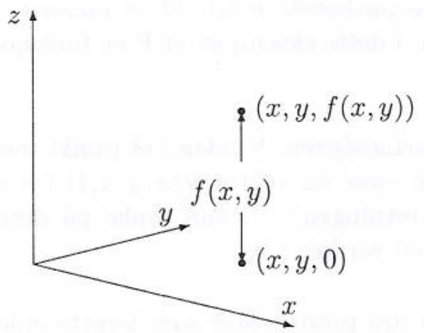
Grafisk fremstilling

Ett av de viktigste verktøyene når man studerer funksjoner av én variabel er grafisk fremstilling. Dessverre er det ikke så lett å gi realistiske grafiske fremstillinger av funksjoner fra $\mathbb{R}^n$ til $\mathbb{R}^m$. Noen ganger kan vi imidlertid ha stor glede av mer stiliserte fremstillinger som på figuren nedenfor. Den illustrerer hvordan en funksjon $\mathbf{F}$ fra $\mathbb{R}^n$ til $\mathbb{R}^m$ sender alle punkter i en mengde $A \subset \mathbb{R}^n$ på en mengde $B \subset \mathbb{R}^m$. Mengden B kalles *bildet av A under F* og betegnes ofte med $B = \mathbf{F}(A)$. Du vil se mange figurer av denne typen utover i kapitlet.



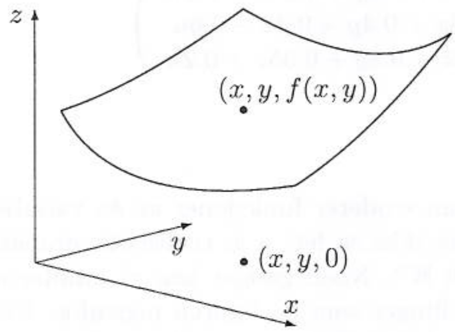
Figur 1: Funksjon fra $\mathbb{R}^n$ til $\mathbb{R}^m$

Det er imidlertid ett tilfelle der det går an å lage gode, realistiske fremstillinger av en funksjon av flere variable, og det er når funksjonen går fra $\mathbb{R}^2$ til $\mathbb{R}$. Vi ser altså på en funksjon $z = f(x, y)$. For å tegne funksjonsgrafen lager vi først et tre-dimensjonalt koordinatsystem som vist på figur 2.



Figur 2: Plotting av skalarfelt

Gitt variabelverdier x og y, finner vi punktet $(x, y, 0)$ i xy-planet. Vi flytter oss nå loddrett (dvs. parallelt med z-aksen) til vi finner punktet $(x, y, f(x, y))$. Dette er det første punktet på funksjonsgrafen vår.



Figur 3: Grafisk fremstilling av skalarfelt

Gjentar vi denne prosedyren for stadig flere variabelverdier (x, y), vokser grafen etterhvert frem som en flate i rommet (se figur 3).

Selv om denne prosedyren på en grei måte forklarer hva grafen til et skalarfelt er, så er den i praksis ubrukelig som en oppskrift på hvordan man tegner grafen. Prøver du den, selv på en enkel funksjon, oppdager du fort at du helt mister romfølelsen i bildet. I neste kapittel skal vi se på mer praktiske metoder for å tegne slike funksjonsgrafer — det som er viktig i dette kapitlet, er at du vet hvordan du kan tenke på grafen til f som en flate.

MATLAB-kommentar: MATLAB er et utmerket verktøy for grafisk fremstilling av funksjoner av flere variable, men vi utsetter dette temaet til seksjon 3.7 og 3.8.

Oppgaver til seksjon 2.1

1. Finn definisjonsområdet til funksjonen:

a) $f(x,y) = \frac{1}{x^2+4y^2}$
b) $f(x,y) = \frac{1}{x^2-y^2}$
c) $f(x,y) = \ln(x+y)$
d) $f(x,y) = \tan(x-y)$
e) $f(x,y,z) = \frac{1}{x^2+y^2+z^2-25}$

2.2 Kontinuerlige funksjoner

I seksjon 5.1 i *Kalkulus* studerte vi ϵ-δ-definisjonen av kontinuitet. Denne definisjonen er sannsynligvis ikke like populær blant alle, men den har mange fordeler, blant annet at den lett kan generaliseres til nye situasjoner. I denne seksjonen skal vi se hvordan den kan generaliseres til funksjoner av flere variable. Før vi begynner, trenger vi å vite litt om avstander og kuler i $\mathbb{R}^n$.

Akkurat som i planet og rommet lar vi *avstanden* mellom to punkter $\mathbf{a}$ og $\mathbf{x}$ i $\mathbb{R}^n$ være lik lengden til vektoren $\mathbf{x}-\mathbf{a}$ som forbinder dem, det vil si at avstanden er

$$|\mathbf{x}-\mathbf{a}| = \sqrt{(x_1-a_1)^2+(x_2-a_2)^2+\cdots+(x_n-a_n)^2}$$

Mengden

$$B(\mathbf{a},r) = \{\mathbf{x}\in\mathbb{R}^n \ : \ |\mathbf{x}-\mathbf{a}|<r\}$$

B for ball!

består av de punktene i $\mathbb{R}^n$ som har avstand mindre enn r til punktet $\mathbf{a}$. Vi kaller $B(\mathbf{a},r)$ *kulen om* $\mathbf{a}$ *med radius* r. Legg merke til at i $\mathbb{R}^3$ er dette virkelig en (åpen) kule i tradisjonell forstand, mens det i $\mathbb{R}^2$ er en (åpen) sirkelskive og i $\mathbb{R}$ et åpent intervall. Vi velger å bruke "kule" som et fellesord i alle dimensjoner, selv om det til å begynne med kan virke litt uvant når vi arbeider i planet eller på tallinjen. De fleste illustrasjonene våre vil være i planet, og der vil kuler fremstå som sirkler.

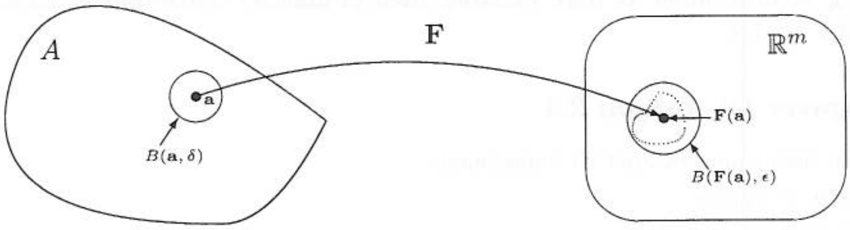
Vi kan nå definere kontinuitet for funksjoner av flere variable:

for

Definisjon 2.2.1 *Anta at $A \subset \mathbb{R}^n$, og at $\mathbf{a} \in A$. En funksjon $\mathbf{F} : A \to \mathbb{R}^m$ er* kontinuerlig *i* $\mathbf{a}$ *dersom det til enhver $\epsilon > 0$ finnes en $\delta > 0$ slik at*

$$|\mathbf{F}(\mathbf{x}) - \mathbf{F}(\mathbf{a})| < \epsilon \text{ for alle } \mathbf{x} \in A \text{ slik at } |\mathbf{x} - \mathbf{a}| < \delta$$

Figuren nedenfor illustrerer definisjonen: Gitt en kule $B(\mathbf{F}(\mathbf{a}), \epsilon)$ om punktet $\mathbf{F}(\mathbf{a})$, kan vi finne en kule $B(\mathbf{a}, \delta)$ om punktet $\mathbf{a}$ slik at bildet av $B(\mathbf{a}, \delta)$ (markert med den stiplede kurven på figuren) ligger helt inni $B(\mathbf{F}(\mathbf{a}), \epsilon)$.



Figur 1: Kontinuitet i punktet $\mathbf{a}$

Siden kontinuitet er definert på akkurat samme måte som for funksjoner av én variabel, har vi de samme reglene med (nesten) de samme bevisene.

F/G?

Setning 2.2.2 *Anta at $A \subset \mathbb{R}^n$, og at funksjonene $\mathbf{F}, \mathbf{G} : A \to \mathbb{R}^m$ er kontinuerlige i $\mathbf{a} \in A$. Da er $\mathbf{F} + \mathbf{G}$, $\mathbf{F} - \mathbf{G}$ og $\mathbf{F} \cdot \mathbf{G}$ kontinuerlige i $\mathbf{a}$. Det er også $\frac{\mathbf{F}}{\mathbf{G}}$, forutsatt at $\mathbf{F}$ og $\mathbf{G}$ tar verdier i $\mathbb{R}$ (slik at divisjon gir mening) og $\mathbf{G}(\mathbf{a}) \neq 0$.*

Bevis: Bevisene er akkurat som for funksjoner av én variabel (se setning 5.1.5 i *Kalkulus*), den eneste ekstra komplikasjonen er at vi noen steder må appellere til trekantulikheten (setning 1.2.4) og Schwarz' ulikhet (setning 1.2.3) for å få argumentene til å gå opp. Vi tar $\mathbf{F}+\mathbf{G}$ og $\mathbf{F}\cdot\mathbf{G}$ som eksempler — det første av disse bevisene er ganske enkelt, det andre er mer komplisert:

For å vise at $\mathbf{F}+\mathbf{G}$ er kontinuerlig i $\mathbf{a}$, må vi vise at for hver $\epsilon > 0$ finnes det en $\delta > 0$ slik at hvis $\mathbf{x} \in A$ og $|\mathbf{x} - \mathbf{a}| < \delta$, så er

$$\big|(\mathbf{F}(\mathbf{x}) + \mathbf{G}(\mathbf{x})) - (\mathbf{F}(\mathbf{a}) + \mathbf{G}(\mathbf{a}))\big| < \epsilon$$

Vi stokker litt om på leddene slik at vi får $\mathbf{F}$ og $\mathbf{G}$ hver for seg, og bruker deretter trekantulikheten:

$$\big|(\mathbf{F}(\mathbf{x}) + \mathbf{G}(\mathbf{x})) - (\mathbf{F}(\mathbf{a}) + \mathbf{G}(\mathbf{a}))\big| = \big|(\mathbf{F}(\mathbf{x}) - \mathbf{F}(\mathbf{a})) + (\mathbf{G}(\mathbf{x}) + \mathbf{G}(\mathbf{a}))\big| \leq$$
$$\leq \big|\mathbf{F}(\mathbf{x}) - \mathbf{F}(\mathbf{a})\big| + \big|\mathbf{G}(\mathbf{x}) + \mathbf{G}(\mathbf{a})\big|$$

Siden $\mathbf{F}$ er kontinuerlig i $\mathbf{a}$, finnes det en $\delta_1 > 0$ slik at $|\mathbf{F}(\mathbf{x}) - \mathbf{F}(\mathbf{a})| < \frac{\epsilon}{2}$ når $\mathbf{x} \in A$ og $|\mathbf{x} - \mathbf{a}| < \delta_1$. Siden $\mathbf{G}$ er kontinuerlig i $\mathbf{a}$, finnes det tilsvarende en $\delta_2 > 0$ slik at $|\mathbf{G}(\mathbf{x}) - \mathbf{G}(\mathbf{a})| < \frac{\epsilon}{2}$ når $\mathbf{x} \in A$ og $|\mathbf{x} - \mathbf{a}| < \delta_2$. Lar vi δ være det minste av de to tallene δ_1 og δ_2, ser vi at når $\mathbf{x} \in A$ og $|\mathbf{x} - \mathbf{a}| < \delta$, så er

$$\big|(\mathbf{F}(\mathbf{x}) + \mathbf{G}(\mathbf{x})) - (\mathbf{F}(\mathbf{a}) + \mathbf{G}(\mathbf{a}))\big| \leq$$

$$|\mathbf{F}(\mathbf{x}) - \mathbf{F}(\mathbf{a})| + |\mathbf{G}(\mathbf{x}) + \mathbf{G}(\mathbf{a})| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

Dermed har vi vist at $\mathbf{F} + \mathbf{G}$ er kontinuerlig i $\mathbf{a}$.

La oss så vise kontinuitet av $\mathbf{F} \cdot \mathbf{G}$. Gitt en $\epsilon > 0$, må vi finne en $\delta > 0$ slik at hvis $\mathbf{x} \in A$ og $|\mathbf{x} - \mathbf{a}| < \delta$, så er $|\mathbf{F}(\mathbf{x}) \cdot \mathbf{G}(\mathbf{x}) - \mathbf{F}(\mathbf{a}) \cdot \mathbf{G}(\mathbf{a})| < \epsilon$. Vi legger til og trekker fra $\mathbf{F}(\mathbf{a}) \cdot \mathbf{G}(\mathbf{x})$, og bruker deretter trekantulikheten og Schwarz' ulikhet:

$$|\mathbf{F}(\mathbf{x}) \cdot \mathbf{G}(\mathbf{x}) - \mathbf{F}(\mathbf{a}) \cdot \mathbf{G}(\mathbf{a})| =$$

$$|\mathbf{F}(\mathbf{x}) \cdot \mathbf{G}(\mathbf{x}) - \mathbf{F}(\mathbf{a}) \cdot \mathbf{G}(\mathbf{x}) + \mathbf{F}(\mathbf{a}) \cdot \mathbf{G}(\mathbf{x}) - \mathbf{F}(\mathbf{a}) \cdot \mathbf{G}(\mathbf{a})| \leq$$

$$\leq |\mathbf{F}(\mathbf{x}) \cdot \mathbf{G}(\mathbf{x}) - \mathbf{F}(\mathbf{a}) \cdot \mathbf{G}(\mathbf{x})| + |\mathbf{F}(\mathbf{a}) \cdot \mathbf{G}(\mathbf{x}) - \mathbf{F}(\mathbf{a}) \cdot \mathbf{G}(\mathbf{a})| \leq$$

$$\leq |\mathbf{F}(\mathbf{x}) - \mathbf{F}(\mathbf{a})||\mathbf{G}(\mathbf{x})| + |\mathbf{F}(\mathbf{a})||\mathbf{G}(\mathbf{x}) - \mathbf{G}(\mathbf{a})|$$

Vi vil være i mål dersom vi kan vise at vi kan få begge uttrykkene $|\mathbf{F}(\mathbf{x}) - \mathbf{F}(\mathbf{a})||\mathbf{G}(\mathbf{x})|$ og $|\mathbf{F}(\mathbf{a})||\mathbf{G}(\mathbf{x}) - \mathbf{G}(\mathbf{a})|$ mindre enn $\frac{\epsilon}{2}$ ved å velge $\mathbf{x}$ tilstrekkelig nær $\mathbf{a}$. Det siste leddet er det enkleste, så vi begynner med det. Hvis $\mathbf{F}(\mathbf{a}) = \mathbf{0}$, er leddet lik 0 og derfor mindre enn $\frac{\epsilon}{2}$. Vi kan derfor anta at $\mathbf{F}(\mathbf{a}) \neq \mathbf{0}$. Siden $\mathbf{G}$ er kontinuerlig i $\mathbf{a}$, finnes det en $\delta_1 > 0$ slik at hvis $|\mathbf{x} - \mathbf{a}| < \delta_1$, så er $|\mathbf{G}(\mathbf{x}) - \mathbf{G}(\mathbf{a})| < \frac{\epsilon}{2|\mathbf{F}(\mathbf{a})|}$. Dermed er

bare

$$|\mathbf{F}(\mathbf{a})||\mathbf{G}(\mathbf{x}) - \mathbf{G}(\mathbf{a})| \leq |\mathbf{F}(\mathbf{a})|\frac{\epsilon}{2|\mathbf{F}(\mathbf{a})|} = \frac{\epsilon}{2}$$

Vi tar så for oss leddet $|\mathbf{F}(\mathbf{x}) - \mathbf{F}(\mathbf{a})||\mathbf{G}(\mathbf{x})|$. Observer først at siden $\mathbf{G}$ er kontinuerlig i $\mathbf{a}$, så finnes det en $\delta_2 > 0$ slik at hvis $|\mathbf{x} - \mathbf{a}| < \delta_2$, så er $|\mathbf{G}(\mathbf{x}) - \mathbf{G}(\mathbf{a})| < 1$. Dette betyr at dersom $|\mathbf{x} - \mathbf{a}| < \delta_2$, så er (vi bruker trekantulikheten igjen)

$$|\mathbf{G}(\mathbf{x})| = |\mathbf{G}(\mathbf{x}) - \mathbf{G}(\mathbf{a}) + \mathbf{G}(\mathbf{a})| \leq |\mathbf{G}(\mathbf{x}) - \mathbf{G}(\mathbf{a})| + |\mathbf{G}(\mathbf{a})| \leq 1 + |\mathbf{G}(\mathbf{a})|$$

Siden $\mathbf{F}$ er kontinuerlig i $\mathbf{a}$, finnes det en $\delta_3 > 0$ slik at hvis $|\mathbf{x} - \mathbf{a}| < \delta_3$, så er $|\mathbf{F}(\mathbf{x}) - \mathbf{F}(\mathbf{a})| \leq \frac{\epsilon}{2(1+|\mathbf{G}(\mathbf{a})|)}$. Er $|\mathbf{x} - \mathbf{a}|$ mindre enn både δ_2 og δ_3, er dermed

$$|\mathbf{F}(\mathbf{x}) - \mathbf{F}(\mathbf{a})||\mathbf{G}(\mathbf{x})| \leq \frac{\epsilon}{2(1 + |\mathbf{G}(\mathbf{a})|)}(1 + |\mathbf{G}(\mathbf{a})|) = \frac{\epsilon}{2}$$

Velger vi δ til å være det minste av tallene δ_1, δ_2 og δ_3, ser vi dermed at når $\mathbf{x} \in A$ og $|\mathbf{x} - \mathbf{a}| < \delta$, så er

$$|\mathbf{F}(\mathbf{x}) \cdot \mathbf{G}(\mathbf{x}) - \mathbf{F}(\mathbf{a}) \cdot \mathbf{G}(\mathbf{a})| \leq |\mathbf{F}(\mathbf{x}) - \mathbf{F}(\mathbf{a})||\mathbf{G}(\mathbf{x})| + |\mathbf{F}(\mathbf{a})||\mathbf{G}(\mathbf{x}) - \mathbf{G}(\mathbf{a})| <$$

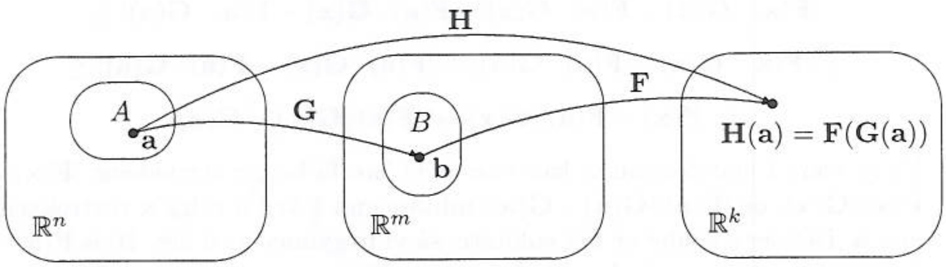
$$< \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

og beviset er fullført. □

Det neste resultatet kjenner du også igjen fra teorien for funksjoner av én variabel:

Setning 2.2.3 *Anta at vi har to mengder $A \subset \mathbb{R}^n$, $B \subset \mathbb{R}^m$, og to funksjoner $\mathbf{G} : A \to B$, $\mathbf{F} : B \to \mathbb{R}^k$ (se figuren nedenfor). Dersom $\mathbf{G}$ er kontinuerlig i punktet $\mathbf{a}$, og $\mathbf{F}$ er kontinuerlig i punktet $\mathbf{b} = \mathbf{G}(\mathbf{a})$, så er den sammensatte funksjonen $\mathbf{H}(x) = \mathbf{F}(\mathbf{G}(x))$ kontinuerlig i $\mathbf{a}$.* □

Beviset er akkurat som for funksjoner av én variabel (se setning 5.1.7 i *Kalkulus*) for hjelp). Figur 2 viser hvordan $\mathbf{G}, \mathbf{F}$ og $\mathbf{H}$ virker.



Figur 2: Sammensetning av funksjoner

Når vi skal bruke reglene ovenfor til å vise at en funksjon $\mathbf{F} : \mathbb{R}^n \to \mathbb{R}^m$ er kontinuerlig, lønner det seg ofte å skrive den på komponentform:

$$\mathbf{F}(\mathbf{x}) = (F_1(\mathbf{x}), F_2(\mathbf{x}), \ldots, F_m(\mathbf{x}))$$

Det neste resultatet forteller oss nemlig at $\mathbf{F}$ er kontinuerlig hvis og bare hvis hver komponent F_i er kontinuerlig.

Setning 2.2.4 *Anta at $\mathbf{F} : A \to \mathbb{R}^m$ er en funksjon av n variable med komponenter*

$$\mathbf{F}(\mathbf{x}) = (F_1(\mathbf{x}), F_2(\mathbf{x}), \ldots, F_m(\mathbf{x}))$$

for $1 \le i \le m$ *Da er $\mathbf{F}$ kontinuerlig i et punkt $\mathbf{a} \in A$ hvis og bare hvis hver komponent F_i er kontinuerlig i $\mathbf{a}$,*

el. $i \in \{1, \ldots, m\}$

Bevis: La oss først anta at alle komponentene $F_1, F_2, \ldots, F_m$ er kontinuerlige i $\mathbf{a}$. Vi må vise at for enhver $\epsilon > 0$, finnes det en $\delta > 0$ slik at hvis $|\mathbf{x} - \mathbf{a}| < \delta$, så er $|\mathbf{F}(\mathbf{x}) - \mathbf{F}(\mathbf{a})| < \epsilon$. Siden hver komponent F_i er kontinuerlig i $\mathbf{a}$, finnes det en $\delta_i > 0$ slik at $|F_i(\mathbf{x}) - F_i(\mathbf{a})| < \frac{\epsilon}{\sqrt{m}}$ når $|\mathbf{x} - \mathbf{a}| < \delta_i$. Vi velger δ til å være det minste av tallene $\delta_1, \delta_2, \ldots, \delta_m$. Hvis $|\mathbf{x} - \mathbf{a}| < \delta$, har vi da x

$$|\mathbf{F}(\mathbf{x}) - \mathbf{F}(\mathbf{a})| = \sqrt{(F_1(\mathbf{x}) - F_1(\mathbf{a}))^2 + \cdots + (F_m(\mathbf{x}) - F_m(\mathbf{a}))^2} \le \quad <$$

$$\leq \sqrt{\left(\frac{\epsilon}{\sqrt{m}}\right)^2 + \cdots + \left(\frac{\epsilon}{\sqrt{m}}\right)^2} = \sqrt{m\left(\frac{\epsilon^2}{m}\right)} = \epsilon$$

Dette viser at $\mathbf{F}$ er kontinuerlig i $\mathbf{a}$.

Anta nå omvendt at $\mathbf{F}$ er kontinuerlig i $\mathbf{a}$, og at vi skal vise at komponenten F_i er kontinuerlig i $\mathbf{a}$. Gitt $\epsilon > 0$, må vi finne en $\delta > 0$ slik at $|F_i(\mathbf{x}) - F_i(\mathbf{a})| < \epsilon$ når $|\mathbf{x} - \mathbf{a}| < \delta$. Siden $\mathbf{F}$ er kontinuerlig i $\mathbf{a}$, finnnes det en $\delta > 0$ slik at $|\mathbf{F}(\mathbf{x}) - \mathbf{F}(\mathbf{a})| < \epsilon$ når $|\mathbf{x} - \mathbf{a}| < \delta$. Men dermed er

$$|F_i(\mathbf{x}) - F_i(\mathbf{a})| \leq \sqrt{(F_1(\mathbf{x}) - F_1(\mathbf{a}))^2 + \cdots + (F_m(\mathbf{x}) - F_m(\mathbf{a}))^2} =$$

$$= |\mathbf{F}(\mathbf{x}) - \mathbf{F}(\mathbf{a})| < \epsilon$$

når $|\mathbf{x} - \mathbf{a}| < \delta$. Dette viser at F_i er kontinuerlig i $\mathbf{a}$. □

De enkleste funksjonene av flere variable (bortsett fra konstantfunksjonene) er de som bare gir oss en av variablene som funksjonsverdi, f.eks.

$$f(x, y, z, u) = y \quad \text{eller} \quad g(x, y, z) = z$$

Generelt kaller vi

$$k_i(x_1, x_2, \ldots, x_n) = x_i$$

den i-te *koordinatfunksjonen*. Det er lett å se at disse koordinatfunksjonene er kontinuerlige, og det neste eksemplet forteller oss hvordan vi kan bruke dem og resultatene ovenfor til å vise at mer kompliserte funksjoner er kontinuerlige.

Eksempel 1: Vis at funksjonen $\mathbf{F} : \mathbb{R}^3 \to \mathbb{R}^2$ definert ved

$$\mathbf{F}(x, y, z) = \left(x^2 \sin(yz), \frac{e^{x^2+z}}{y^2+1}\right)$$

er kontinuerlig i punktet $\mathbf{a} = (1, 0, -1)$. Ifølge setning 2.2.4 er det nok å vise at begge funksjonene

$$F_1(x, y, z) = x^2 \sin(yz) \quad \text{og} \quad F_2(x, y, z) = \frac{e^{x^2+z}}{y^2+1}$$

er kontinuerlig i $\mathbf{a}$. Vi begynner med $F_1(x, y, z)$. Siden koordinatfunksjonene $k_2(x, y, z) = y$ og $k_3(x, y, z) = z$ er kontinuerlige, er produktet av dem $f(x, y, z) = yz$ også kontinuerlig ifølge setning 2.2.2. Siden sinus er en kontinuerlig funksjon, må den sammensatte funksjonen $\sin(yz)$ også være kontinuerlig ifølge setning 2.2.3. På tilsvarende måte ser vi at funksjonen $g(x, y, z) = x^2$ er kontinuerlig (fordi den er lik $k_1(x, y, z) \cdot k_1(x, y, z)$

som er et produkt av to kontinuerlige funksjoner), og dermed er produktet $F_1(x,y,z) = x^2 \sin(yz)$ kontinuerlig.

Vi behandler F_2 på samme måte: Funksjonen $h(x,y,z) = x^2 + z$ er kontinuerlig siden den er bygget opp fra kontinuerlige koordinatfunksjoner ved hjelp av multiplikasjon og addisjon. Siden eksponentialfunksjonen er kontinuerlig, er da $p(x,y,z) = e^{x^2+z}$ kontinuerlig. Ved et tilsvarende resonnement ser vi at nevneren y^2+1 er kontinuerlig, og siden den er forskjellig fra 0, må brøken

$$F_2(x,y,z) = \frac{e^{x^2+z}}{y^2+1}$$

være kontinuerlig. ♣

Ved hjelp av teknikken i dette eksemplet er det lett å vise at funksjoner bygget opp ved hjelp av potensfunksjoner, eksponentialfunksjoner, logaritmer, trigonometriske funksjoner og arcusfunksjoner er kontinuerlige (men husk å sjekke at de er definert i punktet du er interessert i!)

Hittil har teorien for kontinuerlige funksjoner av flere variable vært forbløffende lik teorien for kontinuerlige funksjoner av én variabel. På grunn av den rikere geometrien finnes det imidlertid fenomener for flervariable funksjoner som ikke har noen motsvarighet for funksjoner av én variabel. Her er et eksempel.

Eksempel 2: Funksjonen $f : \mathbb{R}^2 \to \mathbb{R}$ er gitt ved

$$f(x,y) = \begin{cases} \frac{x^2y}{x^4+y^2} & \text{for } (x,y) \neq (0,0) \\ \\ 0 & \text{for } (x,y) = (0,0) \end{cases}$$

V i

Vi skal undersøke om f er kontinuerlig $(0,0)$ dersom $f(x,y)$ nærmer seg $f(0,0)$ når (x,y) nærmer seg $(0,0)$. La oss derfor se hva som skjer når (x,y) nærmer seg $(0,0)$ på forskjellige måter.

seg

La oss først se på hva som skjer når (x,y) nærmer ser origo langs en skrålinje $y = cx$ der $c \neq 0$; dvs. vi har punkter (x, cx) der x går mot null. Setter vi inn i funksjonsuttrykket, får vi

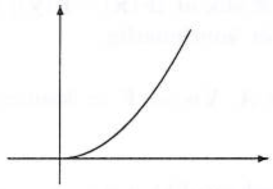
$$f(x,y) = f(x,cx) = \frac{x^2(cx)}{x^4+(cx)^2} = \frac{cx}{x^2+c^2} \longrightarrow \frac{0}{c^2} = 0$$

Tilsvarende ser vi at når (x,y) nærmer seg $(0,0)$ langs x-aksen, så er punktene på formen $(x,0)$ der $x \to 0$, og vi får

$$f(x,y) = f(x,0) = \frac{x^2 \cdot 0}{x^4+0^2} = 0$$

Akkurat det samme skjer om vi lar (x,y) nærme seg $(0,0)$ langs y-aksen.

Alt dette tyder på at f er kontinuerlig i $(0,0)$, men det finnes jo andre måter å nærme seg et punkt på enn å følge en rett linje. La oss prøve å nærme oss $(0,0)$ langs parabelen $y = x^2$ (se figuren). Vi ser altså på punkter (x, x^2) der x går mot 0.



Figur 3. Punkter som nærmer seg $(0,0)$ langs parabelen $y = x^2$

Setter vi inn i funksjonsuttrykket, får vi

$$f(x,y) = f(x,x^2) = \frac{x^2 \cdot x^2}{x^4 + (x^2)^2} = \frac{1}{2}$$

$\frac{x^4}{2x^4} =$

Dette viser at f *ikke* er kontinuerlig i $(0,0)$ siden vi har punkter med verdien $\frac{1}{2}$ så nær $(0,0)$ vi måtte ønske. Funksjonen virker altså å være kontinuerlig så lenge vi beveger oss langs rette linjer, men er det likevel ikke! ♣

Eksemplet ovenfor viser at det kan være vanskelig å få god oversikt over hvordan en funksjon av flere variable oppfører seg, og at det kanskje finnes flere geometriske muligheter enn det vi kan forestille oss. Det er i slike sammenhenger vi virkelige får nytte av abstrakte definisjoner av ϵ-δ-typen; X de gir oss muligheten til å føre vanntette bevis selv i tilfeller der vi ikke er sikre på om vi har fått full oversikt over alle geometriske snurrepiperier!

Hittil har vi bare snakket om kontinuitet i et punkt. Vi avslutter denne seksjonen med definisjonen av en kontinuerlig funksjon — den er helt tilsvarende definisjonen for funksjoner av én variabel. ett

Definisjon 2.2.5 *En funksjon* **F** *kalles* kontinuerlig *dersom den er kontinuerlig i alle punkter i sitt definisjonsområde.*

Oppgaver til seksjon 2.2

1. Vis at funksjonen f er kontinuerlig

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

b) $f(x,y) = x^2y + y$

c) $f(x,y) = \frac{xy}{1+x^2+y^2}$

d) $f(x,y) = e^{-x}\sin(x+y)$

e) $f(x,y,z) = x^2 + y^2 + z^2$

2. Vis at funksjonene er kontinuerlige:

a) $\mathbf{F}(x,y,z) = (x^2z + y, x^2 \sin xyz, x^3)$

b) $\mathbf{G}(x,y,z,u) = (e^{xu+z^2}, z\cos xy^2u)$

c) $\mathbf{H}(x,y,z,u) = (x^y e^{xz^2}, z + u^2, x^2 + 3yzu)$

d) $\mathbf{K}(x, y, z, u, v) = (\sin(xy + z^2v), 2uv)$

3. Vis at koordinatfunksjonene $k_i(x_1, x_2, \ldots, x_n) = x_i$ er kontinuerlige ved å bruke definisjonen av kontinuitet.

4. a) Anta det finnes en konstant $M \in \mathbb{R}$ slik at $|\mathbf{F}(\mathbf{x}) - \mathbf{F}(\mathbf{y})| \leq M|\mathbf{x} - \mathbf{y}|$ for alle $\mathbf{x}, \mathbf{y}$ i definisjonsområdet $D_{\mathbf{F}}$. Vis at $\mathbf{F}$ er kontinuerlig.

b) Anta at $\mathbf{F}(\mathbf{x}) = A\mathbf{x}$ for en matrise A. Vis at $\mathbf{F}$ er kontinuerlig. (*Hint:* Husk setning 1.6.3.)

5. I denne oppgaven har du bruk for trekantulikheten som sier at hvis $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, så er $|\mathbf{x} + \mathbf{y}| \leq |\mathbf{x}| + |\mathbf{y}|$.

a) Vis at $|\,|\mathbf{x}| - |\mathbf{y}|\,| \leq |\mathbf{x} - \mathbf{y}|$ for alle $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$.

b) La $\mathbf{a} \in \mathbb{R}^n$. Vis at funksjonen $f(\mathbf{x}) = |\mathbf{x} - \mathbf{a}|$ er kontinuerlig.

c) Vis at funksjon $g(\mathbf{x}) = \frac{1}{|\mathbf{x} - \mathbf{a}|}$ er kontinuerlig der den er definert.

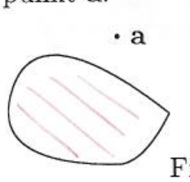
6. Bevis resten av setning 2.2.2.

7. Bevis setning 2.2.3.

2.3 Grenseverdier

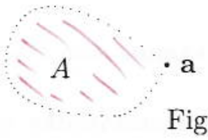
I teorien for funksjoner av én variabel opererer vi både med ensidige og tosidige grenser. Ensidige grenser er blant annet nyttig når vi skal avgjøre om en funksjon er kontinuerlig i enden av sitt definisjonsområde — vi bruker grenseverdien $\lim_{x \to a^+} f(x)$ for å undersøke om f er kontinuerlig i det venstre endepunktet a av intervallet $[a, b]$, og grenseverdien $\lim_{x \to b^-} f(x)$ for å undersøke om f er kontinuerlig i det høyre endepunktet b. Med flere variable er det så mange måter å nærme seg et randpunkt på at vi ikke kan ha ett grensebegrep for hver måte — vi må finne frem til et felles begrep som dekker alle tilfeller.

La oss først se litt på hvilke punkter det er naturlig å regne ut grenseverdier i. Vi ser på en funksjon $\mathbf{F} : A \to \mathbb{R}^m$ definert på en delmengde A av $\mathbb{R}^n$. Figur 1 viser én mulighet; her består A av en sammenhengende mengde pluss et isolert punkt $\mathbf{a}$.



Figur 1

Siden vi ikke kan nærme oss $\mathbf{a}$ innenfor definisjonsmengden til $\mathbf{F}$, er det ikke rimelig å definere grenseverdien til $\mathbf{F}$ i punktet $\mathbf{a}$. Figur 2 viser en annen situasjon.



Figur 2

På denne figuren er omkretsen ("randen") til A stiplet for å markere at punktene der ikke hører med til mengden. Dermed er $\mathbf{a}$ ikke med i definisjonsmengden A, men det går fint an å nærme seg $\mathbf{a}$ fra A. I dette tilfellet *er* det rimelig å definere grenseverdien til $\mathbf{F}$ i punktet $\mathbf{a}$. Konklusjonen på disse observasjonene må bli at det er rimelig å definere grenseverdien til $\mathbf{F}$ i punktet $\mathbf{a}$ dersom det går an å nærme seg $\mathbf{a}$ med punkter ($\neq \mathbf{a}$) fra definisjonsmengden til $\mathbf{F}$, uansett om punktet $\mathbf{a}$ selv ligger i definisjonsmengden eller ikke. Slike punkter kaller vi *opphopningspunkter* for A. Her er den presise definisjonen:

Definisjon 2.3.1 *La A være en delmengde av $\mathbb{R}^n$. Et punkt $\mathbf{a} \in \mathbb{R}^n$ kalles et* opphopningspunkt *for A dersom enhver kule $B(\mathbf{a}, r)$ om $\mathbf{a}$ inneholder uendelig mange punkter fra A.*

med radius $r > 0$

Vi er nå klare til å definere grenseverdien i et opphopningspunkt:

Definisjon 2.3.2 *La $\mathbf{F} : A \to \mathbb{R}^m$ være en funksjon av n variable og anta at $\mathbf{a} \in \mathbb{R}^n$ er et opphopningspunkt for A. Vi sier at $\mathbf{b} \in \mathbb{R}^m$ er* grenseverdien *for $\mathbf{F}$ i punktet $\mathbf{a}$ dersom det for hver $\epsilon > 0$ finnes en $\delta > 0$ slik at*

$$|\mathbf{F}(\mathbf{x}) - \mathbf{b}| < \epsilon \textit{ for alle } \mathbf{x} \in A \textit{ slik at } 0 < |\mathbf{x} - \mathbf{a}| < \delta$$

$\vec{x}$

Vi skriver $\lim_{x \to \mathbf{a}} \mathbf{F}(\mathbf{x}) = \mathbf{b}$. Dersom $\mathbf{a}$ ikke er et opphopningspunkt for A, er grenseverdien ikke definert.

(Grenseverdien eksisterer ikke alltid, selv om $\vec{a}$ er et opphopningspkt for A)

Denne definisjonen ligner på den du finner for funksjoner av én variabel i definisjon 5.4.1 i *Kalkulus*, men skiller seg på et viktig punkt: Vi insisterer ikke lenger på at $\mathbf{F}$ skal være definert i alle punkter i nærheten av $\mathbf{a}$, men kompenserer for dette ved bare å kreve at $|\mathbf{F}(\mathbf{x}) - \mathbf{b}| < \epsilon$ skal holde for punkter $\mathbf{x}$ som er med i definisjonsområdet A til $\mathbf{F}$.

Akkurat som kontinuitet kan grenseverdier studeres komponentvis:

Setning 2.3.3 *La $\mathbf{F} : A \to \mathbb{R}^m$ være en funksjon av n variable, og anta at $\mathbf{a} \in \mathbb{R}^n$ er et opphopningspunkt for A. Anta at komponentene til $\mathbf{F}$ er gitt ved*

$$\mathbf{F}(\mathbf{x}) = (F_1(\mathbf{x}), F_2(\mathbf{x}), \ldots, F_m(\mathbf{x}))$$

og la $\mathbf{b}$ være en vektor med komponenter $\mathbf{b} = (b_1, b_2, \ldots, b_m)$. Da er

$$\lim_{x \to \mathbf{a}} \mathbf{F}(\mathbf{x}) = \mathbf{b}$$

$\vec{x}$

hvis og bare hvis

$$\lim_{x\to \mathbf{a}} F_i(\mathbf{x}) = b_i \text{ for alle } i$$

for alle i.

□

Beviset overlates til leserne. Ideen er den samme som i beviset for setning 2.2.4.

Vi har også de vanlige regnereglene for grenseverdier til summer, differenser, produkter og brøker:

skalar

Setning 2.3.4 (Regneregler for grenseverdier) *Anta at* $\mathbf{F}, \mathbf{G} : A \to \mathbb{R}^m$ *er to funksjoner av* n *variable og at* $\mathbf{a} \in A$ *er et opphopningspunkt for* A. *Dersom* $\lim_{x\to\mathbf{a}} \mathbf{F}(\mathbf{x}) = \mathbf{A}$ *og* $\lim_{x\to\mathbf{a}} \mathbf{G}(\mathbf{x}) = \mathbf{B}$, *så er:*
a) $\lim_{x\to\mathbf{a}}(\mathbf{F}(\mathbf{x}) + \mathbf{G}(\mathbf{x})) = \mathbf{A} + \mathbf{B}$.
b) $\lim_{x\to\mathbf{a}}(\mathbf{F}(\mathbf{x}) - \mathbf{G}(\mathbf{x})) = \mathbf{A} - \mathbf{B}$.
c) $\lim_{x\to\mathbf{a}}(\mathbf{F}(\mathbf{x}) \cdot \mathbf{G}(\mathbf{x})) = \mathbf{A} \cdot \mathbf{B}$.
d) $\lim_{x\to\mathbf{a}} \frac{\mathbf{F}(\mathbf{x})}{\mathbf{G}(\mathbf{x})} = \frac{\mathbf{A}}{\mathbf{B}}$, *forutsatt at* $\mathbf{F}$ *og* $\mathbf{G}$ *tar verdier i* $\mathbb{R}$ *og* $\mathbf{B} \neq 0$. □

Også disse bevisene overlates til leserne.

Det neste resultatet tar seg av sammenhengen mellom grenseverdier og kontinuitet. Legg merke til at det bare gjelder for kontinuitet i opphopningspunkter (men i et isolerte punkt er en funksjon alltid kontinuerlig, så der er det ikke så mye å vise!)

Setning 2.3.5 *La* $\mathbf{F} : A \to \mathbb{R}^m$ *være en funksjon av* n *variable, og anta at* $\mathbf{a} \in A$ *er et opphopningspunkt for* A. *Da er* $\mathbf{F}$ *kontinuerlig i* $\mathbf{a}$ *hvis og bare hvis* $\lim_{\mathbf{x}\to\mathbf{a}} \mathbf{F}(\mathbf{x}) = \mathbf{F}(\mathbf{a})$

Bevis: Sammenlign definisjonen av kontinuitet med definisjonen av grenseverdi. □

Resultatet ovenfor er nyttig når vi skal regne ut enkle grenseverdier:

Eksempel 1: Finn grenseverdien til

$$F(x, y) = (x^2 y, e^{-xy} \sin(\pi x))$$

når $(x, y) \to (1, -2)$.

Vi ser at funksjonen er kontinuerlig i $(1, -2)$, så

$$\lim_{(x,y)\to(1,-2)} F(x, y) = F(1, -2) = (1^2 \cdot (-2), e^{-1\cdot(-2)} \sin(\pi \cdot 1)) = (-2, 0)$$

♣

I noen eksempler må vi forenkle uttrykket før vi går til grensen:

Eksempel 2: Finn grenseverdien til

$$f(x,y) = \frac{x^2 - y^2}{x - y}$$

når $(x,y) \to (0,0)$.

Siden $x^2 - y^2 = (x+y)(x-y)$ har vi

$$\lim_{(x,y)\to(0,0)} f(x,y) = \lim_{(x,y)\to(0,0)} \frac{x^2 - y^2}{x - y} =$$

$$= \lim_{(x,y)\to(0,0)} \frac{(x+y)(x-y)}{x-y} = \lim_{(x,y)\to(0,0)} (x+y) = 0$$

Gjelder i D_f der $x \neq y$

♣

Oppgaver til seksjon 2.3

1. Finn grenseverdiene

a) $\lim_{(x,y)\to(2,3)} (x^3 + 2xy)$

b) $\lim_{(x,y)\to(1,\frac{\pi}{2})} x^2 \sin(xy)$

c) $\lim_{(x,y)\to(1,0)} \frac{e^{x+y}}{x^2+3y}$

d) $\lim_{(x,y)\to(0,0)} \frac{\sin(xy)}{xy} \cdot \cos(x+y)$

2. Anta at $A \subset \mathbb{R}^n$ og at $\mathbf{a} \in \mathbb{R}^n$. Anta at enhver kule $B(\mathbf{a}, \epsilon)$ om $\mathbf{a}$ inneholder minst ett element fra A forskjellig fra $\mathbf{a}$. Vis at $\mathbf{a}$ er et opphopningspunkt for A. $\varepsilon \longleftrightarrow r$?

3. Bevis setning 2.3.3.

4. Bevis setning 2.3.4

5. Bevis setning 2.3.5

2.4 Derivasjon av skalarfelt

Vi skal nå begynne å se på derivasjon av funksjoner av flere variable. For at det ikke skal være altfor mange komponenter å holde styr på, skal vi først derivere skalarfelt, dvs. funksjoner $f : \mathbb{R}^n \to \mathbb{R}$ der verdiene $f(\mathbf{x})$ er tall. Før vi setter igang for alvor, trenger vi en definisjon: Hvis A er en delmengde av $\mathbb{R}^n$, kalles $\mathbf{a} \in A$ et *indre punkt* i A dersom det finnes en $\epsilon > 0$ slik at $B(\mathbf{a}, \epsilon) \subset A$. Dette betyr at dersom vi starter i $\mathbf{a}$, kan vi gå et lite stykke i en hvilken som helst retning uten å forlate A.

For en funksjon $y = f(x)$ av én variabel forteller den deriverte $f'(x)$ oss hvor fort funksjonen vokser i punktet x — går vi et lite skritt med lengde h langs x-aksen, vil funksjonsverdien øke med (omtrent) $f'(x)h$. For funksjoner av flere variable er situasjonen mer komplisert; vi har flere akser å bevege oss langs, og vi kan ikke regne med at funksjonen stiger like mye uansett hvilken retning vi går i. Før vi regner ut stigningstallet til funksjonen, må vi derfor spesifisere hvilken retning vi er interessert i. Dette er idéen bak begrepet retningsderivert:

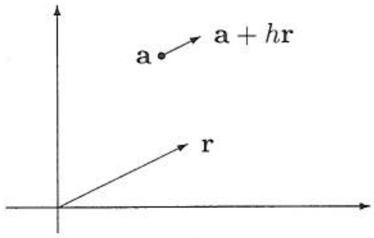
Definisjon 2.4.1 *Anta at funksjonen $f : A \to \mathbb{R}$ er definert på en delmengde A av $\mathbb{R}^n$ og at $\mathbf{a}$ er et indre punkt i A. Tenk på $\mathbf{r} \in \mathbb{R}^n$ som en vektor. Den* retningsderiverte til f i punktet $\mathbf{a}$ og retningen $\mathbf{r}$ *er gitt ved*

Merk notasjonen!

$$f'(\mathbf{a};\mathbf{r}) = \lim_{h \to 0} \frac{f(\mathbf{a} + h\mathbf{r}) - f(\mathbf{a})}{h}$$

forutsatt at denne grensen eksisterer.

Figur 1 viser ideen bak definisjonen.



Figur 1

Punktene $\mathbf{a} + h\mathbf{r}$ er de punktene vi kommer til hvis vi starter i $\mathbf{a}$ og går i retning $\mathbf{r}$. Differansen $f(\mathbf{a} + h\mathbf{r}) - f(\mathbf{a})$ forteller oss hvor mye funksjonen øker når vi beveger oss i denne retning, og brøken

-en

$$\frac{f(\mathbf{a} + h\mathbf{r}) - f(\mathbf{a})}{h}$$

er økningen per lengdeenhet når vi bruker $|\mathbf{r}|$ som måleenhet.

Legg merke til at vi bare har definert retningsderiverte i indre punkter. Det garanterer at uttrykket $f(\mathbf{a} + h\mathbf{r}) - f(\mathbf{a})$ alltid gir mening bare vi velger h liten nok. I punkter som ikke er indre, kan man ofte definere retningsderiverte i noen retninger, men ikke i andre. Vi skal ikke komme inn på dette her, men konsentrere oss om indre punkter.

Eksempel 1: La $f(x, y) = x^2 + xy$. Vi skal beregne den retningsderiverte $f'(\mathbf{a};\mathbf{r})$ når $\mathbf{a} = (1, 0)$, $\mathbf{r} = (2, 1)$. Først observerer vi at

$$\mathbf{a} + h\mathbf{r} = (1, 0) + h(2, 1) = (1 + 2h, h),$$

som gir $f(\mathbf{a} + h\mathbf{r}) = (1 + 2h)^2 + (1 + 2h)h = 1 + 5h + 6h^2$. Tilsvarende er

$f(\mathbf{a}) = f(1, 0) = 1^2 + 1 \cdot 0 = 1$. Vi får

$$f'(\mathbf{a}; \mathbf{r}) = \lim_{h \to 0} \frac{f(\mathbf{a} + h\mathbf{r}) - f(\mathbf{a})}{h} = \lim_{h \to 0} \frac{(1 + 5h + 6h^2) - 1}{h}$$
$$= \lim_{h \to 0} \frac{5h + 6h^2}{h} = \lim_{h \to 0} (5 + 6h) = 5.$$

Hva betyr dette resultatet? Legg merke til at lengden til vektoren $\mathbf{r}$ er $|\mathbf{r}| = \sqrt{2^2 + 1^2} = \sqrt{5}$. Dersom vi går et lite stykke $h\sqrt{5}$ i retningen til vektoren $\mathbf{r} = (2, 1)$, vil funksjonsverdien stige med (omtrent) $f'(\mathbf{a}; \mathbf{r}) \cdot h = 5 \cdot h$. ♣

Det er lettest å forstå hva den retningsderiverte er dersom vektoren $\mathbf{r}$ har lengde 1 – da er $f'(\mathbf{a}; \mathbf{r})$ rett og slett stigningstallet til funksjonen i retning $\mathbf{r}$ når vi måler med vanlige enheter. Ut i fra dette kan det være fristende å forutsette at $|\mathbf{r}| = 1$ når vi regner med retningsderiverte, men dette viser seg å være upraktisk, blant annet fordi enhetsvektorer ofte inneholder stygge kvadratrøtter.

Så langt kan det se ut som om vi må bygge opp en ny derivasjonsteori helt fra bunnen av for å kunne beregne retningsderiverte til funksjoner av flere variable. Det er heldigvis ikke nødvendig; ved hjelp av såkalte partiellderiverte kan vi føre mye av teorien tilbake til vanlig derivasjon av funksjoner av én variabel. Før vi definerer partiellderiverte, er det lurt å bli enig om litt notasjon.

Den *i-te enhetsvektoren* $\mathbf{e}_i$ i $\mathbb{R}^n$ er vektoren

$$\mathbf{e}_i = (0, 0, \ldots, 0, \underset{\substack{\uparrow \\ i\text{-te plass}}}{1}, 0, \ldots, 0)$$

langs den i-te koordinataksen.

Definisjon 2.4.2 *La $f: A \to \mathbb{R}$ være en funksjon av n variable, og la $\mathbf{a}$ være et indre punkt i A. Den i-te* partiellderiverte $\frac{\partial f}{\partial x_i}(\mathbf{a})$ *er den retningsderiverte av f i retning av den i-te enhetsvektoren $\mathbf{e}_i$; det vil si*

$$\frac{\partial f}{\partial x_i}(\mathbf{a}) = f'(\mathbf{a}; \mathbf{e}_i)$$

Her er $(x_1, \ldots, x_n)$ notasjonen for et punkt i A.

Andre notasjoner for $\frac{\partial f}{\partial x_i}(\mathbf{a})$ er $D_i f(\mathbf{a})$ og $f_{x_i}(\mathbf{a})$. De partiellderiverte er altså stigningstallene til funksjonen parallelt med koordinataksene. Skriver vi ut definisjonen i detalj, ser vi at

$$\frac{\partial f}{\partial x_i}(\mathbf{a}) = f'(\mathbf{a}; \mathbf{e}_i) = \lim_{h \to 0} \frac{f(\mathbf{a} + h\mathbf{e}_i) - f(\mathbf{a})}{h} =$$
$$= \lim_{h \to 0} \frac{f(a_1, a_2, \ldots, a_i + h, \ldots, a_n) - f(a_1, a_2, \ldots, a_i, \ldots, a_n)}{h}$$

(For $n = 1$ er $\frac{\partial f}{\partial x} = \frac{df}{dx}$.)

Det siste uttrykket har en slående likhet med definisjonen av vanlig derivert. Underslår vi de variablene hvor det ikke skjer noen endring, ser vi at

$$\frac{\partial f}{\partial x_i}(\mathbf{a}) = \lim_{h \to 0} \frac{f(a_i + h) - f(a_i)}{h}$$

Dette betyr at vi kan finne den partiellderiverte $\frac{\partial f}{\partial x_i}$ ved å derivere uttrykket $y = f(x_1, \ldots, x_i, \ldots, x_n)$ med hensyn på x_i mens vi later som om alle de andre variablene er konstanter.

Eksempel 2: Finn de partiellderiverte $\frac{\partial f}{\partial x}$ og $\frac{\partial f}{\partial y}$ til funksjonen

X (konsekvent)

$$f(x, y) = x^2 + xy^3 + \sin(xy).$$

For å finne $\frac{\partial f}{\partial x}$ deriverer vi uttrykket med hensyn på x mens vi later som om y er en konstant:

$$\frac{\partial f}{\partial x}(x, y) = 2x + y^3 + \cos(xy) \cdot y$$

For å finne $\frac{\partial f}{\partial y}$ deriverer vi med hensyn på y mens vi holder x konstant:

$$\frac{\partial f}{\partial y}(x, y) = 0 + 3xy^2 + \cos(xy)x = 3xy^2 + x\cos(xy)$$

♣

En funksjon $f(x_1, x_2, \ldots, x_n)$ av n variable har n partiellderiverte $\frac{\partial f}{\partial x_1}$, $\frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_n}$. Vi kan sette sammen disse til en vektor

$$\left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_n}\right)$$

Denne vektoren er så viktig at den har fått sitt eget navn og sitt eget symbol.

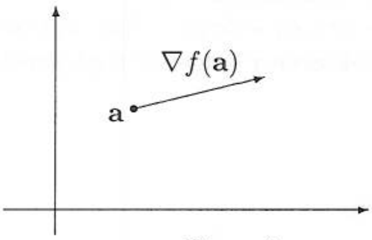
Definisjon 2.4.3 *Anta at de partiellderiverte til f eksisterer i punktet $\mathbf{a} \in \mathbb{R}^n$. Da kalles*

$$\nabla f(\mathbf{a}) = \left(\frac{\partial f}{\partial x_1}(\mathbf{a}), \frac{\partial f}{\partial x_2}(\mathbf{a}), \ldots, \frac{\partial f}{\partial x_n}(\mathbf{a})\right)$$

∇ : nabla

gradienten til f i punktet $\mathbf{a}$.

Det er ofte lurt å tenke på gradienten $\nabla f(\mathbf{a})$ som en vektor som starter i punktet $\mathbf{a}$ slik som vist på figuren nedenfor. Som vi snart skal se (setning 2.4.7), får gradienten da en geometrisk betydning — den peker i den retningen hvor funksjonen vokser raskest, og lengden $|\nabla f(\mathbf{a})|$ er lik stigningstallet i denne retningen.



Figur 2

Eksempel 3: Finn gradienten til

$$f(x, y, z) = x^2 y e^{xz}$$

i punktet $\mathbf{a} = (1, -2, 0)$.

Vi må først finne de partiellderiverte. Deriverer vi mhp. x som om y og z er konstanter, får vi

$$\frac{\partial f}{\partial x} = 2xye^{xz} + x^2 y e^{xz} z = xye^{xz}(2 + xz)$$

Deriverer vi mhp. y som om x og z er konstanter, får vi tilsvarende

$$\frac{\partial f}{\partial y} = x^2 e^{xz}$$

Til slutt deriverer vi mhp. z som om x og y er konstanter:

$$\frac{\partial f}{\partial z} = x^2 y e^{xz} x = x^3 y e^{xz}$$

Gradienten i et generelt punkt er dermed

$$\nabla f(x, y, z) = \left(xye^{xz}(2 + xz), x^2 e^{xz}, x^3 y e^{xz}\right)$$

I vårt punkt $\mathbf{a} = (1, -2, 0)$ får vi

$$\nabla f(1, -2, 0) = \left(1 \cdot (-2) \cdot e^{1 \cdot 0} \cdot (2 + 1 \cdot 0), 1^2 \cdot e^{1 \cdot 0}, 1^3 \cdot (-2) \cdot e^{1 \cdot 0}\right) =$$
$$= (-4, 1, -2)$$

♣

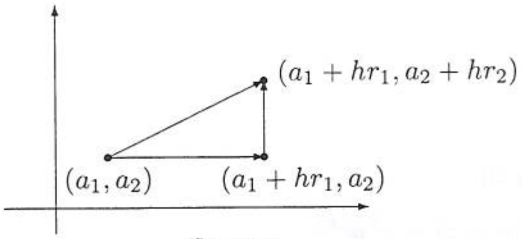
Vi har nå sett hvordan vi kan bruke våre vanlige derivasjonsregler til å regne ut partiellderiverte — vi bare deriverer som om de andre variablene var konstanter. Neste post på programmet er å vise hvordan vi kan regne ut retningsderiverte ved hjelp av partiellderiverte.

La oss begynne med å se på hva som skjer i to dimensjoner. Anta at vi ønsker å derivere f i punktet $\mathbf{a} = (a_1, a_2)$ i retningen $\mathbf{r} = (r_1, r_2)$. Dersom h er en liten størrelse, vet vi at

$$f(a_1 + hr_1, a_2 + hr_2) - f(a_1, a_2) = f'(\mathbf{a}; \mathbf{r}) \cdot h \qquad (2.4.1)$$

når vi ser bort fra en størrelse som er liten sammenlignet med h.

Vi kan også beregne differansen $f(a_1 + hr_1, a_2 + hr_2) - f(a_1, a_2)$ på en annen måte. Istedenfor å gå direkte langs vektoren $\mathbf{r}$ velger vi å gå parallelt med koordinataksene som vist på figur 2.



Figur 3

Vi ser at

$$
\begin{aligned}
f(a_1 + hr_1, a_2 + hr_2) - f(a_1, a_2) &= \\
&= \overbrace{f(a_1 + hr_1, a_2 + hr_2) - f(a_1 + hr_1, a_2)}^{\text{økning parallelt med } y\text{-aksen}} + \\
&\quad + \overbrace{f(a_1 + hr_1, a_2) - f(a_1, a_2)}^{\text{økning parallelt med } x\text{-aksen}}
\end{aligned}
$$

Ser vi bort fra feil som er små sammenlignet med h, får vi videre

$\vec{a} + h\vec{e}_1$?

$$
\begin{aligned}
f(a_1 + hr_1, a_2) - f(a_1, a_2) &= \tfrac{\partial f}{\partial x}(\mathbf{a}) \cdot hr_1 \\
f(a_1 + hr_1, a_2 + hr_2) - f(a_1 + hr_1, a_2) &= \tfrac{\partial f}{\partial y}(\mathbf{a}) \cdot hr_2
\end{aligned}
$$

Dette betyr at

$$
f(a_1 + hr_1, a_2 + hr_2) - f(a_1, a_2) = \frac{\partial f}{\partial x}(\mathbf{a})hr_1 + \frac{\partial f}{\partial y}(\mathbf{a})hr_2 \tag{2.4.2}
$$

(fortsatt med en feil som er liten sammenlignet med h). Sammenligner vi de to uttrykkene (2.4.1) og (2.4.2) vi nå har for $f(a_1+hr_1, a_2+hr_2) - f(a_1, a_2)$, ser vi at

$$
f'(\mathbf{a}; \mathbf{r}) = \frac{\partial f}{\partial x}(\mathbf{a})r_1 + \frac{\partial f}{\partial y}(\mathbf{a})r_2
$$

Gjennomfører vi et tilsvarende argument i n variable, får vi formelen

$$
f'(\mathbf{a}; \mathbf{r}) = \frac{\partial f}{\partial x_1}(\mathbf{a})r_1 + \frac{\partial f}{\partial x_2}(\mathbf{a})r_2 + \ldots + \frac{\partial f}{\partial x_n}(\mathbf{a})r_n
$$

der $\mathbf{r} = (r_1, r_2, \ldots, r_n)$. Husker vi at gradienten til f er gitt ved

$$
\nabla f(\mathbf{a}) = \Big(\frac{\partial f}{\partial x_1}(\mathbf{a}), \frac{\partial f}{\partial x_2}(\mathbf{a}), \ldots, \frac{\partial f}{\partial x_n}(\mathbf{a})\Big),
$$

kan vi skrive formelen ovenfor som et skalarprodukt

$$f'(\mathbf{a};\mathbf{r}) = \nabla f(\mathbf{a}) \cdot \mathbf{r} \tag{2.4.3}$$

Denne formelen gir oss en effektiv måte å regne ut retningsderiverte på; vi finner alle de partiellderiverte, setter dem sammen til en gradient, og tar skalarproduktet mellom gradienten og vektoren $\mathbf{r}$. Beregningen som ledet oss frem til (2.4.3), er imidlertid ingen streng utledning i matematisk forstand (vi har skrevet eksakt likhet $=$ en rekke steder der vi bare hadde omtrentlig likhet), og det viser seg at det finnes funksjoner f som ikke oppfyller (2.4.3) til tross for at både den retningsderiverte og de partiellderiverte eksisterer (et eksempel på dette er funksjonen i eksempel 2 i seksjon 2.2, se oppgave 7). Disse funksjonene oppfører seg imidlertid så merkelig at vi ønsker å utelukke dem fra teorien vår. Vi skal derfor innføre et begrep *deriverbar funksjon* som fanger opp funksjoner med den oppførselen vi ønsker oss. Utgangspunktet for definisjonen er at vi ønsker at $\nabla f(\mathbf{a}) \cdot \mathbf{r}$ skal være en god tilnærming til funksjonsdifferansen $f(\mathbf{a}+\mathbf{r}) - f(\mathbf{a})$ når $\mathbf{r}$ er liten. Mer presist ønsker vi at "feilleddet"

$$\sigma(\mathbf{r}) = f(\mathbf{a}+\mathbf{r}) - f(\mathbf{a}) - \nabla f(\mathbf{a}) \cdot \mathbf{r}$$

skal bli mindre og mindre sammenlignet med størrelsen til $\mathbf{r}$, dvs. at

$$\lim_{\mathbf{r}\to\mathbf{0}} \frac{\sigma(\mathbf{r})}{|\mathbf{r}|} = 0$$

Vi får altså denne definisjonen:

Definisjon 2.4.4 *Anta at $f\colon A \to \mathbb{R}$ er definert på en delmengde A av $\mathbb{R}^n$ og at $\mathbf{a}$ er et indre punkt i A. Anta videre at alle de partiellderiverte til f eksisterer i punktet $\mathbf{a}$. Vi sier at f er* deriverbar *i $\mathbf{a}$ dersom funksjonen*

$$\sigma(\mathbf{r}) = f(\mathbf{a}+\mathbf{r}) - f(\mathbf{a}) - \nabla f(\mathbf{a}) \cdot \mathbf{r}$$

går mot 0 hurtigere enn $|\mathbf{r}|$, dvs.

$$\lim_{\mathbf{r}\to\mathbf{0}} \frac{\sigma(\mathbf{r})}{|\mathbf{r}|} = 0$$

Kommentar: Man kan lure på hvorfor vi ikke rett å slett sier at f er deriverbar i $\mathbf{a}$ dersom $f'(\mathbf{a};\mathbf{r}) = \nabla f(\mathbf{a}) \cdot \mathbf{r}$ holder for alle $\mathbf{r}$. Det viser seg imidlertid at definisjonen ovenfor gir en glattere teori hvor delene passer bedre sammen. Det neste resultatet sier dessuten at vi får likheten $f'(\mathbf{a};\mathbf{r}) = \nabla f(\mathbf{a}) \cdot \mathbf{r}$ uansett.

Setning 2.4.5 *Anta at $f\colon A \to \mathbb{R}$ er deriverbar i $\mathbf{a}$. Da er $f'(\mathbf{a};\mathbf{r}) = \nabla f(\mathbf{a}) \cdot \mathbf{r}$ for alle $\mathbf{r}$.*

r

Bevis: Vi har

$$f(\mathbf{a};\mathbf{r}) = \lim_{h\to 0}\frac{f(\mathbf{a}+h\mathbf{r}) - f(\mathbf{a})}{h} = \lim_{h\to 0}\frac{\nabla f(\mathbf{a})\cdot(h\mathbf{r}) + \sigma(h\mathbf{r})}{h} =$$

$$= \lim_{h\to 0}\left(\nabla f(\mathbf{a})\cdot\mathbf{r} + |\mathbf{r}|\frac{\sigma(h\mathbf{r})}{h|\mathbf{r}|}\right) = \nabla f(\mathbf{a})\cdot\mathbf{r}$$

siden $\lim_{h\to 0}\frac{\sigma(h\mathbf{r})}{h|\mathbf{r}|} = 0$. $\square$

Vi tar med et eksempel på hvordan formel (2.4.3) kan brukes til å beregne retningsderiverte.

Eksempel 4: La oss anta at funksjonen

$$f(x,y,z) = x^2y + e^{-yz}$$

er deriverbar. Vi skal finne den retningsderiverte $f'(\mathbf{a};\mathbf{r})$ der $\mathbf{a} = (1,1,1)$ og $\mathbf{r} = (1,-1,1)$. La oss først finne gradienten til f:

$$\frac{\partial f}{\partial x} = 2xy; \quad \frac{\partial f}{\partial y} = x^2 - ze^{-yz}; \quad \frac{\partial f}{\partial z} = -ye^{-yz}$$

Dette gir $\nabla f = (2xy, x^2 - ze^{-yz}, -ye^{-yz})$ og $\nabla f(\mathbf{a}) = \nabla f(1,1,1) = (2, 1-e^{-1}, -e^{-1})$. Følgelig er

$$f'(\mathbf{a};\mathbf{r}) = \nabla f(\mathbf{a})\cdot\mathbf{r} = (2, 1-e^{-1}, -e^{-1})\cdot(1,-1,1) = 1$$

♣

For å kunne bruke formel (2.4.3) trenger vi å vite at funksjonene våre er deriverbare. Det neste resultatet gir oss den informasjonen vi vanligvis trenger. Beviset er ganske langt og komplisert, og egner seg nok best for de ivrigste og flittigste. Et lite ord om terminologi: Vi sier at en funksjon er definert i en *omegn* om $\mathbf{a}$ dersom det finnes en kule $B(\mathbf{a},\epsilon)$ om $\mathbf{a}$ der funksjonen er definert (den kan godt være definert på et større område — poenget er at vi i hvert fall vil sikre oss at den er definert for alle punkter tilstrekkelig nær $\mathbf{a}$).

Teorem 2.4.6 *La $f : A \to \mathbb{R}^m$ være en funksjon av n variable. Anta at alle de partiellderiverte $\frac{\partial f}{\partial x_i}$ er definert i en omegn om $\mathbf{a} \in A$, og at de er kontinuerlige i $\mathbf{a}$. Da er f deriverbar i $\mathbf{a}$.*

Bevis: For at ikke notasjonen skal bli for overveldende, skal vi nøye oss med å bevise setningen for en funksjon $f(x_1, x_2)$ av to variable. Beviset er en litt oppstrammet variant av det argumentet som ledet oss til formel (2.4.3).

Vi velger $\mathbf{r}$ så liten at $\mathbf{a}+\mathbf{r}$ ligger innenfor ~~den~~ kulen der vi vet at de partiellderiverte eksisterer. Vi lar $\mathbf{a}=(a_1,a_2)$, $\mathbf{r}=(r_1,r_2)$ og observerer at

$$\begin{aligned} f(\mathbf{a}+\mathbf{r})-f(\mathbf{a}) &= f(a_1+r_1,a_2+r_2)-f(a_1,a_2) \\ &= f(a_1+r_1,a_2+r_2)-f(a_1+r_1,a_2)+ \\ &\quad +f(a_1+r_1,a_2)-f(a_1,a_2) \end{aligned}$$

Dette er samme type omskrivning som vi foretok da vi regnet oss frem til (2.4.3). Hvis vi tenker på $x\to f(x,a_2)$ som en funksjon av én variabel, forteller middelverdisetningen (se *Kalkulus*, seksjon 6.2) oss at det finnes et punkt c mellom a_1 og a_1+r_1 slik at

$$f(a_1+r_1,a_2)-f(a_1,a_2)=\frac{\partial f}{\partial x_1}(c,a_2)\,r_1$$

Helt tilsvarende kan vi finne et punkt d mellom a_2 og a_2+r_2 slik at

$$f(a_1+r_1,a_2+r_2)-f(a_1+r_1,a_2)=\frac{\partial f}{\partial x_2}(a_1+r_1,d)\,r_2$$

Kombinerer vi de resultatene vi nå har, ser vi at

$$f(\mathbf{a}+\mathbf{r})-f(\mathbf{a})=\frac{\partial f}{\partial x_1}(c,a_2)\,r_1+\frac{\partial f}{\partial x_2}(a_1+r_1,d)\,r_2$$

Trekker vi fra $\nabla f(\mathbf{a})\cdot\mathbf{r}=\frac{\partial f}{\partial x_1}(a_1,a_2)\,r_1+\frac{\partial f}{\partial x_2}(a_1,a_2)\,r_2$ på begge sider, får vi

$$\begin{aligned} &f(\mathbf{a}+\mathbf{r})-f(\mathbf{a})-\nabla f(\mathbf{a})\cdot r= \\ &=\frac{\partial f}{\partial x_1}(c,a_2)\,r_1+\frac{\partial f}{\partial x_2}(a_1+r_1,d)\,r_2-\frac{\partial f}{\partial x_1}(a_1,a_2)\,r_1-\frac{\partial f}{\partial x_2}(a_1,a_2)\,r_2= \\ &=\left(\frac{\partial f}{\partial x_1}(c,a_2)-\frac{\partial f}{\partial x_1}(a_1,a_2)\right)r_1+\left(\frac{\partial f}{\partial x_2}(a_1+r_1,d)-\frac{\partial f}{\partial x_2}(a_1,a_2)\right)r_2 \end{aligned}$$

Sammenligner vi dette med definisjon 2.4.4, ser vi at

$$\sigma(\mathbf{r})=\left(\frac{\partial f}{\partial x_1}(c,a_2)-\frac{\partial f}{\partial x_1}(a_1,a_2)\right)r_1+\left(\frac{\partial f}{\partial x_2}(a_1+r_1,d)-\frac{\partial f}{\partial x_2}(a_1,a_2)\right)r_2$$

Vår oppgave er å vise at $\lim_{\mathbf{r}\to\mathbf{0}}\frac{1}{|\mathbf{r}|}\boldsymbol{\sigma}(\mathbf{r})=\mathbf{0}$. Siden $|r_1|,|r_2|\le|\mathbf{r}|$, får vi

$$|\sigma(\mathbf{r})|\le\left(\left|\frac{\partial f}{\partial x_1}(c,a_2)-\frac{\partial f}{\partial x_1}(a_1,a_2)\right|+\left|\frac{\partial f}{\partial x_2}(a_1+r_1,d)-\frac{\partial f}{\partial x_2}(a_1,a_2)\right|\right)|\mathbf{r}|$$

Deler vi på på $|\mathbf{r}|$, ser vi at

$$\frac{|\sigma(\mathbf{r})|}{|\mathbf{r}|}\le\left(\left|\frac{\partial f}{\partial x_1}(c,a_2)-\frac{\partial f}{\partial x_1}(a_1,a_2)\right|+\left|\frac{\partial f}{\partial x_2}(a_1+r_1,d)-\frac{\partial f}{\partial x_2}(a_1,a_2)\right|\right)$$

Dette uttrykket går mot 0 fordi de partiellderiverte er kontinuerlige i $\mathbf{a} = (a_1, a_2)$, og (c, a_2) og $(a + r_1, d)$ nærmer seg (a_1, a_2) når $\mathbf{r} \to \mathbf{0}$. Dermed er teoremet bevist. □

Sammen med setning 2.4.5 forteller setningen ovenfor oss at så lenge de partiellderiverte er kontinuerlige, kan vi trygt bruke formelen

$$f'(\mathbf{a}; \mathbf{r}) = \nabla f(\mathbf{a}) \cdot \mathbf{r}.$$

Vi har allerede nevnt den geometriske tolkningen av gradienten — at gradienten i punktet $\mathbf{a}$ peker i den retningen hvor funksjonen vokser hurtigst, og at stigningstallet i denne retningen er lik lengden til gradienten. Vi har nå de redskapene som trengs til å bevise dette.

Setning 2.4.7 *Anta at f er deriverbar i $\mathbf{a}$. Da peker gradienten $\nabla f(\mathbf{a})$ i den retningen hvor f vokser hurtigst i punktet $\mathbf{a}$, og stigningstallet til f i denne retningen er $|\nabla f(\mathbf{a})|$.*

Bevis: Hvis $\mathbf{u}$ er en enhetsvektor (dvs. $|\mathbf{u}| = 1$), så forteller den retningsderiverte $f'(\mathbf{a}; \mathbf{u}) = \nabla f(\mathbf{a}) \cdot \mathbf{u}$ oss hvor fort funksjonen vokser i den retningen $\mathbf{u}$ peker. Funksjonen f vokser derfor hurtigst i den retningen $\mathbf{u}$ hvor $\nabla f(\mathbf{a}) \cdot \mathbf{u}$ er størst. Siden $\mathbf{u}$ er en enhetsvektor, forteller Schwarz' ulikhet (setning 1.2.3) oss at

$$|\nabla f(\mathbf{a}) \cdot \mathbf{u}| \leq |\nabla f(\mathbf{a})|\,|\mathbf{u}| = |\nabla f(\mathbf{a})|$$

med likhet bare hvis $\mathbf{u}$ og $\nabla f(\mathbf{a})$ er parallelle. Dette betyr at $|\nabla f(\mathbf{a}) \cdot \mathbf{u}|$ er størst når $\mathbf{u}$ og $\nabla f(\mathbf{a})$ er parallelle. Nå er det to enhetsvektorer som er parallelle med $\nabla f(\mathbf{a})$ — en som peker samme vei som $\nabla f(\mathbf{a})$, og en som peker motsatt vei. Det er lett å se at $\nabla f(\mathbf{a}) \cdot \mathbf{u} = |\nabla f(\mathbf{a})|$ når $\mathbf{u}$ og $\nabla f(\mathbf{a})$ peker samme vei, og at $\nabla f(\mathbf{a}) \cdot \mathbf{u} = -|\nabla f(\mathbf{a})|$ når $\mathbf{u}$ og $\nabla f(\mathbf{a})$ peker motsatt vei. Altså har f sitt største stigningstall $|\nabla f(\mathbf{a})|$ i den retningen som $\nabla f(\mathbf{a})$ peker. □

Eksempel 5: I hvilken retning vokser funksjonen $f(x, y) = x^3 y \sin(\pi x y)$ hurtigst når vi står i punktet $(1, \frac{1}{2})$?

Vi må først finne gradienten. De partiellderiverte er

$$\frac{\partial f}{\partial x} = 3x^2 y \sin(\pi x y) + x^3 y \cos(\pi x y)(\pi y) = 3x^2 y \sin(\pi x y) + \pi x^3 y^2 \cos(\pi x y)$$

$$\frac{\partial f}{\partial y} = x^3 \sin(\pi x y) + x^3 y \cos(\pi x y)(\pi x) = x^3 \sin(\pi x y) + \pi x^4 y \cos(\pi x y)$$

Setter vi inn $(x, y) = (1, \frac{1}{2})$, får vi

$$\frac{\partial f}{\partial x}\left(1, \frac{1}{2}\right) = 3 \cdot 1^2 \cdot \frac{1}{2} \cdot \sin(\pi \cdot 1 \cdot \frac{1}{2}) + \pi \cdot 1^3 \cdot \left(\frac{1}{2}\right)^2 \cdot \cos(\pi \cdot 1 \cdot \frac{1}{2}) = \frac{3}{2}$$

og

$$\frac{\partial f}{\partial y}\left(1,\frac{1}{2}\right) = 1^3 \cdot \sin(\pi \cdot 1 \cdot \frac{1}{2}) + \pi \cdot 1^4 \cdot \frac{1}{2} \cdot \cos(\pi \cdot 1 \cdot \frac{1}{2}) = 1$$

Dermed er

$$\nabla f\left(1,\frac{1}{2}\right) = \left(\frac{3}{2}, 1\right)$$

Funksjonen vokser altså raskest i retningen $(\frac{3}{2}, 1)$, og stigningstallet i denne retningen er

$$|\nabla f\left(1,\frac{1}{2}\right)| = \sqrt{\left(\frac{3}{2}\right)^2 + 1^2} = \sqrt{\frac{9}{4}+1} = \frac{1}{2}\sqrt{13}$$

♣

Vi tar med et litt mer teoretisk resultat som vi får bruk for siden. For funksjoner av én variabel vet vi at dersom den deriverte eksisterer i et punkt, så er funksjonen kontinuerlig i punktet. For funksjoner av flere variable er sammenhengen litt mer subtil — det kan faktisk hende at alle de retningsderiverte eksisterer i et punkt, men at funksjonen likevel ikke er kontinuerlig i punktet (se oppgave 7). Er funksjonen deriverbar i betydningen vi innførte i definisjon 2.4.4, er vi imidlertid på den sikre siden.

Setning 2.4.8 *Anta at $f : A \to \mathbb{R}$ er en funksjon av n variable. Dersom f er deriverbar i et punkt $\mathbf{a} \in A$, så er f kontinuerlig i $\mathbf{a}$.*

Bevis: Ifølge setning 2.3.5 er det nok å vise at $\lim_{\mathbf{x}\to\mathbf{a}} f(\mathbf{x}) = f(\mathbf{a})$. Setter vi $\mathbf{r} = \mathbf{x} - \mathbf{a}$, er dette det samme som å vise at $\lim_{\mathbf{r}\to\mathbf{0}} f(\mathbf{a}+\mathbf{r}) = f(\mathbf{a})$. Siden f er deriverbar i $\mathbf{a}$, vet vi at

$$f(\mathbf{a}+\mathbf{r}) = f(\mathbf{a}) + \nabla f(\mathbf{a}) \cdot \mathbf{r} + \sigma(\mathbf{r})$$

der $\frac{\sigma(\mathbf{r})}{|\mathbf{r}|} \to 0$ når $\mathbf{r} \to \mathbf{0}$. Dette betyr spesielt at $\sigma(\mathbf{r}) \to 0$, og dermed er

$$\lim_{\mathbf{r}\to\mathbf{0}} f(\mathbf{a}+\mathbf{r}) = \lim_{\mathbf{r}\to\mathbf{0}} \left(f(\mathbf{a}) + \nabla f(\mathbf{a}) \cdot \mathbf{r} + \sigma(\mathbf{r})\right) = f(\mathbf{a}) + 0 + 0 = f(\mathbf{a})$$

□

Helt til slutt skal vi se på eksempel som viser en typisk anvendelse av gradienter og partiellderiverte.

Eksempel 6: Anta at en tynn gass oppbevares i en beholder der vi kan justere volumet V og temperaturen T. Det viser seg da at trykket P er proporsjonalt med temperaturen T og omvendt proporsjonalt med volumet V, dvs. $P = k\frac{T}{V}$ der k er en konstant som blant annet avhenger av hvor mye gass det er i beholderen (dette forutsetter egentlig at temperaturen måles i grader Kelvin og ikke Celsius, men det behøver vi ikke bry oss om her). Vi kan tenke på trykket P som en funksjon av T og V:

$$P(T, V) = k\frac{T}{V}$$

Partiellderiverer vi dette uttrykket, får vi

$$\frac{\partial P}{\partial T} = k\frac{1}{V} \quad \text{og} \quad \frac{\partial P}{\partial V} = -k\frac{T}{V^2}$$

Gradienten er dermed

$$\nabla P(T, V) = \left(k\frac{1}{V}, -k\frac{T}{V^2}\right)$$

Δ : Delta

Det er ofte naturlig å spørre hvor mye trykket endrer seg dersom vi gir temperaturen et lite tillegg ΔT og volumet et lite tillegg ΔV. Det eksakte uttrykket for dette tillegget er selvfølgelig

$$\Delta P = P(T + \Delta T, V + \Delta V) - P(T, V),$$

men denne differansen er tung å arbeide med. Vi vet imidlertid at den har en god tilnærming i skalarproduktet av gradienten ∇P med tilvekstvektoren $(\Delta T, \Delta V)$:

$$\Delta P = P(T + \Delta T, V + \Delta V) - P(T, V) \approx \nabla P(T, V) \cdot (\Delta T, \Delta V) =$$

$$= \frac{\partial P}{\partial T}\Delta T + \frac{\partial P}{\partial V}\Delta V = k\frac{1}{V}\Delta T - k\frac{T}{V^2}\Delta V$$

Ved hjelp av denne formelen er det lett å anslå hvor mye trykket endrer seg når vi regulerer volumet og temperaturen.

La oss nå anta at vi regulerer temperaturen og volumet kontinuerlig, og at endringene ΔP, ΔT og ΔT har foregått i løpet av et lite tidsintervall Δt, Endringen per tidsenhet er da

$$\frac{\Delta P}{\Delta t} \approx k\frac{1}{V}\frac{\Delta T}{\Delta t} - k\frac{T}{V^2}\frac{\Delta V}{\Delta t}$$

med bedre tilnærming dess mindre Δt er. Lar vi $\Delta t \to 0$, får vi

$$P'(t) = k\frac{1}{V}T'(t) - k\frac{T}{V^2}V'(t)$$

Ved hjelp av denne formelen kan vi regne ut hvor fort trykket endrer seg dersom vi kjenner endringshastighetene til temperaturen og volumet. Formelen er en forsmak på *kjerneregelen for funksjoner av flere variable.* Vi skal se nærmere på denne regelen i seksjon 2.7. ♣

Oppgaver til seksjon 2.4

1. Finn de partiellderiverte til f.

a) $f(x,y) = x^3y + 3xy^4$ e) $f(x,y,z) = (x+y)e^{-z}$
b) $f(x,y) = \frac{x^2+x^3}{y}$ f) $f(x,y,z) = \frac{z^2\tan x}{1+y^2}$
c) $f(x,y) = \cos(x+y^2)$ g) $f(x,y,z) = z\arctan(x+y)$
d) $f(x,y) = x^2\ln(xy^2)$ h) $f(x,y,z,u) = (z^2+u)e^{-x+3y}$

2. Finn gradienten til funksjonen:

a) $f(x,y) = x^2y$

b) $f(x,y,z) = x\cos(xy^2z)$

c) $f(u,v,w) = we^{u\cos v}$

d) $f(z_1,z_2,z_3) = z_3\arctan(z_1z_2) + e^{z_3}$

3. Finn den retningsderiverte $f'(\mathbf{a};\mathbf{r})$:

a) $f(x,y) = 3xy + y^2$; $\mathbf{a} = (1,2)$; $\mathbf{r} = (3,-1)$
b) $f(x,y) = \ln(x+y^2)$; $\mathbf{a} = (1,0)$; $\mathbf{r} = (-1,1)$
c) $f(x,y,z) = x^2y + z^2$; $\mathbf{a} = (1,0,1)$; $\mathbf{r} = (1,1,-1)$
d) $f(x,y,z) = z\sin(xy)$; $\mathbf{a} = (\frac{\pi}{2},1,0)$; $\mathbf{r} = (2,0,-1)$

4. I hvilken retning vokser funksjonen hurtigst i det angitte punktet

a) $f(x,y) = -x^2y + 7y^3$, $\mathbf{a} = (4,-3)$
b) $f(x,y,z) = (x^2-y^2)e^z$; $\mathbf{a} = (1,-1,3)$
c) $f(x,y,z,u) = xuz^2 - y^2zu$; $\mathbf{a} = (1,0,-2,3)$

5. Volumet til en sylinder med radius r og høyde h er $V = \pi r^2h$. Når høyden og radien varierer, kan vi tenke på dette som en funksjon i to variable $V(r,h) = \pi r^2h$. Forklar at når radien endrer seg fra r til $r+\Delta r$ og høyden endrer seg fra h til $h+\Delta h$, så er endringen i V tilnærmet gitt ved

$$\Delta V \approx \frac{\partial V}{\partial r}\cdot\Delta r + \frac{\partial V}{\partial h}\cdot\Delta h = 2\pi rh\,\Delta r + \pi r^2\Delta h.$$

Anta at du har en sylinder hvor du vet at radien ligger mellom 2 m og 2.05 m og hvor høyden ligger mellom 5 m og 5.05 m. Bruk formelen ovenfor til å anslå usikkerheten i volumet.

6. BMI (body mass index) er en indikator for undervekt og overvekt. For å finne din BMI tar du vekten din (målt i kilo) og deler på kvadratet av høyden din (målt i meter). Du kan tenke på BMI som en funksjon av to variable

$$f(v,h) = \frac{v}{h^2}$$

a) Vis at dersom Δv og Δh er små endringer i vekt og høyde, så er endringen i BMI gitt ved

$$\Delta f(v,h) \approx \frac{\Delta v}{h^2} - 2\frac{v}{h^3}\Delta h$$

b) En tommelfingerregel sier at for hver ekstra centimeter du har i høyde, kan du "tåle" en ekstra kilo uten at BMI-en din endrer seg særlig. Bruk formelen i a) til å undersøke hvor godt dette passer for personer med forskjellig vekt og høyde.

7. Vi skal se mer på funksjonen $f : \mathbb{R}^2 \to \mathbb{R}$ fra eksempel 2 i seksjon 2.2. Husk at denne funksjonen er gitt ved

$$f(x,y) = \begin{cases} \frac{x^2 y}{x^4+y^2} & \text{for } (x,y) \neq (0,0) \\ 0 & \text{for } (x,y) = (0,0) \end{cases}$$

a) Vis at $\frac{\partial f}{\partial x}(0,0) = 0$ og $\frac{\partial f}{\partial y}(0,0) = 0$. Hva er $\nabla f(0,0)$?

b) Bruk definisjonen av retningsderivert til å vise at $f(\mathbf{0};\mathbf{r}) = \frac{r_1^2}{r_2}$ der $\mathbf{r} = (r_1, r_2)$, $r_2 \neq 0$.

c) Vis at for denne funksjonen gjelder ikke likheten $f(\mathbf{0};\mathbf{r}) = \nabla f(\mathbf{0}) \cdot \mathbf{r}$.

d) Vis at alle de retningsderiverte til f eksisterer i $\mathbf{0}$, men at funksjonen hverken er kontinuerlig eller deriverbar i punktet.

2.5 Partiellderiverte av høyere orden

Fra teorien for funksjoner av én variabel vet vi at det ofte er nyttig eller nødvendig å derivere mer enn én gang. Også i flervariabel teori er det ofte nyttig å arbeide med annenderiverte, tredjederiverte osv. Den store forskjellen er at vi har så mange flere måter å derivere på.

Eksempel 1: La $f(x,y) = x^2y^3 + y^2$. Vi har to partiellderiverte av første orden

$$\frac{\partial f}{\partial x} = 2xy^3 \qquad \text{og} \qquad \frac{\partial f}{\partial y} = 3x^2y^2 + 2y$$

Når vi skal regne ut annenderiverte, har vi mange valg. Vi kan for eksempel derivere $\frac{\partial f}{\partial x}$ med hensyn på x en gang til:

Ny notasjon!

$$\frac{\partial^2 f}{\partial x^2} = \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial x}\right) = \frac{\partial}{\partial x}(2xy^3) = 2y^3.$$

Vi kan også derivere $\frac{\partial f}{\partial x}$ med hensyn på y:

$$\frac{\partial^2 f}{\partial y \partial x} = \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right) = \frac{\partial}{\partial y}(2xy^3) = 6xy^2.$$

I tillegg kan vi derivere $\frac{\partial f}{\partial y}$ med hensyn på både x og y:

$$\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right) = \frac{\partial}{\partial x}(3x^2y^2 + 2y) = 6xy^2$$

$$\frac{\partial^2 f}{\partial y^2} = \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial y}\right) = \frac{\partial}{\partial y}(3x^2y^2 + 2y) = 6x^2y + 2$$

Vi har altså fire annenordens partiellderiverte for denne funksjonen: $\frac{\partial^2 f}{\partial x^2}$, $\frac{\partial^2 f}{\partial y \partial x}$, $\frac{\partial^2 f}{\partial x \partial y}$ og $\frac{\partial^2 f}{\partial y^2}$. ♣

Den generelle notasjonen skulle fremgå av eksemplet ovenfor —

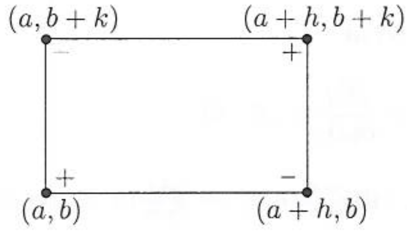
$$\frac{\partial^n f}{\partial x_{i_n} \dots \partial x_{i_2} \partial x_{i_1}}$$

er den funksjonen vi får ved å derivere funksjonen f n ganger, først med hensyn på variabelen x_{i_1}, så med hensyn på variabelen x_{i_2} osv.

I eksemplet ovenfor så vi at de to "blandede" partiellderiverte $\frac{\partial^2 f}{\partial x \partial y}$ og $\frac{\partial^2 f}{\partial y \partial x}$ var like. Dette er ikke en universell regel; det finnes funksjoner f slik at $\frac{\partial^2 f}{\partial x \partial y}$ og $\frac{\partial^2 f}{\partial y \partial x}$ er forskjellige, men for de fleste vi støter på i praksis, vil de blandede partiellderiverte være like. Den neste setningen viser at dette gjelder dersom de annenordens partiellderiverte eksisterer i en omegn rundt punktet $\mathbf{a}$ og er kontinuerlige i $\mathbf{a}$. Beviset er krevende og minner om beviset for teorem 2.4.6. I oppgave 4 finner du et eksempel på en funksjon der de blandede partiellderiverte *ikke* er like.

Setning 2.5.1 *La $f(x_1, \dots, x_n)$ være en funksjon av n variable. Anta at $\frac{\partial^2 f}{\partial x_i \partial x_j}$ og $\frac{\partial^2 f}{\partial x_j \partial x_i}$ eksisterer i en omegn om punktet $\mathbf{a}$ og er kontinuerlige i $\mathbf{a}$. Da er $\frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{a}) = \frac{\partial^2 f}{\partial x_j \partial x_i}(\mathbf{a})$.*

Bevis: For å forenkle notasjonen antar vi at $f(x, y)$ er en funksjon av to variable, og at $\frac{\partial^2 f}{\partial x \partial y}$ og $\frac{\partial^2 f}{\partial y \partial x}$ eksisterer i en omegn om punktet (a, b) og er kontinuerlige i (a, b). Anta at tallene h, k er så små at hele rektangelet i figur 1 ligger i det området der de blandede partiellderiverte eksisterer.



Figur 1

La

$$\Delta(h, k) = f(a+h, b+k) - f(a, b+k) - f(a+h, b) + f(a, b)$$

der vi har kombinert funksjonsverdiene i hjørnene på rektanglet vårt ved å bruke fortegnene vist på figuren. Vi skal vise at grenseverdien

$$\lim_{(h,k)\to(0,0)} \frac{\Delta(h, k)}{h \cdot k}$$

hk ?

er lik både $\frac{\partial^2 f}{\partial x \partial y}(a,b)$ og $\frac{\partial^2 f}{\partial y \partial x}(a,b)$.

Først skritt er å bruke middelverdisetningen på funksjonen

$$g(x) = f(x, b+k) - f(x,b).$$

Vi får

$$g(a+h) - g(a) = g'(c) \cdot h$$

for en c mellom a og $a+h$. Setter vi inn den opprinnelige funksjonen, ser vi at

$$\begin{aligned} f(a+h, b+k) - f(a+h, b) - f(a, b+k) + f(a,b) = \\ = \Big[\frac{\partial f}{\partial x}(c, b+k) - \frac{\partial f}{\partial x}(c,b)\Big] h \end{aligned}$$

Dette kan også skrives

$$\Delta(h,k) = \Big[\frac{\partial f}{\partial x}(c, b+k) - \frac{\partial f}{\partial x}(c,b)\Big] h$$

Neste steg er å bruke middelverdisetningen på funksjonen

$$G(y) = \frac{\partial f}{\partial x}(c, y).$$

Vi får

$$G(b+k) - G(b) = G'(d) \cdot k$$

for en d mellom b og $b+k$. Dette kan også skrives

$$\frac{\partial f}{\partial x}(c, b+k) - \frac{\partial f}{\partial x}(c,b) = \frac{\partial^2 f}{\partial y \partial x}(c,d) \cdot k$$

Kombinerer vi formlene våre, ser vi at

$$\Delta(h,k) = \frac{\partial^2 f}{\partial y \partial x}(c,d) \cdot hk$$

Siden $\frac{\partial^2 f}{\partial y \partial x}$ er kontinuerlig i (a,b), vil $\frac{\partial^2 f}{\partial y \partial x}(c,d) \to \frac{\partial^2 f}{\partial y \partial x}(a,b)$ når $(h,k) \to 0$. Følgelig er

$$\lim_{(h,k)\to(0,0)} \frac{\Delta(h,k)}{hk} = \frac{\partial^2 f}{\partial y \partial x}(a,b)$$

For å vise at også $\lim\limits_{(h,k)\to(0,0)} \frac{\Delta(h,k)}{hk} = \frac{\partial^2 f}{\partial x \partial y}(a,b)$, bytter vi om på rollene til variablene x og y i argumentet ovenfor. Vi starter med å bruke middelverdisetningen på funksjonen

$$\gamma(y) = f(a+h, y) - f(a, y)$$

og fortsetter på akkurat samme måte som ovenfor. Detaljene overlates til leserne. □

At blandede partiellderiverte av annen orden er like, medfører også at blandede partiellderiverte av høyere orden er like dersom de inneholder like mange derivasjoner med hensyn på hver variabel. Dersom f har kontinuerlige fjerdederiverte, kan vi for eksempel vise at

$$\frac{\partial^4 f}{\partial x \partial y \partial z \partial x} = \frac{\partial^4 f}{\partial z \partial y \partial x \partial x}$$

på følgende måte:

$$\frac{\partial^4 f}{\partial x \partial y \partial z \partial x} = \frac{\partial^4 f}{\partial x \partial z \partial y \partial x} = \frac{\partial^4 f}{\partial z \partial x \partial y \partial x} = \frac{\partial^4 f}{\partial z \partial y \partial x \partial x}$$

Overbevis deg selv om at du kan begrunne disse overgangene.

Oppgaver til seksjon 2.5

1. Regn ut de annenordens partiellderiverte til funksjonene:

a) $f(x,y) = 3x^2y + 2y^2x$
b) $f(x,y) = x \sin y$
c) $f(x,y) = x^2e^{x-y}$
d) $f(x,y,z) = x^2z - y^2z^2$

2. Regn ut de partiellderiverte:

a) $\frac{\partial^3 f}{\partial x \partial z \partial x}$ når $f(x,y,z) = x^2ye^{xz}$.

b) $\frac{\partial^4 f}{\partial y \partial z \partial x \partial z}$ når $f(x,y,z) = x^2y^3 \cos xyz$.

3. Gjennomfør den siste delen av beviset for setning 2.5.1 (dvs. at $\lim_{(h,k)\to(0,0)} \frac{\Delta(h,k)}{hk} = \frac{\partial^2 f}{\partial x \partial y}(a,b)$).

4. I denne oppgaven skal vi se på en funksjon $f : \mathbb{R}^2 \to \mathbb{R}$ slik at $\frac{\partial^2 f}{\partial x \partial y}(0,0) \neq \frac{\partial^2 f}{\partial y \partial x}(0,0)$. Funksjonen er gitt ved

$$f(x,y) = \begin{cases} \frac{x^3y - xy^3}{x^2+y^2} & \text{når } (x,y) \neq (0,0) \\ 0 & \text{når } (x,y) = (0,0) \end{cases}$$

a) Vis at $f(x,0) = 0$ for alle x og at $f(0,y) = 0$ for alle y. Bruk dette til å vise at $\frac{\partial f}{\partial x}(0,0) = 0$ og $\frac{\partial f}{\partial y}(0,0) = 0$.

b) Vis at for $(x,y) \neq (0,0)$ er

$$\frac{\partial f}{\partial x}(x,y) = \frac{y(x^4+4x^2y^2-y^4)}{(x^2+y^2)^2}$$

$$\frac{\partial f}{\partial y}(x,y) = -\frac{x(y^4+4x^2y^2-x^4)}{(x^2+y^2)^2}$$

c) Vis at $\frac{\partial^2 f}{\partial y \partial x}(0,0) = -1$ ved å bruke

$$\frac{\partial^2 f}{\partial y \partial x}(0,0) = \lim_{h\to 0} \frac{\frac{\partial f}{\partial x}(0,h) - \frac{\partial f}{\partial x}(0,0)}{h}$$

Vis på tilsvarende måte at $\frac{\partial^2 f}{\partial x \partial y}(0,0) = 1$.

2.6 Derivasjon av vektorvaluerte funksjoner

Hittil har vi sett på derivasjon av funksjoner som tar verdier i $\mathbb{R}$. Vi skal nå utvide teorien vår til også å omfatte funksjoner som tar verdier i $\mathbb{R}^m$ der m er større enn 1, altså til *vektorvaluerte* funksjoner. Husk at en funksjon $\mathbf{F} : \mathbb{R}^n \to \mathbb{R}^m$ kan skrives på komponentform

$$\mathbf{F}(\mathbf{x}) = \begin{pmatrix} F_1(\mathbf{x}) \\ F_2(\mathbf{x}) \\ \vdots \\ F_m(\mathbf{x}) \end{pmatrix}$$

der $F_1, F_2, \ldots, F_m$ er funksjoner med verdier i $\mathbb{R}$ (i denne seksjonen er det lurt å tenke på de fleste vektorer som søylevektorer siden vi skal multiplisere dem med matriser). Dersom alle disse funksjonen lar seg derivere, kan vi samle alle de partiellderiverte i en stor matrise, den såkalte *Jacobi-matrisen* til $\mathbf{F}$:

$$\mathbf{F}'(\mathbf{a}) = \begin{pmatrix} \frac{\partial F_1}{\partial x_1}(\mathbf{a}) & \frac{\partial F_1}{\partial x_2}(\mathbf{a}) & \cdots & \frac{\partial F_1}{\partial x_n}(\mathbf{a}) \\ \frac{\partial F_2}{\partial x_1}(\mathbf{a}) & \frac{\partial F_2}{\partial x_2}(\mathbf{a}) & \cdots & \frac{\partial F_2}{\partial x_n}(\mathbf{a}) \\ \vdots & \vdots & \cdots & \vdots \\ \frac{\partial F_m}{\partial x_1}(\mathbf{a}) & \frac{\partial F_m}{\partial x_2}(\mathbf{a}) & \cdots & \frac{\partial F_m}{\partial x_n}(\mathbf{a}) \end{pmatrix}$$

Legg merke til hvordan denne matrisen er bygget opp: I første linje har vi alle partiellderiverte av førstekomponent F_1, i annen linje alle partiellderiverte av annenkomponent F_2 osv. Sagt på en annen måte: Første linje i Jacobi-matrisen er gradienten til F_1, andre linje er gradienten til F_2 osv. Dersom $\mathbf{F}$ er et skalarfelt (og altså tar verdier i $\mathbb{R}$), har Jacobi-matrisen bare én linje og er identisk med gradienten $\nabla\mathbf{F}$.

Jacobi-matrisen spiller på mange måter den samme rollen for funksjoner av flere variable som den deriverte gjør for funksjoner av én variabel, og vi har derfor valgt en notasjon $\mathbf{F}'(\mathbf{a})$ som ligner på den notasjonen vi er vant til.

Hvis man først har lært seg å partiellderivere, er det ingen kunst (men en del arbeid!) å finne en Jacobi-matrise.

Eksempel 1: Finn Jacobi-matrisen til funksjonen

$$\mathbf{F}(x,y) = \begin{pmatrix} xy^3 \\ e^{x+y^2} \\ 3x^2y \end{pmatrix}$$

I dette tilfellet er

$$F_1(x,y) = xy^3, \quad F_2(x,y) = e^{x+y^2} \quad \text{og} \quad F_3(x,y) = 3x^2y$$

Vi partiellderiverer og får

$$\frac{\partial F_1}{\partial x} = y^3 \quad \text{og} \quad \frac{\partial F_1}{\partial y} = 3xy^2$$

$$\frac{\partial F_2}{\partial x} = e^{x+y^2} \quad \text{og} \quad \frac{\partial F_2}{\partial y} = 2ye^{x+y^2}$$

$$\frac{\partial F_3}{\partial x} = 6xy \quad \text{og} \quad \frac{\partial F_3}{\partial y} = 3x^2$$

Jacobi-matrisen blir dermed

$$\mathbf{F}'(x,y) = \begin{pmatrix} y^3 & 3xy^2 \\ e^{x+y^2} & 2ye^{x+y^2} \\ 6xy & 3x^2 \end{pmatrix}$$

♣

Vi skal nå forsøke å finne ut hvordan vi bør definere deriverbarhet for vektorvaluerte funksjoner $\mathbf{F}$. La oss gå tilbake til den generelle Jacobi-matrisen

$$\mathbf{F}'(\mathbf{a}) = \begin{pmatrix} \frac{\partial F_1}{\partial x_1}(\mathbf{a}) & \frac{\partial F_1}{\partial x_2}(\mathbf{a}) & \cdots & \frac{\partial F_1}{\partial x_n}(\mathbf{a}) \\ \frac{\partial F_2}{\partial x_1}(\mathbf{a}) & \frac{\partial F_2}{\partial x_2}(\mathbf{a}) & \cdots & \frac{\partial F_2}{\partial x_n}(\mathbf{a}) \\ \vdots & \vdots & \cdots & \vdots \\ \frac{\partial F_m}{\partial x_1}(\mathbf{a}) & \frac{\partial F_m}{\partial x_2}(\mathbf{a}) & \cdots & \frac{\partial F_m}{\partial x_n}(\mathbf{a}) \end{pmatrix}$$

Ganger vi denne med en søylevektor

$$\mathbf{r} = \begin{pmatrix} r_1 \\ r_2 \\ \vdots \\ r_n \end{pmatrix}$$

får vi

$$\mathbf{F}'(\mathbf{a})\mathbf{r} = \begin{pmatrix} \frac{\partial F_1}{\partial x_1}(\mathbf{a}) & \frac{\partial F_1}{\partial x_2}(\mathbf{a}) & \cdots & \frac{\partial F_1}{\partial x_n}(\mathbf{a}) \\ \frac{\partial F_2}{\partial x_1}(\mathbf{a}) & \frac{\partial F_2}{\partial x_2}(\mathbf{a}) & \cdots & \frac{\partial F_2}{\partial x_n}(\mathbf{a}) \\ \vdots & \vdots & \cdots & \vdots \\ \frac{\partial F_m}{\partial x_1}(\mathbf{a}) & \frac{\partial F_m}{\partial x_2}(\mathbf{a}) & \cdots & \frac{\partial F_m}{\partial x_n}(\mathbf{a}) \end{pmatrix} \begin{pmatrix} r_1 \\ r_2 \\ \vdots \\ r_n \end{pmatrix} = \begin{pmatrix} \nabla F_1(\mathbf{a}) \cdot \mathbf{r} \\ \nabla F_2(\mathbf{a}) \cdot \mathbf{r} \\ \vdots \\ \nabla F_m(\mathbf{a}) \cdot \mathbf{r} \end{pmatrix}$$

der vi har brukt at første rad i $\mathbf{F}'(\mathbf{a})$ er gradienten til F_1, andre rad er gradienten til F_2 osv. Dersom $F_1, F_2, \ldots, F_n$ er deriverbare funksjoner, er

$$\begin{aligned} \nabla F_1(\mathbf{a}) \cdot \mathbf{r} &\approx F_1(\mathbf{a}+\mathbf{r}) - F_1(\mathbf{a}) \\ \nabla F_2(\mathbf{a}) \cdot \mathbf{r} &\approx F_2(\mathbf{a}+\mathbf{r}) - F_2(\mathbf{a}) \\ \vdots \quad & \vdots \quad \vdots \quad \vdots \\ \nabla F_m(\mathbf{a}) \cdot \mathbf{r} &\approx F_m(\mathbf{a}+\mathbf{r}) - F_m(\mathbf{a}) \end{aligned}$$

med bedre tilnærming dess mindre $\mathbf{r}$ er. Altså er

$$\mathbf{F}'(\mathbf{a})\mathbf{r} \approx \begin{pmatrix} F_1(\mathbf{a}+\mathbf{r}) - F_1(\mathbf{a}) \\ F_2(\mathbf{a}+\mathbf{r}) - F_2(\mathbf{a}) \\ \vdots \\ F_m(\mathbf{a}+\mathbf{r}) - F_m(\mathbf{a}) \end{pmatrix} = \mathbf{F}(\mathbf{a}+\mathbf{r}) - \mathbf{F}(\mathbf{a})$$

Dette gir oss en indikasjon på hvordan vi skal definere deriverbarhet for vektorvaluerte funksjoner — vi ønsker at avviket i tilnærmingen

$$\mathbf{F}(\mathbf{a}+\mathbf{r}) - \mathbf{F}(\mathbf{a}) \approx \mathbf{F}'(\mathbf{a})\mathbf{r}$$

skal være lite sammenlignet med størrelsen til $\mathbf{r}$:

Definisjon 2.6.1 *Anta at $\mathbf{F} : A \to \mathbb{R}^m$ er en funksjon av n variable og at $\mathbf{a}$ er et indre punkt i A. Vi sier at $\mathbf{F}$ er ~~deriverbar~~ deriverbar i $\mathbf{a}$ dersom funksjonen*

$$\boldsymbol{\sigma}(\mathbf{r}) = \mathbf{F}(\mathbf{a}+\mathbf{r}) - \mathbf{F}(\mathbf{a}) - \mathbf{F}'(\mathbf{a})\mathbf{r}$$

(med derivert $\vec{F}'(\vec{a})$)

går mot null fortere enn $|\mathbf{r}|$*, dvs. at*

$$\lim_{\mathbf{r}\to\mathbf{0}} \frac{1}{|\mathbf{r}|}\boldsymbol{\sigma}(\mathbf{r}) = \mathbf{0}$$

(legg merke til at $\boldsymbol{\sigma}$ *nå er en vektorvaluert funksjon med verdier i* $\mathbb{R}^m$*).*

Mange vil kanskje frykte at vi nå må begynne helt forfra med å utvikle en teori for deriverbarhet akkurat som vi gjorde for skalarfelt i seksjon 2.4, men takket være den neste setningen er det unødvendig.

Setning 2.6.2 *En funksjon* $\mathbf{F} : A \to \mathbb{R}^m$ *er deriverbar i et indre punkt* $\mathbf{a} \in A$ *hvis og bare hvis hver komponent* F_i *er deriverbar i* $\mathbf{a}$.

Bevis: Vi har

$$\begin{aligned}
\boldsymbol{\sigma}(\mathbf{r}) &= \mathbf{F}(\mathbf{a}+\mathbf{r}) - \mathbf{F}(\mathbf{a}) - \mathbf{F}'(\mathbf{a})\mathbf{r} = \\
&= \begin{pmatrix} F_1(\mathbf{a}+\mathbf{r}) - F_1(\mathbf{a}) - \nabla \mathbf{F}_1(\mathbf{a})\cdot\mathbf{r} \\ F_2(\mathbf{a}+\mathbf{r}) - F_2(\mathbf{a}) - \nabla \mathbf{F}_2(\mathbf{a})\cdot\mathbf{r} \\ \vdots \\ F_m(\mathbf{a}+\mathbf{r}) - F_m(\mathbf{a}) - \nabla \mathbf{F}_m(\mathbf{a})\cdot\mathbf{r} \end{pmatrix} = \begin{pmatrix} \sigma_1(\mathbf{r}) \\ \sigma_2(\mathbf{r}) \\ \vdots \\ \sigma_m(\mathbf{r}) \end{pmatrix}
\end{aligned}$$

F_1
F_2 (ikke fete)
$\vdots$
F_m

Dermed er

$$\frac{1}{|\mathbf{r}|}\boldsymbol{\sigma}(\mathbf{r}) = \begin{pmatrix} \frac{\sigma_1(\mathbf{r})}{|\mathbf{r}|} \\ \frac{\sigma_2(\mathbf{r})}{|\mathbf{r}|} \\ \vdots \\ \frac{\sigma_m(\mathbf{r})}{|\mathbf{r}|} \end{pmatrix}$$

og vi vet fra setning 2.3.3 at $\frac{1}{|\mathbf{r}|}\boldsymbol{\sigma}(\mathbf{r}) \to \mathbf{0}$ hvis og bare hvis $\frac{\sigma_i(\mathbf{r})}{|\mathbf{r}|} \to 0$ for alle i. Dette betyr at $\mathbf{F}$ er deriverbar hvis og bare hvis hver F_i er deriverbar. □

Kombinerer vi dette resultatet med setning 2.4.6, får vi:

Korollar 2.6.3 *Anta at* $\mathbf{F} : A \to \mathbb{R}^m$ *er en funksjon av n variable og at* $\mathbf{a}$ *er et indre punkt i A. Dersom alle komponentene* $\frac{\partial F_i}{\partial x_j}$ *i Jacobi-matrisen er definert i en omegn rundt* $\mathbf{a}$ *og er kontinuerlige i* $\mathbf{a}$*, så er* $\mathbf{F}$ *deriverbar i* $\mathbf{a}$.

Bevis: Ifølge setning 2.4.6 er hver av komponentene F_i til $\mathbf{F}$ deriverbar i $\mathbf{a}$, og ifølge setningen ovenfor er da $\mathbf{F}$ deriverbar i $\mathbf{a}$. □

Eksempel 2: Vis at funksjonen

$$\mathbf{F}(x,y) = \begin{pmatrix} xy^3 \\ e^{x+y^2} \\ 3x^2y \end{pmatrix}$$

fra eksempel 1 er deriverbar.

Vi har allerede regnet ut Jacobi-matrisen til $\mathbf{F}$, og komponentene er åpenbart kontinuerlige overalt. Altså er $\mathbf{F}$ deriverbar ifølge korollaret. □

Den neste setningen kan se litt underlig ut, men vi kommer til å ha stor glede av den. Setningen sier at ingen annen matrise B kan "stjele jobben" til Jacobi-matrisen $\mathbf{F}'(\mathbf{a})$ — dersom B tilfredsstiller en betingelse av samme type som $\mathbf{F}'(\mathbf{a})$ oppfyller i definisjon 2.6.1, så må B være lik $\mathbf{F}'(\mathbf{a})$. Det finnes også andre tolkninger av setningen som vi skal komme tilbake til senere.

Setning 2.6.4 *La $\mathbf{F} : A \to \mathbb{R}^m$ være en funksjon av n variable og la $\mathbf{a}$ være et indre punkt i A. Anta at det finnes en $m \times n$-matrise B slik at funksjonen*

$$\hat{\sigma}(\mathbf{r}) = \mathbf{F}(\mathbf{a}+\mathbf{r}) - \mathbf{F}(\mathbf{a}) - B\mathbf{r}$$

tilfredsstiller

$$\lim_{\mathbf{r}\to\mathbf{0}} \frac{1}{|\mathbf{r}|}\hat{\sigma}(\mathbf{r}) = \mathbf{0}.$$

Da er $\mathbf{F}$ deriverbar i $\mathbf{a}$ og

$$\mathbf{F}'(\mathbf{a}) = B$$

Bytte $i \leftrightarrow j$?

Bevis: Vi velger $\mathbf{r} = r\mathbf{e}_i$ der

$$\mathbf{e}_i = \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix}$$

er den i-te enhetsvektoren og der $r \in \mathbb{R}$. Da er

$$\mathbf{F}(\mathbf{a}+r\mathbf{e}_i) - \mathbf{F}(\mathbf{a}) = B(r\mathbf{e}_i) + \hat{\sigma}(r\mathbf{e}_i)$$

Deler vi på r og bruker at $|r\mathbf{e}_i| = r$, får vi

$$\frac{\mathbf{F}(\mathbf{a}+r\mathbf{e}_i) - \mathbf{F}(\mathbf{a})}{r} = B\mathbf{e}_i + \frac{\hat{\sigma}(r\mathbf{e}_i)}{|r\mathbf{e}_i|}$$

Når $r \to 0$, vil det siste leddet på høyre side gå mot $\mathbf{0}$ ifølge antagelsen, så høyresiden går mot

$$B\mathbf{e}_i = \begin{pmatrix} b_{1i} \\ b_{2i} \\ \vdots \\ b_{mi} \end{pmatrix}$$

Da må venstresiden konvergere mot det samme, og vi får

$$\lim_{r\to 0}\frac{\mathbf{F}(\mathbf{a}+r\mathbf{e}_i)-\mathbf{F}(\mathbf{a})}{r}=\begin{pmatrix}\lim_{r\to 0}\frac{F_1(\mathbf{a}+r\mathbf{e}_i)-F_1(\mathbf{a})}{r}\\ \lim_{r\to 0}\frac{F_2(\mathbf{a}+r\mathbf{e}_i)-F_2(\mathbf{a})}{r}\\ \vdots\\ \lim_{r\to 0}\frac{F_m(\mathbf{a}+r\mathbf{e}_i)-F_m(\mathbf{a})}{r}\end{pmatrix}=\begin{pmatrix}b_{1i}\\ b_{2i}\\ \vdots\\ b_{mi}\end{pmatrix}$$

Ved definisjonen av partiellderiverte betyr dette at $\frac{\partial F_1}{\partial x_i}=b_{1i}$, $\frac{\partial F_2}{\partial x_i}=b_{2i}$ og så videre. Altså eksisterer alle de partiellderiverte og $\mathbf{F}'(\mathbf{a})=B$. Deriverbarhet følger nå av at $\lim_{\mathbf{r}\to\mathbf{0}}\frac{1}{|\mathbf{r}|}\hat{\boldsymbol{\sigma}}(\mathbf{r})=\mathbf{0}$. □

Kan derfor kalle $F'(a)$ den deriverte til F i a.

Oppgaver til seksjon 2.6

1. Finn Jacobi-matrisen til funksjonene

a) $\mathbf{F}(x,y)=(x^2y, x+y^2)$.

b) $\mathbf{F}(x,y,z)=(e^{x^2y+z}, xyz^2)$.

c) $\mathbf{F}(x,y)=(x\arctan(xy), x\ln y, xy\cos y^2)$.

d) $\mathbf{F}(x,y,z,u)=(xy\sin(xu^2), z^2u)$.

2.7 Kjerneregelen

For funksjoner av én variabel sier kjerneregelen at den deriverte til den sammensatte funksjonen $h(x)=f(g(x))$ er gitt ved $h'(x)=f'(g(x))g'(x)$. Kjerneregelen for funksjoner av flere variable har akkurat samme form; dersom $\mathbf{H}$ er den sammensatte funksjonen $\mathbf{H}(\mathbf{x})=\mathbf{F}(\mathbf{G}(\mathbf{x}))$, så er

$$\mathbf{H}'(\mathbf{x})=\mathbf{F}'(\mathbf{G}(\mathbf{x}))\mathbf{G}'(\mathbf{x}) \tag{2.7.1}$$

Selv om formen er den samme, er innholdet mer komplisert — uttrykkene $\mathbf{H}'(\mathbf{x})$, $\mathbf{F}'(\mathbf{G}(\mathbf{x}))$ og $\mathbf{G}'(\mathbf{x})$ er nå matriser, og formelen sier at Jacobi-matrisen til $\mathbf{H}$ i punktet $\mathbf{x}$ er lik matriseproduktet av Jacobi-matrisen til $\mathbf{F}$ i punktet $\mathbf{G}(\mathbf{x})$ og Jacobi-matrisen til $\mathbf{G}$ i punktet $\mathbf{x}$.

Formel (2.7.1) uttrykker kjerneregelen i en kortfattet form som er lett å huske, men som ikke er så praktisk å bruke i utregninger. Vi skal derfor nærme oss kjerneregelen fra en litt annen synsvinkel og heller komme tilbake til matriseformuleringen i (2.7.1) etter hvert. Vi tenker oss at vi har en funksjon (et skalarfelt) $f(u_1, u_2, \ldots, u_m)$ av m variable. For hver variabel u_i substituerer vi en funksjon $g_i(x_1, x_2, \ldots, x_n)$ slik at vi får en sammensatt funksjon

$$h(x_1, x_2, \ldots, x_n) = f\big(g_1(x_1, x_2, \ldots, x_n), g_2(x_1, x_2, \ldots, x_n), \ldots, g_m(x_1, x_2 \ldots, x_n)\big)$$

av n variable. Spørsmålet er: dersom vi kjenner alle de partiellderiverte til de opprinnelige funksjonene $f, g_1, g_2, \ldots, g_m$, kan vi da finne de partiellderiverte til den sammensatte funksjonen h? Svaret er ja, men før vi skriver opp formelen som viser oss hvordan vi kan regne ut de partiellderiverte til h, er det lurt å komprimere notasjonen litt slik at vi slipper å arbeide med altfor lange uttrykk. Dersom vi bruker vektornotasjon og skriver

$$\mathbf{u} = (u_1, u_2, \ldots, u_m)$$

$$\mathbf{x} = (x_1, x_2, \ldots, x_n)$$

og

$$\mathbf{g}(\mathbf{x}) = \cancel{(}(g_1(x_1, x_2, \ldots, x_n), g_2(x_1, x_2, \ldots, x_n), \ldots, g_m(x_1, x_2 \ldots, x_n)),$$

kan definisjonen av h skrives

$$h(\mathbf{x}) = f(\mathbf{g}(\mathbf{x}))$$

Dersom $\mathbf{g}$ er deriverbar i punktet $\mathbf{x}$ og f er deriverbar i punktet $\mathbf{u} = \mathbf{g}(\mathbf{x})$, viser det seg at de partiellderiverte til h er gitt ved

$$\frac{\partial h}{\partial x_i}(\mathbf{x}) = \frac{\partial f}{\partial u_1}(\mathbf{u})\frac{\partial g_1}{\partial x_i}(\mathbf{x}) + \frac{\partial f}{\partial u_2}(\mathbf{u})\frac{\partial g_2}{\partial x_i}(\mathbf{x}) + \ldots + \frac{\partial f}{\partial u_m}(\mathbf{u})\frac{\partial g_m}{\partial x_i}(\mathbf{x}) \quad (2.7.2)$$

Denne formelen kaller vi *kjerneregelen på komponentform*, mens (2.7.1) kalles *kjerneregelen på matriseform* — vi skal senere se at selv om formen er forskjellig, så uttrykker de to formlene det samme.

Legg merke til at vi i (2.7.2) evaluerer de partiellderiverte til f i punktet $\mathbf{u} = \mathbf{g}(\mathbf{x})$; det betyr at etter at vi har regnet ut de partiellderiverte til f, må vi erstatte u_1 med $g_1(x_1, \ldots, x_n)$, u_2 med $g_2(x_1, \ldots, x_n)$ osv. Legg også merke til at det er et system i hvilke derivasjoner vi gjør i formel (2.7.2); vi deriverer med hensyn på *alle* "mellomvariablene" $u_1, u_2, \ldots, u_m$, men bare med hensyn på én av "grunnvariablene" $x_1, x_2, \ldots, x_n$, nemlig den x_i-en som inngår i den opprinnelige derivasjonen av h.

Ting blir klarere hvis vi ser på et eksempel:

Eksempel 1: Vi skal se på tilfellet der

$$f(u_1, u_2) = 2u_1u_2^2$$

og

$$g_1(x_1, x_2, x_3) = x_1 x_2 \sin x_3$$
$$g_2(x_1, x_2, x_3) = 3x_1^2 x_2 x_3$$

Vi finner den sammensatte funksjonen $h = f(\mathbf{g}(\mathbf{x}))$ ved å substituere $u_1 =$ $h(\vec{x})$
$g_1(x_1, x_2, x_3) = x_1 x_2 \sin x_3$ og $u_2 = g_2(x_1, x_2, x_3) = 3x_1^2 x_2 x_3$ inn i uttrykket for f:

$$h(x_1, x_2, x_3) = f(u_1, u_2) = 2u_1 u_2^2 =$$
$$= 2(x_1 x_2 \sin x_3)(3x_1^2 x_2 x_3)^2 = 18x_1^5 x_2^3 x_3^2 \sin x_3$$

Vi kan selvfølgelig finne de partiellderiverte til h ved å derivere dette uttrykket på vanlig måte, men la oss se hvordan vi kan bruke kjerneregelen isteden. Ifølge (2.7.2) er

$$\frac{\partial h}{\partial x_1}(\mathbf{x}) = \frac{\partial f}{\partial u_1}(\mathbf{u})\frac{\partial g_1}{\partial x_1}(\mathbf{x}) + \frac{\partial f}{\partial u_2}(\mathbf{u})\frac{\partial g_2}{\partial x_1}(\mathbf{x})$$

Regner vi ut de partiellderiverte til f, g_1 og g_2 og setter inn i dette uttrykket, får vi

$$\frac{\partial h}{\partial x_1}(\mathbf{x}) = \left(2u_2^2\right)(x_2 \sin x_3) + (4u_1 u_2)(6x_1 x_2 x_3)$$

Helt til slutt setter vi inn uttrykkene $u_1 = g_1(x_1, x_2, x_3) = x_1 x_2 \sin x_3$ og $u_2 = g_2(x_1, x_2, x_3) = 3x_1^2 x_2 x_3$:

$$\frac{\partial h}{\partial x_1}(\mathbf{x}) = \left(2(3x_1^2 x_2 x_3)^2\right)(x_2 \sin x_3) + \left(4(x_1 x_2 \sin x_3)(3x_1^2 x_2 x_3)\right)(6x_1 x_2 x_3) =$$
$$= 18x_1^4 x_2^3 x_3^2 \sin x_3 + 72x_1^4 x_2^3 x_3^2 \sin x_3 = 90x_1^4 x_2^3 x_3^2 \sin x_3$$

Det er god trening i å finne $\frac{\partial h}{\partial x_2}(\mathbf{x})$ og $\frac{\partial h}{\partial x_3}(\mathbf{x})$ på tilsvarende måte. ♣

Bemerkning: Hvis du ser nøyere på eksemplet ovenfor, vil du oppdage at vi har regnet ut $\frac{\partial h}{\partial x_1}(\mathbf{x})$ på en svært tungvinn måte — det går mye raskere å bare derivere uttrykket $h(x_1, x_2, x_3) = 18x_1^5 x_2^3 x_3^2 \sin x_3$ på vanlig måte! Dette er ganske typisk; kjerneregelen i flere variable er faktisk ikke en særlig effektiv metode for å regne ut partiellderiverte til funksjoner gitt ved formler — regelens styrke ligger isteden i at den er et utmerket redskap til å utlede *generelle* sammenhenger mellom varierende størrelser. Før vi ser nærmere på dette, kan det være lurt å ta med enda et konkret eksempel slik at vi blir litt bedre kjent med kjerneregelen. -nt

Eksempel 2: I eksemplet ovenfor het de variable u_1, u_2 og x_1, x_2, x_3 akkurat som i vår generelle formel (2.7.2). Når man bruker kjerneregelen i andre fag, har de variable vanligvis helt andre navn, og det er viktig at vi da greier å kjenne igjen de forskjellige delene av kjerneregelen. La oss nå tenke oss at vi har en funksjon

$$f(s, t) = e^s t^2$$

og at s og t er funksjoner av underliggende variable u og v

$$s = S(u,v) = 3uv^3 \qquad \text{og} \qquad t = T(u,v) = \cos(uv)$$

Vi vil bruke kjerneregelen til å finne de partiellderiverte til den sammensatte funksjonen

$$h(u,v) = f(S(u,v), T(u,v))$$

I dette tilfellet sier formel (2.7.2)

$$\frac{\partial h}{\partial u}(u,v) = \frac{\partial f}{\partial s}(s,t)\frac{\partial S}{\partial u}(u,v) + \frac{\partial f}{\partial t}(s,t)\frac{\partial T}{\partial u}(u,v)$$

og

$$\frac{\partial h}{\partial v}(u,v) = \frac{\partial f}{\partial s}(s,t)\frac{\partial S}{\partial v}(u,v) + \frac{\partial f}{\partial t}(s,t)\frac{\partial T}{\partial v}(u,v)$$

(husk at vi skal derivere med hensyn på begge "mellomvariablene" s og t, men bare med hensyn på den "opprinnelige" grunnvariablen u eller v). Regner vi ut de partiellderiverte og setter inn $s = S(u,v) = 3uv^3$ og $t = T(u,v) = \cos(uv)$, får vi

$$\frac{\partial h}{\partial u}(u,v) = e^s t^2 3v^3 + 2e^s t(-\sin(uv)v) =$$

$$= e^{3uv^3}(\cos(uv))^2 3v^3 - 2e^{3uv^3}\cos(uv)\sin(uv)v =$$

$$= ve^{3uv^3}\cos(uv)\left(3v^2\cos(uv) - 2\sin(uv)\right)$$

Den andre partiellderiverte $\frac{\partial h}{\partial v}(u,v)$ finner vi på tilsvarende måte. □

Vi skal ta med et par eksempler på hvordan kjerneregelen brukes i andre fag. Før vi gjør det, kan det være greit med noen ord om notasjon. Når vi bruker kjerneregelen i praktiske situasjoner, blir uttrykkene ofte lange og uoversiktlige. For å lette lesbarheten er det vanlig å utelate punktene som funksjonene evalueres i — man skriver altså

$$\frac{\partial h}{\partial x_i} = \frac{\partial f}{\partial u_1}\frac{\partial g_1}{\partial x_i} + \frac{\partial f}{\partial u_2}\frac{\partial g_2}{\partial x_i} + \cdots + \frac{\partial f}{\partial u_m}\frac{\partial g_m}{\partial x_i}$$

istedenfor

$$\frac{\partial h}{\partial x_i}(\mathbf{x}) = \frac{\partial f}{\partial u_1}(\mathbf{u})\frac{\partial g_1}{\partial x_i}(\mathbf{x}) + \frac{\partial f}{\partial u_2}(\mathbf{u})\frac{\partial g_2}{\partial x_i}(\mathbf{x}) + \ldots + \frac{\partial f}{\partial u_m}(\mathbf{u})\frac{\partial g_m}{\partial x_i}(\mathbf{x})$$

og setter først inn verdiene til $\mathbf{x}$ og $\mathbf{u}$ når man har behov for dem.

Eksempel 3: I økonomi er utbyttet til et firma avhengig av forskjellige faktorer. For en bonde kan vi tenke oss at utbyttet er avhengig av tre faktorer: arbeidsinnsatsen a, den løpende kapitalinvesteringen k, og arealet av dyrkbar mark m. Utbyttet U er dermed en funksjon

$$U = F(a, k, m)$$

Vi tenker oss at de tre faktorene varierer med tiden: $a = A(t)$, $k = K(t)$ og $m = M(t)$. Kjerneregelen forteller oss nå hvordan utbyttet U varierer med tiden:

$$\frac{\partial U}{\partial t} = \frac{\partial F}{\partial a}\frac{\partial A}{\partial t} + \frac{\partial F}{\partial k}\frac{\partial K}{\partial t} + \frac{\partial F}{\partial m}\frac{\partial M}{\partial t}$$

der vi må huske at vi skal evaluere de partiellderiverte $\frac{\partial F}{\partial a}$, $\frac{\partial F}{\partial k}$, $\frac{\partial F}{\partial m}$ i punktet $(A(t), K(t), M(t))$. Legg merke til at i denne formelen er U, A, K og M bare avhengige av én variabel t, og vi kan derfor skrive vanlige deriverte istedenfor partiellderiverte:

$$U'(t) = \frac{\partial F}{\partial a}A'(t) + \frac{\partial F}{\partial k}K'(t) + \frac{\partial F}{\partial m}M'(t)$$

I økonomiske modeller antar man ofte at utbyttefunksjonen F er en *Cobb-Douglasfunksjon*, dvs. at den har formen

$$F(a, k, m) = Ca^{\alpha}k^{\beta}m^{\gamma}$$

der C, α, β og γ er positive konstanter. Deriverer vi dette uttrykket, får vi

$$\frac{\partial F}{\partial a} = C\alpha a^{\alpha-1}k^{\beta}m^{\gamma}, \qquad \frac{\partial F}{\partial k} = C\beta a^{\alpha}k^{\beta-1}m^{\gamma}, \qquad \frac{\partial F}{\partial m} = C\gamma a^{\alpha}k^{\beta}m^{\gamma-1}$$

som innsatt i uttrykket ovenfor gir (etter litt opprydning):

$$U'(t) = CA(t)^{\alpha}K(t)^{\beta}M(t)^{\gamma}\left(\alpha\frac{A'(t)}{A(t)} + \beta\frac{K'(t)}{K(t)} + \gamma\frac{M'(t)}{M(t)}\right)$$

I en økonomisk situasjon med begrensede ressurser vil man ofte overveie å styrke noen innsatsområder på bekostning av andre. Bonden vår kan f.eks. overveie å utvide det dyrkede arealet m ved å kjøpe tilleggsjord, men er da tvunget til å redusere kapitalinvesteringen k. Om slike omprioriteringer er lønnsomme eller ikke, viser seg i fortegnet til den deriverte $U'(t)$ av utbyttefunksjonen. Ifølge formelen ovenfor er dette fortegnet bestemt av fortegnet til uttrykket

$$\alpha\frac{A'(t)}{A(t)} + \beta\frac{K'(t)}{K(t)} + \gamma\frac{M'(t)}{M(t)}$$

($CA(t)^{\alpha}K(t)^{\beta}M(t)^{\gamma}$ er alltid positiv). Fortegnet til den deriverte av utbyttefunksjonen er altså bestemt av en kombinasjon av de *relative veksthastighetene* $\frac{A'(t)}{A(t)}$, $\frac{K'(t)}{K(t)}$ og $\frac{M'(t)}{M(t)}$ og Cobb-Douglaseksponentene α, β og γ. For en økonom med trening i å tolke Cobb-Douglaseksponenter gir et slikt uttrykk økonomisk mening. □

Eksempel 4: Når en gass oppbevares i en beholder, er trykket P en funksjon av temperaturen T og volumet V; vi har altså

$$P = f(T, V)$$

Hvilken funksjon f det er naturlig å bruke, avhenger både av hvilken gass vi ser på, og av hvor stor nøyaktighet vi ønsker. Foreløpig lar vi f være uspesifisert slik at regningene våre gjelder for alle modeller. Dersom volumet og temperaturen varierer med tiden t, vil også trykket variere med tiden: !

$$P(t) = f(T(t), V(t))$$

Vi kan finne den deriverte av denne funksjonen ved å bruke kjerneregelen (husk at $P(t)$, $T(t)$ og $V(t)$ bare avhenger av én variabel t slik at vi kan bruke vanlige deriverte $P'(t)$, $T'(t)$ og $V'(t)$ istedenfor partiellderiverte $\frac{\partial P}{\partial t}$, $\frac{\partial T}{\partial t}$ og $\frac{\partial V}{\partial t}$):

$$P'(t) = \frac{\partial f}{\partial T}T'(t) + \frac{\partial f}{\partial V}V'(t)$$

Dersom gassen er tynn og kreftene mellom gasspartiklene er svake (fysikerne kaller dette en *ideell gass*), kan man anta at trykket P er proporsjonalt med temperaturen T og omvendt proporsjonalt med volumet V, dvs. $P = k\frac{T}{V}$ der k er en konstant som blant annet avhenger av hvor mye gass det er i beholderen. Vi har altså

$$P = f(T, V) = k\frac{T}{V}$$

I dette tilfellet er

$$\frac{\partial f}{\partial T} = k\frac{1}{V} \quad \text{og} \quad \frac{\partial f}{\partial V} = -k\frac{T}{V^2}$$

Setter vi dette inn i formelen for $P'(t)$, får vi

$$P'(t) = k\frac{1}{V}T'(t) - k\frac{T}{V^2}V'(t)$$

Vi har tidligere regnet oss frem til denne formelen på en litt annen måte (se eksempel 6 i seksjon 2.4).

Regning med partiellderiverte og kjerneregelen står sentralt i termodynamikken, og utregningene ovenfor er bare et enkelt eksempel. ♣

Vi begynte denne seksjonen med å skrive opp kjerneregelen på matriseform:

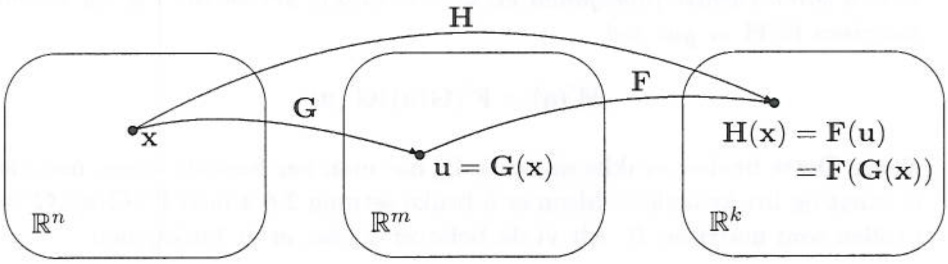
$$\mathbf{H}'(\mathbf{x}) = \mathbf{F}'(\mathbf{G}(\mathbf{x}))\mathbf{G}'(\mathbf{x}),$$

men siden har vi bare arbeidet med regelen på komponentform:

$$\frac{\partial h}{\partial x_i}(\mathbf{x}) = \frac{\partial f}{\partial u_1}(\mathbf{u})\frac{\partial g_1}{\partial x_i}(\mathbf{x}) + \frac{\partial f}{\partial u_2}(\mathbf{u})\frac{\partial g_2}{\partial x_i}(\mathbf{x}) + \ldots + \frac{\partial f}{\partial u_m}(\mathbf{u})\frac{\partial g_m}{\partial x_i}(\mathbf{x})$$

La oss nå se litt på sammenhengen mellom de to formene.

Når vi arbeider med kjerneregelen på matriseform, tenker vi oss at vi starter med to funksjoner $\mathbf{G} : \mathbb{R}^n \to \mathbb{R}^m$, $\mathbf{F} : \mathbb{R}^m \to \mathbb{R}^k$ og at vi danner den sammensatte funksjonen $\mathbf{H} : \mathbb{R}^n \to \mathbb{R}^k$ ved $\mathbf{H}(\mathbf{x}) = \mathbf{F}(\mathbf{G}(\mathbf{x}))$ som vist på figuren.



Skriver vi ut komponentene til $\mathbf{H}'(\mathbf{x})$, $\mathbf{F}'(\mathbf{G}(\mathbf{x}))$, og $\mathbf{G}'(\mathbf{x})$, får kjerneregelen $\mathbf{H}'(\mathbf{x}) = \mathbf{F}'(\mathbf{G}(\mathbf{x}))\mathbf{G}'(\mathbf{x})$ denne formen:

$$\begin{pmatrix} \frac{\partial H_1}{\partial x_1}(\mathbf{x}) & \frac{\partial H_1}{\partial x_2}(\mathbf{x}) & \cdots & \frac{\partial H_1}{\partial x_n}(\mathbf{x}) \\ \frac{\partial H_2}{\partial x_1}(\mathbf{x}) & \frac{\partial H_2}{\partial x_2}(\mathbf{x}) & \cdots & \frac{\partial H_2}{\partial x_n}(\mathbf{x}) \\ \vdots & \vdots & \cdots & \vdots \\ \frac{\partial H_k}{\partial x_1}(\mathbf{x}) & \frac{\partial H_k}{\partial x_2}(\mathbf{x}) & \cdots & \frac{\partial H_k}{\partial x_n}(\mathbf{x}) \end{pmatrix} =$$

$$= \begin{pmatrix} \frac{\partial F_1}{\partial u_1}(\mathbf{G}(\mathbf{x})) & \frac{\partial F_1}{\partial u_2}(\mathbf{G}(\mathbf{x})) & \cdots & \frac{\partial F_1}{\partial u_m}(\mathbf{G}(\mathbf{x})) \\ \frac{\partial F_2}{\partial u_1}(\mathbf{G}(\mathbf{x})) & \frac{\partial F_2}{\partial u_2}(\mathbf{G}(\mathbf{x})) & \cdots & \frac{\partial F_2}{\partial u_m}(\mathbf{G}(\mathbf{x})) \\ \vdots & \vdots & \cdots & \vdots \\ \frac{\partial F_k}{\partial u_1}(\mathbf{G}(\mathbf{x})) & \frac{\partial F_k}{\partial u_2}(\mathbf{G}(\mathbf{x})) & \cdots & \frac{\partial F_k}{\partial u_m}(\mathbf{G}(\mathbf{x})) \end{pmatrix} \begin{pmatrix} \frac{\partial G_1}{\partial x_1}(\mathbf{x}) & \frac{\partial G_1}{\partial x_2}(\mathbf{x}) & \cdots & \frac{\partial G_1}{\partial x_n}(\mathbf{x}) \\ \frac{\partial G_2}{\partial x_1}(\mathbf{x}) & \frac{\partial G_2}{\partial x_2}(\mathbf{x}) & \cdots & \frac{\partial G_2}{\partial x_n}(\mathbf{x}) \\ \vdots & \vdots & \cdots & \vdots \\ \frac{\partial G_m}{\partial x_1}(\mathbf{x}) & \frac{\partial G_m}{\partial x_2}(\mathbf{x}) & \cdots & \frac{\partial G_m}{\partial x_n}(\mathbf{x}) \end{pmatrix}$$

Den ij-te komponenten i den første matrisen er $\frac{\partial H_i}{\partial x_j}(\mathbf{x})$. Denne komponenten må være skalarproduktet av den i-te linjen i den andre matrisen og den j-te søylen i den tredje matrisen, altså: raden

$$\frac{\partial H_i}{\partial x_j}(\mathbf{x}) = \frac{\partial F_i}{\partial u_1}(\mathbf{G}(\mathbf{x}))\frac{\partial G_1}{\partial x_j}(\mathbf{x}) + \frac{\partial F_i}{\partial u_2}(\mathbf{G}(\mathbf{x}))\frac{\partial G_2}{\partial x_j}(\mathbf{x}) + \cdots + \frac{\partial F_i}{\partial u_m}(\mathbf{G}(\mathbf{x}))\frac{\partial G_m}{\partial x_j}(\mathbf{x})$$

Dette er ikke noe annet enn kjerneregelen på komponentform anvendt på funksjonen $H_i(\mathbf{x}) = F_i(\mathbf{G}(\mathbf{x}))$. Vi har dermed vist at kjerneregelen på komponentform rett og slett er det vi får når vi ganger ut matriseproduktet i formelen $\mathbf{H}'(\mathbf{x}) = \mathbf{F}'(\mathbf{G}(\mathbf{x}))\mathbf{G}'(\mathbf{x})$ og sjekker hva som skjer med hver enkelt komponent. Når vi nå skal bevise kjerneregelen, holder det derfor å bevise den på matriseform.

Teorem 2.7.1 (Kjerneregelen på matriseform) *Anta at vi har to mengder $A \subset \mathbb{R}^n$, $B \subset \mathbb{R}^m$ og to funksjoner $\mathbf{G} : A \to B$, $\mathbf{F} : B \to \mathbb{R}^k$. Dersom $\mathbf{G}$ er deriverbar i punktet $\mathbf{a} \in A$, og $\mathbf{F}$ er deriverbar i punktet $\mathbf{b} = \mathbf{G}(\mathbf{a})$, så er den sammensatte funksjonen $\mathbf{H}(\mathbf{x}) = \mathbf{F}(\mathbf{G}(\mathbf{x}))$ deriverbar i $\mathbf{a}$, og Jacobimatrisen til $\mathbf{H}$ er gitt ved*

$$\mathbf{H}'(\mathbf{a}) = \mathbf{F}'(\mathbf{G}(\mathbf{a}))\mathbf{G}'(\mathbf{a})$$

Bevis: Dette beviset er ikke så vanskelig når man har forstått ideen, men det er langt og litt kronglete. Ideen er å bruke setning 2.6.4 med $\mathbf{F}'(\mathbf{G}(\mathbf{a}))\mathbf{G}'(\mathbf{a})$ i rollen som matrisen B. Alt vi da behøver å vise, er at funksjonen

$$\hat{\sigma}(\mathbf{r}) = \mathbf{H}(\mathbf{a}+\mathbf{r}) - \mathbf{H}(\mathbf{a}) - (\mathbf{F}'(\mathbf{G}(\mathbf{a}))\mathbf{G}'(\mathbf{a}))\mathbf{r} \tag{2.7.3}$$

tilfredsstiller kravet

$$\lim_{\mathbf{r}\to\mathbf{0}} \frac{1}{|\mathbf{r}|}\hat{\sigma}(\mathbf{r}) = \mathbf{0} \tag{2.7.4}$$

For å vise dette, må vi bruke at $\mathbf{G}$ og $\mathbf{F}$ er deriverbare. Siden $\mathbf{G}$ er deriverbar i $\mathbf{a}$, vet vi at funksjonen

$$\sigma_1(\mathbf{r}) = \mathbf{G}(\mathbf{a}+\mathbf{r}) - \mathbf{G}(\mathbf{a}) - \mathbf{G}'(\mathbf{a})\mathbf{r} \tag{2.7.5}$$

tilfredsstiller kravet

$$\lim_{\mathbf{r}\to\mathbf{0}} \frac{1}{|\mathbf{r}|}\sigma_1(\mathbf{r}) = \mathbf{0} \tag{2.7.6}$$

Siden $\mathbf{F}$ er deriverbar i $\mathbf{b} = \mathbf{G}(\mathbf{a})$, vet vi tilsvarende at funksjonen

$$\sigma_2(\mathbf{s}) = \mathbf{F}(\mathbf{b}+\mathbf{s}) - \mathbf{F}(\mathbf{b}) - \mathbf{F}'(\mathbf{b})\mathbf{s} = \tag{2.7.7}$$

$$= \mathbf{F}(\mathbf{G}(\mathbf{a})+\mathbf{s}) - \mathbf{F}(\mathbf{G}(\mathbf{a})) - \mathbf{F}'(\mathbf{G}(\mathbf{a}))\mathbf{s}$$

tilfredsstiller kravet

$$\lim_{\mathbf{s}\to\mathbf{0}} \frac{1}{|\mathbf{s}|}\sigma_2(\mathbf{s}) = \mathbf{0} \tag{2.7.8}$$

Vi setter $\mathbf{s} = \mathbf{G}(\mathbf{a}+\mathbf{r}) - \mathbf{G}(\mathbf{a})$. Da er $\mathbf{G}(\mathbf{a}+\mathbf{r}) = \mathbf{G}(\mathbf{a}) + \mathbf{s}$, og fra (2.7.5) ser vi dessuten at $\mathbf{s} = \mathbf{G}'(\mathbf{a})\mathbf{r} + \sigma_1(\mathbf{r})$. Dermed er

$$\mathbf{H}(\mathbf{a}+\mathbf{r}) = \mathbf{F}(\mathbf{G}(\mathbf{a}+\mathbf{r})) = \mathbf{F}(\mathbf{G}(\mathbf{a})+\mathbf{s}) = \mathbf{F}(\mathbf{G}(\mathbf{a})) + \mathbf{F}'(\mathbf{G}(\mathbf{a}))\mathbf{s} + \sigma_2(\mathbf{s})$$

der vi i siste trinn har brukt (2.7.7). Siden $\mathbf{s} = \mathbf{G}'(\mathbf{a})\mathbf{r} + \boldsymbol{\sigma}_1(\mathbf{r})$, gir dette videre

$$\mathbf{H}(\mathbf{a}+\mathbf{r}) = \mathbf{F}(\mathbf{G}(\mathbf{a}))) + \mathbf{F}'(\mathbf{G}(\mathbf{a}))\Big(\mathbf{G}'(\mathbf{a})\mathbf{r} + \boldsymbol{\sigma}_1(\mathbf{r})\Big) + \boldsymbol{\sigma}_2(\mathbf{s}) =$$

$$= \mathbf{H}(\mathbf{a}) + \mathbf{F}'(\mathbf{G}(\mathbf{a}))\mathbf{G}'(\mathbf{a})\mathbf{r} + \mathbf{F}'(\mathbf{G}(\mathbf{a}))\boldsymbol{\sigma}_1(\mathbf{r}) + \boldsymbol{\sigma}_2(\mathbf{s})$$

Sammenligner vi dette med (2.7.3), ser vi at

$$\hat{\boldsymbol{\sigma}}(\mathbf{r}) = \mathbf{F}'(\mathbf{G}(\mathbf{a}))\boldsymbol{\sigma}_1(\mathbf{r}) + \boldsymbol{\sigma}_2(\mathbf{s})$$

Ifølge (2.7.4) er vår oppgave dermed å vise at

$$\lim_{\mathbf{r}\to\mathbf{0}} \frac{1}{|\mathbf{r}|}\hat{\boldsymbol{\sigma}}(\mathbf{r}) = \lim_{\mathbf{r}\to\mathbf{0}} \frac{1}{|\mathbf{r}|}\Big(\mathbf{F}'(\mathbf{G}(\mathbf{a}))\boldsymbol{\sigma}_1(\mathbf{r}) + \boldsymbol{\sigma}_2(\mathbf{s})\Big) = \mathbf{0} \qquad (2.7.9)$$

Vi tar ett ledd av gangen. Fra setning 1.6.3 vet vi at

$$|\mathbf{F}'(\mathbf{G}(\mathbf{a}))\boldsymbol{\sigma}_1(\mathbf{r})| \leq \|\mathbf{F}'(\mathbf{G}(\mathbf{a}))\|\,|\boldsymbol{\sigma}_1(\mathbf{r})|$$

der $\|\mathbf{F}'(\mathbf{G}(\mathbf{a}))\|$ bare er et fast tall (normen til matrisen $\mathbf{F}'(\mathbf{G}(\mathbf{a}))$). Dermed har vi

$$\left|\frac{1}{|\mathbf{r}|}\Big(\mathbf{F}'(\mathbf{G}(\mathbf{a}))\boldsymbol{\sigma}_1(\mathbf{r})\Big)\right| \leq \|\mathbf{F}'(\mathbf{G}(\mathbf{a}))\|\frac{|\boldsymbol{\sigma}_1(\mathbf{r})|}{|\mathbf{r}|} \longrightarrow 0$$

siden $\frac{\boldsymbol{\sigma}_1(\mathbf{r})}{|\mathbf{r}|} \to \mathbf{0}$ når $\mathbf{r} \to \mathbf{0}$. Dette viser at den første delen av uttrykket i (2.7.9) går mot $\mathbf{0}$. Den andre delen

$$\lim_{\mathbf{r}\to\mathbf{0}} \frac{1}{|\mathbf{r}|}\boldsymbol{\sigma}_2(\mathbf{s}) = \mathbf{0}$$

er litt verre. Vi legger først merke til at siden $\boldsymbol{\sigma}_2(\mathbf{0}) = \mathbf{0}$, har vi ingen problemer med tilfellet $\mathbf{s} = \mathbf{0}$. Vi kan derfor anta at $\mathbf{s} \neq \mathbf{0}$, og multiplisere uttrykket $\frac{1}{|\mathbf{r}|}\boldsymbol{\sigma}_2(\mathbf{s})$ med $|\mathbf{s}|$ i teller og nevner:

Litt uklart hvor $\lim_{\vec{r}\to 0}$ dannes ($\vec{s} \neq \vec{0}$) $\vec{G}(\vec{a}) \neq \vec{G}(\vec{a}+r)$

$$\lim_{\mathbf{r}\to\mathbf{0}} \frac{1}{|\mathbf{r}|}\boldsymbol{\sigma}_2(\mathbf{s}) = \lim_{\mathbf{r}\to\mathbf{0}} \frac{|\mathbf{s}|}{|\mathbf{r}|} \cdot \frac{\boldsymbol{\sigma}_2(\mathbf{s})}{|\mathbf{s}|}$$

Når $\mathbf{r} \to \mathbf{0}$, vil $\mathbf{s} = \mathbf{G}(\mathbf{a})\mathbf{r} + \boldsymbol{\sigma}_1(\mathbf{r}) \to \mathbf{0}$, og dette medfører at

$$\lim_{\mathbf{r}\to\mathbf{0}} \frac{\boldsymbol{\sigma}_2(\mathbf{s})}{|\mathbf{s}|} = \lim_{\mathbf{s}\to\mathbf{0}} \frac{\boldsymbol{\sigma}_2(\mathbf{s})}{|\mathbf{s}|} = \mathbf{0}$$

ifølge (2.7.8). Det gjenstår derfor å vise at faktoren $\frac{|\mathbf{s}|}{|\mathbf{r}|}$ er begrenset og ikke kan gå mot uendelig. Bruker vi at $\mathbf{s} = \mathbf{G}'(\mathbf{a})\mathbf{r} + \boldsymbol{\sigma}_1(\mathbf{r})$, får vi

$$\frac{|\mathbf{s}|}{|\mathbf{r}|} = \frac{|\mathbf{G}'(\mathbf{a})\mathbf{r} + \boldsymbol{\sigma}_1(\mathbf{r})|}{|\mathbf{r}|} \leq \frac{|\mathbf{G}'(\mathbf{a})\mathbf{r}|}{|\mathbf{r}|} + \frac{|\boldsymbol{\sigma}_1(\mathbf{r})|}{|\mathbf{r}|} \leq \|\mathbf{G}'(\mathbf{a})\| + \frac{|\boldsymbol{\sigma}_1(\mathbf{r})|}{|\mathbf{r}|}$$

som er begrenset siden $\frac{\sigma_1(\mathbf{r})}{|\mathbf{r}|} \to \mathbf{0}$. Dermed har vi vist at

$$\lim_{\mathbf{r}\to\mathbf{0}} \frac{1}{|\mathbf{r}|}\hat{\boldsymbol{\sigma}}(\mathbf{r}) = \lim_{\mathbf{r}\to\mathbf{0}} \frac{1}{|\mathbf{r}|}\Big(\mathbf{F}'(\mathbf{G}(\mathbf{a}))\boldsymbol{\sigma}_1(\mathbf{r}) + \boldsymbol{\sigma}_2(\mathbf{s})\Big) = \mathbf{0}$$

og beviset er fullført. □

Bemerkning: Legg merke til at kjerneregelen garanterer at $\mathbf{H}$ er deriverbar i $\mathbf{a}$ så lenge $\mathbf{G}$ er deriverbar i $\mathbf{a}$ og $\mathbf{F}$ er deriverbar i $\mathbf{b} = \mathbf{G}(\mathbf{a})$. Siden deriverbarhet kan være vanskelig å sjekke, er denne garantien ofte til stor hjelp i teoretisk arbeid.

Vi skriver også opp den presise formuleringen av kjerneregelen på komponentform:

Teorem 2.7.2 (Kjerneregelen på komponentform) *Anta at vi har to mengder $A \subset \mathbb{R}^n$, $B \subset \mathbb{R}^m$ og to funksjoner $\mathbf{G} : A \to B$, $\mathbf{F} : B \to \mathbb{R}^k$. Dersom $\mathbf{G}$ er deriverbar i punktet $\mathbf{a} \in A$ og $\mathbf{F}$ er deriverbar i punktet $\mathbf{b} = \mathbf{G}(\mathbf{a})$, så er den sammensatte funksjonen $\mathbf{H}(\mathbf{x}) = \mathbf{F}(\mathbf{G}(\mathbf{x}))$ deriverbar i $\mathbf{a}$, og de partiellderiverte til $\mathbf{H}$ er gitt ved*

$$\frac{\partial H_i}{\partial x_j}(\mathbf{a}) = \sum_{p=1}^{m} \frac{\partial F_i}{\partial u_p}(\mathbf{G}(\mathbf{a}))\frac{\partial G_p}{\partial x_j}(\mathbf{a}) =$$

$$= \frac{\partial F_i}{\partial u_1}(\mathbf{G}(\mathbf{a}))\frac{\partial G_1}{\partial x_j}(\mathbf{a}) + \frac{\partial F_i}{\partial u_2}(\mathbf{G}(\mathbf{a}))\frac{\partial G_2}{\partial x_j}(\mathbf{a}) + \cdots + \frac{\partial F_i}{\partial u_m}(\mathbf{G}(\mathbf{a}))\frac{\partial G_m}{\partial x_j}(\mathbf{a})$$

Bevis: Som vi allerede har sett, følger denne formuleringen direkte fra matriseformen når man utfører matrisemultiplikasjonen og sammenligner koeffisientene på begge sider av likhetstegnet. □

Oppgaver til seksjon 2.7

1. La $f(u,v) = u^2 + v$, $g(x,y) = 2xy$, $h(x,y) = x + y^2$. Bruk kjerneregelen til å finne de partiellderiverte av $k(x,y) = f(g(x,y), h(x,y))$.

2. La $f(u,v) = ue^{-v}$, $g(x,y,z) = 2xy + z$, $h(x,y,z) = 2y(z+x)$. Bruk kjerneregelen til å finne de partiellderiverte av $k(x,y,z) = f(g(x,y,z), h(x,y,z))$.

3. Bruk kjerneregelen til å regne ut $\frac{\partial h}{\partial x_2}$ og $\frac{\partial h}{\partial x_3}$ i eksempel 1.

4. Bruk kjerneregelen til å regne ut $\frac{\partial h}{\partial v}$ i eksempel 2.

5. Vi har to funksjoner $\mathbf{G} : \mathbb{R}^2 \to \mathbb{R}^3$ og $\mathbf{F} : \mathbb{R}^3 \to \mathbb{R}^2$. Anta at $\mathbf{G}(1,-2) = (1,2,3)$ og at

$$\mathbf{G}'(1,-2) = \begin{pmatrix} 1 & -2 \\ 3 & 1 \\ 2 & -1 \end{pmatrix}, \qquad \mathbf{F}'(1,2,3) = \begin{pmatrix} 2 & 1 & 4 \\ 0 & 2 & 2 \end{pmatrix}$$

Finn Jacobi-matrisen til den sammensatte funksjonen $\mathbf{H}(\mathbf{x}) = \mathbf{F}(\mathbf{G}(\mathbf{x}))$ i punktet $(1, -2)$.

6. Vi har to funksjoner $\mathbf{G} : \mathbb{R}^3 \to \mathbb{R}^2$ og $\mathbf{F} : \mathbb{R}^2 \to \mathbb{R}^2$. Anta at $\mathbf{G}(-1, -2, 1) = (2, 4)$ og at

$$\mathbf{G}'(-1, -2, 1) = \begin{pmatrix} 4 & -2 & 0 \\ 1 & 3 & -1 \end{pmatrix}, \qquad \mathbf{F}'(2, 4) = \begin{pmatrix} 2 & -3 \\ 0 & 2 \end{pmatrix}$$

Finn Jacobi-matrisen til den sammensatte funksjonen $\mathbf{H}(\mathbf{x}) = \mathbf{F}(\mathbf{G}(\mathbf{x}))$ i punktet $(-1, -2, 1)$.

7. To vareslag konkurrerer om det samme markedet. Etterspørselen E_1 etter det første vareslaget varierer med prisene p_1 og p_2 på begge vareslagene.

Vi har altså en funksjon $E_1 = E_1(p_1, p_2)$. Anta at vi vet hvordan prisene $p_1 = p_1(t)$ og $p_2 = p_2(t)$ varierer med tiden. Vis at etterspørselens variasjon med tiden kan uttrykkes ved

$$\frac{dE_1}{dt} = \frac{\partial E_1}{\partial p_1} p_1'(t) + \frac{\partial E_1}{\partial p_2} p_2'(t)$$

8. Temperaturen T i et område avhenger av posisjonen; vi kan tenke oss at den er gitt som en funksjon $T = f(x, y)$ av to variable der x og y er vanlige koordinater. Vi innfører nå polarkoordinater r og θ på vanlig måte slik at $x = r\cos\theta$, $y = r\sin\theta$. Vi får da temperaturen som en funksjon $T = f(r\cos\theta, r\sin\theta)$ av r og θ.

a) Vis at

$$\frac{\partial T}{\partial r} = \frac{\partial f}{\partial x}\cos\theta + \frac{\partial f}{\partial y}\sin\theta$$

$$\frac{\partial T}{\partial \theta} = -\frac{\partial f}{\partial x} r\sin\theta + \frac{\partial f}{\partial y} r\cos\theta$$

b) En radiomerket fugl beveger seg i området. Radiosignalene viser hvordan avstanden r og vinkelen θ varierer med tiden; vi har $r = g(t)$ og $\theta = h(t)$. Vis at temperaturendringene fuglen opplever er gitt ved

$$T'(t) = \left(\frac{\partial f}{\partial x}\cos\theta + \frac{\partial f}{\partial y}\sin\theta\right) g'(t) + \left(-\frac{\partial f}{\partial x} r\sin\theta + \frac{\partial f}{\partial y} r\cos\theta\right) h'(t)$$

der vi må sette inn $r = g(t)$, $\theta = h(t)$, $x = g(t)\cos h(t)$, $y = g(t)\sin h(t)$.

9. La $f : \mathbb{R}^{n+1} \to \mathbb{R}$ være en deriverbar funksjon av $n + 1$ variable og anta at det finnes en deriverbar funksjon $g : \mathbb{R}^n \to \mathbb{R}$ slik at

$$f(x_1, x_2, \ldots, x_n, g(x_1, x_2, \ldots, x_n)) = 0$$

for alle $x_1, x_2, \ldots, x_n$. (Tenk deg at $y = g(x_1, x_2, \ldots x_n)$ er det uttrykket du får dersom du løser ligningen $f(x_1, x_2, \ldots, x_n, y) = 0$ med hensyn på y).

a) Vis at

$$\frac{\partial g}{\partial x_i}(x_1, x_2, \ldots, x_n) = -\frac{\frac{\partial f}{\partial x_i}(x_1, x_2, \ldots, x_n, g(x_1, x_2, \ldots, x_n))}{\frac{\partial f}{\partial x_{n+1}}(x_1, x_2, \ldots, x_n, g(x_1, x_2, \ldots, x_n))}$$

b) La $f(x,y) = x^2 + y^2 - R^2$ der R er en positiv konstant, og anta at $y = g(x)$ er en deriverbar funksjon slik at $f(x, g(x)) = 0$.Vis at

$$g'(x) = -\frac{x}{g(x)}$$

Gi en geometrisk tolkning av resultatet.

c) La $f(x,y,z) = x^2 + y^2 + z^2 - R^2$ der R er en positiv konstant og anta at $z = g(x,y)$ er en deriverbar funksjon slik at $f(x, y, g(x,y)) = 0$. Vis at

$$\frac{\partial g}{\partial x}(x,y) = -\frac{x}{g(x,y)}$$

og

$$\frac{\partial g}{\partial y}(x,y) = -\frac{y}{g(x,y)}$$

Gi en geometrisk tolkning av resultatet.

2.8 Linearisering

Når vi arbeider med funksjoner $f : \mathbb{R} \to \mathbb{R}$ av én variabel, er vi ofte interessert i tangenten til funksjonen i et punkt a. Tangenten er den rette linjen som ligger tettest opptil funksjonsgrafen i området rundt a. Vi skal nå prøve å generalisere denne ideen til funksjoner $\mathbf{F} : \mathbb{R}^n \to \mathbb{R}^m$, og spørsmålet er: Hva skal egentlig spille rollen til tangenten i dette tilfellet? Siden geometrien er så kompleks at det er vanskelig å holde oversikt, lønner det seg å tenke litt mer abstrakt: En tangent er en rett linje, og en rett linje er grafen til en affinavbildning (husk seksjon 1.10). Det vi gjør når vi finner tangenten til $f : \mathbb{R} \to \mathbb{R}$ i punktet a, er altså å finne den affinavbildningen som ligger nærmest opptil f i området rundt a. Dette er en formulering som også gir mening for funksjoner $\mathbf{F} : \mathbb{R}^n \to \mathbb{R}^m$: Vi er på jakt etter den affinavbildningen som ligger tettest mulig opp til $\mathbf{F}$ i området rundt et punkt $\mathbf{a}$. Denne avbildningen skal vi kalle *lineariseringen* til $\mathbf{F}$ i punktet $\mathbf{a}$, og selv om det sikkert ikke er så lett å se på det nåværende tidspunkt, spiller slike lineariseringer en sentral rolle i mange matematiske resonnementer og anvendelser.

Før vi går løs på lineariseringer, trenger vi å vite litt om hvordan affinavbildinger passer inn i den teorien vi har utviklet i dette kapitlet. Husk at en affinavbildning $\mathbf{F} : \mathbb{R}^n \to \mathbb{R}^m$ er en funksjon på formen $\mathbf{F}(\mathbf{x}) = A\mathbf{x} + \mathbf{c}$ der

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

er en $m \times n$-matrise og

$$\mathbf{c} = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_m \end{pmatrix}$$

er en vektor i $\mathbb{R}^m$. Skriver vi ut definisjonen $\mathbf{F}(\mathbf{x}) = A\mathbf{x} + \mathbf{c}$ på komponentform, får vi

$$\mathbf{F}(\mathbf{x}) = \begin{pmatrix} F_1(\mathbf{x}) \\ F_2(\mathbf{x}) \\ \vdots \\ F_m(\mathbf{x}) \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ldots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_m \end{pmatrix} =$$

$$= \begin{pmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n + c_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n + c_2 \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n + c_m \end{pmatrix}$$

Deriverer vi den i-te komponenten F_i med hensyn på den j-te variabelen x_j, får vi

$$\frac{\partial F_i}{\partial x_j} = \frac{\partial}{\partial x_j}(a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n + c_i) = a_{ij}$$

Dette betyr at Jacobi-matrisen til en affinavbildning $\mathbf{F}$ rett og slett er matrisen A til $\mathbf{F}$.

Setning 2.8.1 *Anta at affinavbildningen $\mathbf{F} : \mathbb{R}^n \to \mathbb{R}^m$ er gitt ved $\mathbf{F}(\mathbf{x}) = A\mathbf{x} + \mathbf{c}$. Da er Jacobi-matrisen til $\mathbf{F}$ lik matrisen A til $\mathbf{F}$.* □

Vi er nå klare til å studere hvordan affinavbildninger kan brukes til å tilnærme mer generelle funksjoner. Husk at hvis $f : \mathbb{R} \to \mathbb{R}$ er en funksjon av én variabel, er tangenten $\cancel{T_af}$ i punktet a gitt ved til grafen til funksjonen T_af gitt ved

$$T_af(x) = f(a) + f'(a)(x - a)$$

Starter vi isteden med en funksjon $\mathbf{F} : \mathbb{R}^n \to \mathbb{R}^m$, kan vi på tilsvarende måte definere en funksjon $T_\mathbf{a}\mathbf{F}$ ved å etterligne uttrykket ovenfor:

$$T_\mathbf{a}\mathbf{F}(\mathbf{x}) = \mathbf{F}(\mathbf{a}) + \mathbf{F}'(\mathbf{a})(\mathbf{x} - \mathbf{a})$$

Ganger vi ut parentesen, får vi

$$T_\mathbf{a}\mathbf{F}(\mathbf{x}) = \mathbf{F}(\mathbf{a}) - \mathbf{F}'(\mathbf{a})\mathbf{a} + \mathbf{F}'(\mathbf{a})\mathbf{x}$$

som viser at $T_a\mathbf{F}$ er en affinavbildning med matrise $\mathbf{F}'(\mathbf{a})$ og konstantledd $\mathbf{F}(\mathbf{a}) - \mathbf{F}'(\mathbf{a})\mathbf{a}$.

Definisjon 2.8.2 *Anta at* $\mathbf{F} : A \to \mathbb{R}^m$ *er en funksjon av* n *variable som er deriverbar i punktet* $\mathbf{a}$. *Affinavbildningen* $T_\mathbf{a}\mathbf{F} : \mathbb{R}^n \to \mathbb{R}^m$ *gitt ved*

$$T_\mathbf{a}\mathbf{F}(\mathbf{x}) = \mathbf{F}(\mathbf{a}) + \mathbf{F}'(\mathbf{a})(\mathbf{x} - \mathbf{a})$$

kalles linneariseringen *til* $\mathbf{F}$ *i punktet* $\mathbf{a}$.

Legg merke til at $T_\mathbf{a}\mathbf{F}(\mathbf{a}) = \mathbf{F}(\mathbf{a})$. Siden vi allerede har observert at matrisen til affinavbildningen $T_a\mathbf{F}$ er $\mathbf{F}'(\mathbf{a})$, vet vi fra setning 2.8.1 at Jacobi-matrisen til $T_\mathbf{a}\mathbf{F}$ er $\mathbf{F}'(\mathbf{a})$. Dette betyr at lineariseringen $T_a\mathbf{F}$ har samme verdi og samme deriverte i punktet $\mathbf{a}$ som den opprinnelige funksjonen $\mathbf{F}$. Dette er akkurat samme egenskap som tangenten har for funksjoner av én variabel. En annen viktig (og nært beslektet) egenskap ved tangenten er at den er linjen som smyger seg tettest inntil funksjonsgrafen til f i nærheten av punktet a. Vi skal snart vise en tilsvarende egenskap for lineariseringen — at den er affinavbildningen som ligger tettest opptil $\mathbf{F}$ i nærheten av $\mathbf{a}$. Men før vi gjør dette, kan det være lurt å kikke på et eksempel.

Eksempel 1: Vi skal finne lineariseringen til funksjonen $\mathbf{F} : \mathbb{R}^3 \to \mathbb{R}^2$ gitt ved

$$\mathbf{F}(x, y, z) = \begin{pmatrix} x^2y \\ 2xyz \end{pmatrix}$$

i punktet $\mathbf{a} = \begin{pmatrix} 1 \\ -2 \\ -1 \end{pmatrix}$. La oss først regne ut Jacobi-matrisen til $\mathbf{F}$:

$$\mathbf{F}'(x, y, z) = \begin{pmatrix} \frac{\partial F_1}{\partial x} & \frac{\partial F_1}{\partial y} & \frac{\partial F_1}{\partial z} \\ \frac{\partial F_2}{\partial x} & \frac{\partial F_2}{\partial y} & \frac{\partial F_2}{\partial z} \end{pmatrix} = \begin{pmatrix} 2xy & x^2 & 0 \\ 2yz & 2xz & 2xy \end{pmatrix}$$

I punktet $\mathbf{a}$ har vi dermed

$$\mathbf{F}'(1, -2, -1) = \begin{pmatrix} 2 \cdot 1 \cdot (-2) & 1^2 & 0 \\ 2 \cdot (-2) \cdot (-1) & 2 \cdot 1 \cdot (-1) & 2 \cdot 1 \cdot (-2) \end{pmatrix} =$$

$$= \begin{pmatrix} -4 & 1 & 0 \\ 4 & -2 & -4 \end{pmatrix}$$

Vi har også

$$\mathbf{F}(1, -2, -1) = \begin{pmatrix} 1^2 \cdot (-2) \\ 2 \cdot 1 \cdot (-2) \cdot (-1) \end{pmatrix} = \begin{pmatrix} -2 \\ 4 \end{pmatrix}$$

Vi kan nå regne ut lineariseringen. Med $\mathbf{x} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$ får vi:

$$T_\mathbf{a}\mathbf{F}(\mathbf{x}) = \mathbf{F}(\mathbf{a}) + \mathbf{F}'(\mathbf{a})(\mathbf{x} - \mathbf{a}) =$$

$$= \begin{pmatrix} -2 \\ 4 \end{pmatrix} + \begin{pmatrix} -4 & 1 & 0 \\ 4 & -2 & -4 \end{pmatrix} \left(\begin{pmatrix} x \\ y \\ z \end{pmatrix} - \begin{pmatrix} 1 \\ -2 \\ -1 \end{pmatrix} \right) =$$

$$= \begin{pmatrix} -2 \\ 4 \end{pmatrix} + \begin{pmatrix} -4x + y \\ 4x - 2y - 4z \end{pmatrix} - \begin{pmatrix} -6 \\ 12 \end{pmatrix} = \begin{pmatrix} -4x + y + 4 \\ 4x - 2y - 4z - 8 \end{pmatrix}$$

Altså er

$$T_{\mathbf{a}}\mathbf{F}(x, y, z) = \begin{pmatrix} -4x + y + 4 \\ 4x - 2y - 4z - 8 \end{pmatrix}$$

♣

La oss vende tilbake til resultatet vi annonserte ovenfor, det som sier at lineariseringen $T_{\mathbf{a}}\mathbf{F}$ er den affinavbildningen som ligger tettest opptil $\mathbf{F}$ i nærheten av $\mathbf{a}$. Et punkt $\mathbf{x}$ i nærheten av $\mathbf{a}$ kan vi skrive $\mathbf{x} = \mathbf{a} + \mathbf{r}$ der $\mathbf{r}$ er liten. Vi ønsker å vise at differensen $\mathbf{F}(\mathbf{a} + \mathbf{r}) - T_{\mathbf{a}}\mathbf{F}(\mathbf{a} + \mathbf{r})$ er liten sammenlignet med $|\mathbf{r}|$ for små $\mathbf{r}$, og at tilnærmingen blir bedre og bedre når $\mathbf{r}$ går mot $\mathbf{0}$. En naturlig måte å formulere dette på er å kreve at

$$\lim_{\mathbf{r}\to\mathbf{0}} \frac{1}{|\mathbf{r}|}\Big(\mathbf{F}(\mathbf{a}+\mathbf{r}) - T_{\mathbf{a}}\mathbf{F}(\mathbf{a}+\mathbf{r})\Big) = \mathbf{0}$$

Bruker vi at

$$T_{\mathbf{a}}\mathbf{F}(\mathbf{a}+\mathbf{r}) = \mathbf{F}(\mathbf{a}) + \mathbf{F}'(\mathbf{a})\big((\mathbf{a}+\mathbf{r}) - \mathbf{a}\big) = \mathbf{F}(\mathbf{a}) + \mathbf{F}'(\mathbf{a})\mathbf{r}$$

ser vi at

$$\mathbf{F}(\mathbf{a}+\mathbf{r}) - T_{\mathbf{a}}\mathbf{F}(\mathbf{a}+\mathbf{r}) = \mathbf{F}(\mathbf{a}+\mathbf{r}) - \mathbf{F}(\mathbf{a}) - \mathbf{F}'(\mathbf{a})\mathbf{r} = \boldsymbol{\sigma}(\mathbf{r})$$

der $\boldsymbol{\sigma}$ er som i definisjon 2.6.1. Nå er det ikke så vanskelig å formulere og bevise resultatet.

Teorem 2.8.3 *Anta at $\mathbf{F} : A \to \mathbb{R}^m$ er en funksjon av n variable som er deriverbar i punktet $\mathbf{a}$, og la $T_{\mathbf{a}}\mathbf{F}$ være lineariseringen til $\mathbf{F}$ i $\mathbf{a}$. Da er*

$$\lim_{\mathbf{r}\to\mathbf{0}} \frac{1}{|\mathbf{r}|}\Big(\mathbf{F}(\mathbf{a}+\mathbf{r}) - T_{\mathbf{a}}\mathbf{F}(\mathbf{a}+\mathbf{r})\Big) = \mathbf{0}$$

Det finnes ingen annen affinavbildning $\mathbf{G} : \mathbb{R}^n \to \mathbb{R}^m$ slik at (enn $T_{\vec{a}}\vec{F}$)

$$\lim_{\mathbf{r}\to\mathbf{0}} \frac{1}{|\mathbf{r}|}\Big(\mathbf{F}(\mathbf{a}+\mathbf{r}) - \mathbf{G}(\mathbf{a}+\mathbf{r})\Big) = \mathbf{0}$$

Bevis: Det følger av definisjonen av deriverbarhet (definisjon 2.6.1) og regningene ovenfor at

$$\lim_{\mathbf{r}\to\mathbf{0}} \frac{1}{|\mathbf{r}|}\Big(\mathbf{F}(\mathbf{a}+\mathbf{r}) - T_{\mathbf{a}}\mathbf{F}(\mathbf{a}+\mathbf{r})\Big) = \lim_{\mathbf{r}\to\mathbf{0}} \frac{1}{|\mathbf{r}|}\boldsymbol{\sigma}(\mathbf{r}) = \mathbf{0}$$

Dermed er første del av teoremet bevist. For å bevise den andre delen, antar vi at $\mathbf{G}(\mathbf{x})$ er en affinavbildning slik at

$$\lim_{\mathbf{r}\to\mathbf{0}} \frac{1}{|\mathbf{r}|}\Big(\mathbf{F}(\mathbf{a}+\mathbf{r}) - \mathbf{G}(\mathbf{a}+\mathbf{r})\Big) = \mathbf{0}$$

Vi må vise at $\mathbf{G} = T_{\mathbf{a}}\mathbf{F}$. Siden $\mathbf{G}$ er en affinavbildning, vet vi at $\mathbf{G}(\mathbf{x}) = A\mathbf{x} + \mathbf{c}$ for en matrise A og en vektor $\mathbf{c}$. Setter vi dette inn i uttrykket ovenfor, får vi

$$\lim_{\mathbf{r}\to\mathbf{0}} \frac{1}{|\mathbf{r}|}\Big(\mathbf{F}(\mathbf{a}+\mathbf{r}) - \mathbf{G}(\mathbf{a}+\mathbf{r})\Big) = \lim_{\mathbf{r}\to\mathbf{0}} \frac{1}{|\mathbf{r}|}\Big(\mathbf{F}(\mathbf{a}+\mathbf{r}) - A\mathbf{a} - A\mathbf{r} - \mathbf{c}\Big) = \mathbf{0}$$

Skal dette være mulig, må $\mathbf{F}(\mathbf{a}+\mathbf{r}) - A\mathbf{a} - A\mathbf{r} - \mathbf{c} \to \mathbf{0}$ når $\mathbf{r} \to \mathbf{0}$, og det er bare tilfellet hvis $\mathbf{F}(\mathbf{a}) = A\mathbf{a} + \mathbf{c}$. Setter vi $A\mathbf{a} + \mathbf{c} = \mathbf{F}(\mathbf{a})$ inn i det siste uttrykket ovenfor, ser vi at

$$\lim_{\mathbf{r}\to\mathbf{0}} \frac{1}{|\mathbf{r}|}\Big(\mathbf{F}(\mathbf{a}+\mathbf{r}) - \mathbf{F}(\mathbf{a}) - A\mathbf{r}\Big) = \mathbf{0}$$

og ifølge setning 2.6.4 er det bare mulig dersom $A = \mathbf{F}'(\mathbf{a})$. Siden vi allerede vet at $\mathbf{F}(\mathbf{a}) = A\mathbf{a} + \mathbf{c}$, følger det at $\mathbf{c} = \mathbf{F}(\mathbf{a}) - A\mathbf{a} = \mathbf{F}(\mathbf{a}) - \mathbf{F}'(\mathbf{a})\mathbf{a}$. Dermed er

$$\mathbf{G}(\mathbf{x}) = A\mathbf{x} + \mathbf{c} = \mathbf{F}'(\mathbf{a})\mathbf{x} + \mathbf{F}(\mathbf{a}) - \mathbf{F}'(\mathbf{a})\mathbf{a} = \mathbf{F}(\mathbf{a}) + \mathbf{F}'(\mathbf{a})(\mathbf{x} - \mathbf{a}) = T_{\mathbf{a}}\mathbf{F}(\mathbf{x})$$

□

Det som er viktigst å ta med seg fra teoremet ovenfor, er at lineariseringen $T_{\mathbf{a}}\mathbf{F}$ er en svært god tilnærming til $\mathbf{F}$ i området rundt $\mathbf{a}$. Dette skal vi få bruk for både når vi studerer Newtons metode i flere variable i kapittel 5, og når vi studerer skifte av variable i multiple integraler (dvs. integraler av funksjoner av flere variable) i kapittel 6.

For en funksjon $f : \mathbb{R} \to \mathbb{R}$ er lineariseringen til f i punktet a det samme som tangenten til f i punktet a (dersom du tenker på tangenten som en funksjon og ikke som en linje). Tilsvarende viser det seg at for en funksjon $f : \mathbb{R}^2 \to \mathbb{R}$ av to variable, så er lineariseringen i punktet $\mathbf{a}$ det samme som tangentplanet i punktet $\mathbf{a}$, altså det planet som smyger seg tettest mulig inntil funksjonsgrafen rundt $\mathbf{a}$. Vi skal komme tilbake til dette i seksjon 3.7.

Bemerkning: Du har kanskje lurt på hvorfor $T_{\mathbf{a}}\mathbf{F}$ kalles en *linearisering* og ikke en *affinisering* — $T_{\mathbf{a}}\mathbf{F}$ er tross alt en affinavbildning og ikke (bortsett fra i helt spesielle tilfeller) en lineæravbildning. Dessverre er bruken av ordparet lineær/affin svært usystematisk og forvirrende; ordene brukes om hverandre på en måte som ikke er lett å få oversikt over. Et annet eksempel på denne vaklingen får du om du betrakter en funksjon $f : \mathbb{R} \to \mathbb{R}$ på formen $f(x) =$

$ax + c$. Når man studerer funksjoner av én variabel, er alle enige om å kalle dette en lineær funksjon, men ifølge terminologien vi nå har innført, er den bare en lineæravbildning hvis $c = 0$; i alle andre tilfeller må vi nøye oss med å kalle den en affinavbildning! Heldigvis er det sjelden at den vaklende terminologibruken skaper misforståelser, men det er greit å vite om den slik at man kan være litt på vakt.

Opppgaver til seksjon 2.8

1. Finn lineariseringen til funksjonen $\mathbf{F} : \mathbb{R}^2 \to \mathbb{R}^2$ gitt ved

$$\mathbf{F}(x, y) = \begin{pmatrix} x^2y \\ xy + x \end{pmatrix}$$

i punktet $\mathbf{a} = (-2, 1)$.

2. Finn lineariseringen til funksjonen $\mathbf{F} : \mathbb{R}^2 \to \mathbb{R}^3$ gitt ved

$$\mathbf{F}(x, y) = \begin{pmatrix} x\sin(xy) \\ xe^y \\ 2x^3 + y \end{pmatrix}$$

i punktet $\mathbf{a} = (2, 0)$.