MAT1110

Sammendrag av forelesningen 19/1-14

Repetisjon

Anta at $f: \mathbb{R}^n \to \mathbb{R}$ er en funksjon av n variable, $y = f(x_1, x_2, \dots, x_n)$. Da er

$$\frac{\partial f}{\partial x_i}(\vec{x}) = \lim_{\Delta x \to 0} \frac{f(x_1, \dots, x_i + \Delta x, \dots, x_n) - f(x_1, \dots, x_i, \dots, x_n)}{\Delta x}$$

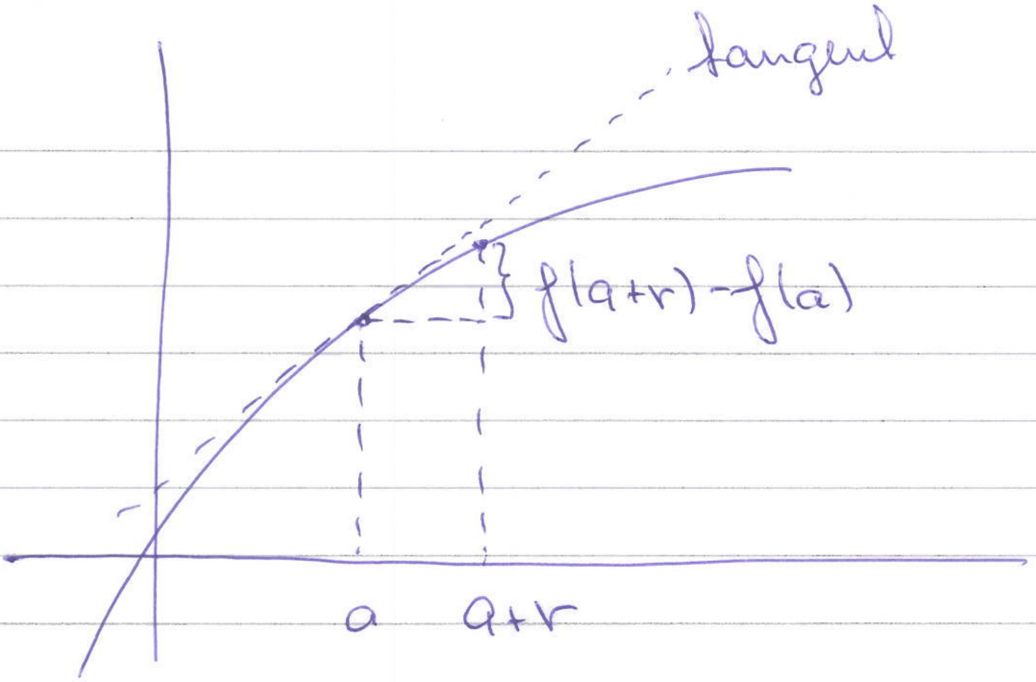
den i-te partiellderiverte til f. Vi regner den ut ved å derivere f m.h.p. x_i som om de andre variablene var konstanter.

Anta så at $\vec{F}: \mathbb{R}^n \to \mathbb{R}^m$ der $\vec{F}(x) = \begin{pmatrix} F_1(\vec{x}) \\ F_2(\vec{x}) \\ \vdots \\ F_m(\vec{x}) \end{pmatrix}$.

Da er Jacobi-matrisen til $\vec{F}$ definert ved:

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

Jacobi-matrisen spiller på mange måter den samme rollen for funksjoner av flere variable som den deriverte gjør for funksjoner av én variabel. Her er det viktigste eksemplet:



Funksjonsdifferansen $f(a+r)-f(a)$ er for små r tilnærmet lik $f'(a)r$ siden $f'(a)r$ er den differansen vi får om vi beveger oss langs tangenten istedenfor langs funksjonsgrafen. Vi har altså

$$f(a+r)-f(a) \approx f'(a)r \quad \text{for små } r.$$

Denne tilnærmingen blir bedre og bedre (også relativt til r) jo mindre r blir.

Vi har samme fenomen for funksjoner av flere variable:

$$\vec{F}(\vec{a}+\vec{r})-\vec{F}(\vec{a}) \approx \vec{F}'(\vec{a})\vec{r} \quad \text{for små } \vec{r}$$

med bedre og ~~[illegible]~~ bedre tilnærming jo mindre $\vec{r}$ blir.

Relasjonene ovenfor forutsetter at funksjonen ikke er for irregulær, og det er dette som fanges opp i definisjonen av deriverbarhet:

<u>Definisjon:</u> En funksjon $\vec{F}: \mathbb{R}^n \to \mathbb{R}^m$ er <u>deriverbar</u> i $\vec{a}$ dersom feilleddet

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

går mot null raskere enn $\vec{r}$, dvs

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

<u>Kjerneregelen på matriseform</u>

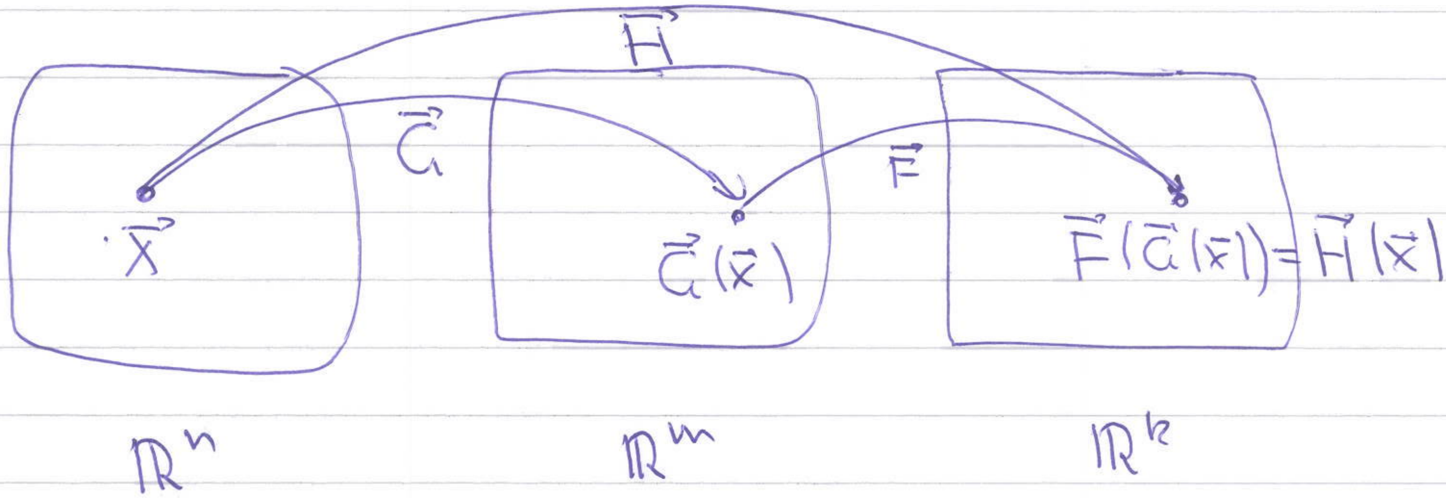
Den vanlige kjerneregelen sier at dersom $h(x) = f(g(x))$, så er den deriverte til h gitt ved

$$h'(x) = f'(g(x))\,g'(x)$$

Vi skal nå se på en tilsvarende formel for funksjoner av flere variable. Husk først at dersom vi er gitt funksjoner

$$\vec{G}: \mathbb{R}^n \to \mathbb{R}^m \quad \text{og} \quad \vec{F}: \mathbb{R}^m \to \mathbb{R}^k,$$

kan vi definere en sammensatt funksjon $\vec{H}: \mathbb{R}^n \to \mathbb{R}^k$ ved $\vec{H}(\vec{x}) = \vec{F}(\vec{G}(\vec{x}))$. Figuren viser hvordan dette fungerer:



Spørsmålet er om vi kan beregne den deriverte til $\vec{H}$ dersom vi kjenner de deriverte til $\vec{F}$ og $\vec{G}$.

<u>Kjerneregelen på matriseform</u>: Dersom $\vec{G}$ er deriverbar i punktet $\vec{x}$ og $\vec{F}$ er deriverbar i punktet $\vec{G}(\vec{x})$, så er $\vec{H}$ deriverbar i punktet $\vec{x}$ og

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

Vi ser at denne kjerneregel har akkurat samme form som den vanlige, men den er likevel adskillig mer komplisert. I uttrykket ovenfor er $\vec{H}'(\vec{x})$ en $k\times n$-matrise som er produktet av $k\times m$-matrisen $\vec{F}'(\vec{G}(\vec{x}))$ og $m\times n$-matrisen $\vec{G}'(\vec{x})$.

<u>Eksempel</u>: Anta at $\vec{G}:\mathbb{R}^2\to\mathbb{R}^3$, $\vec{F}:\mathbb{R}^3\to\mathbb{R}^2$ er deriverbare funksjoner med

$$\vec{G}(1,-2) = (2,3,-4)$$

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

Då er $\vec{H}:\mathbb{R}^2\to\mathbb{R}^2$, $\vec{H}(\vec{x}) = \vec{F}(\vec{G}(\vec{x}))$ deriverbar i $(1,-2)$ og

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

$$= \begin{pmatrix} -4 & 3 \\ -5 & 11 \end{pmatrix}$$

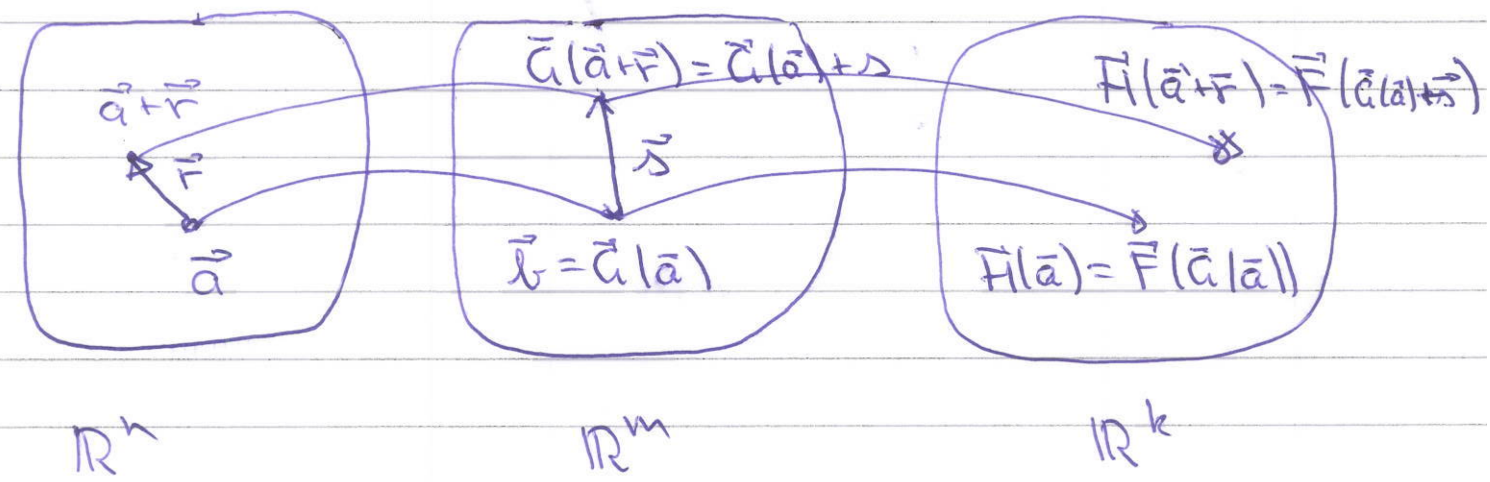
Dette betyr at $\frac{\partial H_1}{\partial x_1}(1,-2) = -4$, $\frac{\partial H_1}{\partial x_2}(1,-2) = 3$, $\frac{\partial H_2}{\partial x_1}(1,-2) = -5$ og $\frac{\partial H_2}{\partial x_2}(1,-2) = 11$.

Jeg skal ikke gi et fullstendig bevis for kjerneregelen her (det står et i boka), men jeg skal prøve å antyde hvorfor den er sann. Vi vet at når $\vec{r}$ og $\vec{s}$ er små, så er

$$\vec{G}(\vec{a}+\vec{r}) - \vec{G}(\vec{a}) \approx \vec{G}'(\vec{a})\vec{r}$$

$$\vec{F}(\vec{b}+\vec{s}) - \vec{F}(\vec{b}) \approx \vec{F}'(\vec{b})\vec{s}.$$

La oss bruke disse formlene med $\vec{b} = \vec{G}(\vec{a})$ og $\vec{s} = \vec{G}(\vec{a}+\vec{r}) - \vec{G}(\vec{a})$ slik figuren viser



Dermed har vi

$$\vec{H}(\vec{a}+\vec{r}) - \vec{H}(\vec{a}) = \vec{F}(\vec{G}(\vec{a})+\vec{s}) - \vec{F}(\vec{G}(\vec{a}))$$

$$\approx \vec{F}'(\vec{G}(\vec{a}))\vec{s} = \vec{F}'(\vec{G}(\vec{a}))(\vec{G}(\vec{a}+\vec{r}) - \vec{G}(\vec{a}))$$

$$\approx \vec{F}'(\vec{G}(\vec{a}))\vec{G}'(\vec{a})\vec{r}$$

som sterkt indikerer at $\vec{H}'(\vec{a}) = \vec{F}'(\vec{G}(\vec{a}))\vec{G}'(\vec{a})$. Dette argumentet kan raffineres til et virkelig bevis, men da må vi holde styr på feilleddene i $\approx$-relasjonene.

Kjerneregelen på koordinatform

Det finnes en annen versjon av kjerneregelen som ofte er enklere å bruke i praksis.

Anta at vi har (skalare) funksjoner $f(u_1, \ldots, u_m)$, $g_1(x_1, \ldots, x_n), \ldots, g_m(x_1, \ldots, x_n)$. Vi kan nå substituere g'ene inn i f og få en sammensatt funksjon

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

Spørsmålet er hvordan vi kan finne de partiellderiverte til h når vi kjenner de partiellderiverte til f og $g_1, g_2, \ldots, g_m$. For å gjøre notasjonen litt mer oversiktlig skriver vi

$$\vec{x} \text{ for } (x_1, x_2, \ldots, x_n)$$
$$\vec{g}(\vec{x}) \text{ for } (g_1(x_1, \ldots, x_n), \ldots, g_m(x_1, \ldots, x_n))$$

Kjerneregelen på komponentform sier nå at

$$\frac{\partial h}{\partial x_i}(\vec{x}) = \frac{\partial f}{\partial u_1}(\vec{g}(\vec{x}))\frac{\partial g_1}{\partial x_i}(\vec{x}) + \frac{\partial f}{\partial u_2}(\vec{g}(\vec{x}))\frac{\partial g_2}{\partial x_i}(\vec{x}) + \cdots + \frac{\partial f}{\partial u_m}(\vec{g}(\vec{x}))\frac{\partial g_m}{\partial x_i}(\vec{x})$$

Vi deriverer altså med hensyn på alle de "ytre" variablene $u_1, u_2, \ldots, u_m$, men alltid med hensyn på den samme "indre" variabelen x_i.

La oss se på et eksempel:

<u>Eksempel:</u> $f(u_1, u_2) = u_1^2 e^{u_2}$

$$u_1 = g_1(x_1, x_2, x_3) = x_1^2 x_2 + x_3$$

$$u_2 = g_2(x_1, x_2, x_3) = x_1 + 2x_2 x_3$$

Vi skal bruke kjerneregelen til å finne den partiellderiverte $\frac{\partial h}{\partial x_1}$ til den sammensatte funksjonen

$$h(x_1, x_2, x_3) = f(g_1(x_1, x_2, x_3), g_2(x_1, x_2, x_3))$$

Vi får

$$\frac{\partial h}{\partial x_1} = \frac{\partial f}{\partial u_1}\frac{\partial g_1}{\partial x_1} + \frac{\partial f}{\partial u_2}\frac{\partial g_2}{\partial x_1}$$

$$= 2u_1 e^{u_2}\, 2x_1 x_2 + u_1^2 e^{u_2} \cdot 1$$

$$= 2(x_1^2 x_2 + x_3) e^{x_1 + 2x_2 x_3} \cdot 2x_1 x_2 + (x_1^2 x_2 + x_3)^2 e^{x_1 + 2x_2 x_3}$$

Legg merke til hvordan vi substituerer for å komme tilbake til de underliggende variablene x_1, x_2, x_3 (svaret burde forøvrig vært ryddet litt opp i!)

Vil vi heller derivere m.h.p. x_3, får vi:

$$\frac{\partial h}{\partial x_3} = \frac{\partial f}{\partial u_1}\frac{\partial g_1}{\partial x_3} + \frac{\partial f}{\partial u_2}\frac{\partial g_2}{\partial x_3} =$$

$$= 2u_1 e^{u_2}\cdot 1 + u_1^2 e^{u_2}\cdot 2x_2$$

$$= 2(x_1^2x_2+x_3)e^{x_1+2x_2x_3} + (x_1^2x_2+x_3)^2 e^{x_1+2x_2x_3}\,2x_2$$

Sammenhengen mellom kjerneregLene

De to formene for kjerneregel kan se svært forskjellige ut, men de er faktisk to versjoner av samme resultat. For å se dette lønner det seg å skrive ut i detalj hva kjerneregelen på matriseform egentlig sier. Skriver vi ut $\vec{H}'(\vec{x}) = \vec{F}'(\vec{G}(\vec{x}))\vec{G}'(\vec{x})$, får vi

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

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

Ganger vi ut, ser vi at det ij-te elementet $\frac{\partial H_i}{\partial x_j}$ i $\vec{H}'$-matrisen, må være

prikkproduktet av den i-te linjen i $\vec{F}'$ og den j-te søylen i $\bar{a}'$, der

$$\frac{\partial H_i}{\partial x_j} = \frac{\partial F_i}{\partial u_1}\frac{\partial a_1}{\partial x_j} + \frac{\partial F_i}{\partial u_2}\frac{\partial a_2}{\partial x_j} + \dots + \frac{\partial F_i}{\partial u_m}\frac{\partial a_m}{\partial x_j}$$

Dette er akkurat det vi får om vi bruker kjerneregelen på komponentform på funksjonen $H_i(\bar{a}(\bar{x}))$. De to versjonene av kjerneregel uttrykker derfor det samme.