Taloustieteen matemaattiset menetelmät - kertausmateriaali Kurssin 1. osa

Topi Hokkanen*

14. lokakuuta 2020

Tämän lyhyen ja epäformaalin materiaalin tavoite on käydä läpi kurssin alkuosalla käytettyjen matemaattisten menetelmien intuitiota, ja käsitellä tiettyjä lisäesimerkkejä soveltuvin osin. Itse materiaali ei ole substituutti kirjalle (Simon & Blume, 1994) eikä luentomateriaaleille, mutta toivottavasti selventää jotain mekaanisia seikkoja käytyjen asioiden taustalla.

1 Yhtälöryhmät ja niiden ratkaisu

Mikäli ratkaistavanamme on yhtälöryhmä, jossa on m yhtälöä ja n muuttujaa, eli seuraava:

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ \vdots &= \vdots \\ a_{m1}x_1 + a_{m2} + \cdots a_{mn}x_n &= b_m \end{aligned}$$

On ensimmäinen toimenpide se, että ilmaisemme yhtälöryhmän tutummassa muodossa, eli kirjoitamme sen muotoon:

$$\mathbf{Ax} = \mathbf{b}$$

Jossa $\mathbf{A}$ on m x n -matriisi, $\mathbf{b}$ on m *x 1* -vektori ja $\mathbf{x}$ tietenkin n *x 1* -vektori. Tämän

*Aalto-yliopiston kauppakorkeakoulu, taloustieteen laitos. Kiitän Lassi Tervosta korjauksista ja hyödyllisistä muokkauksista.

jälkeen voimme pohtia sitä, onko tällaisella järjestelmällä olemassa ratkaisua ja sitä, onko mahdollisesti olemassa oleva ratkaisu yksikäsitteinen. Yllä kuvatulla systeemillä voi olla joko 0, 1 tai ääretön määrä ratkaisuja.

Ratkaisujen lukumäärä riippuu seuraavista asioista (ks. Simon & Blume, 7.3):

1. Kerroinmatriisin $\mathbf{A}$ asteesta (rangista).
2. Yhtälöiden lukumäärästä m.
3. Muuttujien lukumäärästä n.
4. Onko järjestelmä homogeeninen, ts. onko oikean puolen sarakevektori $\mathbf{b}$ nollavektori, vai vain jokin $\mathbf{b} \in \mathbb{R}^m, \mathbf{b} \neq \mathbf{0}$.

1.1 Tuntemattomia enemmän kuin yhtälöitä:

Tällöin $n > m$. Ajatellaan vaikkapa esimerkkiä, jossa meillä on kaksi yhtälöä ja kolme tuntematonta. Riippumatta siitä, mitä kerroinmatriisi $\mathbf{A}$ on syönyt, emme voi millään onnistua määräämään kaikkia tuntemattomia yksikäsitteisesti. Tällöin peruslogiikalla saamme, että:

1. Yhtälöryhmällä $\mathbf{Ax} = \mathbf{0}$ on ääretön määrä ratkaisuja (koska jotkut muuttujat jäävät vapaiksi).
2. Yhtälöryhmällä $\mathbf{Ax} = \mathbf{b}$ on edellisen perusteella joko 0 ratkaisua, tai sitten ratkaisuja on äärettömästi (vrt. vapaat muuttujat)
3. Jos kerroinmatriisin aste on sama kuin yhtälöiden lukumäärä $\text{rank}(\mathbf{A}) = m$, on systeemillä $\mathbf{Ax} = \mathbf{b}$ ääretön määrä ratkaisuja jokaiselle valitsemallemme vektorille $\mathbf{b} \neq \mathbf{0}$. Tällöin yhtälöryhmässä jää aina jokin muuttuja vapaaksi (määräämättömäksi) ja silloin jokaiselle oikean puolen vektorille voidaan aina valita tämä muuttuja äärettömän monella tavalla.

1.2 Yhtälöitä enemmän kuin tuntemattomia:

Tällöin $m > n$. Esimerkiksi kolme yhtälöä ja kaksi tuntematonta. Nyt meillä on yksi "ylimääräinen"yhtälö yhtälöryhmässämme. Silloin:

1. Yhtälöryhmällä $\mathbf{Ax} = \mathbf{0}$ on joko yksikäsitteinen ratkaisu tai ratkaisuja on äärettömästi. Miksi? Yhtälöryhmällä on nyt aina triviaaliratkaisu, eli $\mathbf{x} = \mathbf{0}$, joka toteuttaa

ryhmän. Esimerkkitapauksessamme kolmen yhtälön ja kahden tuntemattoman kanssa, tämä on intuitiivista. Ei voida kuitenkaan olla varmoja siitä, että tämä olisi ainoa ratkaisu. Tämän voi havaita vaikkapa näin: "ylimääräisiä"yhtälöitä on $m-n$ kappaletta. Mikäli kaikki nämä ovat lineaariyhdisteitä jäljelle jäävistä yhtälöistä, saadaan ne "eliminoitua". Jos jäljelle jäävästä n yhtälöstä yksikin on toisten lineaariyhdiste, silloin olemme edellisessä keississä, jossa joku muuttuja jää vapaaksi, ja ratkaisuja on äärettömästi. Täten siis ratkaisuja homogeeniseen yhtälöön on joko 1 tai ääretön määrä.

2. Yhtälöryhmällä $\mathbf{Ax} = \mathbf{b}$ on mille tahansa $\mathbf{b} \in \mathbb{R}^m, \mathbf{b} \neq \mathbf{0}$, 0, 1, tai ääretön määrä ratkaisuja. Joko "ylimääräiset"yhtälöt ovat lineaariyhdisteitä, jolloin "jäljelle jäävän" systeemin ratkaisuja on joko 0 tai 1, ja lisäksi vielä on mahdollista että "jäljelle jäävistä"yhtälöistä joku on muiden lineaariyhdiste, jolloin ratkaisuja on äärettömästi.

3. Jos kerroinmatriisin aste on sama kuin tuntemattomien lukumäärä eli $\text{rank}(\mathbf{A}) = n$, on systeemillä $\mathbf{Ax} = \mathbf{b}$ 0 tai yksi ratkaisua, sillä toteamalla että kerroinmatriisin rangi on n, tulemme samalla myös todenneeksi että lineaarisesti riippumattomien yhtälöiden määrä on n, joka on myös "jäljelle jäävien"yhtälöiden lukumäärä, kun otamme pois "ylimääräiset"yhtälöt. Tällaisella systeemillä on joko 0 tai 1 ratkaisua mille tahansa vektorille $\mathbf{b}$.

1.3 Paras tapaus: Yhtälöitä saman verran kuin tuntemattomia

Tässä tapauksessa kerroinmatriisimme $\mathbf{A}$ on neliömatriisi. Nyt, riippuen siitä millainen matriisi se sattuu olemaan (riippuen sen rangista), meillä on seuraavat tapaukset:

1. Yhtälöryhmällä $\mathbf{Ax} = \mathbf{0}$ on joko yksikäsitteinen ratkaisu, tai sitten ratkaisuja on äärettömästi. Triviaaliratkaisu löytyy aina, mutta nyt voi olla myös että $\mathbf{A}$ ei ole täyttä astetta, ts. joku yhtälö on muiden lineaariyhdiste. Tässä tapauksessa joku muuttuja jää vapaaksi ja ratkaisuja on ääretön määrä.

2. Yhtälöryhmällä $\mathbf{Ax} = \mathbf{b}$ on joko 0, 1, tai äärettömästi ratkaisuja (edellisten kohtien perusteella).

3. **(Mukavin tapaus)** Jos kerroinmatriisi $\mathbf{A}$ on täyttä astetta, ts. jos meillä on n muuttujaa ja matriisin aste on myös n, silloin yhtälöryhmällä $\mathbf{Ax} = \mathbf{b}$ on yksikäsitteinen ratkaisu jokaiselle valitsemallemme $\mathbf{b}$. Tämä johtuu siitä, että tällöin voimme ratkaista yhtälöryhmän seuraavaan tapaan:

$$\mathbf{Ax} = \mathbf{b} \Leftrightarrow \mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$$

Koska $\mathbf{A}$ on kääntyvä, määrää tämä yksikäsitteisen ratkaisun $\mathbf{x}$ jokaiselle valitsemallemme oikean puolen vektorille $\mathbf{b}$.

1.4 Gauss–Jordan

Ratkoessamme yhtälöryhmää $\mathbf{Ax} = \mathbf{b}$, voimme käyttää Gauss–Jordanin menetelmää yhtälön ratkaisemiseen. Ensimmäinen askel tässä on kirjoittaa täydennetty matriisi

$$\mathbf{A} : \mathbf{b}$$

Eli, jos meillä oli vaikkapa (3x3) -kerroinmatriisi $\mathbf{A}$, täydennetty matriisimme $\mathbf{A} : \mathbf{b}$ on (3 x 4) -matriisi, sillä olemme ottaneet kerroinmatriisille kaveriksi oikean puolen sarakevektorin. Gauss–Jordanin menetelmän ideana on käsitellä tätä lisättyä matriisia alkeisrivioperaatioin (lisäämällä ja vähentämällä rivejä toisistaan ja kertomalla niitä reaaliluvuilla) niin, että saamme vihdoin matriisin $\mathbf{A}$ paikalle identiteettimatriisin, jolloin erottamalla systeemin uudestaan, voimme lukea ratkaisun suoraan.

1.4.1 Esimerkki

Olkoon yhtälöryhmämme $\mathbf{Ax} = \mathbf{b}$ matriisimuotoon kirjoitettuna seuraava:

$$\begin{bmatrix} -2 & 1 & 0 & -1 \\ 2 & 3 & 7 & -2 \\ 1 & 1 & 3 & -2 \\ 3 & 5 & 13 & -9 \end{bmatrix} \begin{bmatrix} w \\ x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 9 \\ 0 \\ 1 \end{bmatrix}$$

Täydennetty matriisi $\mathbf{A} : \mathbf{b}$ on nyt siis:

$$\left[\begin{array}{cccc|c} -2 & 1 & 0 & -1 & 0 \\ 2 & 3 & 7 & -2 & 9 \\ 1 & 1 & 3 & -2 & 0 \\ 3 & 5 & 13 & -9 & 1 \end{array}\right]$$

Gauss-Jordan menetelmässä lähdemme tästä liikkeelle, ja muokkaamme alkeisrivio-peraatioilla matriisia niin, että matriisi $\mathbf{A}$ on saatu identiteettimatriisiksi (eli matriisiksi, jossa on lävistäjällä vain alkioita 1, muualla alkiot ovat 0). Silloin voimme lukea ryhmän vastauksen suoraan, sillä "kaverisarakkeemme"on muokkautunut muotoon, joka antaa meille suoraan tuntemattoman $\mathbf{x}$ komponentit (ks. harjoitustehtävä 1).

Assistentin kommentti: Ratkaisemiseen riittää oikeastaan vain muodostaa $\mathbf{A} : \mathbf{b}$ ja saattaa matriisi $\mathbf{A}$ riviporrasmuotoon. Tällöin muuttujan $\mathbf{x}$ komponentit voidaan helposti laskea.

1.5 Cramerin sääntö

Ratkaistaessa yhtälöryhmää $\mathbf{Ax} = \mathbf{b}$ Cramerin säännöllä, voimme etsiä tuntemattomat vain laskemalla matriisien determinantteja. Cramerin sääntö sanoo, että yhtälöryhmästä

$$\mathbf{Ax} = \mathbf{b}$$

voimme ratkaista muuttujan x_i seuraavalla tavalla:

$$x_i = \frac{\det \mathbf{B_i}}{\det \mathbf{A}}$$

missä matriisi $\mathbf{B_i}$ on matriisi, joka on muodostettu ottamalla kerroinmatriisi $\mathbf{A}$ ja vaihtamalla sen sarake i sarakevektoriin $\mathbf{b}$ (alkuperäisen yhtälöryhmän oikean puolen vektori). Ratkaisu vaatii tietenkin sen, että $\det \mathbf{A} \neq 0$, jottemme päädy jakamaan nollalla.

1.5.1 Esimerkki

Käytetään aiemman kohdan Gauss–Jordanin yhtälöryhmää, jossa ryhmä näytti tältä:

$$\begin{bmatrix} -2 & 1 & 0 & -1 \\ 2 & 3 & 7 & -2 \\ 1 & 1 & 3 & -2 \\ 3 & 5 & 13 & -9 \end{bmatrix} \begin{bmatrix} w \\ x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 9 \\ 0 \\ 1 \end{bmatrix}$$

Jos haluaisimme ratkaista tästä vaikkapa muuttujan w käyttäen Cramerin sääntöä, tekisimme seuraavan laskutoimituksen:

$$w = \frac{\det \begin{bmatrix} 0 & 1 & 0 & -1 \\ 9 & 3 & 7 & -2 \\ 0 & 1 & 3 & -2 \\ 1 & 5 & 13 & -9 \end{bmatrix}}{\det \begin{bmatrix} -2 & 1 & 0 & -1 \\ 2 & 3 & 7 & -2 \\ 1 & 1 & 3 & -2 \\ 3 & 5 & 13 & -9 \end{bmatrix}} = -1$$

Riippuen kerroinmatriisin koosta, Cramerin säännön käyttäminen voi joskus olla varsin työlästä[1] (tässä tapauksessa meidän tulee laskea 4×4 -matriisin determinantti neljään otteeseen, joka ei ole kovin mukavaa.).

1.6 Lineaariset yhtälöryhmät: miten toimit

1. Kirjoita ryhmä muotoon $\mathbf{Ax} = \mathbf{b}$.
2. Tarkastele ryhmää, ja ota selvää mistä tapauksesta puhutaan (vrt. kerroinmatriisin $\mathbf{A}$ rangi, tuntemattomien lukumäärä ja yhtälöiden lukumäärä). Käyttäen yllä ollutta luokittelua, tiedät nyt ratkaisujen lukumäärän.
3. Jos ryhmä on ratkaistavissa, ratkaise tuntemattomat. Voit käyttää Gauss–Jordania, kääntää kerroinmatriisin $\mathbf{A}$ tai ratkaista tuntemattomat Cramerin säännöllä.

2 Derivointi

Emme käy läpi yhden muuttujan reaalifunktioiden derivointia. Käsittely löytyy kirjan (Simon & Blume) alusta.

2.1 Gradientti

Funktion gradientti on sen suunnattu derivaatta. Se kertoo suunnan, johon funktio kasvaa nopeiten. Gradientti muodostetaan funktion osittaisderivaatoista. Olkoon meillä

[1] Tämän vuoksi esimerkiksi ohjelmointikielet eivät lähes koskaan käytä tätä menetelmää lineaaristen yhtälöryhmien ratkaisuun, vaan Gauss-Jordanin menetelmää tai vastaavia.

funktio $f(\mathbf{x}) : \mathbb{R}^n \to \mathbb{R}$. Tämän funktion *gradientti* on sen derivaattavektorin transpoosi. Eli siis:

$$\nabla f = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \cdots, \frac{\partial f}{\partial x_n} \right)^T$$

Täten, kun derivaattavektori on rivivektori, on gradientti siis sarakevektori. Gradientilla tulee olemaan paljon käyttöä kurssin toisessa osassa, kun tarkastelemme rajoitettua optimointia. Kirjan luvussa 14.6 on erittäin intuitiivinen kehitys gradientille.

2.2 Jacobin matriisi

Jacobin matriisiksi on derivaatan yleistys siihen tapaukseen, että käsittelemme vektoriarvoista kuvausta. Yleisesti funktiolle $g(\mathbf{x}) : \mathbb{R}^n \to \mathbb{R}^m$, funktion $g(\mathbf{x})$ Jacobin matriisi määritellään seuraavasti:

$$J(g(\mathbf{x})) = \begin{bmatrix} \frac{\partial g_1(\mathbf{x})}{\partial x_1} & \cdots & \frac{\partial g_1(\mathbf{x})}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial g_m(\mathbf{x})}{\partial x_1} & \cdots & \frac{\partial g_m(\mathbf{x})}{\partial x_n} \end{bmatrix}$$

Jossa siis matriisin jokainen alkio on tietyn komponenttifunktion osittaisderivaatta tietyn muuttujan suhteen. Yllä olevassa matriisissa esim. ensimmäisellä rivillä otamme ensimmäisen komponenttifunktion $g_1(\mathbf{x})$ osittaisderivaatat kaikkien muuttujien x_i suhteen. Toisella rivillä taas toisen komponenttifunktion jne. Tässä tapauksessa kuvauksemme Jacobin matriisi on $[m \times n]$ -matriisi. Esimerkiksi, jos funktiomme on kuvaus $f : \mathbb{R}^3 \to \mathbb{R}^2$, tällöin Jacobin matriisi on $[2 \times 3]$ -matriisi.

2.3 Hessen matriisi

Monen muuttujan funktion $r(\mathbf{x}) : \mathbb{R}^n \to \mathbb{R}$ Hessen matriisi voidaan käsittää tämänkaltaisen funktion "toisena derivaattana". Funktion $r(\mathbf{x})$ Hessen matriisi määritellään seuraavasti:

$$H(r(\mathbf{x})) = \begin{bmatrix} \frac{\partial^2 r(\mathbf{x})}{\partial x_1^2} & \cdots & \frac{\partial^2 r(\mathbf{x})}{\partial x_1 \partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 r(\mathbf{x})}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 r(\mathbf{x})}{\partial x_n^2} \end{bmatrix}$$

Täten jokainen alkio tässä matriisissa on alkuperäisen funktion $r(\mathbf{x})$ toinen osittaisderivaatta. Matriisin diagonaalilla ovat toiset osittaisderivaatat saman muuttujan suhteen, kun taas ei-diagonaalialkiot ovat ristiderivaattoja. Matriisin laskemisessa Youngin teoreeman[2] hyödyntäminen pienentää usein työmäärää, mikäli funktio on jatkuvasti derivoituva. Esimerkiksi siis yllä olevan matriisin 1. rivin 3. alkio on

$$H(r(\mathbf{x}))_{1,3} = \frac{\partial^2 r(\mathbf{x})}{\partial x_1 \partial x_3}$$

jne. Jos hahmotus tuntuu vaikealta, voitte ajatella että "ensin derivoidaan rivin mukaan, sitten sarakkeen".

2.3.1 Hessen matriisi ja gradientti

Opimme aiemmin, että gradientti on funktion suunnattu derivaatta. Jos katsomme yllä olevan Hessen matriisin rivejä, huomaamme että jokaisella rivillä i, derivoimme osittaisderivaattaa muuttujan x_i suhteen uudestaan kaikkien muuttujien x suhteen. Näin voimme siis kirjoittaa:

$$H(r(\mathbf{x})) = J(\nabla r(\mathbf{x}))$$

jossa siis $\nabla r(\mathbf{x})$ on funktion r gradientti.

3 Implisiittifunktiolause

Usein taloustieteessä olemme kiinnostuneet komparatiivisesta statiikasta, eli mallin määräämien *endogeenisten muuttujien* muutoksesta silloin kuin jokin *eksogeeninen muuttuja* muuttuu vähän. Joissain tapauksissa voimme ratkaista tämän suoraan eksplisiittisesti mallista, kun taas toisissa tapauksissa emme voi ratkoa endogeenisten muuttujien muutosta suoraan. Implisiittifunktiolause antaa meille työkalut tällaiseen tarkasteluun, ja on siksi hyödyllinen apuväline taloustieteessä.

Käydään läpi luentomateriaalin esimerkki.

Esimerkki 1. *Olkoon* $f : \mathbb{R}^2 \to \mathbb{R} : f(x,y) = bx + ay$. *Analysoidaan mallin käytöstä pisteen* $f(\hat{x}, \hat{y}) = 0$ *ympäristössä. Voimme ratkaista helposti:*

[2] so. ristiderivaattojen symmetrisyyden.

$$\frac{dy}{dx} = -\frac{b}{a}, \quad a \neq 0 \tag{1}$$

Koska nollalla ei ole syytä jakaa, jouduimme aiemmassa määrittelemään ehdon $a \neq 0$. Tämä yksinkertainen ehto kannattaa pitää mielessä, sillä vaatimukset endogeenisten muuttujien kerroinmatriisin asteesta eivät oikeastaan ole mitään muuta kuin tämä sama vaatimus, nyt vain hieman eri ympäristössä ja eri työkaluja käyttäen.

3.1 Miksi ehdot ovat sellaiset kuin ne ovat?

Kehitetään ehdot implisiittifunktiolauseen käyttöön intuitiivisesti. Olkoon meillä joku hyvin käyttäytyvä funktio $F(x, y) : \mathbb{R}^2 \to \mathbb{R}$. Meillä on olemassa joku piste $(\hat{x}, \hat{y})$, joka toteuttaa seuraavan ehdon:

$$F(\hat{x}, \hat{y}) = 0 \tag{2}$$

Jos funktio on sellainen, ettemme helposti saa siitä eksplisiittisesti ratkaistua muuttujaa y muuttujan x funktiona, on seuraavaksi paras vaihtoehto yrittää löytää jonkinnäköinen ympäristö, jossa y voidaan implisiittisesti määritellä muuttujan x funktiona[3]. Oletetaan ensin, että pisteen $(\hat{x}, \hat{y})$ "pienessä"ympäristössä voidaan määritellä

jokaiselle x -muuttujalle yksi ja vain yksi $y(x)$, eli implisiittisesti määräytyvä y niin, että $y(\hat{x}) = \hat{y}$. Tällöin tässä ympäristössä pätee:

$$F(x, y(x)) = 0 \tag{3}$$

Mikäli y on derivoituva funktio x:n suhteen, voimme ottaa näin määräämästämme funktiosta kokonaisdifferentiaalin. Huomatkaa, että ylempänä oleva yhtälö määrittää funktion *tasojoukon*, ts. joukon pisteitä, joille funktiomme F arvo on vakio 0 (ajatelkaa indifferenssikäyriä kuluttajan ongelmassa). Mikäli F:n arvo pysyy vakiona, silloin jos liikumme "pienessä"ympäristössä pisteen $(\hat{x}, \hat{y})$ ympärillä, muuttaen muuttujan x arvoa, on funktion kokonaismuutoksen oltava nolla. Täten voimme kokonaisdifferentiaalia käyttäen kirjoittaa kun muistamme yhdistetyn funktion derivointisäännön:

$$\underbrace{\frac{\partial F}{\partial x}}_{\text{F:n muutos kun x muuttuu}} + \underbrace{\frac{\partial F}{\partial y}}_{\text{F:n muutos kun y muuttuu}} \cdot \underbrace{y'(x)}_{\text{sisäfunktion derivaatta}} = 0 \tag{4}$$

[3]Esimerkkinä tästä vaikkapa työmarkkinoiden etsintämalleissa esiintyvä reservaatiopalkka, jonka lauseke on usein varsin monimutkainen olio. Kuitenkin olisi hyödyllistä tietää, miten jonkun *eksogeenisen* muuttujan, kuten esim. työttömyyskorvauksen muutos vaikuttaa palkkaan. Implisiittifunktiolause antaa meidän tehdä juuri tämän tempun.

Josta voimme ratkaista

$$\frac{dx}{dy} = y'(x) = -\frac{\frac{\partial F}{\partial x}}{\frac{\partial F}{\partial y}} \tag{5}$$

koska nollalla ei ole syytä jakaa, on tämä tulos pätevä ainoastaan silloin, kuin $\frac{\partial F}{\partial y} \neq 0$.

Mikä $\frac{\partial F}{\partial y}$ oikeastaan on? Jos kutsuisimme muuttujaa x *eksogeeniseksi*, eli järjestelmän ulkopuolella määräytyväksi ja muuttujaa y *endogeeniseksi*, olisi tämä termi funktiomme F osittaisderivaatta endogeenisten muuttujien suhteen. Se ei saa olla nolla, koska muuten oletuksemme eivät päde, emmekä voi laskea haluttua implisiittistä derivaattaa. Jos olemme toisenlaisessa ympäristössä, jossa funktiomme F on vaikkapa kuvaus kolmiulotteisesta euklidisesta avaruudesta kaksiulotteiseen avaruteen, eli $F : \mathbb{R}^3 \to \mathbb{R}^2$, tällöin emme enää voi puhua osittaisderivaatoista samaan tapaan, vaan puhumme **Jacobin matriisista**. Siis:

- F vektoriarvoinen: "osittaisderivaatta $\frac{\partial F}{\partial y}$"$\to D_y(F(\mathbf{x}, \mathbf{y}))$, eli funktion F **Jacobin matriisi** endogeenisten muuttujien $\mathbf{y}$ suhteen
- F vektoriarvoinen: "$\frac{\partial F}{\partial y} \neq 0$"$\to$ matriisin $D_y(F(\mathbf{x}, \mathbf{y}))$ pitää olla kääntyvä matriisi.

Palataan aiempaan esimerkkiimme, ja yritetään intuitiivisesti kehittää edellinen kääntyvyysehto. Nyt vektoriarvoisen funktiomme F tapauksessa voimme kirjoittaa:

$$F(\mathbf{x}, H(\mathbf{x})) = \mathbf{0} \tag{6}$$

Derivoimalla molemmat puolet ja pahoinpitelemällä notaatiota hieman, saamme

$$D_x(F(\mathbf{x}, H(\mathbf{x})))D_x\mathbf{x} + D_y(F(\mathbf{x}, H(\mathbf{x})))H'(\mathbf{x}) = \mathbf{0} \tag{7}$$

Koska $D_x\mathbf{x} = \mathbf{I}$, saadaan

$$D_x(F(\mathbf{x}, H(\mathbf{x}))) + D_y(F(\mathbf{x}, H(\mathbf{x})))H'(\mathbf{x}) = \mathbf{0} \tag{8}$$

josta

$$D_y(F(\mathbf{x}, H(\mathbf{x})))H'(\mathbf{x}) = -D_x(F(\mathbf{x}, H(\mathbf{x}))) \tag{9}$$

Ajatellaan tätä aivan samankaltaisena yhtälöryhmänä $\mathbf{Ax} = \mathbf{b}$, joita olemme jo ratkoneet, ja saamme

$$H'(\mathbf{x}) = -\big[D_y(F(\mathbf{x}, H(\mathbf{x})))\big]^{-1} D_x(F(\mathbf{x}, H(\mathbf{x}))) \tag{10}$$

Mutta tällainen ratkaisu on mahdollinen ainoastaan silloin kuin matriisi

$$\big[D_y(F(\mathbf{x}, H(\mathbf{x})))\big]^{-1}$$

on olemassa, so. endogeenisten muuttujien $\mathbf{y}$ Jacobin matriisin pitää olla kääntyvä matriisi. Cramerin säännön soveltaminen muutosten laskemiseen perustuu ainoastaan yhtälön (9) muodostamiseen ja sen ratkaisuun.

3.2 Toimintaohjeet

Miten käytät implisiittifunktiolausetta:

1. Etsi kandidaattipiste, jonka ympärillä funktion muutosta approksimoidaan (tämä tulee yleensä annettuna).

2. Tarkista, että piste toteuttaa implisiittifunktion, ts. sijoita pisteen arvot funktioon, ja tarkasta että yhtälöt pätevät.

3. **Jaottele muuttujat** endogeenisiin (systeemissä määräytyviin) ja eksogeenisiin (systeemin ulkopuolella määräytyviin) muuttujiin. Laskemme siis eksogeenisten muuttujien pienen muutoksen vaikutusta endogeenisiin (esimerkissä 1 siis pienen muutoksen dx vaikutusta muuttujaan y, eli dy.

4. Varmistu, ettet ”jaa nollalla”, eli valitsemiesi endogeenisten muuttujien Jacobin matriisi (”derivaatta”) on **kääntyvä matriisi**. Toisin sanoen, matriisin $D_y f(\cdot, \cdot)$ determinantin on oltava nollasta poikkeava pisteessä $(\hat{x}, \hat{y})$.

5. Laske haluamasi muutos, joko Cramerin säännöllä tai kääntämällä $D_y f(\hat{y})$

3.2.1 Esimerkki

Sovelletaan implisiittifunktiolausetta ja toimintaohjeita seuraavaan ongelmaan:

Olkoon kuvaus

$$F_1(x, y, z) = x + y^2 + \frac{1}{z} - 3$$
$$F_2(x, y, z) = \sqrt{y} - x + 2z - 2$$

Ratkaistaan muuttujien (x, y) muutos kun muuttuja z muuttuu hiukan pisteen $(x = 1, y = 1, z = 1)$ ympäristössä.

- Kohdat 1 ja 2: kandidaattipiste on annettu, ja sijoittamalla sen komponenttifunktioihin näemme, että piste $(1, 1, 1)$ toteuttaa ehdon implisiittifunktiolauseen käyttöön, ts.

$$F_1(1, 1, 1) = 0$$
$$F_2(1, 1, 1) = 0$$

- Kohta 3: **Endogeeniset muuttujat ovat** (x, y), **eksogeeninen muuttuja on** z.

- Kohta 4: Muodostetaan kokonaisdifferentiaali ja varmistetaan endogeenisten muuttujien Jacobin matriisin kääntyvyys. Ensin kokonaisdifferentiaali:

$$J(F(x, y, z)) \begin{bmatrix} dx \\ dy \\ dz \end{bmatrix} = \mathbf{0}$$

eli

$$\begin{bmatrix} 1 & 2y & -z^{-2} \\ -1 & \frac{1}{2}y^{-\frac{1}{2}} & 2 \end{bmatrix} \begin{bmatrix} dx \\ dy \\ dz \end{bmatrix} = \mathbf{0}$$

Sijoitetaan piste $(1, 1, 1)$ ja ryhmä on:

$$\begin{bmatrix} 1 & 2 & -1 \\ -1 & \frac{1}{2} & 2 \end{bmatrix} \begin{bmatrix} dx \\ dy \\ dz \end{bmatrix} = \mathbf{0}$$

kirjoitetaan ryhmä uudestaan, erotellaan endogeeniset ja eksogeeniset muuttujat:

$$\begin{bmatrix} 1 & 2 \\ -1 & \frac{1}{2} \end{bmatrix} \begin{bmatrix} dx \\ dy \end{bmatrix} + \begin{bmatrix} -1 \\ 2 \end{bmatrix} dz = \mathbf{0}$$

Nyt implisiittifunktiolausetta voidaan käyttää muutoksen laskemiseen jos ja vain jos determinantti

$$\det \begin{bmatrix} 1 & 2 \\ -1 & \frac{1}{2} \end{bmatrix} \neq 0$$

Tämä on endogeenisten muuttujien (x, y) Jacobin matriisin determinantti pisteessä $(1, 1, 1)$. Yllä oleva determinantti on

$$\det \begin{bmatrix} 1 & 2 \\ -1 & \frac{1}{2} \end{bmatrix} = \frac{5}{2} \neq 0$$

joten implisiittifunktiolausetta voidaan soveltaa.

- Kohta 5: Lasketaan haluttu muutos, vaikka käyttämällä Cramerin sääntöä. Muokataan ryhmä muotoon

$$\begin{bmatrix} 1 & 2 \\ -1 & \frac{1}{2} \end{bmatrix} \begin{bmatrix} dx \\ dy \end{bmatrix} = - \begin{bmatrix} -1 \\ 2 \end{bmatrix} dz$$

Nyt soveltamalla Cramerin sääntöä voimme laskea

$$dx = \frac{\det \begin{bmatrix} 1 & 2 \\ -2 & \frac{1}{2} \end{bmatrix}}{\det \begin{bmatrix} 1 & 2 \\ -1 & \frac{1}{2} \end{bmatrix}} dz = \frac{\frac{9}{2}}{\frac{5}{2}} = \frac{9}{5} dz$$

muutoksen dy voitte laskea itse samaan tapaan.

4 Matriisien definiittisyys ja kvadraattiset muodot

Neliömatriisien definiittisyydestä puhuttaessa olemme kiinnostuneita siitä, minkä arvon seuraavanlainen termi saa, jota kutsumme kvadraattiseksi muodoksi. Olkoon kiinnostuksemme kohde termi:

$$\mathbf{v}^T \mathbf{A} \mathbf{v}$$

jossa $\mathbf{A}$ on sopivankokoinen neliömatriisi, ja $\mathbf{v}$ sille sopivaulottuvuuksinen vektori (täten yllä oleva termi on skalaari). Sanomme, että:

- $\mathbf{A}$ on positiividefiniitti, jos kaikille $\mathbf{v} \neq \mathbf{0}$, $\mathbf{v}^T \mathbf{A} \mathbf{v} > 0$.
- $\mathbf{A}$ on negatiividefiniitti, jos kaikille $\mathbf{v} \neq \mathbf{0}$, $\mathbf{v}^T \mathbf{A} \mathbf{v} < 0$.

Motivaatio siihen, miksi tämä on tärkeä asia löytyy esim. harjoituksen 3 vastauksista. Matriisien definiittisyys määrätään tarkastelemalla matriisien minoreja. Erityisenä mielenkiinnon kohteenamme ovat matriisin **johtavat pääminorit**.

Määritelmä: $n \times n$ -matriisin $\mathbf{A}$ k -asteen johtava pääminori on sellaisen alimatriisin determinantti, joka saadaan poistamalla matriisista A viimeiset $(n-k)$ riviä ja $(n-k)$ saraketta.

4.1 Esimerkki

Tarkastellaan seuraavan matriisin johtavia pääminoreja ja definiittisyyttä:

$$\mathbf{B} = \begin{bmatrix} 1 & 2 & 4 \\ 5 & 0 & 6 \\ 3 & 1 & 4 \end{bmatrix}$$

Ensimmäinen johtava pääminori saadaan poistamalla $(3-1) = 2$ viimeistä saraketta ja riviä, jolloin saamme

$$M_1 = \det(1) = 1$$

Eli vain matriisin alkio $b_{1,1}$. Toinen johtava pääminori saadaan poistamalla $(3-2) = 1$ riviä ja saraketta, eli

$$M_2 = \det \begin{bmatrix} 1 & 2 \\ 5 & 0 \end{bmatrix} = -10$$

Kolmas johtava pääminori saadaan poistamalla samaan tapaan $(3-3) = 0$ riviä ja saraketta alkuperäisestä matriisista, eli kolmas johtava pääminori on koko matriisin $\mathbf{B}$ determinantti.

$$M_3 = \det \mathbf{B} = 10$$

Nämä työkalut antavat meille keinon tutkia sitä onko matriisi positiivi- vai negatiividefiniitti. Testi, jota käytämme on seuraava:

1. *Matriisi A on positiividefiniitti jos ja vain jos kaikki sen johtavat pääminorit ovat positiivisia.*
2. *Matriisi A on negatiividefiniitti jos ja vain jos sen johtavien pääminorien merkit vaihtelevat seuraavasti:*

$$
\begin{aligned}
M_1 &< 0 \\
M_2 &> 0 \\
M_3 &< 0 \\
&\vdots
\end{aligned}
$$

ja yleisesti asteen k johtavan pääminorin merkki on sama kuin termillä $(-1)^k$.

3. Jos kumpikaan näistä ehdoista ei toteudu, sanomme (tässä vaiheessa) vain että matriisi ei ole positiivi- tai negatiividefiniitti[4]. Näemme, että esimerkkimatriisimme $\mathbf{B}$ ei toteuta ehtoa 1 eikä ehtoa 2. Täten $\mathbf{B}$ ei ole positiivi- eikä negatiividefiniitti matriisi. Olemme kiinnostuneet oppimaan nämä työkalut, sillä tulemme käyttämään niitä monimuuttujafunktioiden optimoinnissa toisen kertaluvun ehtoinamme.

5 Rajoittamaton optimointi ja ääriarvopisteet

5.1 Yhden muuttujan reaaliarvoiset funktiot

Jos meillä on vastassa yhden muuttujan reaalifunktio $f : \mathbb{R} \to \mathbb{R}$, $f(x)$, tiedämme jo aiempien kurssien asioista, miten etsimme tällaisen funktion ääriarvopisteet ja luokittelemme ne. Ensimmäisen kertaluvun välttämätön ehto ääriarvopisteelle, olettaen että f on jatkuva ja derivoituva määrittelyjoukossaan on

$$f'(x^*) = 0$$

Joka voidaan johtaa helposti ääriarvopisteen määritelmästä jatkuvalle ja derivoituvalle funktiolle f. Taylorin approksimaation avulla, voimme helposti myös näyttää toisen kertaluvun riittävät ehdot (lokaalille) minimille tai maksimille (ks. harjoitustehtävä 3). Ne ovat seuraavat:

1. Mikäli $f'(x^*) = 0$, ja $f''(x^*) < 0$, on funktiolla f pisteessä x^* lokaali maksimi.

[4] Myöhemmin kurssilla käsittelemme myös tapauksia, jolloin lievennämme ehtojamme myös siihen tapaukseen, että joku johtavista pääminoreista on 0. Tällöin puhumme positiivi- tai negatiivisemidefiniiteistä matriiseista, mutta kurssin 1. osassa sivuutamme nämä tapaukset.

2. Mikäli $f'(x^*) = 0$, ja $f''(x^*) > 0$, on funktiolla f pisteessä x^* lokaali minimi.

Tämän todistus nojasi Taylor-approksimaatioon ja logiikkaan siitä, että toisen derivaatan merkki kertoo meille mihin suuntaan funktion arvo muuttuu, kun liikumme ääriarvopisteen x^* ympäristössä. Mikäli funktion arvo vähenee, silloin kyseessä on lokaali maksimi, jos se kasvaa, olemme lokaalissa minimissä. Globaalin ääriarvon määrittämiseksi meidän pitää pystyä sanomaan jotakin funktiosta $f(x)$, so. se, onko kyseessä konveksi vai konkaavi funktio (ks. itseopiskelumateriaali kurssisivuilla). Tähän asti asiat ovat tuttuja aiemmista opinnoista (tai ainakin pitäisi olla).

5.2 Monen muuttujan reaaliarvoiset funktiot

Kun siirrymme tarkastelemaan monimuuttujafunktioita, eli kuvauksia $f : \mathbb{R}^n \to \mathbb{R}$, teemme edelleen samat asiat ääriarvojen etsinnässä kuin aiemmin, nyt vain eri työkaluilla. Ensimmäisen kertaluvun ehdot tarkoittavat nyt sitä, että etsimme funktion *gradientin* ja asetamme sen nollaksi. Näin saamme funktion *kriittiset pisteet*, ts. kaikki ne pisteet jotka toteuttavat ensimmäisen kertaluvun välttämättömän ehdon. Toisen kertaluvun ehdot tarkoittavat nyt sitä, että toisen derivaatan sijaan tarkastamme nyt funktion Hessin matriisin definiittisyyden ratkaisemissamme kriittisissä pisteissä.

5.2.1 Hessen matriisi ja toisen kertaluvun riittävät ehdot ääriarvopisteelle

Asioiden ymmärtämistä helpottaa ehkä se, jos vetää mielessään analogian yhden muuttujan funktion ääriarvopisteiden etsimisestä monen muuttujan tapaukseen. Tiedämme jo aiemmin oppimamme perusteella, että etsiessämme jatkuvalle, derivoituvalle funktiolle $f(x) : \mathbb{R} \to \mathbb{R}$ ääriarvoja, löytyvät ne

1. Derivaatan nollakohdasta (-kohdista), eli pisteistä x^*, jotka toteuttavat seuraavan ehdon (ehto voidaan johtaa maksimin tai minimin määritelmästä jatkuvalle ja derivoituvalle f):

$$f'(x^*) = 0$$

2. Mikäli rajaudumme etsimään ääriarvoja jollain välillä $x \in [a, b]$, voi ääriarvopiste sijaita myös välin päätepisteissä.

Toisen kertaluvun ehdoista taas tiedämme yhden muuttujan tapauksessa seuraavaa:

1. Mikäli $f'(x^*) = 0$, ja $f''(x^*) < 0$, on funktiolla f pisteessä x^* lokaali maksimi.

2. Mikäli $f'(x^*) = 0$, ja $f''(x^*) > 0$, on funktiolla f pisteessä x^* lokaali minimi.

Globaalin minimin tai maksimin määrääminen edellyttää, että voimme sanoa jotakin funktiosta f, lähinnä sen onko kyseessä konveksi vai konkaavi funktio. Toisen kertaluvun ehtojen todistusta emme käy tässä läpi, mutta yritämme kehittää intuitiivisen ymmärryksen siitä, miksi saamme yllä olevat ehdot lokaalille minimille tai maksimille. Jotta voisimme määrittää ääriarvopisteen laadun, tulee meillä olla jonkinnäköinen ymmärrys siitä, miten funktiomme käyttäytyy löytämämme ääriarvopisteen ympäristössä. Tähän tutkimiseen voimme käyttää funktion f nk. *Taylorin approksimaatiota* ääriarvopisteen x^* "sopivan pienessä"ympäristössä.

Funktion $f(x)$ Taylor-approksimaatio pisteen x_0 ympäristössä määritellään seuraavasti:

$$f(x) \approx f(x_0) + f'(x_0)h + \frac{1}{2!}f''(x_0)h^2 + \frac{1}{3!}f'''(x_0)h^3 + \cdots \tag{11}$$

jossa $h = x - x_0$.

Muodostetaan nyt Taylor-approksimaatio funktiollemme f löytämämme ääriarvopisteen x^* ympäristössä.

$$f(x) \approx f(x^*) + f'(x^*)h + \frac{1}{2!}f''(x^*)h^2 + \frac{1}{3!}f'''(x^*)h^3 + \cdots \tag{12}$$

Oletetaan, että rajoitumme tarpeeksi pieneen ympäristöön, jolloin kolmannen asteen termit ja siitä ylöspäin ovat häviävän pieniä. Nyt, koska ääriarvopisteessä pitää päteä että:

$f'(x^*) = 0$, funktiomme arvon muutos määräytyy toisen derivaatan merkin mukaan. Jos $f''(x^*) < 0$, silloin tiedämme, että liikkuessamme vähän pisteestä x^* suuntaan tai toiseen, funktion arvo laskee. Tällöin pisteemme on *lokaali maksimi.* Mikäli taas toisinpäin, eli $f''(x^*) > 0$ silloin funktion arvo on aina korkeampi, kun liikumme pisteestä x^* hieman johonkin suuntaan ja silloin löytämämme piste on *lokaali minimi.*

Välikysymys: mitä tällä on tekemistä Hessen matriisin, gradientin ja definiittisyyden kanssa?

Monen muuttujan funktioiden maailmassa, joudumme käyttämään hieman erilaisia työkaluja. Perusintuitio on silti pitkälti sama kuin yhden muuttujan tapauksessa. Olkoon meillä nyt monen muuttujan funktio $f(\mathbf{x})$. Jos meidän pitää etsiä tälle funktiolle ääriarvot ja ääriarvon laatu, aloitamme ensimmäisen kertaluvun ehdoista, eli asettamalla *funktion gradientin nollaksi*. Tätä voi ajatella samoin kuin derivaatan nollakohtaa aiemmin yhden muuttujan funktion kanssa, mutta nyt vain funktiossamme on monta muuttujaa, ja terminologia ja työkalut näyttävät hieman erilaisilta.

Hyvä, siispä asetamme funktion gradientin nollaksi:

$$\nabla f(\mathbf{x}^*) = \mathbf{0}$$

Tästä saamme funktion *kriittiset pisteet* $\mathbf{x}^*$. Olemme siis nyt etsineet mahdolliset kandidaatit pisteille, missä funktio saa ääriarvonsa. Nyt pitäisi tarkastella ääriarvon laatua. Aiemmin tutkimme funktion toista derivaattaa, ja sen tarkoituksenmukaisuuden näytimme käyttämällä Taylorin approksimaatiota ääriarvopisteen ympärillä. Teemme nyt täysin saman asian. Monimuuttujafunktion $f(\mathbf{x})$ Taylor-approksimaatio pisteen $\mathbf{x}^*$ ympäristössä voidaan kirjoittaa seuraavalla tavalla:

$$f(\mathbf{x}) \approx f(\mathbf{x}^*) + (\mathbf{x} - \mathbf{x}^*)^T D(f(\mathbf{x}^*)) + \frac{1}{2!}(\mathbf{x} - \mathbf{x}^*)^T D^2(f(\mathbf{x}^*))(\mathbf{x} - \mathbf{x}^*) + \cdots \quad (13)$$

Jossa nyt joudumme käyttämään eri terminologiaa, sillä puhumme funktiosta, jolla on argumenttina vektori. Ylläolevassa muodossa kuitenkin termit $D(f(\mathbf{x}^*))$ ja $D^2(f(\mathbf{x}^*))$ paljastuvat asioiksi, jotka meille on jo opetettu. Miksi näin? Siksi, koska tässä muodossa

- $D(f(\mathbf{x}^*)) = \nabla f(\mathbf{x}^*)$, ja
- $D^2(f(\mathbf{x}^*)) = H(\mathbf{x}^*)$

Eli ensimmäisen asteen termi on funktion gradientti, ja toisen asteen termi sen *Hessen matriisi* kyseisessä pisteessä. Tiedämme jo, että välttämätön ehto kriittiselle pisteelle on se, että gradientti on nollassa. Täten Taylor-approksimaatiossa

$$D(f(\mathbf{x}^*)) = \mathbf{0}$$

Joten jäljelle jää vain ongelmallinen toisen asteen termi, joka sisältää funktion *Hessen matriisin* $H(\mathbf{x}^*)$ (jos siis jälleen rajoitumme "pieneen"ympäristöön, eikä meidän tarvitse välittää korkeamman asteen termeistä). Olemme juuri oppineet jotakin matriisien definiittisyydestä ja kvadraattisista muodoista, joten sovelletaan oppimaamme tässä. Tiedämme, että jos meillä on vektori $\mathbf{v} \in \mathbb{R}^n$, ja myös [n x n] -matriisi W, sanomme että matriisi W on positiividefiniitti, jos kaikille $\mathbf{v} \neq \mathbf{0}$:

$$\mathbf{v}^T W \mathbf{v} > 0.$$

ja negatiividefiniitti, jos

$$\mathbf{v}^T W \mathbf{v} < 0.$$

Tutkaillaan lähemmin jäljelle jäänyttä Taylor-approksimaatiota funktiosta $f(\mathbf{x})$ pisteen $\mathbf{x}^*$ ympäristössä. Gradientin ollessa nollassa jäljelle jäi vain:

$$f(\mathbf{x}) \approx f(\mathbf{x}^*) + \frac{1}{2!} \underbrace{(\mathbf{x} - \mathbf{x}^*)^T}_{\mathbf{v}^T} \underbrace{D^2(f(\mathbf{x}^*))}_{W} \underbrace{(\mathbf{x} - \mathbf{x}^*)}_{\mathbf{v}} \tag{14}$$

Ja niin kauan kuin $(\mathbf{x} - \mathbf{x}^*) \neq \mathbf{0}$ (se on, koska tarkastelemme ympäristöä!), huomaamme siis olevamme täysin analogisessa tilanteessa kuin yhden muuttujan funktion kanssa, mutta vain käyttäen eri terminologiaa. Siis jos funktion Hessen matriisi $H(\mathbf{x}^*)$ on:

1. Negatiividefiniitti, silloin myös $(\mathbf{x}-\mathbf{x}^*)^T D^2(f(\mathbf{x}^*))(\mathbf{x}-\mathbf{x}^*) < 0$. Täten liikuttaessa johonkin suuntaan kriittisestä pisteestä $\mathbf{x}^*$ funktion arvo aina vähenee. Tällöin meillä on kyseessä **lokaali maksimi**.

2. Positiividefiniitti, silloin $(\mathbf{x} - \mathbf{x}^*)^T D^2(f(\mathbf{x}^*))(\mathbf{x} - \mathbf{x}^*) > 0$, liikuttaessa johonkin suuntaan kriittisestä pisteestä $\mathbf{x}^*$ funktion arvo kasvaa, ja kyseessä on **lokaali minimi**.

Ja mikäli Hessen matriisi on ääriarvopisteessä indefiniitti (ei positiivi- tai negatiividefiniitti, tai -semidefiniitti), on silloin kyseessä **satulapiste**, tällöin termin $(\mathbf{x} - \mathbf{x}^*)^T D^2(f(\mathbf{x}^*))(\mathbf{x} - \mathbf{x}^*)$ merkki riippuu siitä, mihin suuntaan lähdemme etenemään, ts. millaisessa ympäristössä liikumme. Funktio voi olla kasvava yhteen suuntaan ja vähenevä toiseen suuntaan[5].

Jos siis meillä on joku reaaliarvoinen funktio $f(\mathbf{x})$, jossa $\mathbf{x} \in \mathbb{R}^n$, ja onnistumme löytämään jonkin pisteen $\mathbf{x}^*$, jossa

$$\nabla f(\mathbf{x}^*) = \mathbf{0}$$

On tämä funktion *kriittinen piste*. Määrittääksemme ääriarvopisteen laadun, meidän pitää tarkastaa funktion f Hessen matriisi kyseisessä pisteessä. Mikäli funktion Hessen matriisi $H(\mathbf{x}^*)$ on:

1. Negatiividefiniitti, silloin liikuttaessa johonkin suuntaan kriittisestä pisteestä $\mathbf{x}^*$ funktion arvo aina vähenee. Tällöin meillä on kyseessä **lokaali maksimi**.
2. Positiividefiniitti, silloin liikuttaessa johonkin suuntaan kriittisestä pisteestä $\mathbf{x}^*$ funktion arvo kasvaa, ja kyseessä on **lokaali minimi**.
3. Jos Hessen matriisi ei täytä kumpaakaan ehtoa, silloin piste ei ole selkeästi funktion lokaali minimi tai maksimi.

Lisäys: Jos funktion Hessen matriisi ei täytä ehtoja negatiivi- tai positiividefiniittisyydelle, silloin tiedämme, että toisen kertaluvun *riittävät ehdot* eivät toteudu tässä pisteessä. *Välttämätön ehto*, eli negatiivi- tai positiivi**semi**definiittisyys saattaa silti täyttyä, mutta mikäli törmäämme kurssin tässä vaiheessa tämän muotoiseen matriisiin rajoittamattomassa optimoinnissa, tyydymme toteamaan, että kriittinen piste ei ole yksikäsitteisesti minimi eikä maksimi.

[5]Muistakaa, että vektori $(\mathbf{x}-\mathbf{x}^*)$ voidaan esimerkiksi määritellä etenemisenä vain yhden ulottuvuuden yksikkövektorin $\mathbf{e}_i$ suuntaan, jolloin on ehkä helpompi käsittää mitä satula- tai käännepiste tarkoittaa.

5.3 Toimintaohjeet, reaaliarvoisen monimuuttujafunktion rajoittamaton optimointi

1. Muodosta tavoitefunktio, mikäli sitä ei ole suoraan annettu (vrt. harjoitustehtävien voittofunktiot ym.).
2. Ota ensimmäisen kertaluvun ehdot, eli laske funktion gradientti.
3. Aseta funktion gradientti nollaksi.
4. Ratkaise saamasi yhtälöt ja saat funktion *kriittiset pisteet*.
5. Tarkasta jokaiselle saamallesi pisteelle toisen kertaluvun ehdot, eli laske funktion Hessen matriisi jokaisessa kriittisessä pisteessä.
6. Määrää Hessen matriisin definiittisyys jokaisessa kriittisessä pisteessä.
7. Luokittele saamasi kriittiset pisteet yllä olevan taksonomian mukaan, riippuen Hessen matriisin definiittisyydestä.

5.3.1 Esimerkki

Optimoidaan nyt esimerkkinä monimuuttujafunktio $f(x,y) = 3x^4 + 3x^2y - y^3$ käyttäen yllä olevia toimintaohjeita. Tavoitefunktio on yllä oleva funktio, joten voimme hypätä kohdan 1 yli.

- Otamme ensimmäisen kertaluvun ehdot. Muodostetaan funktion gradientti (kohta 2):

$$\nabla f(x,y) = \begin{bmatrix} \frac{\partial f(x,y)}{\partial x} \\ \frac{\partial f(x,y)}{\partial y} \end{bmatrix} = \begin{bmatrix} 12x^3 + 6xy \\ 3x^2 - 3y^2 \end{bmatrix}$$

- Asetamme gradientin nollaksi (kohta 3), saamme:

$$\nabla f(x,y) = \mathbf{0} \Leftrightarrow \begin{bmatrix} 12x^3 + 6xy \\ 3x^2 - 3y^2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

- Ratkaistaan *kriittiset pisteet* (kohta 4):

Gradientin alemmasta komponentista (eli osittaisderivaatasta y:n suhteen) saamme

$$3x^2 - 3y^2 = 0 \Leftrightarrow 3(x+y)(x-y) = 0$$

joten $y = \pm x$. Sijoitetaan ensimmäinen vaihtoehto, $y = x$ gradientin ensimmäiseen komponenttiin (osittaisderivaattaan x:n suhteen) ja saamme:

$$12x^3 + 6x^2 = 0 \Leftrightarrow 6x^2(2x+1) = 0$$

josta voidaan ratkaista että $x = 0$ tai $x = -\frac{1}{2}$. Koska $y = x$, on meillä nyt seuraavat kriittiset pisteet:

$$(0,0) \quad \text{ja} \quad \left(-\frac{1}{2}, -\frac{1}{2}\right)$$

Muistetaan vielä, että pitää käydä vielä läpi tapaus $y = -x$, ja teemme samat vaiheet uudestaan. Sijoittamalla tämä ensimmäiseen komponenttiin, meillä on nyt ratkaistavana yhtälö

$$12x^3 - 6x^2 = 0 \Leftrightarrow 6x^2(2x-1) = 0$$

Josta saadaan yksi sama kriittinen piste kun aiemmin, piste $(0,0)$, ja toinen kandidaatti $x = \frac{1}{2}$, $y = -\frac{1}{2}$. Löytämämme kriittiset pisteet ovat siis:

$$(0,0) \quad , \left(-\frac{1}{2}, -\frac{1}{2}\right) \quad \text{ja} \quad \left(\frac{1}{2}, -\frac{1}{2}\right)$$

- **Kohta 5**. Tarkastamme löytämiemme ääriarvopisteiden laadun laskemalla funktion Hessen matriisin ja tarkastamalla sen definiittisyyden näissä pisteissä.

  Funktiomme Hessen matriisi on:

$$H(f(x,y)) = \begin{bmatrix} \frac{\partial^2 f(x,y)}{\partial x^2} & \frac{\partial^2 f(x,y)}{\partial x \partial y} \\ \frac{\partial^2 f(x,y)}{\partial y \partial x} & \frac{\partial^2 f(x,y)}{\partial y^2} \end{bmatrix} = \begin{bmatrix} 36x^2 + 6y & 6x \\ 6x & -6y \end{bmatrix}$$

Pisteessä $(0, 0)$,

$$H(f(0,0)) = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

Ensimmäinen johtava pääminori M_1 on $\det[0] = 0$, ja toinen johtava pääminori on $M_2 = \det(H) = 0$. Täten Hessen matriisi ei ole positiivi- eikä negatiividefiniitti, ja löytämämme piste **ei ole yksiselitteisesti lokaali minimi eikä maksimi** (kohta 6).

Pisteessä $\left(-\frac{1}{2}, -\frac{1}{2}\right)$:

$$H\left(f\left(-\frac{1}{2}, -\frac{1}{2}\right)\right) = \begin{bmatrix} 6 & -3 \\ -3 & 3 \end{bmatrix}$$

Ensimmäinen johtava pääminori on $M_1 = \det[6] = 6 > 0$. Toinen johtava pääminori on $M_2 = \det(H) = 18 - 9 = 9 > 0$. Koska kummatkin johtavat pääminorit ovat positiivisia, on Hessen matriisi positiividefiniitti. Täten tämä piste on **lokaali minimi** (kohta 6).

Pisteessä $\left(\frac{1}{2}, -\frac{1}{2}\right)$:

$$H\left(f\left(\frac{1}{2}, -\frac{1}{2}\right)\right) = \begin{bmatrix} 6 & 3 \\ 3 & 3 \end{bmatrix}$$

Ensimmäinen johtava pääminori on $M_1 = 6 > 0$, ja toinen johtava pääminori on $M_2 = \det(H) = 18 - 9 = 9 > 0$. Johtavat pääminorit ovat positiivisia, ja Hessen matriisi tässä pisteessä on myös positiividefiniitti. Täten myös tämä piste on **lokaali minimi** (kohta 6).

- Kerätään saamamme tulokset:

  Piste $(0,0)$: **Ei minimi eikä maksimi**, Hessen matriisi ei neg. eikä pos. definiitti.

  Piste $\left(-\frac{1}{2},-\frac{1}{2}\right)$: **lokaali minimi**, Hessen matriisi positiividefiniitti.

  Piste $\left(\frac{1}{2},-\frac{1}{2}\right)$: **lokaali minimi**, Hessen matriisi positiividefiniitti.