Lineaarialgebra ja matriisilaskenta III

Pekka Pankka

29. huhtikuuta 2020

Lukijalle

Näiden luentomuistiinpanojen tarkoituksena on antaa tarvittava teoreettinen pohja niille lineaarialgebran menetelmille, joita käytetään jatkuvasti erityisesti tilastotieteellisissä sovelluksissa. Näitä ovat *pienimmän neliösumman menetelmä*, *ositetut matriisit* ja *singulaariarvohajotelma*. Menetelmien taakse kätkeytyy lineaarialgebran edistyneen teorian kova ydin: *sisätuloavaruuksien*, *ominaisarvojen* ja *lineeaaristen operaattorien* teoria.

Monissa yleisesityksissä tästä aiheesta keskitytään joko teoriaan tai sovelluksiin. Tämä tarkoittaa sitä, että käsittely keskittyy joko vektoriavaruuksiin ja lineaarikuvauksiin tai sarakkeisiin ja matriiseihin. Tästä jaosta seuraa kaksi ongelmaa: joko sovelluksissa tarvittavat muotoilut ja korollaarit jäävät pois tai teorian taustalla olevaa geometriaa ei käsitellä.

Näissä luentomuistiinpanoissa olen yrittänyt kattaa molemmat lähestymistavat. Aiheita lähestytään ensin toiselta kannalta ja käsittelyn aikana vaihdetaan näkökulmaa. Käsittelytapa on valittu aiheen luonnollisen tulokulma mukaisesti. Esimerkiksi isometrioiden käsittely aloitetaan geometrisesi.

Luentomuistiinpanojen pyrkimys käsitellä aihetta kahdesta suunnasta tarkoittaa kuitenkin myös sitä, että tekstille on kertynyt tavallista enemmän pituutta. Luentomuistiinpanot sisältävät myös runsaasti materiaalia, joka ei ole kurssin kannalta keskeistä, vaan on tarkoitettu kiinnostuneelle lukijalle lisämateriaaliksi. Nämä aiheet on erotettu omaksi osakseen luentomuistiinpanojen liitteisiin.

Nämä luentomuistiinpanot eivät syntyneet tyhjästä. Ne perustuvat Jyrki Möttösen luentomuistiinpanoille [4] samalta kurssilta vuodelta 2019. Suuret kiitokset Jyrkille lähdemateriaalista! Tekstin muina lähteinä on käytetty David C. Layn kirjaa [3] ja Sheldon Axlerin loistavaa kirjaa *Linear algebra done right* [1], jota voi kutsua moderniksi klassikoksi lineaarialgebran esittämisessä. Lukijalta oletetaan *Lineaarialgebra ja matriisilaskenta I & II* kursseja vastaavat tiedot, kuten ne on esitetty Häsän, Oinosen ja Rämön luentomonisteissa [2] ja [5].

Helsingissä
Pekka Pankka

Kirjallisuutta

[1] S. Axler. *Linear algebra done right.* Undergraduate Texts in Mathematics. Springer, Cham, third edition, 2015.

[2] J. Häsä, L. Oinonen, and J. Rämö. Johdatus lineaarialgebraan – osa I. 2017.

[3] D. C. Lay. *Linear Algebra and its applications.* Addison-Wesley, third edition, 2006.

[4] J. Möttönen. Lineaari algebra ja matriisilaskenta III. 2019.

[5] L. Oinonen and J. Rämö. Johdatus lineaarialgebraan – osa II. 2016.

Sisältö

Luku 1

Kertausta ja täydennystä

Näissä luentomuistiinpanoissa oletamme tunnetuiksi Lineaarialgebra ja matriisilaskenta I ja II kurssin käsittelemät asiat. Erityisesti oletetaan, että *(äärellisulotteisen) vektoriavaruuden*, *linaarikuvauksen*, *kannan*, *dimension*, *sarakevektorin* ja *matriisin* käsitteet ovat lukijan hallussa.

Merkintöjä

Lineaarikuvauksen $L\colon V \to W$ *kuva* $\operatorname{im} L$ on aliavaruus

$$\operatorname{im} L = \{L(v) \in W : v \in V\} \subset W$$

ja *ydin* $\ker L$ on aliavaruus

$$\ker L = \{v \in V : L(v) = 0\} \subset V.$$

Lineaarikuvausta $\mathrm{id}\colon V \to V$, $v \mapsto v$, kutsutaan *identtiseksi kuvaukseksi.*

1.1 Matriisit ja lineaarikuvaukset

Tässä luvussa kerrataan joitakin asioita merkintöjen kiinnittämiseksi ja joidenkin uusien käsitteiden esittelemiseksi. Luvun tärkein (uusi) käsite on *sarakeavaruus* $\mathbb{R}^{n\times 1}$, jota käytetään jatkossa avaruuden $\mathbb{R}^n$ sijaan matriisien yhteydessä. Muutos on pitkälti formaali, joten aloitetaan motivoinnilla.

Motivointi

Aloitetaan tutusta euklidisesta avaruudesta $\mathbb{R}^n$. Yleensä ajatellaan, että avaruuden $\mathbb{R}^n$ alkiot ovat reaalilukujen jonoja $(x_1, \ldots, x_n)$, missä siis $x_i \in \mathbb{R}$ jokaisella $i \in \{1, \ldots, n\}$. Jono $(x_1, \ldots, x_n)$ taas puolestaan on formaalimmin kuvaus $x\colon \{1, \ldots, n\} \to \mathbb{R}$, missä $x(i) = x_i$ jokaisella $i \in \{1, \ldots, n\}$.

Vektoriavaruuden $\mathbb{R}^n$ alkiot ovat siis itseasiassa kuvauksia ja niiden yhteenlasku on määritelty kuvausten yhteenlaskun avavulla seuraavasti. Olkoot $x\colon \{1, \ldots, n\} \to \mathbb{R}$ ja

$y\colon \{1,\dots,n\} \to \mathbb{R}$ funktiota ja olkoon $a \in \mathbb{R}$. Tällöin funktioiden yhteenlaskun nojalla $x + ay\colon \{1,\dots,n\} \to \mathbb{R}$ on funktio, joka on määritelty kaavalla

$$(x + ay)(i) = x(i) + ay(i)$$

kaikilla $i \in \{1,\dots,n\}$. Jos kirjoitetaan $x_i = x(i)$ ja $y_i = y(i)$ jokaisella $i \in \{1,\dots,n\}$ ja merkitään funktioita x ja y lyhyesti $(x_1,\dots,x_n)$ ja $(y_1,\dots,y_n)$, niin tämä yhteenlasku vastaa täsmälleen tapausta

$$(x_1,\dots,x_n) + a(y_1,\dots,y_n) = (x_1 + ay_1,\dots,x_n + ay_n).$$

1.1.1 Matriisit ja sarakkeet formaalisti

Monissa lähteissä sanotaan, että $(m \times n)$-matriisi on reaalilukujen taulukko

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}$$

eikä käsitteelle “taulukko” anneta tulkintaa. Yleensä merkitään $A = [a_{ji}] \in \mathbb{R}^{m\times n}$ tai $A = (a_{ji}) \in \mathbb{R}^{m\times n}$. Jatkossa voidaan myös merkitä A_{ji} tarkoittaen alkiota a_{ji}. Tämä on hyödyllinen merkintä erityisesti matriisitulon yhteydessä.

Kuten avaruuden $\mathbb{R}^n$ vektoreiden tapauksessa, $(m \times n)$-matriisi A on itseasiassa funktio $A\colon \{1,\dots,m\} \times \{1,\dots,n\} \to \mathbb{R}$, $(j,i) \mapsto a_{ji}$. Tällöin $A(j,i) = a_{ji}$ kaikilla $j \in \{1,\dots,m\}$ ja $i \in \{1,\dots,n\}$.

Näin määriteltäessä matriisien yhteenlasku ja reaalisella vakiolla voidaan määritellä suoraan funktioiden vastaavien laskutoimitusten avulla. Tämä tarkoittaa seuraavaa. Olkoot $A\colon \{1,\dots,m\} \times \{1,\dots,n\} \to \mathbb{R}$ ja $B\colon \{1,\dots,m\} \times \{1,\dots,n\} \to \mathbb{R}$ $(m \times n)$-matriiseja (tässä mielessä) ja $c \in \mathbb{R}$. Tällöin funktio $A + cB\colon \{1,\dots,m\} \times \{1,\dots,n\} \to \mathbb{R}$ on $(m \times n)$-matriisi, joka on määritelty kaavalla

$$(A + cB)(j,i) = A(j,i) + cB(j,i)$$

kaikilla $j \in \{1,\dots,m\}$ ja $i \in \{1,\dots,n\}$. Yleensä tämä laskutoimitus kirjoitetaan lyhyemmin

$$\begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} + c \begin{bmatrix} b_{11} & \cdots & b_{1n} \\ \vdots & & \vdots \\ b_{m1} & \cdots & b_{mn} \end{bmatrix} = \begin{bmatrix} a_{11} + cb_{11} & \cdots & a_{1n} + cb_{1n} \\ \vdots & & \vdots \\ a_{m1} + cb_{m1} & \cdots & a_{mn} + cb_{mn} \end{bmatrix}.$$

Jatkossa ei matriiseja tulkita funktioiksi vaan käytetään tuttua taulukkoesitystä. Funktiotulkinta on kuitenkin hyödyllinen sarakeavaruuden $\mathbb{R}^{m\times 1}$ määritelmässä.

Reaalisten $(m \times n)$-matriisien avaruutta merkitään $\mathbb{R}^{m\times n}$, eli

$$\mathbb{R}^{m\times n} = \{A\colon A \text{ on } (m \times n)\text{-matriisi}\}.$$

Erikoistapauksessa $n = 1$ avaruutta $\mathbb{R}^{m\times 1}$ kutsutaan *sarakeavaruudeksi*. Sarakeavaruuden $\mathbb{R}^{m\times 1}$ alkiot ovat siis matriiseja

$$\begin{bmatrix} a_{11} \\ \vdots \\ a_{m1} \end{bmatrix}.$$

Koska matriisien yhteenlasku ja skalaarilla kertominen tekee jokaisesta matriisiavaruudesta $\mathbb{R}^{m\times n}$ vektoriavaruuden, niin myös $\mathbb{R}^{m\times 1}$ on vektoriavaruus. Tämä avaruus on itseasiassa isomorfinen avaruuden $\mathbb{R}^m$. Halutun isomorfismin antaa

$$\Phi\colon \mathbb{R}^m \to \mathbb{R}^{m\times 1}, (x_1, \ldots, x_m) \mapsto \begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix}.$$

Tämä isomorfismi on niin luonnollinen, että yleensä avaruus $\mathbb{R}^m$ ja sarakeavaruus $\mathbb{R}^{m\times 1}$ samaistetaan ilman mitään kommenttia.

Matriisitulo on syy, miksi sarakeavaruudet $\mathbb{R}^{m\times 1}$ kannattaa erottaa euklidisista avaruuksista $\mathbb{R}^m$. Palauteutaan mieleen, että mikäli $A \in \mathbb{R}^{m\times n}$ ja $B \in \mathbb{R}^{n\times k}$, niin $AB \in \mathbb{R}^{m\times k}$ on matriisi, jonka alkiot $(AB)_{ji}$ määräytyvät kaavasta

$$(AB)_{ji} = \sum_{\ell=1}^{n} A_{j\ell} B_{\ell i}.$$

Näin ollen matriisin $A \in \mathbb{R}^{m\times n}$ tulo sarakkeen $x \in \mathbb{R}^{n\times 1}$ kanssa on määritelty ja tulo Ax on sarakevektori avaruudessa $\mathbb{R}^{m\times 1}$. Kaava $x \mapsto Ax$ määrittelee siis kuvauksen $f_A\colon \mathbb{R}^{n\times 1} \to \mathbb{R}^{m\times 1}$. Koska

$$A(x + cy) = Ax + cAy$$

kaikilla $x, y \in \mathbb{R}^{n\times 1}$ ja $c \in \mathbb{R}$, niin kuvaus f_A on itseasiassa lineaarikuvaus.

Huomautus 1.1.1. *Jatkossa kuvaus f_A tarkoittaa aina matriisin $A \in \mathbb{R}^{m\times n}$ määräämää lineaarikuvausta $f_A\colon \mathbb{R}^{n\times 1} \to \mathbb{R}^{m\times 1}$, $x \mapsto Ax$. Tätä kuvausta voi kutsua (vaikka sitä emme tee) matriisin A* kanoniseksi lineaarikuvauseksi*.*

Huomautus 1.1.2. *Matriisi $A = [a_{ji}] \in \mathbb{R}^{m\times n}$ määrittelee tietysti myös lineaarikuvauksen $f\colon \mathbb{R}^n \to \mathbb{R}^m$ kaavalla*

$$f(x_1, \ldots, x_n) = \left(\sum_{i=1}^{n} a_{1i} x_i, \ldots, \sum_{i=1}^{n} a_{mi} x_i \right).$$

Formaalisti f ja f_A vastaavat toisiaan kaavalla $f = \Phi^{-1} \circ f_A \circ \Phi$, joten yleensä vaan kirjoitetaan, että

$$f(x_1, \ldots, x_n) = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

ja matriisitulon antama avaruuden $\mathbb{R}^{m\times 1}$ sarakevektori samaistetaan avaruuden $\mathbb{R}^m$ vektorin kanssa ilman kommenttia.

Todetaan vielä tämän luvun lopuksi, että sarakkeet $v_1, \dots, v_n \in \mathbb{R}^{m\times 1}$ määrittelevät luonnollisella tavalla $(m \times n)$-matriisin kaavalla

$$A = \begin{bmatrix} v_{11} & \cdots & v_{1n} \\ \vdots & & \vdots \\ v_{m1} & \cdots & v_{mn} \end{bmatrix}$$

missä

$$v_i = \begin{bmatrix} v_{1i} \\ \vdots \\ v_{mi} \end{bmatrix}$$

jokaisella $i \in \{1, \dots, n\}$. Matriisia A merkitään lyhyesti

$$A = [v_1|\cdots|v_n]$$

tai

$$A = [v_1 \cdots v_n]\,.$$

Tämä merkintä on hyödyllinen seuraavassa mielessä. Olkoon

$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^{n\times 1}.$$

Tällöin matriisitulon ja sarakeavaruuden yhteenlaskun nojalla pätee

$$Ax = [v_1 \cdots v_n] \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = x_1 v_1 + \cdots + x_n v_n.$$

(Harjoitustehtävä.)

1.1.2 Neliömatriisit

Palautetaan myös mieleen, että avaruuden $\mathbb{R}^{n\times n}$ matriiseja kutsutaan *neliömatriiseiksi*. Neliömatriisia $I = [I_{ji}] \in \mathbb{R}^{n\times n}$, jolle pätee

$$I_{ji} = \begin{cases} 1, & j = i \\ 0, & j \neq i, \end{cases}$$

kutsutaan *identtiseksi matriisiksi*. Jos halutaan korostaa, että identtinen matriisi on $(n \times n)$-matriisi, merkitään $I_{n\times n}$.

Huomautus 1.1.3. *Identtisen matriisin I kertoimet I_{ji} vastaavat niin sanotun* Kroneckerin deltan *arvoja. Yleisesti, jos S on joukko, niin* Kroneckerin delta *on funktio $\delta\colon S \times S \to \{0,1\}$, joka määritellään kaavalla*

$$\delta(s,t) = \begin{cases} 1, & s = t \\ 0, & s \neq 0 \end{cases},$$

kaikilla $s, t \in S$, ja jonka yhteydessä käytetään merkintää $\delta_{st} = \delta(s,t)$ kaikilla $s, t \in S$.

Indenttisen matriisin $I_{n\times n}$ sarakkeet $e_1, \ldots, e_n$ muodostavat avaruuden $\mathbb{R}^{n\times 1}$ kannan, niin kutsutun *standardikannan.* Huomaa, että e_i on siis vektori

$$e_i = \begin{bmatrix} e_{1i} \\ \vdots \\ e_{ni} \end{bmatrix},$$

missä $e_{ji} = I_{ji}$ kaikilla $j \in \{1, \ldots, n\}$ ja $i \in \{1, \ldots n\}$, eli

$$e_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, e_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \ldots, e_n = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}.$$

Identtinen matriisi on erikoistapaus diagonaalimatriisista.

Määritelmä 1.1.4. *Matriisi $D \in \mathbb{R}^{n\times n}$ on* diagonaalimatriisi*, jos $D_{ji} = 0$ kaikilla $j \neq i$.*

Esimerkki 1.1.5. *Muita esimerkkejä diagonaalimatriiseista ovat*

$$\begin{bmatrix} -2 & 0 \\ 0 & 3 \end{bmatrix} \quad \text{ja} \quad \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}.$$

Neliomatriisi $A \in \mathbb{R}^{n\times n}$ on *kääntyvä* (tai *säännöllinen*), jos on olemassa sellainen matriisi $B \in \mathbb{R}^{n\times n}$, että

$$AB = BA = I_{n\times n}.$$

Tällöin merkitään $A^{-1} = B$. Palautetaan myös mieleen tulos, että *neliömatriisi A on kääntyvä, jos ja vain jos sen determinantti* $\det(A)$ *ei ole nolla.*

Huomautus 1.1.6. *Huomaa, että determinantti ja kääntyvyys on määritelty ainoastaan neliömatriiseille.*

1.1.3 Kolmiomatriisit

Mikäli halutaan ratkaista useita yhtälöryhmiä

$$Ax_i = y_i$$

vektoreille $y_1, y_2, \ldots$ on mielekästä yrittää palauttaa ongelma yhtälöryhmiin

$$Tz_i = y_i,$$

missä T on ylä- tai alakolmiomatriisi.

Määritelmä 1.1.7. *Matriisi $U = [u_{ji}] \in \mathbb{R}^{m\times n}$ on* yläkolmiomatriisi*, jos $u_{ji} = 0$ kaikilla $j > i$. Matriisi $L = [l_{ji}] \in \mathbb{R}^{m\times n}$ on* alakolmiomatriisi*, jos $l_{ji} = 0$ kaikilla $j < i$.*

Syy miksi yhtälöryhmän ratkaisu halutaan palauttaa näihin tapauksiin käsitellään tarkemmin Choleskyn hajotelman yhteydessä luvussa 8.5.

Tehdään kaksi havaintoa. Ensimmäinen havainto on, että neliömatriisi $A \in \mathbb{R}^{n\times n}$ on sekä ylä- että alakolmiomatriisi, jos ja vain jos A on diagonaalimatriisi. Toinen havainto on, että yläkolmiomatriisien (tai vastaavasti alakolmiomatriisien) tulo on yläkolmiomatriisi (vastaavasti alakolmiomatriisi). Havainto seuraa suoraan matriisitulon määritelmästä ja jätetään lukijalle. Huomautettakoon kuitenkin, että tästä seuraa, että kolmiomatriisin käänteismatriisi on myös samaa tyyppiä oleva kolmiomatriisi. Tarkka väite jätetään lukijan mietittäväksi.

1.1.4 Vektoriavaruuden ja sarakeavaruuden samastaminen

Äärellisulotteiset vektoriavaruudet, vektoriavaruuden kanta, virittäminen ja vapaus on määritelty aiemmilla kursseilla ja niitä ei tässä kerrata. Palautetaan kuitenkin mieleen, miten äärellisulotteinen vektoriavaruus V voidaan samaistaa sarakeavaruuden kanssa valitsemalla avaruudelle V kanta.

Olkoon V äärellisulotteinen vektoriavaruus ja $(v_1, \dots, v_n)$ sen kanta. Tällöin kuvaus

$$\Phi_V \colon \mathbb{R}^{n\times 1} \to V, \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \mapsto x_1 v_1 + \cdots + x_n v_n,$$

on lineaarinen isomorfismi. Tämä on helpointa osoittaa havaitsemalla, että Φ_V on lineaarikuvaus, joka kuvaa avaruuden $\mathbb{R}^{n\times 1}$ standardikannan $(e_1, \dots, e_n)$ avaruuden V kannaksi $(v_1, \dots, v_n)$, eli $\Phi_V(e_j) = v_j$ jokaisella $j = 1, \dots, n$. Näin ollen Φ_V on sekä injektiivinen että surjektiivinen ja siten isomorfismi.

1.1.5 Lineaarikuvauksen ja sen esitysmatriisin samastaminen

Tarkastellaan nyt kuinka lineaariaarikuvauksen $L \colon V \to W$ esitysmatriisi $A \in \mathbb{R}^{m\times n}$ määräytyy, kun avaruudet V ja W samastetaan sarakeavaruuksien kanssa. Palautetaan ensin mieleen kuinka sarakeavaruuksien väliset lineaarikuvaukset vastaavat (luonnollisella tavalla) matriisilla kertomista.[1] Tällä tarkoitetaan seuraavaa.

Palautetaan käsittelyä varten mieleen lineaarikuvaustan kantalause.

Lause 1.1.8 (Kantalause)**.** *Olkoot V ja W äärellisulotteisia vektoriavaruuksia, olkoon $(v_1, \dots, v_n)$ avaruuden V kanta ja olkoot $w_1, \dots, w_n \in W$ vektoreita. Tällöin on olemassa täsmälleen yksi sellainen lineaarikuvaus $f \colon V \to W$, että $f(v_i) = w_i$ kaikilla $i \in \{1, \dots, n\}$.*

Kantalause kertoo, että lähtöavaruuden kannan kuvat määräävät lineaarikuvauksen yksikäsitteisesti. Tämän vuoksi avaruuksien V ja W välinen lineaarikuvaus $f \colon V \to W$ voidaan määritellä lyhyesti kaavalla $v_i \mapsto w_i$, mikäli $(v_1, \dots, v_n)$ on avaruuden V kanta ja $w_1, \dots, w_n$ ovat avaruuden W vektoreita.

[1] Tässä termi "luonnollisella tavalla" tarkoittaa, että vastaavuuden antava isomorfismi määritellään ilman kannan valintaa.

Sarakeavaruuksien väliset lineaarikuvaukset

Olkoon $A \in \mathbb{R}^{m\times n}$ matriisi. Kuten edellä on käsitelty matriisitulon avulla määritelty kuvaus

$$f_A\colon \mathbb{R}^{n\times 1} \to \mathbb{R}^{m\times 1}, x \mapsto Ax,$$

on lineaarikuvaus.

Olkoon nyt $L\colon \mathbb{R}^{n\times 1} \to \mathbb{R}^{m\times 1}$ lineaarikuvaus. Nyt standarikannan $e_1, \ldots, e_n$ kuvat $L(e_1), \ldots, L(e_n)$ kuvauksessa L ovat sarakevektoreita avaruudessa $\mathbb{R}^{m\times 1}$. Olkoon nyt $A \in \mathbb{R}^{m\times n}$ se matriisi, jonka sarakkeet ovat $L(e_1), \ldots, L(e_n)$, eli

$$A = \Big[a_1|\cdots|a_n\Big] = \Big[L(e_1)|\cdots|L(e_n)\Big].$$

Nyt standardikannan ja matriisitulon määritelmien perusteella

$$L(e_i) = a_i = Ae_i$$

jokaisella $i = 1, \ldots, n$. Näin ollen kantalauseen perusteella pätee $L(x) = Ax$ kaikilla $x \in \mathbb{R}^{n\times 1}$, eli

$$L = f_A.$$

Tämän vuoksi matriisia A kutsutaan *lineaarikuvauksen L kanoniseksi esitysmatriisiksi* ja sarakeavaruuksien väliset lineaarikuvaukset vastaavat täsmälleen matriisilla kertomista.

Mikäli avaruuksille $\mathbb{R}^{n\times 1}$ ja $\mathbb{R}^{m\times 1}$ valitaan jotkin muut kannat kuin standardikannat, niin lineaarikuvauksen L matriisi muuttuu vastaavasti. Tätä käsitellään seuraavaksi.

Äärellisulotteisten vektoriavaruuksien väliset lineaarikuvaukset

Olkoot V ja W äärellisulotteisia vektoriavaruuksia ja olkoon $L\colon V \to W$ niiden välinen lineaarikuvaus. Kuvauksella L ei ole yksikäsitteistä esitysmatriisia vaan matriisi määräytyy kannan valinnan mukaan.

Olkoot lisäksi $(v_1, \ldots, v_n)$ avaruuden V kanta ja $(w_1, \ldots, w_m)$ avaruuden W kanta. Määritellään kuvaus $g_L\colon \mathbb{R}^{n\times 1} \to \mathbb{R}^{m\times 1}$ kaavalla

$$g_L = \Phi_W^{-1} \circ L \circ \Phi_V,$$

missä isomorfismit Φ_V ja Φ_W ovat kuin luvussa 1.1.4, eli $\Phi_V\colon \mathbb{R}^{n\times 1} \to V$ on isomorfismi $e_i \mapsto v_i$ ja $\Phi_W\colon \mathbb{R}^{m\times 1} \to W$ on isomorfismi $e_i \mapsto w_i$. Tässä on käytetty kantalauseen antamaa mahdollisuutta määritellä Φ_V ja Φ_W kanta-alkioiden avulla.

Kuvaus g_L on sarakeavaruuksien välinen lineaarikuvaus, koska se on yhdiste lineearikuvauksista. Näin ollen sillä on esitysmatriisi $A \in \mathbb{R}^{m\times n}$, joka toteuttaa ehdon

$$g_L(x) = Ax$$

kaikilla $x \in \mathbb{R}^{n\times 1}$. Erityisesti siis pätee

$$(\Phi_W^{-1} \circ L \circ \Phi_V)(x) = Ax \tag{1.1}$$

kaikilla $x \in \mathbb{R}^{n\times 1}$. Tämän vuoksi sanotaan, että A on lineaarikuvauksen $L\colon V \to W$ esitysmatriisina kantojen $(v_1, \ldots, v_n)$ ja $(w_1, \ldots, w_n)$ suhteen.

Huomautus 1.1.9. *Yhtälö* (1.1) *antaa hieman implisiittisen tavan tarkastella kuvauksen L matriisia. Lasketaan tämän vuoksi vielä auki, kuinka matriisin A kertoimet tulevat esiin vektoreiden koordinaattiesityksissä. Merkitään*

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}.$$

Olkoot nyt $v = x_1v_1 + \cdots + x_nv_n \in V$ ja $x \in \mathbb{R}^{n\times 1}$ sarakevektori, jolle pätee $\Phi_V(x) = v$, eli

$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}.$$

Tällöin

$$\begin{aligned} L(x_1v_1 + \cdots + x_nv_n) &= L(v) = L(\Phi_V(x)) = \Phi_W(g_L(x)) = \Phi_W(Ax) \\ &= \Phi_W\left(\begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}\right) \\ &= \Phi_W\left(\begin{bmatrix} a_{11}x_1 + \cdots + a_{1n}x_n \\ \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n \end{bmatrix}\right) \\ &= (a_{11}x_1 + \cdots + a_{1n}x_n)\, w_1 + \cdots + (a_{m1}x_1 + \cdots + a_{mn}x_n)\, w_m \end{aligned}$$

Näin ollen

$$L(v_j) = L(\Phi_V(e_j)) = \Phi_W(A_{g_L}e_j) = a_{1j}w_1 + \cdots + a_{mj}w_m$$

jokaisella $j = 1, \ldots, n$. Tämä antaa toisen tulkinnan kuinka matriisi A esittää lineaarikuvausta $L\colon V \to W$.

Kannanvaihtomatriisi

Tarkasteellaan vielä kuinka kannanvaihtomatriisi syntyy tässä formalismissa, eli miten lineaarikuvausta esittävät eri matriisit suhtautuvat toisiinsa.

Olkoon V äärellisulottinen vektoriavaruus sekä olkoot $(v_1, \ldots, v_n)$ ja $(v'_1, \ldots, v'_n)$ avaruuden V kantoja. Merkitään $\Phi_V\colon \mathbb{R}^{n\times 1} \to V$ isomorfismia, jolle pätee $e_i \mapsto v_i$ kaikilla $i = 1, \ldots, n$, ja merkitään $\Phi'_V\colon \mathbb{R}^{n\times 1} \to V$ isomorfismia, jolle pätee $e_i \mapsto v'_i$ kaikilla $i = 1, \ldots, n$.

Huomautus 1.1.10. *Huomaa, että*

$$(\Phi'_V \circ \Phi_V^{-1})(v_i) = \Phi'_V(\Phi_V^{-1}(v_i)) = \Phi'_V(e_i) = v'_i$$

jokaisella $i \in \{1, \ldots, n\}$, *eli että lineaarikuvaus*

$$\Phi'_V \circ \Phi_V^{-1} \colon V \to V$$

vie avaruuden V *kannan* $(v_1, \ldots, v_n)$ *kannaksi* $(v'_1, \ldots, v'_n)$.

Lineaarikuvaus $(\Phi'_V)^{-1} \circ \Phi_V \colon \mathbb{R}^{n\times 1} \to \mathbb{R}^{n\times 1}$ on avaruuden $\mathbb{R}^{n\times 1}$ isomorfismi. Näin ollen on olemassa matriisi $P \in \mathbb{R}^{n\times n}$ jolle pätee

$$\left((\Phi'_V)^{-1} \circ \Phi_V\right)(x) = Px$$

kaikilla $x \in \mathbb{R}^{n\times 1}$. Koska kuvaus $(\Phi'_V)^{-1} \circ \Phi_V$ on isomorfismi (eli sillä on käänteiskuvaus), niin myös matriisi P on kääntyvä (eli sillä on käänteismatriisi $P^{-1} \in \mathbb{R}^{n\times n}$).

Olkoon nyt W (toinen) äärellisulotteinen vektoriavaruus. Olkoot $(w_1, \ldots, w_m)$ ja $(w'_1, \ldots, w'_m)$ avaruuden W kantoja ja olkoot lisäksi $\Phi_W \colon \mathbb{R}^{m\times 1} \to W$ ja $\Phi'_W \colon \mathbb{R}^{m\times 1} \to W$ niitä vastaavat isomorfismit kantojen $(w_1, \ldots, w_m)$ ja $(w'_1, \ldots, w'_n)$ suhteen kuten yllä. Olkoon nyt $Q \in \mathbb{R}^{m\times m}$ isomorfismin

$$(\Phi'_W)^{-1} \circ \Phi_W \colon \mathbb{R}^{m\times 1} \to \mathbb{R}^{m\times 1}$$

matriisi. Myös Q on kääntyvä.

Osoitetaan nyt, että matriisit P ja Q ovat haluamamme kannanvaihtomatriisit. Tarkastellaan tämän vuoksi lineaarikuvausta $L \colon V \to W$. Kuvauksella L on esitysmatriisi $A \in \mathbb{R}^{m\times n}$ kannoissa $(v_1, \ldots, v_n)$ ja $(w_1, \ldots, w_n)$ sekä esitysmatriisi $A' \in \mathbb{R}^{m\times n}$ kannoissa $(v'_1, \ldots, v'_n)$ ja $(w'_1, \ldots, w'_m)$. Käyttäen yllä esitettyjä merkintöjä saadaan, että matriisit A ja A' toteuttavat kaikilla $x \in \mathbb{R}^{n\times 1}$ yhtälöt

$$(\Phi_W^{-1} \circ L \circ \Phi_V)(x) = Ax$$

ja

$$((\Phi'_W)^{-1} \circ L \circ \Phi'_V)(x) = A'x.$$

Koska $\Phi_W \circ \Phi_W^{-1} = \mathrm{id}_W$ ja $\Phi_V \circ \Phi_V^{-1} = \mathrm{id}_V$, niin kaikilla $x \in \mathbb{R}^{n\times 1}$ pätee

$$\begin{aligned}
A'x &= ((\Phi'_W)^{-1} \circ L \circ \Phi'_V)(x) \\
&= ((\Phi'_W)^{-1} \circ \mathrm{id}_W \circ L \circ \mathrm{id}_V \circ \Phi'_V)(x) \\
&= ((\Phi'_W)^{-1} \circ \Phi_W \circ \Phi_W^{-1} \circ L \circ \Phi_V \circ \Phi_V^{-1} \circ \Phi'_V)(x) \\
&= \left(((\Phi'_W)^{-1} \circ \Phi_W) \circ (\Phi_W^{-1} \circ L \circ \Phi_V) \circ (\Phi_V^{-1} \circ \Phi'_V)\right)(x) \\
&= \left(((\Phi'_W)^{-1} \circ \Phi_W) \circ (\Phi_W^{-1} \circ L \circ \Phi_V) \circ ((\Phi'_V)^{-1} \circ \Phi_V)^{-1})\right)(x) \\
&= (QAP^{-1})x.
\end{aligned}$$

Näin ollen

$$A' = QAP^{-1}.$$

Tämän vuoksi matriiseja Q ja P kutsutaan *kannanvaihtomatriiseiksi*.

1.2 Aliavaruuksien summa ja suora summa

Määritelmä 1.2.1. *Vektoriavaruuden V aliavaruuksien W_1 ja W_2* summa *on avaruuden V aliavaruus*

$$W_1 + W_2 = \{w_1 + w_2 \in V : w_1 \in W_1,\ w_2 \in W_2\}.$$

Mikäli $W_1 \cap W_2 = \{0\}$, niin aliavaruutta $V + V'$ kutsutaan suoraksi summaksi *ja merkitään $W_1 \oplus W_2$.*

Suoran summan määritelmä voidaan kirjoittaa myös toisella tavalla lineaarisen riippumattomuuden avulla.

Lemma 1.2.2. *Olkoon V vektoriavaruus ja olkoot W_1 ja W_2 aliavaruuksia. Tällöin summa $W_1 + W_2$ on suora, jos ja vain jos aliavaruuksien W_1 ja W_2 vektorit ovat riippumattomia toisistaan, eli että kaikilla nollasta poikkeavilla vektoreilla $w_1 \in W_1$ ja $w_2 \in W_2$ yhtälön $aw_1 + bw_2 = 0$ ainoa ratkaisu on $a = b = 0$.*

Todistus. Oletetaan, että summa $W_1 + W_2$ on suora. Olkoot $w_1 \in W_1$ ja $w_2 \in W_2$ nollasta poikkeavia vektoreita. Oletetaan, että $aw_1 + bw_2 = 0$. Tällöin $aw_1 = -bw_2$. Koska $aw_1 \in W_1$ ja $-bw_2 \in W_2$, niin $aw_1 \in W_1 \cap W_2$. Näin ollen $aw_1 = 0$, joten $a = 0$. Tällöin $-bw_2 = 0$, joten myös $b = 0$.

Oletetaan nyt, että kaikilla nollasta poikkeavilla vektoreilla $w_1 \in W_1$ ja $w_2 \in W_2$ yhtälön $aw_1 + bw_2 = 0$ ainoa ratkaisu on $a = b = 0$. Osoitetaan, että $W_1 \cap W_2 = \{0\}$. Tehdään vastaoletus, että on olemassa nollasta poikkeava $w \in W_1 \cap W_2$. Valitaan $w_1 = w_2 = w$. Tällöin $w_1 \in W_1$, $w_2 \in W_2$ ja $w_1 + (-1)w_2 = w - w = 0$. Tämä on ristiriita, joten $w = 0$. □

Korollaari 1.2.3. *Olkoot V vektoriavaruus, W_1 ja W_2 aliavaruuksia. Olkoon lisäksi $(w_1, \ldots, w_k)$ aliavaruuden W_1 kanta ja $(w'_1, \ldots, w'_\ell)$ aliavaruuden W_2 kanta. Jos aliavaruuksien W_1 ja W_2 summa on suora, niin $(v_1, \ldots, v_k, v'_1, \ldots, v'_\ell)$ on aliavaruuden $W_1 \oplus W_2$ kanta.*

Todistus. Selvästi $(v_1, \ldots, v_k, v'_1, \ldots, v'_\ell)$ virittää aliavaruuden $W_1 \oplus W_2$, sillä jokaisella $w \in W_1 \oplus W_2$ on olemassa sellaiset $w_1 \in W_1$ ja $w_2 \in W_2$, että $w_1 \in \mathrm{Sp}\{v_1, \ldots, v_k\}$ ja $w_2 \in \mathrm{Sp}\{v'_1, \ldots, v'_\ell\}$. Osoitetaan nyt, että $(v_1, \ldots, v_k, v'_1, \ldots, v'_\ell)$ on lineaarisesti riippumaton. Olkoot $a_1, \ldots, a_k, b_1, \ldots, b_\ell \in \mathbb{R}$ sellaisia, että

$$a_1 v_1 + \cdots + a_k v_k + b_1 v'_1 + \cdots + b_\ell v'_\ell = 0.$$

Olkooot $w_1 = a_1 v_1 + \cdots + a_k v_k$ ja $w_2 = b_1 v'_1 + \cdots + b_\ell v'_\ell$. Koska $w_1 = -w_2$, niin ehdosta $w_1 \neq 0$ seuraa $w_2 \neq 0$ ja $W_1 \cap W_2 \neq \{0\}$. Näin ollen $w_1 = 0$ ja $w_2 = 0$. Koska $(v_1, \ldots, v_k)$ ja $(v'_1, \ldots, v'_\ell)$ ovat kantoja, niin $a_1 = \cdots = a_k = b_1 = \cdots = b_\ell = 0$. □

Korollaarin seuraus on, että aliavaruuden $W_1 \oplus W_2$ dimensio on aliavaruuksien W_1 ja W_2 dimensioiden summa. Yleisemmin pätee seuraava tulos, jonka ainoastaan palautamme mieleen.

Korollaari 1.2.4. *Olkoon V vektoriavaruus ja olkoot W_1 ja W_2 äärellisulotteisia aliavaruuksia. Tällöin*

$$\dim(W_1 + W_2) + \dim(W_1 \cap W_2) = \dim W_1 + \dim W_2.$$

1.3 Sisätulo

Pythagoraasta sisätuloon

Tavallisissa euklidisissa avaruuksissa $\mathbb{R}^2$ ja $\mathbb{R}^3$ kohtisuoruus on euklidisesta geometriasta tuttu käsite, joka yleensä ymmärretään intuitiivisesti kuvan perusteella. Yleisemmissä vektoriavaruuksissa kohtisuoruuden ja kulman käsitteet eivät ole perusteltavissa kuvalla tai intuitiolla. Konseptuaalisesti oikea käsite kulman ja kohtisuoruuden ymmärätämiseen on sisätulo. Motivoidaan tämä kommentti seuraavalla esimerkillä.

Esimerkki 1.3.1. *Olkoot $v = (x_1, x_2)$ ja $w = (y_1, y_2)$ vektoreita vektoriavaruudessa $\mathbb{R}^2$. Oletetaan, että kumpikaan vektoreista ei ole nollavektori 0. Tarkastellaan kolmiota jonka kärkipisteet ovat v, w ja origo sekä vektoreiden v ja w välistä kulmaa $\theta \in [0, 2\pi[$ origossa. (Piirrä kuva!) Merkitään vektoreiden v, w ja $w - v$ pituuksia symboleilla a, b ja c. Tällöin euklidisen tasogeometrian kosinilause sanoo, että*

$$c^2 = a^2 + b^2 - 2ab\cos\theta.$$

(Huomaa, että jos v ja w ovat kohtisuorassa, niin $\theta = \pi/2$ ja $\cos\theta = 0$. Tällöin kosinilause palautuu Pythagoraan lauseeksi.)

Toisaalta Pythagoraan lauseen perusteella tiedetään, että vektoreiden v, w ja $v + w$ ovat

$$a = \sqrt{x_1^2 + x_2^2}, \quad b = \sqrt{y_1^2 + y_2^2}, \quad \text{ja} \quad c = \sqrt{(y_1 - x_1)^2 + (y_2 - x_2)^2}.$$

Näin ollen tiedetään, että

$$(y_1 - x_1)^2 + (y_2 - x_2)^2 = x_1^2 + y_1^2 + x_2^2 + y_2^2 - 2ab\cos\theta.$$

Sieventämisen jälkeen saadaan

$$-2x_1x_2 - 2y_1y_2 = -2ab\cos\theta$$

eli

$$\cos\theta = \frac{x_1x_2 + y_1y_2}{ab} = \frac{x_1x_2 + y_1y_2}{\sqrt{x_1^2 + y_1^2}\sqrt{x_2^2 + y_2^2}}.$$

Vektoreiden v ja w välinen kulma θ riippuu siis funktion $\cdot\colon \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$,

$$((x_1, y_1), (x_2, y_2)) \mapsto x_1x_2 + y_1y_2,$$

antaman arvon suhteesta vektoreiden v ja w pituuteen. Funktiota $\cdot\colon \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ kutsutaan pistetuloksi avaruudessa $\mathbb{R}^2$ *ja se on yksi esimerkki vektoriavaruuden sisätulosta.*

Määritelmä 1.3.2. *Funktio* $\langle\cdot,\cdot\rangle\colon V\times V\to\mathbb{R}$ *on* sisätulo vektoriavaruudessa V*, jos*

- *kaikilla* $a,b\in\mathbb{R}$ *ja* $v,v',w\in V$ *pätee*
$$\langle av+bv',w\rangle=a\langle v,w\rangle+b\langle v',w\rangle, \tag{ST1}$$
- *kaikilla* $v,w\in V$ *pätee*
$$\langle v,w\rangle=\langle w,v\rangle \tag{ST2}$$
ja
- *kaikilla* $v\in V$*,* $v\neq 0$*, pätee*
$$\langle v,v\rangle>0. \tag{ST3}$$

Paria $(V,\langle\cdot,\cdot\rangle)$*, jossa* V *on vektoriavaruus ja* $\langle\cdot,\cdot\rangle\colon V\times V\to\mathbb{R}$ *on sisätulo, kutsutaan* sisätuloavaruudeksi.

Esimerkki 1.3.3. *Pistetulo* $\cdot\colon\mathbb{R}^n\times\mathbb{R}^n\to\mathbb{R}$*,*
$$((x_1,\ldots,x_n),(y_1,\ldots,y_n))\mapsto x_1y_1+\cdots+x_ny_n,$$
on sisätulo avaruudessa $\mathbb{R}^n$*. Vastaavalla kaavalla voidaan määritellä pistetulo avaruuksiin* $\mathbb{R}^{n\times 1}$ *ja* $\mathbb{R}^{1\times n}$*. Itseasiassa funktio* $\cdot\colon\mathbb{R}^{m\times n}\times\mathbb{R}^{m\times n}\to\mathbb{R}$*,*
$$\left(\begin{bmatrix}a_{11}&\cdots&a_{1n}\\ \vdots&&\vdots\\ a_{m1}&\cdots&a_{mn}\end{bmatrix},\begin{bmatrix}b_{11}&\cdots&b_{1n}\\ \vdots&&\vdots\\ b_{m1}&\cdots&b_{mn}\end{bmatrix}\right)\mapsto\sum_{i=1}^{n}\sum_{j=1}^{m}a_{ij}b_{ij},$$
määrittelee sisätulon (jota kutsutaan pistetuloksi*) matriisiavaruuteen* $\mathbb{R}^{m\times n}$*.*

Sisätulon ehtoa (ST1) kutsutaan *lineaarisuudeksi*, ehtoa (ST2) kutsutaan *symmetrisyydeksi* ja ehtoa (ST3) *positiividefiniittisyydeksi*. Muutama huomatus on paikallaan.

Huomautus 1.3.4. *Sisätulon lineaarisuus kirjoitetaan usein yhtäpitävästi kahdessa osassa: kaikilla* $a\in\mathbb{R}$ *ja* $v,v',w\in V$ *pätee sekä*
$$\langle v+v',w\rangle=\langle v,w\rangle+\langle v',w\rangle$$
että
$$\langle av,w\rangle=a\langle v,w\rangle.$$
(Ehtojen yhtäpitävyys on harjoitustehtävä.)

Huomautus 1.3.5. *Sisätulon symmetrisyydestä yhdessä lineaarisuuden kanssa seuraa, että sisätulo on* bilineaarinen *eli lineaarinen molempien muuttujen suhteen: kaikilla* $a,b\in\mathbb{R}$ *ja* $v,w,w'\in V$ *seuraa*
$$\langle v,aw+bw'\rangle=a\langle v,w\rangle+b\langle v,w'\rangle.$$
(Perustelu jätetään harjoitustehtäväksi.)

Huomautus 1.3.6. *Sisätulon positiivideﬁniittisyydestä ja lineaarisuudesta seuraa, että vektorille $v \in V$ pätee $\langle v, v\rangle = 0$, jos ja vain jos $v = 0$. (Perustelu jätetään jälleen harjoitustehtäväksi.)*

Ennen siirtymistä eteenpäin, tehdään vielä helppo huomio, että sisätulon rajoittuma vektoriavaruuden aliavaruuteen on sisätulo. Näin sisätuloavaruuden aliavaruudella on luonnollinen sisätulo.

Lemma 1.3.7. *Olkoon $(V, \langle\cdot,\cdot\rangle)$ sisätuloavaruus ja $W \subset V$ aliavaruus. Tällöin sisätulon $\langle\cdot,\cdot\rangle$ rajoittuma $\langle\cdot,\cdot\rangle|_{W\times W} \colon W\times W \to \mathbb{R}$ on sisätulo avaruudessa W, eli $(W, \langle\cdot,\cdots\rangle|_{W\times W})$ on sisätuloavaruus.*

Todistus. Koska W on vektoriavaruus ja koska ehdot (ST1), (ST2) ja (ST3) pätevät kaikilla $v, v', w \in V$, niin ne pätevät myös kaikille $v, v', w \in W$, sillä $W \subset V$. $\square$

Vektorin pituus eli normi

Sisätuloavaruuden $(V, \langle\cdot,\cdot\rangle)$ vektorin $v \in V$ pituus määritellään kaavalla

$$\|v\| = \sqrt{\langle v, v\rangle}. \tag{1.2}$$

Tämä pituuden määritelmä perustuu Pythagoraan lauseeseen ja on yleistys pistetulon ja pituuden välisestä yhteydestä avaruudessa $\mathbb{R}^n$. Tämän yhteyden voi havaita seuraavasti.

Esimerkki 1.3.8. *Olkoon $\cdot \colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ (tavallinen) pistetulo*

$$((x_1, \dots, x_n), (y_1, \dots, y_n)) \to x_1y_1 + \cdots + x_ny_n.$$

Tällöin jokaisella $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ pätee

$$\|x\|^2 = x \cdot x = x_1x_1 + \cdots + x_nx_n = x_1^2 + \cdots + x_n^2$$

eli

$$\|x\| = \sqrt{x_1^2 + \cdots x_n^2}.$$

Erityisesti siis avaruuksien $\mathbb{R}^2$ ja $\mathbb{R}^3$ vektoreille $(x, y) \in \mathbb{R}^2$ ja $(x, y, z) \in \mathbb{R}^3$ pätee

$$\|(x, y)\| = \sqrt{x^2 + y^2} \quad \text{ja} \quad \|(x, y, z)\| = \sqrt{x^2 + y^2 + z^2}.$$

(Yhteys Pythagoraan lauseeseen seuraa piirtämällä kuva.)

Sisätulon $\langle\cdot,\cdot\rangle \colon V \times V \to \mathbb{R}$ määräämää funktiota $\|\cdot\| \colon V \to [0, \infty[$ kutsutaan *sisätulon induksoimaksi (tai määräämäksi) normiksi.* Tämä on erikoistapaus yleisemmästä normin käsitteestä.

Määritelmä 1.3.9. *Vektoriavaruudessa V määritelty funktio $\|\cdot\| \colon V \to [0, \infty[$ on* normi*, jos seuraavat ehdot toteutuvat:*

- *kaikilla* $v, w \in V$ *pätee*
$$\|v + w\| \leq \|v\| + \|w\|, \tag{N1}$$
- *kaikilla* $v \in V$ *ja* $a \in \mathbb{R}$ *pätee*
$$\|av\| = |a| \, \|v\| \tag{N2}$$
ja
- *kaikilla* $v \in V \setminus \{0\}$ *pätee*
$$\|v\| > 0. \tag{N3}$$

Paria $(V, \|\cdot\|)$*, missä* V *on vektoriavaruus ja* $\|\cdot\|$ *on normi, kutsutaan* normiavaruudeksi.

Huomautus 1.3.10. *Ehtoa* (N1) *kutsutaan* kolmioepäyhtälöksi*, koska ehdosta seuraa, että kolmen vektorin* $u, v, w \in V$ *määräämän kolmion sivut toteuttavat ehdon, että jokainen sivu on pituudeltaan korkeintaan kahden muun sivun pituuksien summa.*

Sisätulon määräämä normi on yleisen normin erikoistapaus siinä mielessä, että kolmioepäyhtälön lisäksi normi toteuttaa Pythagoraan lauseen: jos vektorit $v, w \in V$ *ovat kohtisuorassa toisiaan vastaan, niin*

$$\|v + w\|^2 = \langle v + w, v + w\rangle = \langle v, v\rangle + 2\langle v, w\rangle + \langle w, w\rangle = \|v\|^2 + \|w\|^2.$$

Huomautus 1.3.11. *Ehdoilla* (N2) *ja* (N3) *ei ole vakiintuneita nimiä. Ehto* (N3) *voidaan antaa useilla ekvivalenteilla tavoilla. Huomaa, että ehdot* (N2) *ja* (N3) *yhdessä antavat, että* $\|v\| = 0$*, jos ja vain jos* $v = 0$*. (Harjoitustehtävä)*

Osoitetaan, että sisätulon $\langle\cdot, \cdot\rangle \colon V \times V \to \mathbb{R}$ määräämä funktio $\|\cdot\| \colon V \to [0, \infty[$ todella on normi.

Lause 1.3.12. *Olkoon* $(V, \langle\cdot, \cdot\rangle)$ *sisätuloavaruus. Tällöin funktio* $\|\cdot\| \colon V \to [0, \infty[$*,* $v \mapsto \sqrt{\langle v, v\rangle}$*, on avaruuden* V *normi.*

Lauseen todistuksen vaativin kohta on kolmioepäyhtälön todistus. Se perustuu Schwarzin lemmaan.

Lemma 1.3.13 (Schwarzin lemma[2])**.** *Olkoon* $(V, \langle\cdot, \cdot\rangle)$ *sisätuloavaruus ja* $\|\cdot\| \colon V \to \mathbb{R}$ *funktio* $v \mapsto \sqrt{\langle v, v\rangle}$*. Tällöin kaikilla* $v, w \in V$ *pätee*

$$|\langle v, w\rangle| \leq \|v\|\|w\|.$$

Todistus. Voidaan olettaa, että $v \neq 0$ ja $w \neq 0$. Tarkastellaan suoraa

$$\ell = \{v + tw \in V \colon t \in \mathbb{R}\}$$

avaruudessa V ja etsitään vektori $v + t_0 w \in \ell$, joka on kohtisuorassa suoran ℓ suuntavektoria w vastaan, eli ratkaistaan yhtälö

$$\langle v + t_0 w, w\rangle = 0.$$

[2]Tunnetaan myös nimillä *Schwarzin epäyhtälö* ja *Cauchy–Schwarzin epäyhtälö*.

Koska
$$\langle v + t_0 w, w\rangle = \langle v, w\rangle + t_0\langle w, w\rangle = \langle v, w\rangle + t_0\|w\|^2,$$
niin saadaan
$$t_0 = -\frac{\langle v, w\rangle}{\|w\|^2}.$$
Huomaa, että $\|w\| > 0$, koska $w \neq 0$.

Tarkastellaan nyt pisteen $v + t_0 w \in \ell$ etäisyyttä origosta. Koska
$$\begin{aligned}
0 \leq \|v + t_0 w\|^2 &= \langle v + t_0 w, v + t_0 w\rangle = \|v\|^2 + 2t_0\langle v, w\rangle + t_0^2\|w\|^2 \\
&= \|v\|^2 + t_0\left(2\langle v, w\rangle + t_0\|w\|^2\right) = \|v\|^2 + t_0\left(2\langle v, w\rangle - \langle v, w\rangle\right) \\
&= \|v\|^2 + t_0\langle v, w\rangle = \|v\|^2 - \frac{\langle v, w\rangle}{\|w\|^2}\langle v, w\rangle = \|v\|^2 - \left(\frac{\langle v, w\rangle}{\|w\|}\right)^2,
\end{aligned}$$
niin
$$\left|\frac{\langle v, w\rangle}{\|w\|}\right| \leq \|v\|.$$
Näin ollen
$$|\langle v, w\rangle| \leq \|v\|\|w\|.$$
□

Huomautus 1.3.14. *Todistus osoittaa, että Schwarzin lemman taustalla on geometrinen havainto, että affiinin suoran (todistuksessa suora ℓ) etäisyys origosta saadaan etsimällä se suoran piste (tässä $v + t_0 w$), jonka suuntavektori on kohtisuorassa affiinin suoran suuntavektoria (tässä w) vastaan. Tätä havaintoa ei suoraan käytetty todistuksessa, mutta seuraa suoraan laskemalla pisteen $v + tw \in \ell$ etäisyys origosta neliöksi täydentämällä, eli*
$$\begin{aligned}
\|v + tw\|^2 &= \langle v + tw, v + tw\rangle = \|v\|^2 + 2t\langle v, w\rangle + t^2\|w\| \\
&= \|v\|^2 + 2t\|w\|\frac{\langle v, w\rangle}{\|w\|} + (t\|w\|)^2 \\
&= \|v\|^2 - \left(\frac{\langle v, w\rangle}{\|w\|}\right)^2 + \left(\frac{\langle v, w\rangle}{\|w\|} + t\|w\|\right)^2 \\
&\geq \|v\|^2 - \left(\frac{\langle v, w\rangle}{\|w\|}\right)^2 = \|v + t_0 w\|^2.
\end{aligned}$$
Näin ollen piste $v + t_0 w \in \ell$ antaa suoran ℓ etäisyyden origosta.

Huomautus 1.3.15. *Mikäli sekä v että w ovat nollasta poikkeavia vektoreita, Schwarzin epäyhtälö voidaan kirjoittaa muodossa*
$$\left|\left\langle \frac{v}{\|v\|}, \frac{w}{\|w\|}\right\rangle\right| \leq 1,$$

missä $v/\|v\|$ ja $w/\|w\|$ ovat yksikkövektoreita. Eristyisesti Schwarzin lemma siis sanoo, että yksikkövektoreiden sisätulo on itseisarvoltaan korkeintaan yksi.

Lisäksi vektoreiden v ja w välinen kulma $\angle(v,w) \in [0, 2\pi[$ voidaan määritellä kaavalla

$$\cos \angle(v,w) = \left\langle \frac{v}{\|v\|}, \frac{w}{\|w\|} \right\rangle$$

Schwarzin lemma siis osoittaa, että sisätulo on pistetulon luonnollinnen yleistys.

Todistetaan nyt lause 1.3.12, eli että sisätulon $\langle \cdot, \cdot \rangle$ määräämä funktio $\| \cdot \|$ toteuttaa normin ehdot.

Lauseen 1.3.12 todistus. Osoitetaan ensin kolmioepäyhtälö. Olkoot $v, w \in V$. Tällöin

$$\begin{aligned}
\|v+w\|^2 &= \langle v+w, v+w \rangle = \|v\|^2 + 2\langle v, w \rangle + \|w\|^2 \\
&\leq \|v\|^2 + 2|\langle v, w \rangle| + \|w\|^2 \\
&\leq \|v\|^2 + 2\|v\|\|w\| + \|w\|^2 \\
&= (\|v\| + \|w\|)^2 .
\end{aligned}$$

Näin ollen

$$\|v+w\| \leq \|v\| + \|w\|.$$

Kolmioepäyhtällö on näin osoitettu.

Normin ehto (N2) seuraa lähes välittömästi. Olkoot $v \in V$ ja $a \in \mathbb{R}$. Tällöin

$$\|av\| = \sqrt{\langle av, av \rangle} = \sqrt{a^2 \langle v, v \rangle} = |a|\|v\|.$$

Osoitetaan vielä normin ehto (N3). Olkoon $v \in V \setminus \{0\}$. Koska $\langle v, v \rangle > 0$, niin

$$\|v\| = \sqrt{\langle v, v \rangle} > 0.$$

□

1.3.1 Kohtisuoruus

Mahdollisesti tärkein sisätulon antama vektoreiden ominaisuus on kohtisuoruus.

Määritelmä 1.3.16. *Sisätuloavaruuden $(V, \langle \cdot, \cdot \rangle)$ vektorit v ja w ovat* kohtisuorassa (eli ortogonaalisia) toisiaan vastaan*, jos*

$$\langle v, w \rangle = 0.$$

Kirjataan kaksi tämän määritelmän helppoa, mutta konseptuaalisesti tärkeää, seurausta.

Lemma 1.3.17. *Olkoon $(V, \langle \cdot, \cdot \rangle)$ sisätuloavaruus. Jos vektori $v \in V$ on kohtisuorassa kaikkia avaruuden V vektoreita vastaan, niin $v = 0$.*

Todistus. Koska oletuksen nojalla $\langle v, v \rangle = 0$, niin $v = 0$ sisätulon määritelmän kohdan (ST3) nojalla. $\square$

Lemma 1.3.18. *Olkoon $(V, \langle \cdot, \cdot \rangle)$ äärellisulotteinen sisätuloavaruus ja $(v_1, \ldots, v_n)$ avaruuden V kanta. Jos vektori $v \in V$ on kohtisuorassa kaikki kantavektoreita $v_1, \ldots, v_n$ vastaan, niin $v = 0$.*

Todistus. Koska $(v_1, \ldots, v_n)$ on avaruuden V kanta, niin on olemassa yksikäsitteiset reaaliluvut $a_1, \ldots, a_n \in \mathbb{R}$, joille pätee $v = a_1 v_1 + \cdots + a_n v_n$. Näin ollen

$$\langle v, v \rangle = \langle v, a_1 v_1 + \cdots + a_n v_n \rangle = \sum_{i=1}^{n} a_i \langle v, v_i \rangle = 0$$

ja siten $v = 0$. $\square$

1.3.2 Ortonormaali kanta

Sisätuloavaruuden luonnollisimmat kannat ovat sellaisia, joiden alkiot ovat kohtisuorassa toisiaan vastaan olevia yksikkövektoreita.

Määritelmä 1.3.19. *Sisätuloavaruuden $(V, \langle \cdot, \cdot \rangle)$ kanta $(v_1, \ldots, v_n)$ on* ortonormaali*, jos*

- $\|v_i\| = 1$ *jokaisella* $i = 1, \ldots, n$ *ja*
- *vektorit* v_i *ja* v_j *ovat kohtisuorassa toisiaan vastaan kaikilla* $i, j \in \{1, \ldots, n\}$, $i \neq j$.

Huomautus 1.3.20. *Yhtäpitävästi sisätuloavaruuden $(V, \langle \cdot, \cdot \rangle)$ kanta $(v_1, \ldots, v_n)$ on ortonormaali, jos*

$$\langle v_i, v_j \rangle = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}$$

kaikilla $i, j \in \{1, \ldots, n\}$. Tässä yhteydessä käytetään usein merkintää $\langle v_i, v_j \rangle = \delta_{ij}$, missä

$$\delta_{ij} = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases} .$$

Lukua δ_{ij} kutsutaan Kroneckerin deltaksi.

Esimerkki 1.3.21. *Avaruuden $\mathbb{R}^n$ standardikanta $(e_1, \ldots, e_n)$ on ortonormaali pistetulon $\cdot \colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ suhteen.*

Ortonormaalilla kannalla on ominaisuus, että vektorin koordinaatit ja pituus määräytyvät sisätuloista vastaavia kannan alkiota vasten. Kirjataan tämä tärkeä huomio seuraavasti.

Lause 1.3.22. *Olkoon $(V, \langle \cdots, \cdots \rangle)$ äärellisulotteinen sisätuloavaruus ja $(v_1, \ldots, v_n)$ ortonormaalikanta. Tällöin jokaisella $v \in V$ pätee*

$$v = \langle v, v_1 \rangle v_1 + \cdots + \langle v, v_n \rangle v_n \tag{1.3}$$

ja

$$\|v\| = \sqrt{\langle v, v_1 \rangle^2 + \cdots + \langle v, v_n \rangle^2}. \tag{1.4}$$

Huomautus 1.3.23. *Esityskaava* (1.3) *sanoo olleellisesti, että vektorin koordinaatteja ortonormaalin kannan suhteen ei tarvitse ratkaista yhtälöryhmästä vaan, että ne voidaan laskea suoraan. Vektorin pituuden lauseke* (1.4) *on puolestaan jälleen yksi versio Pythagoraan lauseesta.*

Lauseen 1.3.22 todistus. Osoitetaan ensin (1.3) todistamalla, että erotusvektori

$$w = v - (\langle v, v_1 \rangle v_1 + \cdots + \langle v, v_n \rangle v_n)$$

on nollavektori. Lemman 1.3.18 perusteella riittää osoittaa, että w on kohtisuorassa jokaista kannan $(v_1, \ldots, v_n)$ vektoria vastaan. Olkoon $i \in \{1, \ldots, n\}$. Tällöin

$$\begin{aligned}
\langle w, v_i \rangle &= \langle v - (\langle v, v_1 \rangle v_1 + \cdots + \langle v, v_n \rangle v_n), v_i \rangle \\
&= \langle v, v_i \rangle - \langle (\langle v, v_1 \rangle v_1 + \cdots + \langle v, v_n \rangle v_n), v_i \rangle \\
&= \langle v, v_i \rangle - \sum_{j=1}^{n} \langle \langle v, v_j \rangle v_j, v_i \rangle = \langle v, v_i \rangle - \sum_{j=1}^{n} \langle v, v_j \rangle \langle v_j, v_i \rangle \\
&= \langle v, v_i \rangle - \sum_{j=1}^{n} \langle v, v_j \rangle \delta_{ij} = \langle v, v_i \rangle - \langle v, v_i \rangle = 0.
\end{aligned}$$

Näin ollen $w = 0$ ja väite seuraa.

Osoitetaan nyt (1.4). Olkoon $v \in V$. Tällöin esityksen (1.3) perusteella

$$\begin{aligned}
\|v\|^2 = \langle v, v \rangle &= \left\langle \sum_{i=1}^{n} \langle v, v_i \rangle v_i, \sum_{j=1}^{n} \langle v, v_j \rangle v_j \right\rangle \\
&= \sum_{i,j=1}^{n} \langle v, v_i \rangle \langle v, v_j \rangle \langle v_i, v_j \rangle = \sum_{i,j=1}^{n} \langle v, v_i \rangle \langle v, v_j \rangle \delta_{ij} \\
&= \sum_{i=1}^{n} \langle v, v_i \rangle \langle v, v_i \rangle = \sum_{i=1}^{n} \langle v, v_i \rangle^2.
\end{aligned}$$

□

Korollaari 1.3.24. *Olkoon* $(V, \langle \cdot, \cdot \rangle)$ *äärellisulotteinen sisätuloavaruus ja* $(v_1, \ldots, v_n)$ *ortonormaalikanta. Tällöin vektoria* $v \in V$ *vastaava sarakevektori* $x \in \mathbb{R}^{n \times 1}$ *on*

$$x = \begin{bmatrix} \langle v, v_1 \rangle \\ \vdots \\ \langle v, v_n \rangle \end{bmatrix}.$$

1.3.3 Ortonormaalin kannan olemassaolo

Lauseesta 1.3.22 seuraa, että äärellisulotteisella sisätuloavaruudella on aina ortonormaalikanta.

Lause 1.3.25 (Gram–Schmidt)**.** *Äärellisulotteisella sisätuloavaruudella* $(V, \langle\cdot,\cdot\rangle)$ *on ortonormaalikanta.*

Todistusta varten tarvitaan aputulos.

Lemma 1.3.26. *Olkoon* $(V, \langle\cdot,\cdot\rangle)$ *sisätuloavaruus,* $W \subset V$ *aliavaruus ja* $(w_1, \ldots, w_n)$ *aliavaruuden* W *ortonormaalikanta. Olkoon* $P_W : V \to W$ *lineaarikuvaus*

$$v \mapsto \langle v, w_1\rangle w_1 + \cdots + \langle v, w_n\rangle w_n.$$

Tällöin jokaisella $v \in V$ *vektorit* $P_W(v)$ *ja* $v - P_W(v)$ *ovat kohtisuorassa toisiaan vastaan. Lisäksi* $P_W|_W = \mathrm{id}$ *eli* $P_W(w) = w$ *jokaisella* $w \in W$*.*

Todistus. Koska sisätulo on lineaarinen ensimmäisen muuttujan suhteen, niin P_W on lineaarikuvaus.

Olkoon $v \in V$ ja olkoon $i \in \{1, \ldots, n\}$. Tällöin

$$\begin{aligned}
\langle v - P_W(v), w_i\rangle &= \langle v, w_i\rangle - \langle P_W(v), w_i\rangle = \langle v, w_i\rangle - \left\langle \sum_{j=1}^{n} \langle v, w_j\rangle w_j, w_i \right\rangle \\
&= \langle v, w_i\rangle - \sum_{j=1}^{n} \langle v, w_j\rangle \langle w_j, w_i\rangle = \langle v, w_i\rangle - \sum_{j=1}^{n} \langle v, w_j\rangle \delta_{ij} \\
&= \langle v, w_i\rangle - \langle v, w_i\rangle = 0.
\end{aligned}$$

Näin ollen $v - P_W(v)$ on kohtisuorassa aliavaruutta W vastaan. □

Huomautus 1.3.27. *Kuvaus* P_W *on itseasiassa niin sanottu ortonormaaliprojektio aliavaruudelle* W*. Tähän aiheeseen palataan luvussa 4.*

Gram–Schmidtin todistus. Koska V on äärellisulotteinen vektoriavaruus, sillä on kanta $(w_1, \ldots, w_n)$. Muodostetaan ortonormaali kanta $(v_1, \ldots, v_n)$ induktiolla seuraavasti. Määritellään ensin $W_k = \mathrm{Sp}\{w_1, \ldots, w_k\}$ jokaisella $k \in \{1, \ldots, n\}$.

Koska $w_1 \neq 0$, niin voidaan valita $v_1 = w_1/\|w_1\|$. Oletetaan nyt, että $k \in \{1, \ldots, n-1\}$ on sellainen luku, että on valittu sellaiset vektorit $v_1, \ldots, v_k$, että $\mathrm{Sp}\{v_1, \ldots, v_k\} = W_k$ ja että $\langle v_i, v_j\rangle = \delta_{ij}$ kaikilla $i, j \in \{1, \ldots, k\}$.

Olkoon nyt $P_k = P_{W_k} : V \to W_k$ kohtisuora projektio aliavaruudelle W_k. Koska $w_{k+1} \notin W_k$, niin $w_{k+1} \neq P_k(w_{k+1})$ ja voidaan määritellä

$$v_{k+1} = \frac{w_{k+1} - P_k(w_{k+1})}{\|w_{k+1} - P_k(w_{k+1})\|}.$$

Koska $\mathrm{Sp}\{v_1, \ldots, v_k\} = W_k$ ja $w_{k+1} \in \mathrm{Sp}\{v_1, \ldots, v_k, v_{k+1}\}$, niin $\mathrm{Sp}\{v_1, \ldots, v_{k+1}\} = W_{k+1}$. Selvästi $\|v_{k+1}\| = 1$ ja lisäksi lemma 1.3.26 nojalla vektori $w_{k+1} - P_k(w_{k+1})$ on kohtisuorassa aliavaruuden W_k vektoreita vastaan. Näin ollen $(v_1, \ldots, v_{k+1})$ on aliavaruuden W_{k+1} ortonormaalikanta. □

1.3.4 Sisätulon ja pistetulon samastus

Ortonormaalin kannan valinta samaistaa[3] sisätuloavaruuden $(V, \langle\cdot,\cdot\rangle)$ sarakeavaruuden $(\mathbb{R}^{n\times 1}, \cdot)$ kanssa, missä $n = \dim V$ ja $\cdot$ on tavallinen pistetulo. Tämä tarkoittaa sitä, että on olemassa isomorfismi näiden avaruuksien välillä, joka muuttaa sisätulon $\langle\cdot,\cdot\rangle$ pistetuloksi. Kirjataan tämä havainto tarkasti.

Lause 1.3.28. *Olkoon $(V, \langle\cdot,\cdot\rangle)$ ja $(v_1, \dots, v_n)$ ortonormaalikanta. Tällöin isomorfismille $\Psi_V : \mathbb{R}^{n\times 1} \to V$, $e_i \mapsto v_i$ pätee*

$$\langle \Phi_V(x), \Phi_V(y)\rangle = x \cdot y$$

kaikilla $x, y \in \mathbb{R}^{n\times 1}$.

Todistus. Olkoot

$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^{n\times 1} \quad \text{ja} \quad y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \in \mathbb{R}^{n\times 1}.$$

Koska kanta $(v_1, \dots, v_n)$ on ortonormaali, niin

$$\langle \Phi_V(x), \Phi_V(y)\rangle = \langle x_1v_1 + \cdots + x_nv_n, y_1v_1 + \cdots + y_nv_n\rangle = x_1y_1 + \cdots + x_ny_n = x \cdot y.$$

$\square$

Lause voidaan muotoilla yhtäpitävästi käänteiskuvausten avulla.

Korollaari 1.3.29. *Olkoon $(V, \langle\cdot,\cdot\rangle)$ ja $(v_1, \dots, v_n)$ ortonormaalikanta. Tällöin isomorfismille $\Psi_V : \mathbb{R}^{n\times 1} \to V$, $e_i \mapsto v_i$ pätee*

$$\langle v, w\rangle = \Phi_V^{-1}(v) \cdot \Phi_V^{-1}(w)$$

kaikilla $v, w \in V$.

1.3.5 Aliavaruuden kohtisuorakomplementti

Määritelmä 1.3.30. *Sisätuloavaruuden $(V, \langle\cdot,\cdot\rangle)$ aliavaruuden W* kohtisuorakomplementti *on joukko*

$$W^\perp = \{w \in V : \langle w, v\rangle = 0 \text{ kaikilla } v \in W\}.$$

Sisätulon bilineaarisuudesta seuraa, että kohtisuorakomplementti on aina aliavaruus.

Lemma 1.3.31. *Olkoot $(V, \langle\cdot,\cdot\rangle)$ sisätuloavaruus ja $W \subset V$ aliavaruus. Tällöin $W^\perp$ on aliavaruus.*

[3] Suomen kielen lautakunnan tulkinnan mukaan voidaan perustellusti sekä "samastaa" että "samaistaa". Molemmat muodot esiintyvät tässä prujussa.

Todistus. Olkoot $w, w' \in W^\perp$, $a \in \mathbb{R}$ ja $v \in W$. Tällöin

$$\langle w + aw', v\rangle = \langle w, v\rangle + a\langle w', v\rangle = 0.$$

Näin ollen $w + aw' \in W^\perp$. Näin ollen $W^\perp$ on aliavaruus. $\square$

Kohtisuoralla komplementilla on tärkeä ominaisuus, että aliavaruuden ja sen kohtirsuoran komplementin summa on avaruuden suora summa. Tämäkin seuraa suoraan sisätulon ominaisuuksista. Muotoillaan tämä tulos tarkemmin seuraavasti.

Lemma 1.3.32. *Olkoot $(V, \langle\cdot,\cdot\rangle)$ sisätuloavaruus ja $W \subset V$ aliavaruus. Tällöin $V = W \oplus W^\perp$.*

Todistus. Osoitetaan ensin, että $W \cap W^\perp = \{0\}$. Olkoon $w \in W \cap W^\perp$. Tällöin $w \in W^\perp$ ja $w \in W$, eli

$$\langle w, w\rangle = 0.$$

Näin ollen $w = 0$, eli $W \cap W^\perp = \{0\}$.

Osoitetaan nyt, että $W + W^\perp = V$. Olkoon $P\colon V \to W$ kohtisuora projektio ja $v \in V$. Tällöin $P(v) \in W$ ja kohtisuoran projektion määritelmän mukaan $\langle v - P(v), w\rangle = 0$ kaikilla $w \in W$. Näin ollen $v - P(v) \in W^\perp$ ja

$$v = P(v) + (v - P(v)) \in W + W^\perp.$$

$\square$

1.4 Matriisin transpoosi ja lineaarikuvauksen adjungaatti

Transpoosi on karakterisesti matriiseihin liittyvä käsite.

Määritelmä 1.4.1. *Matriisin $A = (a_{ji}) \in \mathbb{R}^{m\times n}$* transpoosi *on matriisi $A^T = (b_{ij}) \in \mathbb{R}^{n\times m}$, jonka rivit ovat matriisin A sarakkeet, eli kaikilla $\in \{1, \ldots, n\}$ ja $j \in \{1, \ldots, m\}$ pätee $b_{ij} = a_{ji}$.*

Huomautus 1.4.2. *Transpoosille on useita merkintöjä. Usein matriisin A transpoosia merkitään $t(A)$ tai A'. Näissä luentomuistiinpanoissa käytetään merkintää A^T.*

Mikäli siis

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

niin

$$A^T = \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & \ddots & & a_{m2} \\ \vdots & & \ddots & \vdots \\ a_{1n} & a_{2n} & & a_{mn} \end{bmatrix}.$$

Esimerkki 1.4.3. *Konkreettiisia esimerkkejä transpoosista ovat esimerkiksi*

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^T = \begin{bmatrix} a & c \\ b & d \end{bmatrix} \quad ja \quad \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 4 \\ 0 & 0 & 1 \end{bmatrix}^T = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 3 & 4 & 1 \end{bmatrix}.$$

Tärkein syy kiinnostuksestamme transpoosiin liittyy huomioon, että pistetulo voidaan kirjoittaa transpoosin avulla. Vaikka väite seuraa suoraan matriisitulon määritelmästä ja jätetään todistamatta, kirjataan se kuitenkin lemmaksi sen tärkeyden vuoksi.

Lemma 1.4.4. *Olkoon* $\cdot\colon \mathbb{R}^{n\times 1} \times \mathbb{R}^{n\times 1} \to \mathbb{R}$ *avaruuden* $\mathbb{R}^{n\times 1}$ *pistetulo. Tällöin*

$$x \cdot y = x^T y$$

kaikilla $x, y \in \mathbb{R}^{n\times 1}$.

Suoraan transpoosin määritelmästä on helppo johtaa laskusääntöjä matriisin transpoosille. Kirjataan niistä ylös muutamia. Väitteiden todistus jätetään harjoitustehtäväksi.

Lemma 1.4.5. *Olkoot* $A, B \in \mathbb{R}^{m\times n}$ *ja* $a \in \mathbb{R}$. *Tällöin* $(A + B)^T = A^T + B^T$, $(cA)^T = cA^T$. *Lisäksi, jos* $C \in \mathbb{R}^{n\times k}$, *niin* $(AC)^T = C^T A^T$. *Erityisesti, jos* $E \in R^{n\times n}$ *on kääntyvä matriisi, niin* $(E^{-1})^T = (E^T)^{-1}$.

Koska matriiseilla on tärkeä rooli lineaarikuvausten teoriassa, herää luonnollinen kysymys, mikä on matriisin transpoosin tulkinta lineaarikuvausten tapauksessa. Tähän on kaksi vastausta. Ensimmäinen vastaus liittyy vektoriavaruuksien duaaliavaruuksiin ja duaaliavaruuksien välille indusoituihin lineaarikuvauksiin. Tätä tulkintaa käsitellään liitteessä C. Käsitellään nyt lähemmin kurssin materiaaliin liittyvää tulkintaa eli lineaarikuvauksen adjungaattia. Tässä tulkinnassa aloitetaan vektoriavaruuden sijaan sisätuloavaruudesta.

Määritelmä 1.4.6. *Olkoot* $(V, \langle\cdot,\cdot\rangle_V)$ *ja* $(W, \langle\cdot,\cdot\rangle_W)$ *sisätuloavaruuksia. Lineaarikuvauksen* $f\colon V \to W$ adjungaatti *on lineaarikuvaus* $f^*\colon W \to V$, *joka toteuttaa ehdon* $\langle f(v), w\rangle_W = \langle v, f^*(w)\rangle_V$ *kaikilla* $v \in V$ *ja* $w \in W$.

Jokaisella lineaarikuvauksella sisätuloavaruuksien välillä on adjungaatti ja se löydetään kiinnittämällä avaruuksiin ortonormaalit kannat ja transponoimalla näissä kannoissa oleva lineaarikuvauksen esitysmatriisi. Tarkemmin tämä tulos muotoillaan seuraavasti.

Lause 1.4.7. *Olkoot* $(V, \langle\cdot,\cdot\rangle_V)$ *ja* $(W, \langle\cdot,\cdot\rangle_W)$ *sisätuloavaruuksia ja* $f\colon V \to W$ *lineaarikuvaus. Tällöin lineaarikuvauksella* f *on adjungaatti* $f^*\colon W \to V$. *Lisäksi, jos* $(v_1, \ldots, v_n)$ *ja* $(w_1, \ldots, w_m)$ *ovat avaruuksien* V *ja* W *ortonormaaleja kantoja ja* $A \in \mathbb{R}^{m\times n}$ *on kuvauksen* f *esitysmatriisi näissä kannoissa, niin kuvauksen* f^* *matriisi näissä kannoissa on* A^T.

Lauseen todistuksen varsinainen ydin on seuraavassa havainnossa, joka sanoo, että lineaarikuvauksen $f_A\colon \mathbb{R}^{n\times 1} \to \mathbb{R}^{m\times 1}$, $x \mapsto Ax$, adjungaatti on lineaarikuvaus $f_{A^T}\colon \mathbb{R}^{m\times 1} \to \mathbb{R}^{n\times 1}$, $y \mapsto A^T y$.

Lemma 1.4.8. *Olkoon $A \in \mathbb{R}^{m\times n}$. Tällöin*

$$f_A(x) \cdot y = x \cdot f_{A^T}(y)$$

kaikilla $x \in \mathbb{R}^{n\times 1}$ ja $y \in \mathbb{R}^{m\times 1}$.

Todistus. Olkoot $x \in \mathbb{R}^{n\times 1}$ ja $y \in \mathbb{R}^{m\times 1}$. Tällöin

$$x \cdot f_{A^T}(y) = x \cdot (A^T y) = x^T A^T y = (Ax)^T y = (Ax) \cdot y = f_A(x) \cdot y.$$

□

Lauseen 1.4.7 todistus. Olkoot $(v_1, \ldots, v_n)$ ja $(w_1, \ldots, w_m)$ avaruuksien V ja W ortonormaaleja kantoja sekä olkoot $\Psi_V : \mathbb{R}^{n\times 1} \to V$, $e_i \to v_i$, ja $\Psi_W : \mathbb{R}^{m\times 1} \to W$, $e_i \mapsto w_i$, isomorfismeja. Tällöin kaikilla $v, v' \in V$ ja $w, w' \in W$ pätee

$$\langle v, v' \rangle_V = (\Phi_V)^{-1}(v) \cdot (\Phi_V)^{-1}(v') \tag{1.5}$$

ja

$$\langle w, w' \rangle_V = (\Phi_W)^{-1}(w) \cdot (\Phi_W)^{-1}(w'), \tag{1.6}$$

missä $\cdot$ on pistetulo avaruuksissa $\mathbb{R}^{n\times 1}$ ja $\mathbb{R}^{m\times 1}$.

Olkoon nyt $A \in \mathbb{R}^{m\times n}$ lineaarikuvauksen matriisi näissä kannoissa. Tällöin $f = \Phi_W \circ f_A \circ \Phi_V^{-1}$, missä $f_A : \mathbb{R}^{n\times 1} \to \mathbb{R}^{m\times 1}$ on tuttu kuvaus $x \mapsto Ax$.

Määritellään lineaarikuvaus $g : W \to V$ kaavalla $g = \Phi_V \circ f_{A^T} \circ (\Phi_W)^{-1}$. Olkoot nyt $v \in V$ ja $w \in W$. Merkitään $x = \Phi_V^{-1}(v)$ ja $y = \Phi_W^{-1}(w)$. Tällöin kaavan (1.5) ja kuvauksen g määritelmän nojalla pätee

$$\langle v, g(w) \rangle_V = (\Phi_V)^{-1}(v) \cdot (\Phi_V)^{-1}(g(w)) = x \cdot f_{A^T}(\Phi_W^{-1}(w)) = x \cdot f_{A^T}(y).$$

Toisaalta vastaavasti

$$\langle f(v), w \rangle_W = (\Phi_W)^{-1}(f(v)) \cdot (\Phi_W)^{-1}(w) = f_A(\Phi_V^{-1}(v)) \cdot y = f_A(x) \cdot y.$$

Lemman 1.4.8 perusteella $f_A(x) \cdot y = x \cdot f_{A^T}(y)$. Näin ollen

$$\langle v, g(w) \rangle_V = \langle f(v), w \rangle_W$$

kaikilla $v \in V$ ja $w \in W$. Kuvaus g on siis kuvauksen f adjungaatti f^*. Lisäksi kuvauksen f^* esitysmatriisi on A^T kuten haluttiin. □

Huomautus 1.4.9. *Huomaa, että lauseen 1.4.7 todistuksessa on olennaista, että jonot $(v_1, \ldots, v_n)$ ja $(w_1, \ldots, w_m)$ ovat ortonormaaleja kantoja. Tätä oletusta tarvittiin siihen, että A^T on kuvauksen f^* esitysmatriisi. Kantojen ortonormaaliuden vaatimus on luonnollinen, koska adjungaatti määritellään sisätulon avulla.*

1.5 Matriisin sarakeavaruus

Matriisin $A \in \mathbb{R}^{m\times n}$ sarakeavaruus on sama kuin lineaarikuvauksen $f_A\colon \mathbb{R}^{n\times 1} \to \mathbb{R}^{m\times 1}$ kuva.

Määritelmä 1.5.1. *Matriisin* $A = [v_1 \cdots v_n] \in \mathbb{R}^{m\times n}$ sarakeavaruus *on avaruuden* $\mathbb{R}^{m\times 1}$ *aliavaruus*

$$\mathrm{Col}(A) = \{x_1 v_1 + \cdots + x_n v_n \in \mathbb{R}^{m\times 1}\colon x_1, \ldots, x_n \in \mathbb{R}\}.$$

Suoraan matriisitulon määritelmästä havaitaan, että

$$\mathrm{Col}(A) = \{Ax \in \mathbb{R}^{m\times 1}\colon x \in \mathbb{R}^{n\times 1}\}.$$

Näin ollen suoraan kuvauksen $f_A\colon \mathbb{R}^{n\times 1} \to \mathbb{R}^{m\times 1}$, $x \mapsto Ax$, määritelmästä havaitaan, että

$$\mathrm{Col}(A) = \{Ax \in \mathbb{R}^{m\times 1}\colon x \in \mathbb{R}^{n\times 1}\} = \{f_A(x) \in \mathbb{R}^{m\times 1}\colon x \in \mathbb{R}^{n\times 1} = \mathrm{im}(f_A),$$

eli että $\mathrm{Col}(A)$ on itseasiassa kuvauksen f_A kuva.

1.5.1 Matriisin sarakeavaruuden kanta

Usein on tärkeää etsiä matriisin A sarakeavaruuden $\mathrm{Col}(A)$ kanta. Aloitetaan havainnosta.

Olkoon $A = \begin{bmatrix} v_1 \\ \cdots \\ v_n \end{bmatrix} \in \mathbb{R}^{m\times n}$ matriisi. Tällöin sarakkeet $v_1, \ldots, v_n \in \mathbb{R}^{m\times 1}$ virittävät sarkeavaruuden $\mathrm{Col}(A)$, eli

$$\mathrm{Col}(A) = \mathrm{Sp}\{v_1, \ldots, v_n\}.$$

Vektoriavaruuksien yleisen teorian perusteella jono $(v_1, \ldots, v_n)$ sisältää avaruuden $\mathrm{Col}(A)$ kannan $(v_{i_1}, \ldots, v_{i_k})$, missä $1 \le k \le n$ ja $1 \le i_1 < i_2 < \cdots < i_k \le n$. Kanta voidaan löytää seuraavalla algoritmilla.

Ensimmäisen indeksin i_1 valinta. Olkoon $i_1 \in \{1, \ldots, n\}$ pienin sellainen indeksi $1 \le i \le n$, että vektori v_i ei ole nolla-vektori, eli $v_{i_1} \neq 0$ ja $v_i = 0$ kaikilla $i = 1, \ldots, i_1 - 1$.

Induktioaskel. Oletetaan, että on valittu indeksit $1 \le i_1 < \cdots < i_\ell$. Jos vektorit $v_{i_1}, \ldots, v_{i_\ell}$ eivät viritä aliavaruutta $\mathrm{Col}(A)$, niin voidaan valita pienin sellainen indeksi $i_\ell < i_{\ell+1} \le n$, että

$$v_{i_{\ell+1}} \notin \mathrm{Sp}\{v_{i_1}, \ldots, v_{i_\ell}\}.$$

Tätä induktioaskelta voi toistaa alle n kertaa, joten induktio päättyy ja löytyy sellainen $1 \le k \le n$, että

$$\mathrm{Col}(A) = \mathrm{Sp}\{v_{i_1}, \ldots, v_{i_k}\}.$$

Määritellään

$$P = \begin{bmatrix} v_{i_1} & \cdots & v_{i_k} \end{bmatrix} \in \mathbb{R}^{m\times k}.$$

Koska $(v_{i_1}, \ldots, v_{i_k})$ on avaruuden $\mathrm{Col}(A)$ kanta, niin

$$\mathrm{Col}(P) = \mathrm{Col}(A).$$

1.5.2 Matriisi sarakeavaruuden kantamatriisin ja yläkolmiomatriisin tulona

Matriisi A voidaan luonnollisella tavalla kirjoittaa tulona matriisista P ja yläkolmiomatriisista.

Aloitetaan havainnosta. Olkoon $1 \leq j \leq n$. Tällöin on olemassa suurin sellainen indeksi $1 \leq \ell \leq k$, että $i_\ell \leq j$. Kannan $(v_{i_1}, \ldots, v_{i_k})$ valinnan perusteella $v_j \in \mathrm{Sp}\{v_{i_1}, \ldots, v_{i_\ell}\}$. Näin ollen

$$v_j = t_{1j} v_{i_1} + \cdots + t_{\ell j} v_{i_\ell}.$$

Huomautus 1.5.2. *Huomaa, että jos $i_\ell = j$, niin tällöin $t_{j\ell} = 1$ ja muut kooordinaatit $t_{j1}, \ldots, t_{j(\ell-1)}$ ovat nollia. Huomataan myös, että koska $i_\ell \leq j$, niin $\ell \leq j$.*

Olkoon

$$t_j = \begin{bmatrix} t_{1j} \\ \vdots \\ t_{\ell j} \\ 0 \\ \vdots \\ 0 \end{bmatrix} \in \mathbb{R}^{k \times 1}$$

jokaisella $1 \leq j \leq n$ ja olkoon

$$T = \begin{bmatrix} t_1 & \cdots & t_n \end{bmatrix} \in \mathbb{R}^{k \times n}.$$

Matriisi T on yläkolmiomatriisi.

Jäljellä on osoittaa, että

$$A = PT.$$

Aloitetaan jälleen havainnosta. Olkoon $1 \leq j \leq n$. Tällöin

$$\begin{aligned} v_j &= t_{1j} v_{i_1} + \cdots + t_{\ell j} v_{i_\ell} \\ &= t_{1j} v_{i_1} + \cdots + t_{\ell j} v_{i_\ell} + 0 v_{i_\ell + 1} + \cdots + 0 v_{i_k} \\ &= \begin{bmatrix} v_{i_1} & \cdots & v_{i_\ell} & v_{i_\ell + 1} & \cdots & v_{i_k} \end{bmatrix} \begin{bmatrix} t_{1j} \\ \vdots \\ t_{\ell j} \\ 0 \\ \vdots \\ 0 \end{bmatrix} = P t_j. \end{aligned}$$

Näin ollen matriisitulon määritelmän nojalla

$$A = \begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix} = \begin{bmatrix} v_{i_1} & \cdots & v_{i_k} \end{bmatrix} \begin{bmatrix} t_1 & \cdots & t_n \end{bmatrix} = PT.$$

Helppo esimerkki tästä matriisitulosta on seuraava.

Esimerkki 1.5.3. *Olkoon*

$$A = \begin{bmatrix} v_1 & v_2 & v_3 & v_4 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 & 2 \\ 2 & 4 & 1 & 6 \end{bmatrix}.$$

Algoritmi antaa kannan (v_{i_1}, v_{i_2}), *missä*

$$v_{i_1} = \begin{bmatrix} 0 \\ 2 \end{bmatrix} \quad \textit{ja} \quad v_{i_2} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

Matriisi P *on siis*

$$P = \begin{bmatrix} v_{i_1} & v_{i_2} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 2 & 1 \end{bmatrix}.$$

Matriisi T *löydetään ratkaisemalla yhtälöt*

$$Pt_j = v_j$$

kaikilla $j = 1, \dots, 4$, *eli ratkaisemalla matriisiyhtälöryhmä*

$$\left[\begin{array}{cc|cccc} 0 & 1 & 0 & 0 & 1 & 2 \\ 2 & 1 & 2 & 4 & 1 & 6 \end{array}\right].$$

Rivioperaatiot antavat

$$\left[\begin{array}{cc|cccc} 2 & 0 & 2 & 4 & 0 & 4 \\ 0 & 1 & 0 & 0 & 1 & 2 \end{array}\right]$$

eli

$$\left[\begin{array}{cc|cccc} 1 & 0 & 1 & 2 & 0 & 2 \\ 0 & 1 & 0 & 0 & 1 & 2 \end{array}\right].$$

Näin ollen

$$T = \begin{bmatrix} 1 & 2 & 0 & 2 \\ 0 & 0 & 1 & 2 \end{bmatrix}.$$

1.6 Matriisin nolla-avaruus

Kuvauksen f_A ydin $\ker(f_A)$ vastaa puolestaan matriiisin A nolla-avaruutta.

Määritelmä 1.6.1. *Matriisin* $A \in \mathbb{R}^{m \times n}$ nolla-avaruus *on avaruuden* $\mathbb{R}^{n \times 1}$ *aliavaruus*

$$\mathrm{Null}(A) = \{x \in \mathbb{R}^{n \times 1} \colon Ax = 0\}.$$

Yksi lineaarikuvausten perustuloksista on, että äärellisulotteisten vektoriavaruuksien väliselle lineaarikuvauselle $f \colon V \to W$ pätee

$$\dim V = \ker(f) + \mathrm{im}(f).$$

Matriiseille tulkittuna tämä tarkoittaa, että matriisille $A \in \mathbb{R}^{m \times n}$ pätee

$$n = \dim(\mathrm{Col}(A)) + \dim(\mathrm{Null}(A)).$$

Matriisin sarakeavaruuden dimensiota kutstuun matriisin asteeksi.

Määritelmä 1.6.2. *Matriisin* $A \in \mathbb{R}^{n\times n}$ aste $\operatorname{rank}(A)$ *on sen sarakeavaruuden dimensio, eli* $\operatorname{rank}(A) = \dim \operatorname{Col}(A)$.

Matriisin asteen avulla voidaan helposti muotoilla matriisin kääntyvyys ja siihen liittyviä tuloksia. Seuraava lemma seuraa suoraan vastaavista lineaarikuvausten tuloksista ja todistus jätetään lukijalle.

Lemma 1.6.3. *Olkoon* $A \in \mathbb{R}^{m\times n}$. *Tällöin seuraavat ehdot ovat yhtäpitäviä:*

1. *matriisi* A *on kääntyvä,*
2. *matriisin* A *sarekkeet ovat lineaarisesti riippumattomia,*
3. $\operatorname{rank}(A) = n$,
4. $\dim(\operatorname{Null}(A)) = 0$.

1.7 Matriisin riviavaruus

Määritelmä 1.7.1. *Matriisin*

$$A = \begin{bmatrix} a_1 \\ \vdots \\ a_m \end{bmatrix} \in \mathbb{R}^{m\times n},$$

missä $a_1, \ldots, a_m \in \mathbb{R}^{1\times n}$, riviavaruus *on avaruuden* $\mathbb{R}^{1\times n}$ *aliavaruus*

$$\operatorname{Row}(A) = \{x_1 a_1 + \cdots + x_m a_m \in \mathbb{R}^{1\times n} \colon x_1, \ldots, x_m \in \mathbb{R}\}.$$

On syvällinen fakta, että $\dim \operatorname{Col}(A) = \dim \operatorname{Row}(A)$.

Lause 1.7.2. *Olkoon* $A \in \mathbb{R}^{m\times n}$. *Tällöin* $\dim \operatorname{Col}(A) = \dim \operatorname{Row}(A)$.

Todistetaan ensin vastaava väite sisätuloavaruuksien välisille lineaarikuvauksille.

Lause 1.7.3. *Olkoot* $(V, \langle\cdot,\cdot\rangle_V)$ *ja* $(W, \langle\cdot,\cdot\rangle_W)$ *äärellisulotteisia sisätuloavaruuksia ja* $L\colon V \to W$ *lineaarikuvaus. Tällöin adjungaatille* $L^*\colon W \to V$ *pätee*

$$\dim \operatorname{im} L = \dim \operatorname{im} L^*.$$

Lauseen 1.7.3 todistuksen ytimessä on kuvauksen L^* ytimen tulkinta kohtisuoran komplementin avulla.

Lemma 1.7.4. *Olkoot* $(V, \langle\cdot,\cdot\rangle_V)$ *ja* $(W, \langle\cdot,\cdot\rangle_W)$ *äärellisulotteisia sisätuloavaruuksia ja* $L\colon V \to W$ *lineaarikuvaus. Tällöin*

$$(\operatorname{im} L)^\perp = \ker L^*.$$

Todistus. Olkoon $w \in (\operatorname{im} L)^{\perp}$. Tällöin

$$\langle v, L^*(w) \rangle_V = \langle L(v), w \rangle_W = 0$$

kaikilla $v \in V$. Näin ollen $L^*(w) = 0$, eli $w \in \ker L^*$.

Olkoon nyt $w \in \ker L^*$. Tällöin jokaisella $v \in V$ pätee

$$\langle L(v), w \rangle_W = \langle v, L^* w \rangle_V = \langle v, 0 \rangle_V = 0.$$

Näin ollen $w \in (\operatorname{im} L)^{\perp}$. □

Lauseen 1.7.3 todistus. Koska

$$\dim \operatorname{im} L^* + \dim \ker L^* = \dim W = \dim \operatorname{im} L + \dim (\operatorname{im} L)^{\perp} = \dim \operatorname{im} L + \dim \ker L^*,$$

niin $\dim \operatorname{im} L^* = \dim \operatorname{im} L$. □

Matriiseille tulkittuna lause 1.7.3 sanoo seuraavaa.

Korollaari 1.7.5. *Olkoon $A \in \mathbb{R}^{m \times n}$. Tällöin* $\dim \operatorname{Col}(A) = \dim \operatorname{Col}(A^T)$.

Lauseen 1.7.2 todistus. Havaitaan aluksi, että matriisin A^T sarakkeet vastaavat matriisin A rivejä. Olkoon nyt $\tau \colon \mathbb{R}^{1 \times m} \to \mathbb{R}^{m \times 1}$, $x \mapsto x^T$. Transpoosin ominaisuuksien perusteella τ on lineaarikuvaus ja kuvaus $\mathbb{R}^{m \times 1} \to \mathbb{R}^{1 \times m}$, $y \mapsto y^T$, on kuvauksen τ käänteiskuvaus. Näin ollen τ on isomorfismi. Koska $\tau(\operatorname{Row}(A)) = \operatorname{Col}(A^T)$, niin $\dim \operatorname{Row}(A) = \dim \operatorname{Col}(A^T)$. Näin ollen korollaarin 1.7.5 perusteella

$$\dim \operatorname{Col}(A) = \dim \operatorname{Col}(A^T) = \dim \operatorname{Row}(A).$$

□

Luku 2

Ositetut matriisit

Joissain tilanteissa on merkintöjen kannalta hyödyllistä kirjoittaa matriisi $M \in \mathbb{R}^{m\times n}$ muodossa

$$M = \begin{bmatrix} A & C \\ B & D \end{bmatrix}, \tag{2.1}$$

missä $A \in \mathbb{R}^{k\times\ell}$, $B \in \mathbb{R}^{(m-k)\times\ell}$, $C \in \mathbb{R}^{k\times(n-\ell)}$ ja $D \in \mathbb{R}^{(m-k)\times(n-\ell)}$ ovat matriiseja, eli

$$M = \begin{bmatrix} a_{11} & \cdots & a_{1\ell} & c_{11} & \cdots & c_{1(n-\ell)} \\ \vdots & & \vdots & \vdots & & \vdots \\ a_{k1} & \cdots & a_{k\ell} & c_{k1} & \cdots & c_{k(n-\ell)} \\ b_{11} & \cdots & b_{1\ell} & d_{11} & \cdots & d_{1(n-\ell)} \\ \vdots & & \vdots & \vdots & & \vdots \\ b_{(m-k)1} & \cdots & b_{(m-k)\ell} & d_{(m-k)1} & \cdots & d_{(m-k)(n-\ell)} \end{bmatrix}.$$

Matriiseja A, B, C ja D kutsutaan jatkossa matriisin M *alimatriiseiksi.*

Huomaa, että yhtälön 2.1 vasemman puolen matriisia ei siis varsinaisesti ajatella 'matriisien matriisina' vaan ainoastaan tapana merkitä matriisin A' alkioita matriisien A, B, C ja D alkioiden avulla. Yhtälön (2.1) vasenta puolta kutsutaan *matriisin M ositetuksi matriisiksi.* Tässä luvussa tarkastellaan, miten ositetut matriisit liittyvät lineaarikuvauksiin ja miten ne syntyvät luonnollisella tavalla aliavaruuksien suorista summista.

Ositettu matriisi syntyy, kun sarakeavaruuden $\mathbb{R}^{n\times 1}$ standardikanta $e_1, \ldots, e_n$ ryhmitellään kahteen (tai useampaan) osaan kuvauksen $f_M \colon \mathbb{R}^{n\times 1} \to \mathbb{R}^{m\times 1}$ lähtö- ja maalipuolella. Tarkemmin, jos $1 \le \ell \le n$ ja $1 \le k \le m$ ja

$$M = \begin{bmatrix} A & C \\ B & D \end{bmatrix},$$

missä $A \in \mathbb{R}^{k\times\ell}$, $C \in R^{k\times(n-\ell)}$, $B \in \mathbb{R}^{(m-k)\times\ell}$ ja $D \in \mathbb{R}^{(m-k)\times(n-\ell)}$, niin jokaisella $1 \le i \le \ell$ pätee

$$Me_i = \begin{bmatrix} A & C \\ B & D \end{bmatrix} e_i = \begin{bmatrix} A \\ B \end{bmatrix} e_i = \begin{bmatrix} Ae_i \\ Be_i \end{bmatrix},$$

missä $Ae_i \in \mathbb{R}^{k\times 1}$ ja $Be_i \in \mathbb{R}^{(m-k)\times 1}$.

Vastaavasti jokaisella $\ell + 1 \le i \le n$ pätee

$$Me_i = \begin{bmatrix} A & C \\ B & D \end{bmatrix} e_i = \begin{bmatrix} C \\ D \end{bmatrix} e_{i-\ell} = \begin{bmatrix} Ce_{i-\ell} \\ De_{i-\ell} \end{bmatrix}.$$

Huomautus 2.0.1. *Ositetussa matriisissa alimatriisit vastaavat luonnollisella tavalla alivaruuksien välisiä lineaarikuvauksia. Tätä on käsitelty tarkemmin liitteessä B.*

2.1 Laskentoa ositetuilla matriiseilla

Ositettuja matriiseja käytetään usein laskuissa, joten kirjataan niiden peruslaskusäännöt. Todistukset ovat suoraviivaisia laskuja, joten ne jätetään harjoitustehtäviksi. Aloitetaan yhteenlaskusta.

Lemma 2.1.1. *Olkoot*

$$M_1 = \begin{bmatrix} A_1 & C_1 \\ B_1 & D_1 \end{bmatrix} \in \mathbb{R}^{m\times n} \quad \textit{ja} \quad M_2 = \begin{bmatrix} A_2 & C_2 \\ B_2 & D_2 \end{bmatrix} \in \mathbb{R}^{m\times n}.$$

Jos matriisiyhteenlaskut $A_1 + A_2$ ja $D_1 + D_2$ ovat hyvin määriteltyjä, niin

$$aM_1 + M_2 = a \begin{bmatrix} A_1 & C_1 \\ B_1 & D_1 \end{bmatrix} + \begin{bmatrix} A_2 & C_2 \\ B_2 & D_2 \end{bmatrix} = \begin{bmatrix} aA_1 + A_2 & aC_1 + C_2 \\ aB_1 + B_2 & aD_1 + D_2 \end{bmatrix} \in \mathbb{R}^{m\times n}$$

kaikilla $a \in \mathbb{R}$.

Huomautus 2.1.2. *Edellisessä lemmassa käytettiin huomiota, että jos matriisi yhteenlaskut $A_1 + A_2$ ja $D_1 + D_2$ ovat hyvin määriteltyjä (eli matriiseilla A_1 ja A_2 on sama määrä rivejä ja sarakkeita kuten myös matriiseilla D_1 ja D_2), niin silloin myös yhteenlaskut $B_1 + B_2$ ja $C_1 + C_2$ ovat hyvin määriteltyjä. Tämänkin havainnon toteaminen jätetään osaksi lemman todistusta. Huomaa, että on helppo osittaa matriisit M_1 ja M_2 sellaisella tavalla, että yhteenlaskua ei voi suorittaa alimatriisien avulla.*

Ositettujen matriisien kertolaskun kohdalla saadaan vastaava tulos.

Lemma 2.1.3. *Olkoot*

$$M_1 = \begin{bmatrix} A_1 & C_1 \\ B_1 & D_1 \end{bmatrix} \in \mathbb{R}^{m\times n} \quad \textit{ja} \quad M_2 = \begin{bmatrix} A_2 & C_2 \\ B_2 & D_2 \end{bmatrix} \in \mathbb{R}^{n\times k}.$$

Jos matriisikertolaskut A_1A_2 ja D_1D_2 ovat hyvin määriteltyjä, niin

$$M_1M_2 = \begin{bmatrix} A_1 & C_1 \\ B_1 & D_1 \end{bmatrix} \begin{bmatrix} A_2 & C_2 \\ B_2 & D_2 \end{bmatrix} = \begin{bmatrix} A_1A_2 + C_1B_2 & A_1C_2 + C_1D_2 \\ B_1A_2 + D_1B_2 & B_1C_2 + D_1D_2 \end{bmatrix} \in \mathbb{R}^{m\times k}.$$

2.2 Ositetun matriisin käänteismatriisi

Ositetun matriisin

$$M = \begin{bmatrix} A & C \\ B & D \end{bmatrix} \in \mathbb{R}^{n \times n}$$

käänteismatriisille ei ole yleistä kaavaa, joka perustuisi matriiseihin A, B, C ja D. Tähän on monia syitä. Yksi sellainen on, että yhdenkään alimatriisin ei tarvitse olla kääntyvä vaikka M olisi kääntyvä. Tälläinen tapaus on matriisi

$$M = \begin{bmatrix} A & C \\ B & D \end{bmatrix} = \left[\begin{array}{cc|cc} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \hline 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right].$$

Näin ollen on luonnollista tarkastella erikoistapauksia, joissa jokin alimatriiseista on kääntyvä. Tarkastellaan tarkemmin tapausta, että alimatriisi A on kääntyvä. Aloitetaan erikoistapauksista. Merkintöjen yksinkertaistamiseksi käytetään jatkossa identtisestä matriisista $I_{\ell \times \ell} \in \mathbb{R}^{\ell \times \ell}$ lyhyempää merkitää I.

Lemma 2.2.1. *1. Jos neliömatriisit A ja D ovat kääntyviä, niin*

$$\begin{bmatrix} A & 0 \\ 0 & D \end{bmatrix}^{-1} = \begin{bmatrix} A^{-1} & 0 \\ 0 & D^{-1} \end{bmatrix}.$$

2. Jos neliömatriisit B ja C ovat kääntyviä, niin

$$\begin{bmatrix} 0 & C \\ B & 0 \end{bmatrix}^{-1} = \begin{bmatrix} 0 & C^{-1} \\ B^{-1} & 0 \end{bmatrix}.$$

3. Jokaisella matriisilla $B \in \mathbb{R}^{(n-k) \times k}$ pätee

$$\begin{bmatrix} I & 0 \\ B & I \end{bmatrix}^{-1} = \begin{bmatrix} I & 0 \\ -B & I \end{bmatrix}.$$

4. Jokaisella matriisilla $C \in \mathbb{R}^{k \times (n-k)}$ pätee

$$\begin{bmatrix} I & C \\ 0 & I \end{bmatrix}^{-1} = \begin{bmatrix} I & -C \\ 0 & I \end{bmatrix}.$$

Todistus. Jokainen kohta voidaan todistaa suoralla laskulla tarkistamalla, että annettu matriisi on käänteismatriisi. Kohta (3) on mielenkiitoisin, joten tehdään siihen liityvät laskut. Lemman 2.1.3 perusteella pätee

$$\begin{bmatrix} I & 0 \\ -B & I \end{bmatrix} \begin{bmatrix} I & 0 \\ B & I \end{bmatrix} = \begin{bmatrix} I & 0 \\ -B + B & I \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix}.$$

Vastaavasti lasketaan

$$\begin{bmatrix} I & 0 \\ B & I \end{bmatrix} \begin{bmatrix} I & 0 \\ -B & I \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix}.$$

Kohdan (3) väite on siis todistettu. □

Huomautus 2.2.2. *Lemmalla 2.2.1 on erikoistapauksien kirjaamista syvempi merkitys. Lemmassa tarkastellut matriisit ovat kannanvaihtomatriiseja tapauksissa, joissa sarakeavaruuden $\mathbb{R}^{n\times n}$ standardikanta $(e_1, \ldots, e_k, e_{k+1}, \ldots, e_n)$ on ositettu kahteen osaan $(e_1, \ldots, e_k)$ ja $(e_{k+1}, \ldots, e_n)$ ja näitä osia käsitellään eri tavoin. Esimerkiksi kohdassa 3 kanta $(e_1, \ldots, e_k, e_{k+1}, \ldots, e_n)$ vaihdetaan kannaksi $(e_1 + Be_{k+1}, \ldots, e_k + Be_n, e_{k+1}, \ldots, e_n)$.*

Annetaan nyt välttämätön ja riittävä ehto ositetunmatriisin kääntyvyydelle erikoistapauksessa, että alimatriisi A on kääntyvä. Yleinen tapaus seuraa tästä erikoistapauksesta kannanvaihdoilla. Lauseen ytimessä on ositetun matriisin hajoittaminen tuloksi, joka tehdään ensin.

Lemma 2.2.3. *Olkoon*

$$M = \begin{bmatrix} A & C \\ B & D \end{bmatrix} \in \mathbb{R}^{n\times n}$$

ositettumatriisi, jonka alimatriisi $A \in \mathbb{R}^{k\times k}$ on kääntyvä neliömatriisi. Tällöin

$$M = \begin{bmatrix} A & C \\ B & D \end{bmatrix} = \begin{bmatrix} I & 0 \\ BA^{-1} & D - BA^{-1}C \end{bmatrix} \begin{bmatrix} I & C \\ 0 & I \end{bmatrix} \begin{bmatrix} A & 0 \\ 0 & I \end{bmatrix}$$

Huomautus 2.2.4. *Väite voidaan todistaa suoralla laskulla (kuten alla) tai tekemällä kertomalla matriisia M ensin matriisilla*

$$\begin{bmatrix} A^{-1} & 0 \\ 0 & I \end{bmatrix}$$

ja sitten matriisilla

$$\begin{bmatrix} I & -C \\ 0 & I \end{bmatrix}$$

löytäen näin puuttuvan matriisin.

Todistus. Suora laskulla saadaan

$$\begin{aligned} \begin{bmatrix} I & 0 \\ BA^{-1} & D - BA^{-1}C \end{bmatrix} \begin{bmatrix} I & C \\ 0 & I \end{bmatrix} \begin{bmatrix} A & 0 \\ 0 & I \end{bmatrix} &= \begin{bmatrix} I & 0 \\ BA^{-1} & D - BA^{-1}C \end{bmatrix} \begin{bmatrix} A & C \\ 0 & I \end{bmatrix} \\ &= \begin{bmatrix} A & C \\ BA^{-1}A & (BA^{-1}C + D - BA^{-1}C) \end{bmatrix} \\ &= \begin{bmatrix} A & C \\ B & D \end{bmatrix}. \end{aligned}$$

□

Lause 2.2.5. *Olkoon*
$$M = \begin{bmatrix} A & C \\ B & D \end{bmatrix} \in \mathbb{R}^{n\times n}$$
ositettumatriisi, jonka alimatriisi $A \in \mathbb{R}^{k\times k}$ *on kääntyvä neliömatriisi. Tällöin matriisi* M *on kääntyvä, jos ja vain jos matriisi* $E = D - BA^{-1}C$ *on kääntyvä. Lisäksi tällöin*
$$M^{-1} = \begin{bmatrix} (A^{-1} + A^{-1}CE^{-1}BA^{-1}) & (-A^{-1}CE^{-1}) \\ (-E^{-1}BA^{-1}) & E^{-1} \end{bmatrix}. \tag{2.2}$$

Todistus. Olkoon
$$M' = \begin{bmatrix} I & 0 \\ BA^{-1} & D - BA^{-1}C \end{bmatrix}.$$
Tällöin Lemman 2.2.3 mukaan
$$M = \begin{bmatrix} I & 0 \\ BA^{-1} & D - BA^{-1}C \end{bmatrix} \begin{bmatrix} I & C \\ 0 & I \end{bmatrix} \begin{bmatrix} A & 0 \\ 0 & I \end{bmatrix} = M' \begin{bmatrix} I & C \\ 0 & I \end{bmatrix} \begin{bmatrix} A & 0 \\ 0 & I \end{bmatrix},$$
missä matriisit
$$\begin{bmatrix} I & C \\ 0 & I \end{bmatrix} \quad \text{ja} \quad \begin{bmatrix} A & 0 \\ 0 & I \end{bmatrix}$$
ovat kääntyviä lemman 2.2.1 perusteella.

Oletetaan ensin, että matriisi $D - BA^{-1}C$ on kääntyvä. Tällöin matriisi M' on lemman 2.2.1 kohdan 3 perusteella kääntyvä. Näin ollen matriisi M on kääntyvien matriisien tulona kääntyvä.

Oletetaan nyt, että matriisi M on kääntyvä. Koska matriisit
$$\begin{bmatrix} I & C \\ 0 & I \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} A & 0 \\ 0 & I \end{bmatrix}$$
ovat kääntyviä, niin matriisi M' on kääntyvä. Koska matriisi M' on kääntyvä, niin sen sarakkeet ovat lineaarisesti riippumattomia. Näin ollen neliömatriisin $D - BA^{-1}C$ sarakkeet ovat lineaarisesti riippumattomia, eli matriisi $D - BA^{-1}C$ on kääntyvä.

Osoitetaan vielä, että käänteismatriisilla M^{-1} on haluttu kaava (2.2). Koska matriisi M on kääntyvä, niin myös $E = D - BA^{-1}C$ on kääntyvä. Näin ollen
$$\begin{aligned} M^{-1} = \begin{bmatrix} A & C \\ B & D \end{bmatrix}^{-1} &= \begin{bmatrix} A & 0 \\ 0 & I \end{bmatrix}^{-1} \begin{bmatrix} I & C \\ 0 & I \end{bmatrix}^{-1} \begin{bmatrix} I & 0 \\ BA^{-1} & D - BA^{-1}C \end{bmatrix}^{-1} \\ &= \begin{bmatrix} A^{-1} & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} I & -C \\ 0 & I \end{bmatrix} \begin{bmatrix} I & 0 \\ BA^{-1} & E \end{bmatrix}^{-1}, \end{aligned}$$

missä

$$\begin{aligned}
\begin{bmatrix} I & 0 \\ BA^{-1} & E \end{bmatrix}^{-1} &= \left(\begin{bmatrix} I & 0 \\ BA^{-1} & I \end{bmatrix} \begin{bmatrix} I & 0 \\ 0 & E \end{bmatrix} \right)^{-1} \\
&= \begin{bmatrix} I & 0 \\ 0 & E \end{bmatrix}^{-1} \begin{bmatrix} I & 0 \\ BA^{-1} & I \end{bmatrix}^{-1} \\
&= \begin{bmatrix} I & 0 \\ 0 & E^{-1} \end{bmatrix} \begin{bmatrix} I & 0 \\ -BA^{-1} & I \end{bmatrix} \\
&= \begin{bmatrix} I & 0 \\ -E^{-1}BA^{-1} & E^{-1} \end{bmatrix}.
\end{aligned}$$

Näin ollen pätee

$$\begin{aligned}
M^{-1} &= \begin{bmatrix} A^{-1} & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} I & -C \\ 0 & I \end{bmatrix} \begin{bmatrix} I & 0 \\ BA^{-1} & E \end{bmatrix}^{-1} \\
&= \begin{bmatrix} A^{-1} & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} I & -C \\ 0 & I \end{bmatrix} \begin{bmatrix} I & 0 \\ -E^{-1}BA^{-1} & E^{-1} \end{bmatrix} \\
&= \begin{bmatrix} A^{-1} & -A^{-1}C \\ 0 & I \end{bmatrix} \begin{bmatrix} I & 0 \\ -E^{-1}BA^{-1} & E^{-1} \end{bmatrix} \\
&= \begin{bmatrix} (A^{-1} + A^{-1}CE^{-1}BA^{-1}) & -A^{-1}CE^{-1} \\ -E^{-1}BA^{-1} & E^{-1} \end{bmatrix}
\end{aligned}$$

□

Huomautus 2.2.6. *Edellisestä todistuksesta sanottakoon, että mikäli olettaa matriisin E kääntyvyyden, niin matrisiin M kääntyvyyden voi helposti todistaa suoralla kertolaskulla. Tämä ei kuitenkaan juurikaan selitä, miksi väite on totta ja miksi matriisin E kääntyvyys on välttämätön ehto matriisin M kääntyvyydelle.*

Huomautus 2.2.7. *Kuten ennen lauseen 2.2.5 väitettä mainittiin, mikäli ositetunmatriisin*

$$M = \begin{bmatrix} A & C \\ B & D \end{bmatrix} \in \mathbb{R}^{n \times n}$$

jokin toinen alimatriisi B, C tai D on kääntyvä, niin voidaan kysymys matriisin M kääntyvyydestä palauttaa lauseen 2.2.5 tilanteeseen sopivalla kannanvaihdolla. Tällöin tietysti lauseessa esiintyvä matriisi E muuntuu kannanvaihtoa vastaavalla tavalla. (Harjoitustehtävä)

2.3 Sovellus: Matriisiavaruuksien välisen lineaarikuvausten matriisi

Tähän mennessä on käsitelty sarakeavaruuksien välisiä lineaarikuvauksia $f_A \colon \mathbb{R}^{n\times 1} \to \mathbb{R}^{m\times 1}$ ja niiden matriiseja $A \in \mathbb{R}^{m\times n}$. Matriisitulon ominaisuuksien nojalla sama matriisi $A \in \mathbb{R}^{m\times n}$ määrittelee jokaisella $k \in \mathbb{N}$ lineaarikuvauksen $F_A \colon \mathbb{R}^{m\times k} \to \mathbb{R}^{n\times k}$ kaavalla

$X \mapsto AX$. Myös tällä lineaarikuvauksella on luonnollinen matriisi, joka liittyy matriisiin A. Koska avaruus $\mathbb{R}^{m\times k}$ on mk-ulotteinen ja avaruus $\mathbb{R}^{n\times k}$ on nk-ulotteinen, on lineaarikuvauksen esitysmatriisin siis oltava $(mk) \times (nk)$-matriisi. Näin ollen se ei voi olla matriisi A itse! Paljastuu, että avaruuksien $\mathbb{R}^{m\times k}$ ja $\mathbb{R}^{n\times k}$ luonollisissa kannoissa lineaarikuvauksen F_A matriisi on sellainen ositettumatriisi, jonka alimatriisit ovat matriisin A kopioita. Tehdään tämä nyt tarkasti.

Tarkastellaan ensin avaruuden $\mathbb{R}^{m\times k}$ kantaa. Palautetaan mieleen Lineaarialgebra ja matriisilaskenta I kurssilta alkeismatriisit $E_{ji} = (e^{ji}_{j'i'} \in \mathbb{R}^{m\times k}$, joiden kertoimet on määritelty kaavalla

$$e^{ji}_{j'i'} = \delta_{jj'}\delta_{ii'} = \begin{cases} 1, & i = i' \text{ ja } j = j' \\ 0, & \text{muutoin} \end{cases}$$

kaikilla $j, j' \in \{1, \ldots, m\}$ ja $i, i' \in \{1, \ldots, k\}$. Matriisi E_{ji} on siis se matriisi jonka i:s sarake on standardikannan sarakevektori e_j. Selvästi jokainen avaruuden $A = (a_{ji})\mathbb{R}^{m\times k}$ alkio voidaan esittää yksikäsitteiseti matriisien E_{ji} lineaarikombinaationa kaavalla

$$A = \sum_{j=1}^{m}\sum_{i=1}^{k} a_{ji}E_{ji}.$$

(Harjoitustehtävä.) Näin ollen $E_{11}, \ldots, E_{m1}, E_{21}, \ldots, E_{mk}$ on vektoriavaruuden $\mathbb{R}^{m\times k}$ kanta. Järjestetään nyt nämä kantaalkiot jonoksi $E_1, \ldots, E_{mk}$ kaavalla

$$E_{m(i-1)+j} = E_{ji} \in \mathbb{R}^{m\times k}$$

missä $j \in \{1, \ldots, m\}$ ja $i \in \{1, \ldots, k\}$. Huomaa, että avaruuden $\mathbb{R}^{m\times 1}$ tapauksessa tämä järjestys tuottaa standardikannan $e_1, \ldots, e_m$.

Määritelmä 2.3.1. *Kantaa $E_1, \ldots, E_{mk}$ kutsutaan avaruuden $\mathbb{R}^{m\times k}$* standardikannaksi.

Huomautus 2.3.2. *Huomaa, että lineaarikuvaus $\mathbb{R}^{mk} \to \mathbb{R}^{m\times k}$, $e_\ell \mapsto E_\ell$, on (tuttu?) isomorfismi, joka samaistaa matriisiavaruuden $\mathbb{R}^{m\times k}$ sarakeavaruuden $\mathbb{R}^{mk}$ kanssa.*

On tärkeää huomata, että kantavektoreiden $E_1, \ldots, E_{mk}$ järjestyksellä on väliä ja että tätä kantaa kannattaa kutsua standardikannaksi ainoastaan kantavektoreiden tässä järjestyksesä. Syy tähän selviää tarkasteltassa lineaarikuvauksen $X \mapsto AX$ matriisiesitystä. Aloitetaan esimerkillä.

Esimerkki 2.3.3. *Olkoot $m = n = k = 2$ eli tarkastellaan matriisia*

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \in \mathbb{R}^{2\times 2}.$$

avaruuden $\mathbb{R}^{2\times 2}$ kannassa

$$E_1 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, E_2 = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, E_3 = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \textit{ ja } E_4 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}.$$

Suoralla laskulla havaitaan, että

$$AE_1 = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} a_{11} & 0 \\ a_{21} & 0 \end{bmatrix} = a_{11}E_1 + a_{21}E_2,$$

$$AE_2 = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} a_{12} & 0 \\ a_{22} & 0 \end{bmatrix} = a_{12}E_1 + a_{22}E_2,$$

$$AE_3 = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & a_{11} \\ 0 & a_{21} \end{bmatrix} = a_{11}E_3 + a_{21}E_4$$

ja

$$AE_4 = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & a_{12} \\ 0 & a_{22} \end{bmatrix} = a_{12}E_3 + a_{22}E_4.$$

Näin ollen lineaarikuvauksen $F_A\colon \mathbb{R}^{2\times 2} \to \mathbb{R}^{2\times 2}$, $X \mapsto AX$, *matriisi kannassa* $E_1, \dots, E_4$ *on*

$$M_A = \begin{bmatrix} a_{11} & a_{12} & 0 & 0 \\ a_{21} & a_{22} & 0 & 0 \\ 0 & 0 & a_{11} & a_{12} \\ 0 & 0 & a_{21} & a_{22} \end{bmatrix} = \begin{bmatrix} A & 0 \\ 0 & A \end{bmatrix}.$$

Yleisessä matriisin $A \in \mathbb{R}^{m\times n}$ tilanteessa todistus toimii aivan samoin kuin yllä käsitellyssä erikoistapauksessa. Seuraavan lauseen todistus jätetään siis harjoitustehtäväksi.

Lause 2.3.4. *Olkoon* $A \in \mathbb{R}^{m\times n}$ *ja* $k \in \mathbb{N}$. *Tällöin lineaarikuvauksen* $F_A\colon \mathbb{R}^{n\times k} \to \mathbb{R}^{m\times k}$, $X \mapsto AX$, *matriisi* $M_A \in \mathbb{R}^{(mk)\times(nk)}$ *avaruuksien* $\mathbb{R}^{m\times k}$ *ja* $\mathbb{R}^{n\times k}$ *standardikannoissa on ositettumatriisi*

$$M_A = \begin{bmatrix} A & 0 & \cdots & 0 & 0 \\ 0 & A & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & A & 0 \\ 0 & 0 & \cdots & 0 & A \end{bmatrix}.$$

Luku 3

Isometriat

3.1 Isometriset lineaarikuvaukset

Sisätuloavaruuksien välillä on luonnollista tarkastella lineaarikuvauksia, jotka säilyttävät pisteiden väliset etäisyydet.

Määritelmä 3.1.1. *Sisätuloavaruuksien* $(V, \langle\cdot,\cdot\rangle_V)$ *ja* $(W, \langle\cdot,\cdot\rangle_W)$ *välinen lineaarikuvas* $L\colon V \to W$ *on* pituuden säilyttä *eli* isometria*, jos*

$$\|L(v) - L(v')\|_W = \|v - v'\|_V$$

kaikilla $v, v' \in V$*.*

Huomautus 3.1.2. *Koska* $L\colon V \to W$ *on lineearikuvaus, niin etäisyyden säilyttäminen on yhtäpitävää vektorin pituuden säilyttämisen kanssa, eli että kaikilla* $v \in V$ *pätee*

$$\|L(v)\|_W = \|v\|_V.$$

Nimittäin, jos f *on pituuden säilyttävä lineearikuvaus, niin kaikilla* $v, v' \in V$ *pätee*

$$\|f(v) - f(v')\|_W = \|f(v - v')\|_W = \|v - v'\|_V.$$

Toisaalta, jos f *on etäisyyden säilyttävä lineaarikuvaus, niin silloin kaikilla* $v \in V$ *pätee*

$$\|f(v)\|_W = \|f(v) - f(0)\|_W = \|v - 0\|_V = \|v\|_V.$$

On erittäin syvällinen tulos (Mazurin ja Ulamin lause), että kaikki normiavaruuksien väliset surjektiiviset etäisyyden säilyttävät kuvaukset ovat affiineeja (eli siirtoa vaille lineaarisia), eli muotoa $v \mapsto L(v) + w$*, missä* L *on lineaarikuvaus ja* w *maaliavaruuden vektori.*

Isometrian määritelmä herättää kysymyksen, että miksi sisätuloavaruuksien yhteydessä ei pikemminkin tarkastella sisätulon säilyttäviä kuvauksia, eli lineaariaarikuvauksia $L\colon V \to W$, joille pätee $\langle L(v), L(v')\rangle_W = \langle v, v'\rangle_V$ kaikilla $v, v' \in V$. Syy on yksinkertainen: nämä ovat samoja kuvauksia.

Lemma 3.1.3. *Olkoot* $(V, \langle \cdot, \cdot \rangle_V)$ *ja* $(W, \langle \cdot, \cdot \rangle_W)$ *sisätuloavaruuksia. Lineaarikuvaus* $L\colon V \to W$ *on isometria, jos ja vain jos* $\langle L(v), L(v') \rangle_W = \langle v, v' \rangle_V$ *kaikilla* $v, v' \in V$.

Todistus. Oletetaan ensin, että $\langle L(v), L(v') \rangle_W = \langle v, v' \rangle_V$ kaikilla $v, v' \in V$. Tälllöin jokaisella $v \in V$ pätee

$$\|L(v)\|_W^2 = \langle L(v), L(v) \rangle_W = \langle v, v \rangle_V = \|v\|_V^2.$$

Oletetaan nyt, että $\|L(v)\|_W = \|v\|_V$ kaikilla $v \in V$. Nyt lemman A.1.1 perusteella

$$\begin{aligned}
\langle L(v), L(v') \rangle_W &= \frac{1}{2}\left(\|L(v) + L(v')\|_W^2 - \|L(v)\|_W^2 - \|L(v')\|_W^2\right) \\
&= \frac{1}{2}\left(\|L(v + v')\|_W^2 - \|L(v)\|_W^2 - \|L(v')\|_W^2\right) \\
&= \frac{1}{2}\left(\|v + v'\|_V^2 - \|v\|_V^2 - \|v'\|_V^2\right) = \langle v, v' \rangle.
\end{aligned}$$

Näin ollen lineaarikuvaus L säilyttää sisätulon ja väite on todistettu. $\square$

Lineaariset isometriat on helppo karakterioida ortonormaalien kantojen avulla.

Lause 3.1.4. *Olkoot* $(V, \langle \cdot, \cdot \rangle_V)$ *ja* $(W, \langle \cdot, \cdot \rangle_W)$ *sisätuloavaruuksia ja* $(v_1, \ldots, v_n)$ *avaruuden* V *ortonormaalikanta. Tällöin lineaarikuvaus* $L\colon V \to W$ *on isometria, jos ja vain jos vektorit* $L(v_1), \ldots, L(v_n)$ *ovat yksikkövektoreita, jotka ovat kohtisuorassa toisiaan vastaan, eli*

$$\langle L(v_i), L(v_j) \rangle_W = \delta_{ij}$$

kaikilla $i, j \in \{1, \ldots, n\}$.

Todistus. Oletetaan ensin, että L on isometria. Tällöin isometrian määritelmän nojalla $\|L(v_i)\|_W = \|v_i\|_V = 1$ kaikilla $i = 1, \ldots, n$. Vektorit $L(v_1), \ldots, L(v_n)$ ovat siis yksikkövektoreita. Olkoot nyt $i, j \in \{1, \ldots, n\}$. Koska lemman 3.1.3 perusteella L säilyttää sisätulon, niin pätee

$$\langle L(v_i), L(v_j) \rangle_W = \langle v_i, v_j \rangle_V = \delta_{ij}.$$

Oletetaan nyt, että vektorit $L(v_1), \ldots L(v_n)$ ovat yksikkövektoreita, jotka ovat kohtisuorassa toisiaan vastaan. Osoitetaan, että L on isometria.

Olkoon $v \in V$. Tällöin

$$L(v) = L(\langle v, v_1 \rangle v_1 + \cdots + \langle v, v_n \rangle v_n)) = \langle v, v_1 \rangle L(v_1) + \cdots + \langle v, v_n \rangle L(v_n) = \sum_{i=1}^{n} \langle v, v_i \rangle L(v_i).$$

Koska

$$\langle L(v_i), L(v_j) \rangle = \delta_{ij},$$

niin

$$\begin{aligned}
\|L(v)\|_W^2 = \langle L(v), L(v) \rangle_W &= \left\langle \sum_{i=1}^{n} \langle v, v_i \rangle_V L(v_i), \sum_{j=1}^{n} \langle v, v_j \rangle_V L(v_j) \right\rangle_W \\
&= \sum_{i,j=1}^{n} \langle v, v_i \rangle_V \langle v, v_j \rangle_V \langle L(v_i), L(v_j)_W \rangle = \sum_{i=1}^{n} \langle v, v_i \rangle_V^2 = \|v\|_V^2.
\end{aligned}$$

$\square$

Edellisen lauseen suorana seurauksena saadaan, että lineaarinen isomorfismi on isometria, jos ja vain jos se kuvaa ortonormaalin kannan ortonormaaliksi kannaksi.

Korollaari 3.1.5. *Olkoon $L\colon V \to W$ lineaarinen isomorfismi sisätuloavaruudesta $(V, \langle\cdot,\cdot\rangle_V)$ sisätuloavaruuteen $(W, \langle\cdot,\cdot\rangle_W)$ ja $(v_1, \ldots, v_n)$ avaruuden V ortonormaalikanta. Tällöin L on isometria, jos ja vain jos $(L(v_1), \ldots, L(v_n))$ on avaruuden W ortonormaalikanta.*

Todistus. Oletetaan ensin, että L on isometria. Koska L on isomorfismi, on $(L(v_1), \ldots, L(v_n))$ kanta. Koska L on isometria ja $(v_1, \ldots, v_n)$ on ortonormaalikanta, niin lauseen 3.1.4 vektorit $L(v_1), \ldots, L(v_n)$ ovat yksikkövektoreita, jotka ovat keskenään kohtisuorassa. Näin ollen $(L(v_1), \ldots, L(v_n))$ on ortonormaalikanta.

Oletetaan nyt, että $(L(v_1), \ldots, L(v_n))$ on avaruuden W ortonormaalikanta. Tällöin L on isometria lauseen 3.1.4 perusteella. $\square$

Koska lineearinen isometria saman ulotteisten vektoriavaruuksien välillä vie ortonormaalin kannan ortonormaaliksi kannaksi, niin erityisesti kuvaus on tällöin isomorfismi. Kirjataan tämä tulos vielä korollaariksi.

Korollaari 3.1.6. *Olkoon $L\colon V \to W$ äärellisulotteisten sisätuloavaruuksien $(V, \langle\cdot,\cdot\rangle_V)$ ja $(W, \langle\cdot,\cdot\rangle_W)$ välinen isometria. Jos $\dim V = \dim W$, niin L on isomorfismi.*

Korollaarin 3.1.5 avulla on myös helppo tunnistaa tuttuja, jettä jotkin tutut isomorfismit ovat isometrioita. Seuraava esimerkki on hyödyllinen tarkasteltaessa isometrioiden matriisiesityksiä.

Esimerkki 3.1.7. *Olkoon $(V, \langle\cdot,\cdot\rangle)$ sisätuloavaruus ja $(v_1, \ldots, v_n)$ sen kanta. Tällöin luvussa 1.1 määritelty isomorfismi $\Phi_V : \mathbb{R}^{n\times 1} \to V$,*

$$\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \mapsto x_1 v_1 + \cdots + x_n v_n,$$

on isometria sisätuloavaruudesta $(\mathbb{R}^{n\times 1}, \cdot)$ sisätuloavaruuteen $(V, \langle\cdot,\cdot\rangle)$. (Harjoitustehtävä)

Osoitetaan yleisen teorian sovelluksena tulos, että aliavaruuksien välinen isometria aina laajenee koko avaruuksien väliseksi isometriaksi.

Lause 3.1.8. *Olkoot $(V, \langle\cdot,\cdot\rangle_V)$ ja $(W, \langle\cdot,\cdot\rangle_W)$ n-ulotteisia sisätuloavaruuksia, olkoot $V_1 \subset V$ ja $W_1 \subset W$ aliavaruuksia ja $f_1\colon V_1 \to W_1$ isometria. Tällöin on olemassa isometria $f\colon V \to W$, jolle pätee $f|_{V_1} = f_1$.*

Todistus. Olkoon $k = \dim V_1$ ja $v_1, \ldots, v_k$ aliavaruuden V_1 ortonormaalikanta. Koska $f_1\colon V_1 \to W_1$ on isometria, niin se on isomorfismi. Erityisesti siis $\dim V_1 = \dim W_1$ ja $f(v_1), \ldots, f(v_k)$ on aliavaruuden W_1 kanta. Merkitään $w_i = f(v_i)$ jokaisella $i \in \{1, \ldots, k\}$. Koska f_1 on isometria, niin $w_1, \ldots, w_k$ on aliavaruuden W_1 ortonormaalikanta.

Koska $\dim V_1^{\perp} = \dim V - \dim V_1 = n - k$ ja $\dim W_1^{\perp} = \dim W - \dim W_1 = n - k$, niin on olemassa aliavaruuksien $V_1^{\perp} \subset V$ ja $W_1^{\perp} \subset W$ ortonormaalit kannat $v_{k+1}, \dots, v_n$ ja $w_{k+1}, \dots w_n$. Tälllöin $v_1, \dots, v_n$ on avaruuden V ortonormaalikanta ja $w_1, \dots, w_n$ on avaruuden W ortonormaalikanta (harjoitustehtävä).

Määritellään nyt $f \colon V \to W$ kaavalla $v_i \mapsto w_i$ jokaisella $i \in \{1, \dots, n\}$. Tällöin f on isometria. Koska $f(v_i) = w_i = f_1(v_i)$ jokaisella $i \in \{1, \dots, k\}$, niin $f|_{V_1} = f_1$. □

3.2 Ortogonaaliset matriisit

Sarakeavaruuksien väliset lineaarikuvaukset vastaavat luonnollisella tavalla matriiseja, eli jokainen lineaarikuvaus $L \colon \mathbb{R}^{n\times 1} \to \mathbb{R}^{m\times 1}$ on muotoa $x \mapsto Ax$, missä $A \in \mathbb{R}^{m\times n}$ on matriisi. Kuvaus L on siis kuvaus $f_A \colon \mathbb{R}^{n\times 1} \to \mathbb{R}^{m\times 1}$, $x \mapsto Ax$, kuten luvussa 1.1.

Lause 3.2.1. *Matriisin $A \in \mathbb{R}^{m\times n}$ määrittelemä lineaarikuvaus $f_A \colon \mathbb{R}^{n\times 1} \to \mathbb{R}^{m\times 1}$ on isometria avaruudesta $(\mathbb{R}^{n\times 1}, \cdot)$ avaruuteen $(\mathbb{R}^{m\times 1}, \cdot)$, jos ja vain jos $A^T A = I_{n\times n}$.*

Huomautus 3.2.2. *Lause saattaa vaikuttaa ensinäkemältä hyvin kryptiseltä. Huomaa kuitenkin, että matriisitulon ja pistetulon määritelmistä seuraa suoraan, että matriisille $A = \left[a_1 | \cdots | a_n\right] \in \mathbb{R}^{m\times n}$ pätee*

$$(A^T A)_{ij} = a_i^T a_j = a_i \cdot a_j \tag{3.1}$$

kaikilla $i, j \in \{1, \dots, n\}$, missä $\cdot$ on avaruuden $\mathbb{R}^{m\times 1}$ pistetulo. Matriisi $I_{n\times n}$ on identtisenkuvauksen $\mathrm{id} \colon \mathbb{R}^{n\times 1} \to \mathbb{R}^{n\times 1}$, *$x \mapsto x$, matriisi, eli matriisi, jolle pätee*

$$(I_{n\times n})_{ij} = \delta_{ij}$$

kaikilla $i, j \in \{1, \dots, n\}$.

Todistus. Havaitaan aluksi, että

$$A = \left[a_1 | \cdots | a_n\right] = \left[f_A(e_1) | \cdots | f_A(e_n)\right],$$

missä $(e_1, \dots, e_n)$ on avaruuden $\mathbb{R}^{n\times 1}$ standardikanta, joka on ortonormaali pistetulon $\cdot \colon \mathbb{R}^{n\times 1} \times \mathbb{R}^{n\times 1} \to \mathbb{R}$ suhteen. Lisäksi matriisitulon ja pistetulon määritelmien perusteella kaikilla $i, j \in \{1, \dots, n\}$ pätee

$$(A^T A)_{ij} = a_i^T a_j = (f_A(e_i))^T f_A(e_j) = f_A(e_i) \cdot f_A(e_j). \tag{3.2}$$

Oletetaan ensin, että f_A on isometria. Tällöin lauseen 3.1.4 nojalla pätee

$$f_A(e_i) \cdot f_A(e_j) = \delta_{ij}$$

kaikilla $i, j \in \{1, \dots, n\}$. Näin ollen yhtälöstä (3.2) seuraa, että

$$(A^T A)_{ij} = \delta_{ij}$$

kaikilla $i, j \in \{1, \dots, n\}$, eli että $A^T A$ on identiteettimatriisi $I_{n\times n}$.

Oletetaan nyt, että $A^T A = I_{n\times n}$. Tällöin yhtälöstä (3.2) puolestaan seuraa, että $f_A(e_i) \cdot f_A(e_j) = \delta_{ij}$ kaikilla $i, j \in \{1, \dots, n\}$. Kuvaus f_A on siis isometria lauseen 3.1.4 perusteella. □

Lauseen 3.2.1 matriiseja kutsutaan *ortogonaalisiksi.*

Määritelmä 3.2.3. *Matriisi* $A \in \mathbb{R}^{m\times n}$ *on* ortogonaalinen, *jos* $A^T A = I_{n\times n}$.

Suoraan määritelmästä havaitaan, että ortogonaaliset neliömatriisit ovat kääntyiviä. Seuraavan lemman todistus on kokonaisuudessaan olellisesti matriisin transpoosin tulo-ominaisuuden $(AB)^T = B^T A^T$ sovellus.

Lemma 3.2.4. *Olkoon* $A \in \mathbb{R}^{n\times n}$ *ortogonaalinen neliömatriisi. Tällöin* A *on kääntyvä,* $A^T = A^{-1}$ *ja käänteismatriisi* A^{-1} *on ortogonaalinen. Jos lisäksi* $B \in \mathbb{R}^{n\times n}$ *on ortogonaalinen, niin matriisi* AB *on ortogonaalinen.*

Todistus. Olkoot $f_A \colon \mathbb{R}^{n\times 1} \to \mathbb{R}^{n\times 1}$, $x \mapsto Ax$, ja $f_{A^T} \colon \mathbb{R}^{n\times 1} \to \mathbb{R}^{n\times 1}$, $y \mapsto A^T y$. Koska $f_{A^T} \circ f_A = \mathrm{id}$, niin kuvaus f_A on injektio. Koska kuvauksen f_A lähtö- ja maaliavaruus ovat sama avaruus $\mathbb{R}^{n\times 1}$ ja $\mathbb{R}^{n\times 1}$ on äärellisulotteinen, niin

$$\dim \operatorname{im} f_A = \dim \mathbb{R}^{n\times 1} - \dim \ker f_A = n - 0 = n.$$

Näin ollen f_A on surjektio. Koska f_A on sekä injektio että surjektio, niin f_A on isomorfismi ja matriisi A on kääntyvä. Näin ollen yhtälöstä $A^T A = I$ seuraa $A^T = A^{-1}$.

Matriisi A^{-1} on ortogonaalinen, koska

$$(A^{-1})^T A^{-1} = (A^T)^T A^{-1} = AA^{-1} = I.$$

Viimeinen väite seuraa myös suoraan transpoosin ominaisuuksista, sillä

$$(AB)^T AB = B^T A^T AB = B^T I_{n\times n} B = B^T B = I.$$

□

Huomautus 3.2.5. *Ensimmäisen päättelyn olisi voinut tehdä toisinkin. Ortogonalisen neliömatriisin* $A \in \mathbb{R}^{n\times n}$ *määrittelemä lineaarikuvaus* $f_A \colon \mathbb{R}^{n\times 1} \to \mathbb{R}^{n\times 1}$ *on isometria ja siten injektio. Itseasiassa kaikki lemman 3.2.4 väitteet voi päätellä käyttämällä isometrioita.*

Huomautus 3.2.6. *Koska identtinen matriisi* $I_{n\times n}$ *on selvästi ortogonaalinen, niin lemma 3.2.4 osoittaa, että ortogonaaliset neliömatriisit*

$$O(n) = \{A \in \mathbb{R}^{n\times n} \colon A^T A = I_{n\times n}\}$$

muodostavat ryhmän, niin sanotun ortogonaaliryhmän $O(n)$. *Ryhmistä lisää algebran kursseilla.*

Yhtälö (3.1) karakterisoi ortogonaaliset matriisit seuraavasti.

Lemma 3.2.7. *Matriisi* $A = \begin{bmatrix} a_1 | \cdots | a_n \end{bmatrix} \in \mathbb{R}^{m\times n}$ *on ortogonaalinen, jos ja vain jos sen sarakkeet* $a_1, \ldots, a_n$ *ovat yksikkövektoreita, jotka ovat kohtisuorassa toisiaan vastaan avaruudessa* $(\mathbb{R}^{m\times 1}, \cdot)$.

Todistus. Harjoitustehtävä. □

Esimerkki 3.2.8. *Kiertomatriisit*

$$R_\theta = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \in \mathbb{R}^{2\times 2}$$

ovat ortogonaalisia jokaisella $\theta \in \mathbb{R}$. *(Harjoitustehtävä: Selvitä miksi näitä matriiseja kutsutaan kiertomatriiseiksi.)*

Lemmojen 3.2.4 ja 3.2.7 avulla on helppo löytää kaikki ortogonaaliset 2×2-matriisit.

Esimerkki 3.2.9. *Olkoon*

$$A = \begin{bmatrix} a & c \\ b & d \end{bmatrix} \in \mathbb{R}^{2\times 2}$$

ortogonaalimatriisi. Koska $a^2 + b^2 = 1$, *niin* $|a| \leq 1$ *ja on olemassa sellainen* $\theta \in \mathbb{R}$, *että* $\cos\theta = a$. *Tällöin joko* $b = \sin\theta$ *tai* $b = -\sin\theta = \sin(-\theta)$. *Koska* $\cos(-\theta) = \cos(\theta)$, *niin voidaan olettaa, että* $\cos\theta = a$ *ja* $\sin\theta = b$.

Olkoon nyt R_θ *kiertomatriisi kuten esimerkkissä 3.2.8. Koska*

$$R_\theta e_1 = \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix} = \begin{bmatrix} a \\ b \end{bmatrix},$$

niin

$$R_\theta^{-1} \begin{bmatrix} a \\ b \end{bmatrix} = e_1.$$

Näin ollen

$$R_\theta^{-1} A = \begin{bmatrix} 1 & c' \\ 0 & d' \end{bmatrix},$$

missä

$$\begin{bmatrix} c' \\ d' \end{bmatrix} = R_\theta^{-1} \begin{bmatrix} c \\ d \end{bmatrix}.$$

Koska lemman 3.2.4 perusteella matriisi

$$\begin{bmatrix} 1 & c' \\ 0 & d' \end{bmatrix}$$

on ortogonaalimatriisi, niin tiedetään sekä $(c')^2 + (d')^2 = 1$ *että* $1 \cdot c' + 0 \cdot d' = 0$. *Näin ollen* $c' = 0$ *ja* $|d'| = 1$ *eli pätee joko*

$$R_\theta^{-1} A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

tai

$$R_\theta^{-1} A = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

Ensimmäisessä tapauksessa

$$A = R_\theta = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}$$

ja jälkimmäisessa tapauksessa

$$A = R_\theta \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{bmatrix} = \begin{bmatrix} a & b \\ b & -a \end{bmatrix}.$$

Ortogonaaliset 2×2*-matriisit voidaan siis luokitella ensimmäisen sarakkeen ja determinantin avulla.*

3.3 Sovellus: Matriisin QR-hajotelma

Ortonormaalin kannan avulla, mikä $m \times n$-matriisi, jonka sarakeavaruuden dimsion on n, voidaan kirjoittaa tulona ortogonaalisesta matriisista ja yläkolmiomatriisista. Tätä kutsutaan matriisin QR-hajotelmaksi. Hajotelma on tärkeä sekä käytännön sovelluksissa että teoreettisesti, kuten tulemme havaitsemaan.

Lause 3.3.1. *Olkoon* $A \in \mathbb{R}^{m\times n}$ *sellainen matriisi, että* $\dim \mathrm{Col}(A) = n$*. Tällöin on olemassa sellainen ortogonaalimatriisi* $Q \in \mathbb{R}^{m\times n}$ *ja sellainen yläkolmiomatriisi* $R \in \mathbb{R}^{n\times n}$*, että*

$$A = QR.$$

Huomautus 3.3.2. *Matriisi* Q *ei ole yksikäsitteinen, mutta ei myöskään mielivaltainen. Matriisit* Q *vastaavat sarakeavaruuden* $\mathrm{Col}(A)$ *ortonormaaleja kantoja. Tarkemmin sanottuna matriisin* Q *sarakkeet muodostavat avaruuden* $\mathrm{Col}(A)$ *ortonormaalin kannan. Todistus osoittaa lisäksi, että matriisin* R *diagonaalielementit ovat nollasta poikkeavia. Se on siis kääntyvä neliömatriisi.*

Todistus. Olkooon $A = \begin{bmatrix} a_1 | \cdots | a_n \end{bmatrix} \in \mathbb{R}^{m\times n}$.

Koska sarakeavaruuden $\mathrm{Col}(A)$ dimensio on matriisin A sarakkeiden lukumäärä, niin matriisin A sarakevektorit $a_1, \ldots, a_n$ muodostavat sarakeavaruuden $\mathrm{Col}(A)$ kannan.

Hyödynnetään nyt avaruuden $\mathrm{Col}(A)$ luonnollista sisätuloa, eli pistetulon $\cdot\colon \mathbb{R}^{m\times 1} \times \mathbb{R}^{m\times 1}$ rajoittumaa aliavaruuteen $\mathrm{Col}(A)$. Koska $a_1, \ldots, a_n$ on avaruuden $\mathrm{Col}(A)$ kanta, niin Gram–Schmidt ortonormeerausprosessi antaa avaruuden $\mathrm{Col}(A)$ ortonormaalin kannan $v_1, \ldots, v_n$. Olkoon nyt Q näiden sarakevektoreiden muodostama matriisi $Q \in \mathbb{R}^{m\times n}$, eli $Q = \begin{bmatrix} v_1 | \cdots | v_n \end{bmatrix}$.

Yläkolmiomatriisi $R \in \mathbb{R}^{n\times n}$ syntyy nyt kannanvaihtomatriisina uudesta kannasta vanhaan kantaan. Havaitaan ensin, että Gram–Schmidt prosessi on sellainen, että $\mathrm{Sp}\{a_1, \ldots, a_j\} = \mathrm{Sp}\{v_1, \ldots, v_j\}$ jokaisella $j \in \{1, \ldots, n\}$, eli jokaisella $j \in \{1, \ldots, n\}$ on olemassa sellaiset $r_{j,1}, \ldots, r_{j,j} \in \mathbb{R}$, että

$$a_j = r_{j,1}v_1 + \cdots + r_{j,j}v_j = r_{j,1}v_1 + \cdots + r_{j,j}v_j + 0v_{j+1} + \cdots + 0v_n,$$

missä $r_{j,j} \neq 0$.

Olkoon nyt $r_j \in \mathbb{R}^{n\times 1}$ vektori

$$r_j = \begin{bmatrix} r_{j,1} \\ \vdots \\ r_{j,j} \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

Tällöin

$$a_j = Qr_j$$

Muodostetaan nyt sarakevektoreista $r_1, \ldots, r_k$ matriisi $R \in \mathbb{R}^{k\times n}$, eli asetetaan

$$R = \Big[r_1 | \cdots | r_k \Big].$$

Nyt matriisitulon laskusääntöjen perusteella pätee, että

$$A = \left[\ a_1 \mid \cdots \mid a_n\ \right] = \left[\ Qr_1 \mid \cdots \mid Qr_n\ \right] = Q \left[\ r_1 \mid \cdots \mid r_n\ \right] = QR.$$

□

Huomautus 3.3.3. *QR-hajotelma on hyödyllinen myös sellaisissa tapauksissa, missä matriisi A ei ole neliömatriisi. Tälläisessä tapauksessa QR-hajotelma voidaan muodostaa havaitsemalla, että matriisin A sarakkeet virittävät avaruuden* $\mathrm{Col}(A)$*, joten sarakkeiden joukosta voidaan valita avaruuden* $\mathrm{Col}(A)$ *kanta.*

3.3.1 Yleinen tapaus

Tarkastellaan nyt yleistä matriisia $A \in \mathbb{R}^{m\times n}$, joka ei siis ole täyttä sarakeastetta, eli $\operatorname{rank} A < n$.

Palautetaan mieleen luvusta 1.5.2, että on olemassa sellaiset täyttä sarakeastetta oleva matriisi $P \in \mathbb{R}^{m\times k}$ ja yläkolmiomatriisi $T \in \mathbb{R}^{k\times n}$, että

$$A = PT.$$

Koska P on täyttä sarakeastetta, niin sillä on lauseen 3.3.1 mukainen QR-hajotelma, eli on olemassa sellaiset ortogonaalimatriisi $Q \in \mathbb{R}^{m\times k}$ ja yläkolmiomatriisi $R \in \mathbb{R}^{k\times n}$, että

$$P = QR.$$

Koska yläkolmiomatriisien tulo on yläkolmiomatriisi, niin matriisi $Y = RT$ on yläkolmiomatriisia. Nyt

$$A = PT = QRT = QY.$$

Luku 4

Projektiot

4.1 Yleiset projektiot

Projektiolla tarkoitetaan useimmiten ortogonaalista projektiota, mutta projektio voidaan määritellä myös täysin yleisesti ilman sisätuloa.

Määritelmä 4.1.1. *Lineaarikuvaus $P\colon V \to V$ on* projektio*, jos $P \circ P = P$.*

Koska projektio $P\colon V \to V$ on lineaarikuvaus, niin sen kuva $\operatorname{im} P = \{Pv\colon v \in V\}$ on avaruuden V aliavaruus. Näin ollen voidaan määritellä, että jos W on aliavaruus ja $W = \operatorname{im} P$, niin $P\colon V \to V$ on *projektio aliavaruudelle W*.

Seuraava havainto on projektion määritelmän uudelleen karakterisointi.

Lemma 4.1.2. *Jos $P\colon V \to V$ on projektio aliavaruudelle W, niin $P|_W = \mathrm{id}_W$, eli $P(w) = w$ kaikilla $w \in W$, ja $V = \ker P \oplus W$. Toisaalta, jos $P\colon V \to V$ on lineaarikuvaus, jolle pätee $P|_W = \mathrm{id}_W$ ja $V = \ker P \oplus W$, niin P on projektio aliavaruudelle W.*

Todistus. Olkoon $w \in W$. Koska $W = \operatorname{im} P$, niin on olemassa sellainen $v \in V$, että $P(v) = w$. Tällöin $P(w) = P(P(v)) = (P \circ P)(v) = P(v) = w$. Tämä todistaa ensimmäisen ominaisuuden. Toinen ominaisuus seuraa kahdesta havainnosta. Havaitaan ensin, että jokaisella $v \in V$ pätee $P(v - P(v)) = P(v) - P(P(v)) = P(v) - P(v) = 0$, eli $v - P(v) \in \ker P$. Näin ollen $v = v - P(v) + P(v) = \ker P + \operatorname{im} P = \ker P + W$, eli $V = \ker P + W$. Toiseksi havaitaan, että $\ker P \cap \operatorname{im} P = \{0\}$, sillä jokaisella $v \in \ker P \cap \operatorname{im} P$ pätee $P(v) = v$ ensimmäisen kohdan nojalla. Näin ollen $V = \ker P \oplus W$.

Osoitetaan nyt toinen väite. Olkoon $P\colon V \to V$ sellainen lineaarikuvaus, että $P|_W = \mathrm{id}_W$ ja $V = \ker P \oplus W$. Osoitetaan, että $P \circ P = P$. Olkoon $v \in V$. Tällöin on olemassa yksikäsitteiset vektorit $v' \in \ker P$ ja $w \in W$, joille pätee $v = v' + w$. Nyt $P(v) = P(v' + w) = P(v') + P(w) = P(w)$, koska $v' \in \ker P$. Koska $P|_W = \mathrm{id}_W$, niin $P(w) = w$ ja saadaan $P(P(v)) = P(P(w)) = P(w) = w = P(v)$. $\square$

Lemman 4.1.2 avulla on helppo määritellä projektioita käyttäen kantoja.

Esimerkki 4.1.3. *Olkoon V äärellisulotteinen vektoriavaruus ja $(v_1, \dots, v_n)$ avaruuden V kanta. Valitaan $1 \le k \le n$ ja määritellään lineaarikuvaus $P\colon V \to V$ kanta-alkiota käyttäen kaavalla*

$$v_i \mapsto \begin{cases} v_i, & i \le k \\ 0, & i > k \end{cases}$$

Tällöin jokaisella $i \in \{1, \dots, k\}$ pätee $(P \circ P)(v_i) = P(P(v_i)) = P(v_i)$ ja jokaisella $i \in \{k+1, \dots, n\}$ pätee $(P \circ P)(v_i) = P(P(v_i)) = P(0) = 0 = P(v_i)$. Näin ollen $P \circ P = P$ eli P on projektio. Tässä tapauksessa projektion P kuva on aliavaruus $W = \mathrm{Sp}\{v_1, \dots, v_k\}$. Tämän projektion matriisiesitys, eli kuvauksen P matriisi A on ositettumatriisi

$$A = \begin{bmatrix} I_{k \times k} & 0 \\ 0 & 0 \end{bmatrix}.$$

Esimerkki 4.1.4. *Edellisen esimerkin projektio $P\colon V \to V$ on yleistys usemman muuttujan analyysistä tutusta* koordinaattiprojektiosta *$\pi_i\colon \mathbb{R}^n \to \mathbb{R}$, $(x_1, \dots, x_n) \mapsto x_i$, kun avaruudeksi V valitaan euklidinen avaruus $\mathbb{R}^n$ ja kannaksi standardikanta $(e_1, \dots, e_n)$. Huomaa, että koordinaattiprojektiot eivät ole formaalisti projektioita määritelmän 4.1.1 mielessä, koska niillä on eri lähtö- ja maaliavaruus.*

4.2 Ortogonaaliprojektiot

Sisätuloavaruudessa voidaan tarkastella tutumpia ortogonaaliprojektioita.

Määritelmä 4.2.1. *Sisätuloavaruuden $(V, \langle \cdot, \cdot \rangle)$ projektio $P\colon V \to V$ on* ortogonaaliprojektio, *jos jokaisella $v \in V$ erotusvektori $v - P(v)$ on kohtisuorassa kuva-avaruutta $\operatorname{im} P$ vastaan, eli kaikilla $v, v' \in V$ pätee $\langle v - P(v), P(v') \rangle = 0$.*

Huomautus 4.2.2. *Määritelmän voi kirjoittaa toisinkin käyttäen lemman 4.1.2 tietoa, että jokaisella projektiolle $P\colon V \to V$ pätee $V = \ker P \oplus \operatorname{im} P$, josta seuraa, että $v - P(v) \in \ker P$. Näin ollen projektio P on ortogonaalinen, jos ja vain jos sen ydin $\ker P$ on kohtisuorassa kuva-avaruutta $\operatorname{im} P$ vastaan, eli $V = \ker P \oplus \operatorname{im} P$ on ortogonaalisten avaruuksin suora summa.*

Ortogonaaliprojektion määritelmä voidaan antaa yhtäpitävästi myös toisin. Kirjataan tämäkin lauseeksi. Asiaan palataan vielä myöhemmin toisessa yhteydessä.

Lause 4.2.3. *Olkoon $(V, \langle \cdot, \cdot \rangle)$ sisätuloavaruus ja $P\colon V \to V$ projektio. Tällöin P on ortogonaaliprojektio, jos ja vain jos P on itseadjungoitu, eli kaikilla $v, v' \in V$ pätee*

$$\langle P(v), v' \rangle = \langle v, P(v') \rangle. \tag{4.1}$$

Todistus. Oletetaan ensin, että P on ortogonaaliprojektio. Olkoot $v, v' \in V$. Tällöin

$$\langle P(v), v' \rangle = \langle P(v), v' - P(v') \rangle + \langle P(v), P(v') \rangle = \langle P(v), P(v') \rangle.$$

Vastaavasti $\langle v, P(v') \rangle = \langle P(v), P(v') \rangle$. Väite seuraa.

Oletetaan nyt, että projektio P toteuttaa ehdon (4.1) kaikilla $v, v' \in V$. Olkoot $v, v' \in V$. Tällöin

$$\langle v - P(v), P(v')\rangle = \langle P(v - P(v)), v'\rangle = \langle P(v) - P(P(v)), v'\rangle = \langle P(v) - P(v), P(v')\rangle = 0.$$

□

Koska ortogonaaliprojektio on itseadjungoitu, niin sen kuva ja ydin ovat kohtisuorassa toisiaan vastaan. Kirjataan tämä tärkeä huomio lauseeksi.

Lause 4.2.4. *Olkoon $(V, \langle\cdot,\cdot\rangle)$ sisätuloavaruus ja $P\colon V \to V$ ortogonaaliprojektio. Tällölin aliavaruudet $\ker P$ ja $\operatorname{im} P$ ovat kohtisuorassa toisiaan vastaan, eli $\langle v, w\rangle = 0$ kaikilla $v \in \ker P$ ja $w \in \operatorname{im} P$. Erityisesti $\ker P = (\operatorname{im} P)^\perp$.*

Todistus. Olkoot $v \in \ker P$ ja $w \in \operatorname{im} P$. Tällöin $w = P(w)$. Näin ollen

$$\langle v, w\rangle = \langle v, P(w)\rangle = \langle P(v), w\rangle = \langle 0, w\rangle = 0.$$

Osoitetaan nyt toinen väite. Edellisen kohdan nojalla $\ker P \subset (\operatorname{im} P)^\perp$. Olkoon nyt $v \in (\operatorname{im} P)^\perp$. Tällöin $\langle v, P(v)\rangle = 0$. Koska P on ortogonaaliprojektio, niin $\langle P(v) - v, P(v)\rangle = -\langle v - P(v), P(v)\rangle = 0$. Koska

$$\langle P(v), P(v)\rangle = \langle P(v) - v, P(v)\rangle + \langle v, P(v)\rangle = 0,$$

niin $P(v) = 0$ ja $v \in \ker P$. Näin ollen $(\operatorname{im} P)^\perp \subset \ker P$. □

Esimerkin 4.1.3 projektio $P\colon V \to V$ on ortogonaaliprojektio, jos kanta $(v_1, \ldots, v_n)$ on ortonormaali. Kirjataan tämä tärkeä havainto lemmaksi.

Lemma 4.2.5. *Olkoon $(V, \langle\cdot,\cdot\rangle)$ äärellisulotteinen sisätuloavaruus, $(v_1, \ldots, v_n)$ avaruuden V ortonormaalikanta ja $1 \le k \le n$. Tällöin kuvaus $P\colon V \to V$, joka on määritelty kannan avulla kaavalla*

$$v_i \mapsto \begin{cases} v_i, & i \le k \\ 0, & i > k \end{cases}$$

on ortogonaaliprojektio.

Todistus. Esimerkin 4.1.3 perusteella lineaarikuvaus P on projektio. Riittää siis osoittaa, että $v - P(v)$ on kohtisuorassa kuva-avaruutta $\operatorname{im} P$ vastaan kaikilla $v \in V$.

Olkoon $v = x_1 v_1 + \cdots + x_n v_n \in V$. Tällöin

$$\begin{aligned} v - P(v) &= x_1 v_1 + \cdots + x_n v_n - P(x_1 v_1 + \cdots + x_n v_n) \\ &= x_1 v_1 + \cdots x_k v_k + x_{k+1} v_{k+1} + \cdots + x_n v_n - (x_1 v_1 + \cdots + x_k v_k) \\ &= x_{k+1} v_{k+1} + \cdots + x_n v_n. \end{aligned}$$

Näin ollen jokaisella $j \in \{1, \ldots, k\}$ pätee

$$\langle v - P(v), v_j\rangle = \langle x_{k+1} v_{k+1} + \cdots + x_n v_n, v_j\rangle = \sum_{i=k+1}^{n} x_i \langle v_i, v_j\rangle = 0.$$

Vektori $v - P(v)$ on siis kohtisuorassa kuva-avaruutta $\operatorname{im} P$ vastaan. □

Edellinen lemma sanoo, että ortogonaaliprojektiota on helppo löytää ortonormaalijen kantojen avulla. Seuraava lause sanoo, että ortogonaaliprojektio kiinnitettylle aliavaruudelle on yksikäsitteinen ja tämä ortogonaaliprojektio voidaan ilmaista sisätulon avulla. Tämä tarkoittaa sitä, että eri menetelmät projektion löytämiseen johtavat aina samaan lopputulokseen, eli lemman 4.2.5 projektioon.

Lause 4.2.6. *Olkoon $(V, \langle\cdot,\cdot\rangle)$ äärellisulotteinen sisätuloavaruus ja $W \subset V$ aliavaruus. Tällöin on olemassa täsmälleen yksi ortogonaaliprojektio $P\colon V \to V$ aliavaruudelle W. Lisäksi, jos $(v_1, \ldots, v_k)$ on aliavaruuden W ortonormaalikanta, niin*

$$P(v) = \sum_{j=1}^{k} \langle v_j, v\rangle v_j \tag{4.2}$$

kaikilla $v \in V$.

Todistus. Osoitetaan ensin, että kaavalla (4.2) annettu lineaarikuvaus $P\colon V \to V$ on projektio avaruudelle P. Olkoon $1 \leq i \leq k$. Tällöin $P(v_i) = \sum_{j=1}^{k}\langle v_j, v_i\rangle v_j = \sum_{j=1}^{k} \delta_{ij} v_j = v_i$. Olkoon nyt $k < i \leq n$. Tällöin $P(v_i) = \sum_{j=1}^{k}\langle v_j, v_i\rangle v_j = 0$. Näin ollen P on lemman 4.2.5 lineaarikuvaus ja siten ortogonaaliprojektio avaruudelle W.

Olkoon nyt $Q\colon V \to V$ jokin ortogonaaliprojektio avaruudelle W ja osoitetaan, että pätee $Q = P$. Havaitaan aluksi, että $\operatorname{im} Q = W = \operatorname{im} P$ ja $Q|_W = \mathrm{id} = P|_W$. Toisaalta lauseen 4.2.4 perusteella pätee $\ker Q = W^\perp = \ker P$. Osoitetaan, että näistä havainnoista seuraa $Q = P$. Olkoon $v \in V$. Koska $V = W \oplus W^\perp$, niin on olemassa sellaiset $w \in \operatorname{im} P$ ja $w' \in W^\perp$, että $v = w + w'$. Näin ollen

$$Q(v) = Q(w + w') = Q(w) + Q(w') = Q(w) = P(w) = P(w) + P(w') = P(v).$$

□

4.3 Sovellus: Pisteen etäisyys aliavaruudesta

Kohtisuoruus on sisäänkirjoitettu ortogonaaliprojektion määritelmään, mutta sillä on syvempi merkitys: jos $P\colon V \to V$ on projektio aliavaruudelle W, niin kuvavektori $P(v)$ minimoi vektorin v etäisyyden projektion kuvaan W. Kirjataan tämä tulos lauseeksi sen tärkeyden vuoksi.

Lause 4.3.1. *Olkoon $(V, \langle\cdot,\cdot\rangle)$ ja $P\colon V \to V$ ortogonaaliprojektio aliavaruudelle W. Tällöin kaikilla $v \in V$ ja $w \in W \setminus \{P(v)\}$ pätee*

$$\|v - P(v)\| < \|v - w\|.$$

Erityisesti kaikilla $v \in V$ pätee

$$\|v - P(v)\| = \min\{\|v - w\| \colon w \in W\}.$$

Todistus. Olkoon $v \in V$ ja $w \in W$. Koska $W = \operatorname{im} P$, niin on olemassa sellainen $v' \in V$, että $P(v') = w$. Näin ollen

$$\langle v - P(v), w \rangle = \langle v - P(v), P(v') \rangle = 0$$

ja

$$\begin{aligned}
\|v - w\|^2 &= \|v - P(v) + P(v) - w\|^2 = \|v - P(v) + P(v) - P(v')\|^2 \\
&= \|v - P(v) + P(v - v')\|^2 = \langle v - P(v) + P(v - v'), v - P(v) + P(v - v') \rangle \\
&= \|v - P(v)\|^2 + 2\langle v - P(v), P(v - v') \rangle + \|P(v - v')\|^2 \\
&= \|v - P(v)\|^2 + \|P(v - v')\|^2 = \|v - P(v)\|^2 + \|P(v) - w\|^2 > \|v - P(v)\|^2.
\end{aligned}$$

□

4.4 Ortogonaaliprojektion matriisi

Käytännössä projektiokuvaus $P \colon V \to V$ aliavaruudelle W halutaan aina esittää matriisimuodossa. Yleisin käyttötilanne on, että on annettu aliavaruuden $W \subset \mathbb{R}^{n \times 1}$ kanta $a_1, \ldots, a_k$ ja halutaan kirjoittaa projektio P avaruuden $\mathbb{R}^{n \times 1}$ standardikannassa $(e_1, \ldots, e_n)$ matriisin $A = \begin{bmatrix} a_1 | \cdots | a_k \end{bmatrix}$ avulla. Jos kanta $a_1, \ldots, a_k$ on ortonormaali, haluttu matriisi löytyy helposti lauseen 4.2.6 avulla. Jos taas kanta $a_1, \ldots, a_k$ ei ole ortonormaali, niin asiaa kannattaa lähestyä toisella tavalla. Tämän vuoksi käsittely jaetaan näihin kahteen tapaukseen.

4.4.1 Ortonormaali kanta

Seuraava lause antaa matriisiesityksen projektiolle ortonormaalin kannan tapauksessa. Huomaa, että sarakeavaruudessa on sisätulona aina pistetulo ellei toisin mainita.

Lause 4.4.1. *Olkoon $A = \begin{bmatrix} a_1 | \cdots | a_k \end{bmatrix} \in \mathbb{R}^{n \times k}$ ortogonaalimatriisi. Tällöin $P \colon \mathbb{R}^{n \times 1} \to \mathbb{R}^{n \times 1}$,*

$$x \mapsto AA^T x,$$

on ortogonaaliprojektio sarakkeiden virittämälle avaruudelle

$$\operatorname{Col}(A) = \operatorname{Sp}\{a_1, \ldots, a_k\}.$$

Todistus. Koska vektorit $a_1, \ldots, a_k$ ovat ortogonaalisia keskenään, ovat ne lineaarisesti riippumattomia, eli $(a_1, \ldots, a_k)$ on aliavaruuden $\operatorname{Col}(A)$ kanta ja vieläpä ortonormaalikanta. Lauseen 4.2.6 perusteella ortogonaaliprojektiolle $P \colon \mathbb{R}^{n \times 1} \to \mathbb{R}^{n \times 1}$ pätee

$$P(x) = \sum_{j=1}^{k} (a_j \cdot x) a_j$$

kaikilla $x \in \mathbb{R}^{n\times 1}$. Olkoon $x \in \mathbb{R}^{n\times 1}$. Matriisituloa käyttäen saadaan

$$P(x) = \sum_{j=1}^{k} (a_j \cdot x) a_j = \sum_{j=1}^{k} (a_j^T x) a_j = \sum_{j=1}^{k} a_j (a_j^T x) = A \begin{bmatrix} a_1^T x & \cdots & a_k^T x \end{bmatrix} = AA^T x.$$

□

4.4.2 Yleinen kanta

Projektion matriisin löytäminen on yleissä tapauksessa haastavampaa kuin ortonormaalin kannan tapauksessa. Ratkaisun ytimessä on seuraava yleisiin sisätuloihin perustuva havainto.

Lause 4.4.2. *Matriisin $A \in \mathbb{R}^{m\times n}$ sarakkeet ovat lineaarisesti riippumattomia, jos ja vain jos matriisi $A^T A$ on kääntyvä.*

Todistus. Oletetaan ensin, että matriisin A sarakkeet ovat lineaarisesti riippumattomia. Määritellään $\langle \cdot, \cdot \rangle \colon \mathbb{R}^{n\times 1} \times \mathbb{R}^{n\times 1} \to \mathbb{R}$ kaavalla $\langle x, y \rangle = (Ax) \cdot (Ay)$ kaikilla $x, y \in \mathbb{R}^{n\times 1}$. Tällöin $\langle \cdot, \cdot \rangle$ on sisätulo avaruudessa $\mathbb{R}^{n\times 1}$ (Harjoitustehtävä). Koska $A^T A \in \mathbb{R}^{n\times n}$, niin se on kääntyvä, jos yhtälöllä $A^T Ax = 0$ on ainoastaan triviaaliratkaisu $x = 0$. Olkoon $x \in \mathbb{R}^{n\times 1}$ yhtälön $A^T Ax = 0$ ratkaisu. Tällöin

$$0 = x^T (A^T Ax) = (Ax)^T Ax = (Ax) \cdot (Ax) = \langle x, x \rangle,$$

eli $x = 0$. Näin ollen $A^T A$ on kääntyvä.

Oletetaan nyt, että $A^T A$ on kääntyvä. Osoitetaan, että f_A on injektio. Olkoon $x \in \mathbb{R}^{n\times 1}$ sellainen, että $Ax = 0$. Tällöin $A^T Ax = 0$. Koska $A^T A$ on kääntyvä, niin $x = 0$. □

Huomautus 4.4.3. *Itseasiassa hieman enemmän on totta: Olkoon $A \in \mathbb{R}^{m\times n}$. Tällöin $Ax = 0$, jos ja vain jos $A^T Ax = 0$. (Harjoitustehtävä.)*

Huomautus 4.4.4. *Lineaarikuvausten terminologiaa käyttäen voidaan tulos muotoilla seuraavasti: Matriisin $A \in \mathbb{R}^{m\times n}$ määräämä lineaarikuvaus $f_A \colon \mathbb{R}^{n\times 1} \to \mathbb{R}^{m\times 1}$, $x \mapsto Ax$, on injektio, jos ja vain jos lineaarikuvaus $f_{A^T} \circ f_A \colon \mathbb{R}^{n\times 1} \to \mathbb{R}^{n\times 1}$, $x \mapsto A^T Ax$, on isomorfismi, missä $f_{A^T} \colon \mathbb{R}^{m\times 1} \to \mathbb{R}^{n\times 1}$ on kuvauksen $f_A \colon \mathbb{R}^{n\times 1} \to \mathbb{R}^{m\times 1}$ adjungaatti f_A^*. Huomaa myös, että mikäli matriisin A sarakkeet ovat kohtisuorassa toisiaan vastaan olevia yksikkövektoreita, niin $A^T A = I_{n\times n}$.*

Lause 4.4.5. *Olkoon $A \in \mathbb{R}^{m\times n}$ sellainen matriisi, jonka sarakkeet ovat lineaarisesti riippumattomia. Tällöin ortogonaaliprojektio $P \colon \mathbb{R}^{m\times 1} \to \mathbb{R}^{m\times 1}$ aliavaruudelle* $\mathrm{Col}(A)$ *on lineaarikuvaus*

$$y \mapsto A(A^T A)^{-1} A^T y.$$

Todistus. Lauseen 4.4.2 perusteella matriisi $A^T A$ on kääntyvä. Osoitetaan, että kuvauksella P on matriisi $A(A^T A)^{-1} A^T$. Olkoot $a_1, \dots, a_n \in \mathbb{R}^{m\times 1}$ matriisin A sarakkeet, eli $A = \Big[a_1 | \cdots | a_n \Big]$.

Olkoon $y \in \mathbb{R}^{m\times 1}$. Koska $y - P(y)$ on kohtisuorassa aliavaruutta $\mathrm{Col}(A)$ vastaan, niin $(y-P(y))\cdot a_i = 0$ kaikilla $i \in \{1,\ldots,n\}$, eli $a_i^T(y-P(y)) = 0$ kaikilla $i \in \{1,\ldots,n\}$. Näin ollen $A^T(y - P(y) = 0$ eli $A^Ty = A^TP(y)$.

Koska $P(y) \in \mathrm{Col}(A)$ eli $P(y) \in \mathrm{im}\, f_A$, niin on olemassa sellainen $x \in \mathbb{R}^{n\times 1}$, että $P(y) = Ax$. Näin ollen

$$A^TAx = A^Ty.$$

Koska matriisi A^TA on kääntyvä, niin

$$x = (A^TA)^{-1}A^Ty.$$

Näin ollen

$$P(y) = Ax = A(A^TA)^{-1}A^Ty.$$

Väite on todistettu. □

Huomautus 4.4.6. *Jos olettaa matriisin A^TA kääntyvyyden, niin on helppo havaita suoralla laskulla, että annettu kaava määrittelee projektion aliavaruudelle $\mathrm{Col}(A)$: riittää kertoa matriisi $A(A^TA)^{-1}A^T$ itsellään. Sopivalla avaruuden $\mathbb{R}^{m\times 1}$ ortonormaalin kannan valinnalla pystyy myös helposti osoittamaan, että kaava määrittelee ortogonaaliprojektion. Vaikeus on siis lopulta matriisin A^TA kääntyvyydessä ja matriisin $A(A^TA)^{-1}A^T$ löytämisessä.*

Matriisin A QR-hajotelman avulla voidaan kohtisuoran projektion matriisia analysoida pidemmälle ja löytää geometrinen syy projektiomatriisin muodolle. Olkoon $A = QR$, missä $Q \in \mathbb{R}^{m\times n}$ on ortogonaalimatriisi ja $R \in \mathbb{R}^{n\times n}$ yläkolmiomatriisi. Tällöin

$$A^TA = (QR)^TQR = R^TQ^TQR = R^TR.$$

Koska matriisi R on kääntyvä neliömatriisi, niin saadaan

$$\det(A^TA) = \det(R^TR) = \det(R^T)\det(R) = \det(R)^2 \neq 0.$$

Näin ollen matriisin A^TA kääntyvyys voidaan havaita suoraan QR-hajotelmasta.

Koska matriisit R ja R^T ovat kääntyviä, saadaan lisäksi

$$A(A^TA)^{-1}A^T = QR(R^TR)^{-1}R^TQ^T = QRR^{-1}(R^T)^{-1}R^TQ^T = QQ^T.$$

Huomaa, että tämä tieto seuraa epäsuorasti jo lauseista 4.4.1 ja 4.2.6, koska matriisin Q sarakkeet muodostavat avaruuden $\mathrm{Col}(A)$ ortonormaalin kannan. QR-hajotelma antaa kuitenkin suoran yhteyden projektion eri matriisiesitysten välille.

4.4.3 Yleinen tapaus

Tarkastellaan vielä projektiota avaruudelle $\mathrm{Col}(A)$ tilanteessa, että matriisi $A \in \mathbb{R}^{m\times n}$ ei ole täyttä sarakeastetta, eli $k = \mathrm{rank}\, A < n$. Tällöin voidaan hyödyntää matriisin A QR-hajotelmaa $A = QR$, missä $Q \in \mathbb{R}^{m\times k}$ on ortogonaalimatriisi ja $R \in \mathbb{R}^{k\times n}$ on yläkolmiomatriisi.

Matriisi Q on täyttä sarakeastetta ja sille pätee

$$\mathrm{Col}(A) = \mathrm{Col}(Q).$$

Näin ollen haluttu ortonormaaliprojektio $P\colon \mathbb{R}^{m\times 1} \to \mathbb{R}^{m\times 1}$ sarakeavaruudelle $\mathrm{Col}(A)$ on itseasiassa lineaarikuvaus $y \mapsto QQ^T y$.

Huomautus 4.4.7. *Palautetaan mieleen, että yleinen QR-hajotelma perustuu havainnolle, että $A = BT$, missä $B \in \mathbb{R}^{m\times k}$ on täyttä sarakeastetta ja $T \in \mathbb{R}^{k\times n}$ on yläkolmiomatriisi. Näin ollen ortonormaaliprojektio $P\colon \mathbb{R}^{m\times 1} \to \mathbb{R}^{m\times 1}$ saadaan myös kaavasta $y \mapsto B(B^TB)^{-1}B^Ty$.*

4.5 Sovellus: Pienimmän neliösumman ratkaisu

Sovelletaan nyt projektioita pienimmän neliösumman ratkaisun olemassa oloon. Esitellään ensin ratkaistavana oleva kysymys.

Klassisesti lineaarisen yhtälöryhmän ratkaiseminen vastaa seuraavaa kysymystä.

Ongelma. *Olkoon $y \in \mathbb{R}^{m\times 1}$ ja $A \in \mathbb{R}^{m\times n}$. Ratkaise yhtälö*

$$Ax = y.$$

Jos vektori ei kuulu sarakeavaruuteen $\mathrm{Col}(A)$ eli jos y ei ole lineaarikuvauksen $x \mapsto Ax$ kuvassa, niin yhtälöllä ei ole ratkaisua. Yleisesti siis yhtälöllä $Ax = y$ ei voi olettaa olevan ratkaisuja ellei lineaarikuvaus $x \mapsto Ax$ ole surjektio.

Jos $x \mapsto Ax$ ei ole surjektio, niin onkin luonnollista kysyä, mikä vektori $x \in \mathbb{R}^{n\times 1}$ antaa parhaimman *approksimatiivisen ratkaisun*, eli millä vektorilla $\hat{x} \in \mathbb{R}^{n\times 1}$ vektoreiden y ja $A\hat{x}$ etäisyys on mahdollisimman pieni. Uusi ongelma on siis seuraava:

Ongelma. *(Pienimmän neliösumman ongelma) Olkoon $y \in \mathbb{R}^{m\times 1}$ ja $A \in \mathbb{R}^{m\times n}$. Etsi vektori $\hat{x} \in \mathbb{R}^{n\times 1}$, jolle pätee*

$$\|y - A\hat{x}\| = \min\{\|y - Ax\|\colon x \in \mathbb{R}^{n\times 1}\}.$$

Koska etäisyyden $\|y - Ax\|$ minimi saadaan samalla vektorilla kuin etäisyyden neliön $\|y - Ax\|^2$, niin tätä approksimatiivista ratkaisua kutsutaan myös *pienimmän neliösumman ratkaisuksi*[1].

Koska $Ax \in \mathrm{Col}(A)$ kaikilla $x \in \mathbb{R}^{n\times 1}$, niin pienin mahdollinen etäisyys on lauseen 4.3.1 perusteella $\|y - P(y)\|$, missä $P\colon \mathbb{R}^{n\times 1} \to \mathbb{R}^{n\times 1}$ on ortogonaalinen projektio aliavaruudelle $\mathrm{Col}(A)$. Näin ollen haettu approksimatiivinen ratkaisu on siis vektori $\hat{x} \in \mathbb{R}^{n\times 1}$, joka ratkaisee yhtälön

$$A\hat{x} = P(y).$$

Jos matriisin A sarakkeet ovat lineaarisesti riippumattomia, tämä yhtälö osataan ratkaista yksikäsitteisesti projektion P matriisiesityksen avulla. Kirjataan tämä lauseeksi.

[1] Kirjoita etäisyys $\|y - Ax\|^2$ koordinaatien avulla termin “neliösumma” selittämiseksi.

Lause 4.5.1. *Olkoon* $A \in \mathbb{R}^{m\times n}$ *matriisi, jonka sarakkeet ovat lineaarisesti riippumattomia. Tällöin jokaisella* $y \in \mathbb{R}^{m\times n}$ *yhtälöllä*

$$Ax = y$$

on yksikäsitteinen pienimmän neliösumman (tai approksimatiivinen) ratkaisu

$$\hat{x} = (A^T A)^{-1} A^T y \in \mathbb{R}^{n\times 1},$$

eli $\hat{x}$ *on yksikäsitteinen vektori, jolle pätee*

$$\|y - A\hat{x}\| = \min\{\|y - Ax\| : x \in \mathbb{R}^{n\times 1}\}.$$

Huomautus 4.5.2. *Pienimmän neliösumman ratkaisu* $\hat{x}$ *on siis yhtäpitävästi yhtälön*

$$A^T Ax = A^T y$$

ratkaisu. Kirjallisuudessa tätä yhtälöä kutsutaan yhtälön $Ax = y$ normaaliyhtälöksi.

Lauseen 4.5.1 todistus. Olkoon $P \colon \mathbb{R}^{n\times 1} \to \mathbb{R}^{n\times 1}$ ortogonaalinen projektio aliavaruudelle $\mathrm{Col}(A)$. Lauseen 4.4.5 perusteella

$$P(y) = A(A^T A)^{-1} A^T y$$

kaikilla $y \in \mathbb{R}^{m\times 1}$. Lisäksi matriisi $A^T A$ on kääntyvä lauseen 4.4.2 perusteella. Näin ollen vektori

$$\hat{x} = (A^T A)^{-1} A^T y \in \mathbb{R}^{n\times 1}$$

on hyvin määritelty. Lisäksi

$$A\hat{x} = A(A^T A)^{-1} A^T y = P(y).$$

Koska matriisin A sarakkeet ovat lineaarisesti riippumattomia, niin lineaarikuvaus $x \mapsto Ax$ on injektio. Näin ollen $\hat{x}$ on yhtälön

$$Ax = P(y)$$

ainoa ratkaisu.

Lauseen 4.3.1 nojalla

$$\|y - A\hat{x}\| = \|y - P(y)\| = \min\{\|y - w\| : w \in \mathrm{Col}(A)\} = \min\{\|y - Ax\| : x \in \mathbb{R}^{n\times 1}\}.$$

Väite on näin todistettu. $\square$

4.5.1 Yleinen tapaus

Tarkastellaan vielä normaaliyhtälön muotoa tilanneessa, jossa matriisi $A \in \mathbb{R}^{m\times n}$ ei ole täyttä sarakeastetta, eli $k = \operatorname{rank} A < n$. Myös tässä tilanteessa normaaliyhtälöllä

$$A^T A\hat{x} = A^T y \tag{4.3}$$

on ratkaisu. Ratkaisu ei kuitenkaan ole yksikäsitteinen, kuten kohta huomataan.

Tarkastellaan ensin matriisin A^T nollaavaruutta $\mathrm{Null}(A^T) \subset \mathbb{R}^{m\times 1}$, eli lineaarikuvauksen $f_{A^T} \colon \mathbb{R}^{m\times 1} \to \mathbb{R}^{n\times 1}$ ydintä $\ker f_{A^T}$. Koska kuvaus f_{A^T} on kuvauksen f_A adjungaatti, niin yleinen teoria (lemma 1.7.4) paljastaa, että

$$\mathrm{Null}(A^T) = \mathrm{Col}(A)^\perp.$$

Olkoon $P\colon \mathbb{R}^{m\times 1} \to \mathbb{R}^{m\times 1}$ ortogonaaliprojektio aliavaruudelle $\mathrm{Col}(A)$. Tällöin $\ker P = \mathrm{Col}(A)^\perp = \mathrm{Null}(A^T)$.

Olkoon nyt $y \in \mathbb{R}^{m\times 1}$. Tällöin

$$A^T y = A^T P(y) + A^T(y - P(y)).$$

Havaitaan ensin, että koska $P(y) \in \mathrm{Col}(A)$, niin on olemassa sellainen $\hat{x} \in \mathbb{R}^{n\times 1}$, että $A\hat{x} = P(y)$. Huomaa kuitenkin, että $\hat{x}$ ei ole yksikäsitteinen. Havaitaan nyt, että koska $y - P(y) \in \ker P = \mathrm{Null}(A^T)$, niin $A^T(y - P(y)) = 0$. Näin ollen

$$A^T y = A^T A\hat{x},$$

eli yleisellä normaaliyhtälöllä (4.3) on ratkaisu.

Huomautus 4.5.3. *Yllä oleva todistus osoittaa, että yleisen normaaliyhtälön* (4.3) *ratkaisut antavat pienimmän neliösumman ratkaisun yhtelölle $Ax = y$. Tämä seuraa suoraan siitä havainnosta, että tämän yhtälön ratkaisut ovat yhtälön $Ax = P(y)$ ratkaisut, missä $P\colon \mathbb{R}^{m\times 1} \to \mathbb{R}^{m\times 1}$ on ortogonaaliprojektio aliavaruudelle* $\mathrm{Col}(A)$*. Normaaliyhtälön* (4.3) *ovat juuri tämän yhtälön ratkaisut.*

4.6 Sovellus: Lineaarisen mallin sovittaminen

Pienimmän neliösumman menetelmä sopii myös lineaarisen mallin sovittamiseen dataan. Tässä ongelmassa lähtökohta on, että meille on annettu vektoreita $b_1, \dots, b_k \in \mathbb{R}^{n\times 1}$, jotka ajatellaan lähtöarvoiksi ja vektoreita $y_1, \dots, y_k \in \mathbb{R}^{m\times 1}$, jotka ajatellaan mittaustuloksiksi. Lisäksi tehdään oletus, että nämä vektorit riippuvat toisistaan mallin (eli lineaarisen yhtälöryhmän) välityksellä

$$y_i = Xb_i + \varepsilon_i, \tag{4.4}$$

missä $X \in \mathbb{R}^{m\times n}$ on tuntematon matriisi ja $\varepsilon_i \in \mathbb{R}^{m\times 1}$ on häiriövektori. Vektoreita ε_i ajatellaan yleensä satunnaisiksi (eli nousevan jostain jakaumasta) ja matriisia X ajatellaan deterministiseksi. Koska yhtälö (4.4) on yhtäpitävää yhtälön

$$y_i - \varepsilon_i = Xb_i$$

kanssa, niin jatkossa tarkastellaan vain lineaarista ongelmaa

$$y_i = Xb_i, \tag{4.5}$$

missä $i = 1, \dots, k$.

Kirjoitetaan nyt

$$Y = \begin{bmatrix} y_1 & \cdots & y_k \end{bmatrix} \in \mathbb{R}^{m \times k}$$

ja

$$B = \begin{bmatrix} b_i & \cdots & b_k \end{bmatrix} \in \mathbb{R}^{n \times k}.$$

Nyt k kappaletta yhtälöitä (4.5) voidaan esittää yhtenä matriisiyhtälönä

$$Y = XB, \tag{4.6}$$

missä matriisit $Y \in \mathbb{R}^{m \times k}$ ja $B \in \mathbb{R}^{n \times k}$ ovat tunnettuja ja $X \in \mathbb{R}^{m \times n}$ on tuntematon matriisi, joka halutaan ratkaista. Siirtymällä transpoosiin saadaan tämä yhtälö kirjoitettua tutummassa matriisiyhtälö muodossa

$$Y^T = B^T X^T.$$

Tarkastellaan nyt tilannetta tarkemmin. Palautetaan ensin mieleen, että matriisiavaruudessa $\mathbb{R}^{k \times m}$ on sisätulona tavallinen pistetulo $\cdot \colon \mathbb{R}^{k \times m} \times \mathbb{R}^{k \times m} \to \mathbb{R}$,

$$(a_{ij}) \cdot (a'_{ij}) = \sum_{ij=1}^{n} a_{ij} a'_{ij}.$$

Näin ollen voidaan puhua pienimmän neliösumman ratkaisusta (tai sovituksesta) yhtälölle (4.6). Ongelman ratkaisu on formaalilla tasolla sama kuin sarakeavaruuksien tapauksessa.

Lause 4.6.1. *Olkoon $B \in \mathbb{R}^{n \times k}$ matriisi, jonka sarakkeet ovat lineaarisesti riippumattomia. Tällöin yhtälön*

$$BB^T \hat{X}^T = BY$$

ratkaisu $\hat{X} \in \mathbb{R}^{m \times n}$ on yhtälön

$$Y = XB$$

pienimmän neliösumman ratkaisu.

Väite on yksinkertainen esittää, mutta varsin konstikas todistaa formaalisti ilman matriisiavaruuksien välisiä lineaarikuvauksia ja ositettuja matriiseja.

Huomaa, että avaruuden $\mathbb{R}^{k \times m}$ sisätulo on sellainen, että avaruuden $\mathbb{R}^{k \times m}$ standarikanta e_{ji} on ortonormaalikanta. Näin ollen standarikantojen avulla määritelty isomorfismi $\mathbb{R}^{k \times m} \to \mathbb{R}^{km \times 1}$, $e_{ji} \mapsto e_{m(i-1)+j}$, on isometria. Tämä sallii neliösumman ongelman ratkaisemisen sarakeavaruuksissa.

Todistus. Matriisi $B^T \in \mathbb{R}^{k\times n}$ määrittelee lineaarikuvauksen $F_{B^T} : \mathbb{R}^{n\times m} \to \mathbb{R}^{k\times m}$, $A \mapsto B^T A$, matriisiavaruuksien $\mathbb{R}^{n\times m}$ ja $\mathbb{R}^{k\times m}$ välille. Tällöin lineaarikuvauksen F_{B^T} matriisi $M_{B^T} \in \mathbb{R}^{(km)\times(nm)}$ avaruuden $\mathbb{R}^{(km)\times(km)}$ standardikannassa on ositettumatriisi

$$M_{B^T} = \begin{bmatrix} B^T & 0 & \cdots & 0 & 0 \\ 0 & B^T & & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & & B^T & 0 \\ 0 & 0 & \cdots & 0 & B^T \end{bmatrix}.$$

Toisaalta matriisi $Z = Y^T = [z_1 \cdots z_n] \in \mathbb{R}^{k\times m}$ on avaruuden $\mathbb{R}^{k\times m}$ standardikannassa sarakevektori

$$z = \begin{bmatrix} z_1 \\ \vdots \\ z_n \end{bmatrix} \in \mathbb{R}^{mk\times 1}.$$

Vastaavasti matriisin $\hat{X} = [\hat{x}_1 \cdots \hat{x}_n] \in \mathbb{R}^{m\times n}$ transpoosi on matriisiavaruuden $\mathbb{R}^{n\times m}$ standarikannassa sarakevektori

$$v = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} \in \mathbb{R}^{nm\times 1},$$

missä $v_i^T = \hat{x}_i$ jokaisella $i \in \{1, \ldots, n\}$. Näin ollen tarkasteltava pienimmän neliösumman ongelma on

$$y = M_{B^T} v$$

eli

$$\begin{bmatrix} z_1 \\ \vdots \\ z_n \end{bmatrix} = \begin{bmatrix} B^T & 0 & \cdots & 0 & 0 \\ 0 & B^T & & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & & B^T & 0 \\ 0 & 0 & \cdots & 0 & B^T \end{bmatrix} \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}.$$

Yleisen teorian mukaan yhtälön ratkaisu on sama kuin yhtälön

$$M_{B^T}^T z = M_{B^T}^T M_{B^T} v$$

ratkaisu. Koska $M_{B^T}^T = M_B$, niin saadaan yhtälö

$$M_B z = M_{BB^T} v.$$

Tämä matriisiyhtälö vastaa yhtälöryhmää

$$\begin{cases} By_1 &= BB^T \hat{x}_1 \\ \quad\vdots & \\ By_k &= BB^T \hat{x}_k \end{cases}$$

(Yksityiskohdat harjoitustehtävä.) Näin ollen pienimmän neliösumman ratkaisu vastaa yhtälön

$$BY = BB^T\hat{X}$$

ratkaisua. $\square$

Luku 5

Kompleksiset vektoriavaruudet

Tärkein työkalu äärellisulotteisten vektoriavaruuksien lineaarioperaattoreiden tutkimisessä ovat ominaisarvot ja ominaisavaruudet. Tässä luvussa laajennetaan ominaisarvojen ja ominaisavaruuksien käsitteet kompleksikertoimisille avauuruuksille tilanteeseen. Syy tähän laajennukseen yksinkertainen:

> Kompleksikertoimisella matriisilla (ja lineaarikuvauksilla) on aina kompleksinen ominaisarvo.

Väite ei päde reaalikertoimisille matriiseille ja lineaarikuvauksille. Esimerkiksi reaalisilla 2×2-matriiseilla kiertomatriiseilla ei yleensä ole ominaisarvoja, kuten kohta osoitetaan.

Tämän luvun tarkoituksena on siis antaa lyhyt käytännöllinen johdanto kompleksisiin vektoriavaruuksiin, eli vektoriavaruuksiin, jossa reaaliset kertoimet eli kerroinkunta $\mathbb{R}$, on korvatty yleisemmillä kompleksisilla kertoimilla eli kerroinkunnalla $\mathbb{C}$. Tarvittavat tiedot kompleksiluvuista kerrataan.

5.1 Motivointi: kompleksiset ominaisarvot

Tarkastellaan kiertomatriisia

$$A = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix},$$

missä $\theta \in \mathbb{R} \setminus \pi\mathbb{Z}$, eli $\theta \neq k\pi$ kaikilla $k \in \mathbb{Z}$.

Matriisia A vastaava lineaarikuvaus $f_A \colon \mathbb{R}^{2\times 1} \to \mathbb{R}^{2\times 1}$, $x \mapsto Ax$, on perusesimerkki lineaarikuvauksesta, jolla ei ole ominaisvektoreita. Tämän voi jopa havaita konkreettisen geometrisesti kuvasta. Tarkempi perustelu tälle faktalle on, että jos kuvauksella f_A olisi ominaisvektori $x \in \mathbb{R}^{2\times 1}$, niin sillä olisi myös ominaisarvo $\lambda \in \mathbb{R}$. Palautetaan ensin mieleen, miksi ominaisarvoa ei ole.

Merkintöjen helpottamiseksi, merkitään $a = \cos\theta$ ja $b = \sin\theta$. Tarkasteltava matriisi on siis

$$A = \begin{bmatrix} a & -b \\ b & a \end{bmatrix},$$

missä $a^2 + b^2 = 1$. Lisäksi $b \neq 0$, koska $\theta \neq k\pi$, niin $b \neq 0$.

5.1.1 Kiertomatriisilla ei ole reaalisia ominaisarvoja

Jos nollasta poikkeava vektori $x \in \mathbb{R}^{2\times 1}$ on lineaarikuvauksen f_A ominaisvektori ominaisarvolla $\lambda \in \mathbb{R}$, niin $f_A(x) = \lambda x$, eli $Ax = \lambda x$. Näin ollen nollasta poikkeava vektori x toteuttaa yhtälön

$$(A - \lambda I_{2\times 2})\, x = 0$$

eli matriisi $A - \lambda I_{2\times 2}$ ei ole kääntyvä. Tämä tarkoittaa, että

$$\det(A - \lambda I_{2\times 2}) = 0.$$

Koska

$$\begin{aligned}\det(A - \lambda I_{2\times 2}) &= \det \begin{bmatrix} a-\lambda & -b \\ b & (\cos\theta) - \lambda \end{bmatrix} \\ &= (a-\lambda)^2 - (-b^2) \\ &= (a-\lambda)^2 + b^2,\end{aligned}$$

niin saadaan yhtälö

$$(a-\lambda)^2 = -b^2$$

ominaisarvolle λ.

Koska $b \neq 0$, niin $b^2 > 0$. Näin ollen

$$(a-\lambda)^2 < 0,$$

joka on ristiriita, koska λ oletettiin reaaliluvuksi. (Harjoitustehtävä: Millaista matriisia tarkastellaan, jos $b = 0$? Entä jos $a = 0$?)

5.1.2 Kiertomatriisin kompleksiset ominaisarvot

Tarkastellaan nyt yhtälöä

$$(a-\lambda)^2 = -b^2 \tag{5.1}$$

ilman tulkintaa, että yhtälöllä on yhteys ominaisarvoihin. Paljastuu, että tapaukset $b < 0$ ja $b > 0$ ovat samankaltaiset, joten riittää tarkastella tapausta $b > 0$.

Käytetään nyt tietoa, että kompleksilukujen joukossa yhtälöllä $z^2 = -b^2$ on kaksi ratkaisua

$$z = ib$$

ja

$$z = -ib.$$

Näin ollen jos sallitaan, että luku λ on kompleksinen, saadaan yhtälöt

$$a - \lambda = ib \quad \text{ja} \quad a - \lambda = -ib$$

eli

$$\lambda = a - ib \quad \text{ja} \quad \lambda = a + ib.$$

Näin saadaan, että alkuperäisellä yhtälöllä (5.1) on kaksi kompleksista ratkaisua.

Huomautus 5.1.1. *Huomaa, että ratkaisut $a-ib$ ja $a+ib$ ovat toistensa kompleksikonjugaatteja. Tämä ei ole sattumaa. Jatkossa osoitetaan, että reaalikertoimisen matriisin aidosti kompleksiset ominaisarvot esiintyvät pareittain.*

5.1.3 Kiertomatriisin ominaisvektorit?

Edellisen luvun havaintojen perusteella on selvää, että matriisin A kompleksiset ominaisarvot $\lambda = a + ib$ ja $\lambda = a - ib$ eivät vastaa lineaarikuvauksen $f_A \colon \mathbb{R}^{2\times 1} \to \mathbb{R}^{2\times 1}$ reaaliasia ominaisvektoreita.

Koska sarakeavaruus $\mathbb{R}^{2\times 1}$ on oleellisesti $\mathbb{R}^2$, niin tarkastellaankin avaruuden $\mathbb{R}^2$ sijasta avaruutta $\mathbb{C}^2$ laskutoimituksilla

$$(z_1, w_1) + (z_2, w_2) = (z_1 + z_2, w_1 + w_2)$$

ja

$$\alpha(z, w) = (\alpha z, \alpha w)$$

kaikilla $(z_1, w_1), (z_2, w_2), (z, w) \in \mathbb{C}^2$ ja $\alpha \in \mathbb{C}$, sekä tarkastellaan kuvauksen f_A sijasta kuvausta

$$\varphi_A \colon \mathbb{C}^2 \to \mathbb{C}^2, \quad (z, w) \mapsto (az + bw, -bz + aw).$$

Kiinnitetään nyt

$$\lambda = a + ib$$

ja tarkastellaan onko yhtälöllä

$$\varphi_A((z, w)) = \lambda(z, w) \tag{5.2}$$

eli yhtälöllä

$$(az + bw, -bz + aw) = (\lambda z, \lambda w).$$

ratkaisuja (z, w) joukossa $\mathbb{C}^2$.

Koska avaruuden $\mathbb{C}^2$ alkiot (z_1, w_1) ja (z_2, w_2) ovat sama alkio, jos ja vain jos $z_1 = z_2$ ja $w_1 = w_2$, niin saadaan yhtälöpari

$$\begin{cases} az & +bw & = \lambda z \\ -bz & +aw & = \lambda w \end{cases}$$

eli yhtälöpari

$$\begin{cases} (a-\lambda)z & +bw & = 0 \\ -bz & +(a-\lambda)w & = 0. \end{cases}$$

Koska $\lambda = a + ib$, niin oikeasti olemme tarkastelemassa kahta yhtälöparia

$$\begin{cases} -ibz & +bw & = 0 \\ -bz & -ibw & = 0. \end{cases}$$

Koska $b \neq 0$, niin tämä yhtälöpari on ekvivalentti yhtälöparin

$$\begin{cases} -iz & +w & = 0 \\ -z & -iw & = 0 \end{cases}$$

kanssa. Koska $i^2 = -1$, niin yhtälöparin ensimmäinen rivi antaa

$$z = \frac{w}{-i} = \frac{iw}{-i^2} = iw,$$

joka on sama ehto kuin minkä yhtälöparin toinen rivi antaa. Näin ollen yhtälön (5.2) ratkaisut ovat täsmälleen vektorit

$$\{(iw, w) \in \mathbb{C}^2 \colon w \in \mathbb{C}\},$$

Tarkastellaan tätä ratkaisujoukkoa vielä lähemmin. Avaruuden $\mathbb{C}^2$ laskutoimitusten avulla saadaan, että

$$V_\lambda = \{(iw, w) \in \mathbb{C}^2 \colon w \in \mathbb{C}\} = \{w(i, 1) \in \mathbb{C}^2 \colon w \in \mathbb{C}\}.$$

Koska

$$\varphi_A((z, w)) = \alpha(z, w)$$

kaikilla $(z, w) \in V_\lambda$, niin annattujen laskutoimitusten avulla on helppo havaita, että kaikilla $\alpha \in \mathbb{C}$ ja $(z_1, w_1), (z_2, w_2) \in V_\lambda$ pätee $\alpha(z_1, w_1) + (z_2, w_2) \in V_\lambda$.

Näin ollen, voidaan tulkita, että V_λ on itseasiassa avaruuden $\mathbb{C}^2$ (kompleksisesti) 1-ulotteinen aliavaruus ja kuvauksen φ_A ominaisarvoa λ vastaava ominaisavaruus. Tämä onkin on tulkinta, joka seuraavaksi formalisoidaan.

5.2 Määritelmät ja perustulokset

Kompleksiset vektoriavaruudet eli kompleksikertoimiset vektoriavaruudet ovat itseasiassa todella helppo asia:

> Vaihda kaikissa lineaarialgebran tutuissa käsitteissä reaaliluvut kompleksilukuihin. Tutut tulokset ja menetelmät pätevät.

Kirjataan nyt hieman formaalimmin ja tarkemmin tuon yleisperiaatteen seuraukset. Kompleksikertoiminen vektoriavaruus määritellään aivan kuten reaalinen vektoriavaruus.

Määritelmä 5.2.1. *Joukko V laskutoimituksilla $+\colon V \times V \to V$ ja $\cdot\colon \mathbb{C} \times V \to V$ on* kompleksikertoiminen vektoriavaruus*, jos*

1. *kaikilla $v, w, u \in V$ pätee $v + w = w + v$ ja $(v + w) + u = v + (w + u)$,*
2. *on olemassa nollavektori $0 \in V$, jolle pätee $v + 0 = v$ kaikilla $v \in V$,*
3. *jokaisella vektorilla $v \in V$ on vastavektori $-v \in V$, jolle pätee $v + (-v) = 0$,*
4. *kaikilla vektoreilla $v, w \in V$ ja $a \in \mathbb{C}$ pätee $a(v + w) = av + aw$,*
5. *jokaisella $v \in V$ ja kaikilla $a, b \in \mathbb{C}$ pätee $(a + b)v = av + bv$ ja $a(bv) = (ab)v$ ja*

6. $1v = v$ *jokaisella* $v \in V$*.*

Kompleksikertoimisen vektoriavaruuden aliavaruus määritellään aivan kuten reaalisessa tilanteessa.

Määritelmä 5.2.2. *Kompleksikertoiminse vektoriavaruuden* V *osajoukko* $W \subset V$ *on* (kompleksikertoiminen) aliavaruus*, jos* $av - w \in W$ *kaikilla* $v, w \in W$ *ja* $a \in \mathbb{C}$*.*

Myös lineaarikuvaukset kompleksikertoimisten vektoriavaruuksien välillä määritellään kuten reaalisessa tilanteessa.

Määritelmä 5.2.3. *Kuvaus* $f \colon V \to W$ *kompleksikertoimisten vektoriavaruuksien* V *ja* W *välillä on* (kompleksi)lineaarinen *(tai* $\mathbb{C}$-lineaarinen*), jos kaikilla* $v, w \in V$ *ja* $a \in \mathbb{C}$ *pätee*

$$f(av + w) = af(v) + f(w).$$

Esimerkki 5.2.4. *Tärkein esimerkki kompleksisesta vektoriavaruudesta on*

$$\mathbb{C}^n = \{(z_1, \ldots, z_n) \colon z_1, \ldots, z_n \in \mathbb{C}\},$$

joka on reaalisen avaruuden $\mathbb{R}^n$ *kompleksinen vastine. Sarake-,rivi- ja matriisiavaruudet* $\mathbb{C}^{n\times 1}$*,* $\mathbb{C}^{1\times n}$*,* $\mathbb{C}^{m\times n}$ *määritellään kuten reaalisessa tapauksessa.*

5.2.1 Virittäminen, lineaarinen riippumattomuus ja kanta

Virittäminen, lineaarinen riippumattomuus ja kanta määritellään kuten reaalisessa tapauksessa.

Määritelmä 5.2.5. *Olkoon* V *kompleksikertoiminen vektoriavaruus. Joukon* $S \subset V$ virittämä osajoukko $\mathrm{Sp}(S)$ *on*

$$\mathrm{Sp}(S) = \{a_1 v_1 + \cdots + a_n v_n \in V : n \in \mathbb{N}, a_1, \ldots, a_n \in \mathbb{C},\ v_1, \ldots, v_n \in S\}.$$

Kuten reaalisen vektoriavaruuden tapauksessa, osajoukko $\mathrm{Sp}(S)$ on vektoriavaruuden V aliavaruus. (Harjoitustehtävä)

Määritelmä 5.2.6. *Osajoukko* $\{v_1, \ldots, v_n\} \subset V$ *on kompleksikertoimisen vektoriavaruude* V virittäjä joukko*, jos* $V = \mathrm{Sp}(S)$*. Tällöin sanotaan myös, että* V *on joukon* $\{v_1, \ldots, v_n\}$ *virittämä.*

Määritelmä 5.2.7. *Kompleksikertoimisen vektoriavaruuden* V *on äärellinen osajoukko* $\{v_1, \ldots, v_n\}$ *on* lineaarisesti riippumaton *(eli* vapaa*), jos kaikilla* $a_1, \ldots, a_n \in \mathbb{C}$ *ehdosta*

$$a_1 v_1 + \cdots + a_n v_n = 0,$$

seuraa, että $a_1 = \cdots = a_n = 0$*. Ääretön osajoukko* $S \subset V$ *on* lineaarisesti riippumaton*, jos sen jokainen äärellinen osajoukko on lineaarisesti riippumaton.*

Määritelmä 5.2.8. *Äärellinen jono* $(v_1, \ldots, v_n) \in V^n$ *on kompleksikertoimisen vektoriavaruuden* V kanta*, jos joukko* $\{v_1, \ldots, v_n\}$ *on avaruuden* V *vapaa virittäjä joukko.*

Tutut lauseet kantoihin liittyen pätevät. Seuraavat lauseet todistetaan kuten reaalisten vektoriavaruuksien tapauksessa: todistuksissa riittää korvata jokainen $\mathbb{R}$ symbolilla $\mathbb{C}$.

Lause 5.2.9. *Äärellisulotteisen kompleksikertoimisen vektoriavaruuden virittäjä joukko sisältää kannan. Tarkemmin: Olkoon V äärellisulotteinen kompleksikertoiminen vektoriavaruus ja $S = \{v_1, \ldots, v_n\}$ avaruuden V virittäjäjoukko. Tällöin on olemassa $k \in \{1, \ldots, n\}$ ja sellaiset indeksit $1 \leq i_i < \cdots < i_k \leq n$, että $(v_{i_1}, \ldots, v_{i_k})$ on avaruuden V kanta.*

Lause 5.2.10. *Jokainen äärelisulotteisen kompleksikertoimisen vektoriavaruuden lineaarisesti riippumaton osajoukko voidaan laajentaa kannaksi. Tarkemmin: Olkoon V äärellisulotteinen kompleksikeroiminen vektoriavaruus ja $S = \{v_1, \ldots, v_n\}$ lineaarisesti riippumaton osajoukko. Tällöin on olemassa $k \in \mathbb{N}$ ja sellaiset vektorit $v_{n+1}, \ldots, v_{n+k} \in V$, että $(v_1, \ldots, v_{n+k})$ on avaruuden V kanta.*

Korollaari 5.2.11. *Jokaisella äärellisulotteisella kompleksikertoimisella vektoriavaruudella on kanta.*

Lause 5.2.12. *Äärellisulotteisen kompleksikertoimisen vektoriavaruuden jokaisessa kannassa on sama määrä alkioita.*

Määritelmä 5.2.13. *Äärellisulotteisen kompleksikertoimisen vektoriavaruuden V* dimensio $\dim V$ *on avaruuden V mielivaltaisen kannan alkioiden lukumäärä. Avaruuden V dimensiota merkitään merkinnällä* $\dim V$.

Esimerkki 5.2.14. *Avaruuden $\mathbb{C}^n$ standardikanta on $e_1, \ldots, e_n$, missä $e_i = (\delta_{1i}, \ldots, \delta_{ni})$ jokaisella $i \in \{1, \ldots, n\}$. Näin ollen* $\dim \mathbb{C}^n = n$.

Korollaari 5.2.15. *Jokainen äärellisulotteinen kompleksikertoiminen vektoriavaruus on isomorfinen jonkin avaruuden $\mathbb{C}^{n\times 1}$ kanssa, eli tarkemmin: Jos V on n-ulottinen kompleksinen vektoriavaruus, niin on olemassa lineaarinen isomorfismi $\Phi \colon \mathbb{C}^{n\times 1} \to V$.*

5.2.2 Kompleksikertoimiset matriisit ja $\mathbb{C}$-lineaarikuvaukset

Kompleksikertoimisten matriisien laskutoimitukset määritellään kuten reaalisessa tapauksessa. Erityisesti siis matriisitulo ja käänteismatriisi on määritelty kuten aiemmin, kuten myös transpoosi.

Kompleksikertoiminen matriisi $A \in \mathbb{C}^{m\times n}$ määrittelee lineaarikuvauksen $\varphi_A \colon \mathbb{C}^{n\times 1} \to \mathbb{C}^{m\times 1}$ kaavalla $z \mapsto Az$, jossa Az on matriisin A ja sarakevektorin

$$z = \begin{bmatrix} z_1 \\ \vdots \\ z_n \end{bmatrix} \in \mathbb{C}^{n\times 1}$$

matriisitulo.

Kompleksikojugaatin käsite on hyödyllistä laajentaa matriiseille ja sarekevektoreille. Matriisin

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \in \mathbb{C}^{m\times n}$$

kompleksikonjugaatti on matriisi

$$\overline{A} = \overline{\begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}} = \begin{bmatrix} \overline{a_{11}} & \cdots & \overline{a_{1n}} \\ \vdots & & \vdots \\ \overline{a_{m1}} & \cdots & \overline{a_{mn}} \end{bmatrix} \in \mathbb{C}^{m\times n}.$$

Sarakevektorin $z \in \mathbb{C}^{n\times 1}$ kompleksikonjugaatti $\overline{z} \in \mathbb{C}^{n\times 1}$ on tämän määritelmän erikoistapaus.

Reaalikertoimiset matriisit $\mathbb{R}^{m\times n}$ voidaan tulkita luonnollisella tavalla kompleksikertoimisina matriiseina, sillä jokainen $x \in \mathbb{R}$ on myös kompleksiluku $x + i \cdot 0 \in \mathbb{C}$, eli $\mathbb{R} \subset \mathbb{C}$.

Huomautus 5.2.16. *Algebran kursseilla opitaan, että itseasiassa $\mathbb{R}$ on kompleksilukujen kunnan $\mathbb{C}$ alikunta.*

Koska $\mathbb{R} \subset \mathbb{C}$, niin reaalikertoiminen matriisi $A \in \mathbb{R}^{m\times n}$ on samalla myös kompleksikertoiminen matriisi $A \in \mathbb{C}^{m\times n}$, eli $\mathbb{R}^{m\times n} \subset \mathbb{C}^{m\times n}$. Kuten jatkossa huomataan, tätä havaintoa tarvitaan, kun tarkastellaan matriisin ominaisarvoja.

Huomautus 5.2.17. *Huomaa, että matriisin $A \in \mathbb{R}^{m\times n}$ tulkitseminen kompleksikertoimisena tarkoittaa, että ajatttelemme sitä sekä $\mathbb{R}$-lineaarikuvauksen $f_A\colon \mathbb{R}^{n\times 1} \to \mathbb{R}^{m\times 1}$ että $\mathbb{C}$-lineaarikuvauksen $\varphi_A\colon \mathbb{C}^{n\times 1} \to \mathbb{C}^{m\times 1}$ esitysmatriisina.*

Kantalause pätee kompleksilineaarisille kuvauksilla samassa muodossa ja samalla todistuksella kuin reaalisessa tapauksessa.

Lause 5.2.18. *Olkoot V ja W äärellisulotteisia kompleksikertoimisia vektoriavaruuksia sekä olkoon $(v_1, \ldots, v_n)$ avaruuden V kanta ja olkoon $(w_1, \ldots, w_m)$ avaruuden W kanta. Tällöin $\mathbb{C}$-lineaarinen kuvaus $\varphi\colon V \to W$ määräytyy yksikäsitteisesti kannan $(v_1, \ldots, v_n)$ alkioiden kuvista, eli kertoimet $a_{ji} \in \mathbb{C}$, jotka toteuttavat ehdon*

$$\varphi(v_i) = a_{1i}w_1 + \cdots + a_{mi}w_m$$

jokaisella $i \in \{1, \ldots, m\}$ määräävät kuvauksen φ yksikäsitteisesti.

Kantalause siis sanoo, että myös kompleksisten lineearikuvausten tilanteessa, jokaisella lineaarikuvauksella $\varphi\colon V \to W$ on esitysmatriisi $A \in \mathbb{C}^{m\times n}$, joka riippuu kantojen valinnasta.

Huomautus 5.2.19. *Tarkemmin sanoen, jos $(v_1, \ldots, v_n)$ on avaruuden V kanta ja $(w_1, \ldots, w_m)$ on avaruuden W kanta, niin lineaarikuvaukset $\Phi_V\colon \mathbb{C}^{n\times 1} \to V$, $e_i \mapsto v_i$, ja $\Phi_W\colon \mathbb{C}^{m\times 1} \to W$, $e_j \mapsto w_j$, ovat isomorfismeja ja yhdistetty kuvaus $\tilde{\varphi} = \Phi_W^{-1} \circ \varphi \circ \Phi_V\colon \mathbb{C}^{n\times 1} \to \mathbb{C}^{m\times 1}$ on lineaarikuvaus $\varphi_A\colon \mathbb{C}^{n\times 1} \to \mathbb{C}^{m\times 1}$, $z \mapsto Az$, missä $A \in \mathbb{C}^{m\times n}$.*

Kirjataan vielä lineaarikuvausten peruslause kompleksisessa tapauksessa. Jälleen todistus on sama kuin reaalisessa tapauksessa.

Lause 5.2.20. *Olkoot V ja W kompleksisia vektoriavaruuksia ja olkoon $f\colon V \to W$ lineaarikuvaus. Jos V on äärellisulotteinen, niin* $\operatorname{im} f$ *on avaruuden W äärellisulotteinen aliavaruus ja*

$$\dim V = \dim \ker f + \dim \operatorname{im} f.$$

5.3 Kompleksiset sisätuloavaruudet

Sisätulo yleistyy kompleksikertoimiseen avaruuteen lähes sellaisenaan. Ainoa poikkeus on, että kompleksisissa avaruuksissa sisätulo on kompleksiarvoinen ja konjugaattisymmetrinen.

Määritelmä 5.3.1. *Olkoon V kompleksinen vektoriavaruus. Funktio $\langle \cdot, \cdot \rangle V \times V \to \mathbb{C}$ on* avaruuden V sisätulo, *jos kaikilla $u, v, w \in V$ ja $\lambda \in \mathbb{C}$ pätee*

- *(positiivisuus) $\langle v, v \rangle \in \mathbb{R}$ ja $\langle v, v \rangle \geq 0$,*
- *(definiittisyys) $\langle v, v \rangle = 0$, jos ja vain jos $v = 0$,*
- *(additiivisuus) $\langle u + v, w \rangle = \langle u, w \rangle + \langle v, w \rangle$,*
- *(homogeenisuus) $\langle \lambda v, w \rangle = \lambda \langle v, w \rangle$ ja*
- *(konjugaattisymmetrisyys) $\langle w, v \rangle = \overline{\langle v, w \rangle}$.*

Paria $(V, \langle \cdot, \cdot \rangle)$ kutsutaan kompleksiseksi sisätuloavaruudeksi.

Esimerkki 5.3.2. *Funktio $\cdot\colon \mathbb{C}^n \times \mathbb{C}^n \to \mathbb{C}$,*

$$(z_1, \ldots, z_n) \cdot (w_1, \ldots, w_n) = z_1 \overline{w_1} + \cdots + z_n \overline{w_n},$$

on avaruuden $\mathbb{C}^n$ sisätulo. Kuten reaalisessa tapauksessa kutsutaan pistetuloksi. Sarakeavaruuden $\mathbb{C}^{n\times 1}$ pistetulo määritellään vastaavasti.

Kuten reaalisessa tapauksessa, funktio $\|\cdot\|\colon V \to \mathbb{R}$, $v \mapsto \sqrt{\langle v, v \rangle}$, on normi. Näin ollen myös kompleksisissa avaruuksissa voidaan määritellä ortonormaali kanta. Gram–Schmidtin lauseen todistus toimii sellaisenaan myös kompleksisille sisätuloavaruuksille, joten jokaisella äärellisulotteisella kompleksisella sisätuloavaruudella on ortonormaali kanta. Näitä tuloksia ei todisteta näissä luentomuistiinpanoissa. Ne on todistettu yksityiskohtaisiesti Axlerin kirjan [1] luvussa 6.

Ortogonaalimatriisin vastine kompleksisessa tapauksessa on unitaarimatriisi. Määritelmää varten tarvitaan kongugaattitranspoosin käsite.

Määritelmä 5.3.3. *Matriisin $A \in \mathbb{C}^{m\times n}$* konjugaattitranspoosi *on matriisi $A^* \in \mathbb{C}^{n\times m}$, joka on määritelty kaavalla*

$$A^* = \overline{A}^T,$$

eli kaavalla

$$(A^*)_{ji} = \overline{A_{ij}}.$$

kaikilla $j \in \{1, \ldots, n\}$ *ja* $i \in \{1, \ldots, n\}$.

Määritelmä 5.3.4. *Matriisi* $U \in \mathbb{C}^{n\times n}$ *on* unitaarimatriisi, *jos* $U^{-1} = U^*$.

Kuten reaalisessa tapauksessa unitaarimatriisin $U \in \mathbb{C}^{n\times n}$ sarakkeet muodostavat avaruuden $\mathbb{C}^{n\times 1}$ ortonormaalin kannan. Kirjataan tämä lauseeksi, jonka todistus jätetään harjoitustehtäväksi.

Lause 5.3.5. *Olkoon* $A = [a_1|\cdots|a_n] \in \mathbb{C}^{n\times n}$ *matriisi. Tällöin* A *on unitaarinen, jos ja vain jos* $(a_1, \ldots, a_n)$ *on avaruuden* $(\mathbb{C}^{n\times n}, \cdot)$ *ortonormaalikanta.*

5.4 Kommentti

Kuten on varmaan helppo huomata, tässä luvussa on ainoastaan kerrattu tunnettuja tuloksia ja todettu niiden seuraavan kuten reaalikertoimisessa tapauksessa. Syy tähän on se, että algebralliselta rakenteeltaan sekä $\mathbb{R}$ että $\mathbb{C}$ ovat kuntia. Yleisemmin vektoavaruudet voidaan määritellä yleisen kunnan $\mathbb{F}$ avulla samoin kuin on nyt tehty erikseen reaalilukujen ja kompleksilukujen kohdalla. Tässä luvussa käsitellyt lauseet pätevät suoraan yleisen kunnan yli määritellyille vektoriavaruuksille.

Luku 6

Kompleksiset ominaisarvot ja operaattorin yläkolmioesitys

Määritelmä 6.0.1. *Lineaarikuvausta* $L\colon V \to V$ *vektoriavaruudesta* V *itseensä kutsutaan* (lineaari)operaattoriksi*. Mikäli vektoriavaruus on kompleksikertoiminen, kutsutaan operaattoria* kompleksiseksi operaattoriksi*.*

Kompleksisen lineaarioperaattorin $L\colon V \to V$ ominaisarvot ja ominaisavaruudet määritellään kuten reaalisessa tapauksessa; ainoa ero on, että käytetään kompleksilukuja reaalilukujen sijaan.

Määritelmä 6.0.2. *Olkoon* V *kompleksinen vektoriavaruus. Luku* $\lambda \in C$ *on kompleksisen lineaarikuvauksen* $f\colon V \to V$ ominaisarvo*, jos on olemassa sellainen vektori* $v \in V \setminus \{0\}$*, että* $f(v) = \lambda v$*. Ominaisarvoa* λ *vastaava ominaisavaruus on*

$$E(\lambda, f) = \{v \in V\colon f(v) = \lambda v\}.$$

Määritelmä 6.0.3. *Ominaisavaruus* $E(\lambda, f)$ *on epätriviaali avaruuden* V *aliavaruus. Tässä epätriviaali tarkottaa, että* $E(\lambda, f) \neq \{0\}$*. Huomaa, että* $E(\lambda, f)$ *voi olla koko avaruus* V*.*

6.1 Kompleksiset ominaisarvot

Tärkein syy tarkastella kompleksia vektoriavaruuksia on, että jokaisella kompleksisella lineaarikuvauksella $V \to V$ on ominaisarvo.

Lause 6.1.1. *Olkoon* V *äärellisulotteinen kompleksinen vektoriavaruus, jolle* $\dim V \neq 0$*, ja olkoon* $f\colon V \to V$ *lineaarikuvaus. Tällöin on lineaarikuvauksella* f *on olemassa ominaisarvo* $\lambda \in C$*.*

Kuten aiemmin on todettu. Tämä on erittäin syvällinen tulos ja ei päde reaalisten vektoriavaruuksien välisille lineaarikuvauksille. Se perustuu kompleksilukujen ominaisuuteen, että jokaisella kompleksikertoimisella polynomilla on juuri. Tätä tulosta kutsutaan *algebran peruslauseeksi.*

Lause 6.1.2 (Algebran peruslause)**.** *Olkoon* $P: \mathbb{C} \to \mathbb{C}$, $z \mapsto a_n z^n + \cdots + a_1 z + a_0$, *kompleksinen polynomi (eli* $a_1, \ldots, a_n \in \mathbb{C}$*). Jos polynomin* P *aste* $\deg P$ *on vähintään yksi, niin polynomilla* P *on juuri, eli jos* $n \geq 1$ *ja* $a_n \neq 0$*, niin on olemassa sellainen* $\lambda \in \mathbb{C}$ *että* $P(\lambda) = 0$.

Lisäksi tällöin on olemassa astetta $(\deg P) - 1$ *oleva polynomi* $Q: \mathbb{C} \to \mathbb{C}$*, jolle pätee* $P(z) = (z - \lambda)Q(z)$ *kaikilla* $z \in \mathbb{C}$.

Algebran peruslauseesta seuraa välittömästi kompleksisten polynomien tekijöihin jako.

Korollaari 6.1.3. *Olkoon* $P: \mathbb{C} \to \mathbb{C}$ *komplesinen astetta* $n \geq 1$ *oleva polynomi. Tällöin on olemassa* $a \in \mathbb{C}$ *ja sellaiset* $\lambda_1, \ldots, \lambda_n \in \mathbb{C}$*, että*

$$P(z) = a(z - \lambda_1) \cdots (z - \lambda_n) = a \prod_{j=1}^{n} (z - \lambda_j).$$

Todistus. Olkoon P asteen 1 polynomi. Tällöin P on funktio $z \mapsto a_1 z + a_0$, missä $a_1 \neq 0$ ja $a_0 \in \mathbb{C}$. Koska $a_1 z + a_0 = a_1(z - (-a_0/a_1))$ kaikilla $z \in \mathbb{C}$, niin väite pätee, kun $n = 1$.

Oletetaan, että väite pätee asteen $n \in \mathbb{N}$ polynomeille ja olkoon P asteen $n+1$ polynomi. Algebran peruslauseen nojalla on olemassa $\lambda \in C$ ja sellainen asteen n polynomi Q, että $P(z) = (z - \lambda)Q(z)$ kaikilla $z \in \mathbb{C}$. Väite seuraa soveltamalla induktio-oletusta voidaan nyt soveltaa polynomiin Q. $\square$

Huomautus 6.1.4. *Korollaari pätee myös asteen* $n = 0$ *polynomeille, kun funktio*

$$z \mapsto \prod_{j=1}^{0} (z - \lambda_j)$$

tulkitaan vakiofunktioksi $z \mapsto 1$.

Algebran peruslause ja kompleksikertoimisten polynomien teoria otetaan tällä kurssilla annettuna. Aihetta on käsitelty tarkasti esimerkiksi Axlerin erinomaisessa kirjassa [1], jota kirjoittaja suosittelee tässäkin yhteydessä kaikille.

Todistetaan nyt ominaisarvon olemassaoloa kompleksisessa tapauksessa determinanttien avulla. Lukijaa kehotetaan palaamaan kurssin Lineaarialgebra ja matriisilaskenta I matriaaliin ja tarkistamaan, että määritelmät ja todistukset toimivat myös kompleksisessä tilanteessa. Determinantin määritelmä ja tarvittavat tulokset on käsitelty liitteessä E. Todistus ilman determinantteja, joka perustuu kompleksisten polynomien tekijöihin jakoon, on esitetty liitteessä F.

Lauseen 6.1.1 todistus determinanteilla. Tarkastellaan ensin tapausta, että $\varphi: \mathbb{C}^n \to \mathbb{C}^n$ lineaarikuvaus ja $A \in \mathbb{C}^{n \times n}$ sitä vastaava matriisi, eli $\varphi = \varphi_A$. Tarkastellaan funktiota $P: \mathbb{C} \to \mathbb{C}$,

$$\lambda \mapsto \det(A - \lambda I_{n \times n}).$$

Kuten reaalisessakin tapauksessa, determinantin kehityskaavojen perusteella, funktio P on asteen n polynomi

$$\lambda \mapsto (-1)^n \lambda^n + a_{n-1}\lambda^{n-1} + \cdots + a_1\lambda + a_0,$$

missä $a_{n-1}, \ldots, a_0 \in \mathbb{C}$ ja $a_0 = \det A$.

Algebran peruslauseen nojalla polynomilla P on juuri λ, eli

$$\det(A - \lambda I_{n\times n}) = 0.$$

Näin ollen, kuten reaalisten matriisien tapauksessa, matriisi $A - \lambda I_{n\times n}$ ei ole kääntyvä ja yhtälöllä

$$(A - \lambda I_{n\times n})v = 0$$

on epätriviaali ratkaisu $v \in \mathbb{C}^{n\times 1}$. Koska

$$\varphi(v) = Av = \lambda v,$$

niin λ on kuvauksen φ ominaisarvo.

Tarkastellaan nyt yleistä tapausta. Olkoon $\Phi\colon \mathbb{C}^{n\times 1} \to V$ isomorfismi. Tällöin $\tilde{\varphi} = \Phi^{-1} \circ f \circ \Phi\colon \mathbb{C}^{n\times 1} \to \mathbb{C}^{n\times 1}$ on lineaarikuvaus. Todistuksen alkuosan nojalla kuvauksella $\tilde{\varphi}$ on ominaisarvo $\lambda \in \mathbb{C}$ ja ominaisvektori $\tilde{v} \in \mathbb{C}^{n\times 1} \setminus \{0\}$. Olkoon $v = \Phi(\tilde{v})$. Nyt

$$f(v) = f(\Phi(\tilde{v})) = \Phi\left(\Phi^{-1}(\varphi(\Phi(\tilde{v})))\right) = \Phi(\tilde{\varphi}(\tilde{v})) = \Phi(\lambda\tilde{v}) = \lambda\Phi(\tilde{v}) = \lambda v.$$

Näin ollen λ on kuvauksen f ominaisarvo ja v on kuvauksen φ ominaisvektori. □

6.2 Ominaisavaruuksien summa on suora

On helppo huomata, että eri ominaisarvoihin $\lambda \neq \mu$ liittyvät ominaisavaruudet $E(\lambda, f)$ ja $E(\mu, f)$ eivät leikkaa epätriviaalisti, eli että $E(\lambda, f) \cap E(\mu, f) = \{0\}$. Tämä havaitaan seuraavasti. Olkoon $v \in E(\lambda, f) \cap E(\mu, f)$. Tällöin $\lambda v = f(v) = \mu v$. Koska $\lambda \neq \mu$, niin $v = 0$. Väite siis seuraa.

Yleisemmin pätee, että ominaisavaruuksien summa on suora summa. Tämä on suora seuraus seuraavasta lauseesta. Huomaa, että seuraavassa lauseessa ei ole tarkennettu päteekö se reaalisille vai kompleksisille avaruuksille, koska se pätee molemmille.

Lause 6.2.1. *Olkoot V äärellisulotteinen vektoriavaruus, $f\colon V \to V$ operaattori ja $\lambda_1, \ldots, \lambda_m$ operaattorin f erisuuria ominaisarvoja sekä $v_1, \ldots, v_m \in V$ operaattorin f ominaisvektoreita, joille pätee $v_i \in E(\lambda_i, f)$ kaikilla $i \in \{1, \ldots, n\}$. Tällöin jono $(v_1, \ldots, v_n)$ on vapaa.*

Todistus. Tehdään vastaoletus, että jono $(v_1, \ldots, v_n)$ on lineaarisesti riippuva. Koska jono (v_1) on vapaa, niin on olemassa suurin sellainen luku $1 \leq k \leq n$, että jono $(v_1, \ldots, v_{k-1})$ on vapaa. Tällöin $v_k \in \mathrm{Sp}\{v_1, \ldots, v_{k-1}\}$ ja on olemassa sellaiset luvut $a_1, \ldots, a_k \in \mathbb{C}$, jotka kaikki eivät ole nollia, että

$$a_1 v_1 + \cdots + a_k v_k = 0.$$

Koska jono $(v_1, \ldots, v_{k-1})$ on vapaa, niin pätee $a_k \neq 0$. Näin ollen

$$v_k = -\left(\frac{a_1}{a_k}v_1 + \cdots + \frac{a_{k-1}}{a_k}v_{k-1}\right). \tag{6.1}$$

Koska $v_k \in E(\lambda_k, f)$, niin

$$\begin{aligned}\lambda_k v_k = f(v_k) &= -\left(\frac{a_1}{a_k}f(v_1) + \cdots + \frac{a_{k-1}}{a_k}f(v_{k-1})\right)\\ &= -\left(\frac{a_1}{a_k}\lambda_1 v_1 + \cdots + \frac{a_{k-1}}{a_k}\lambda_{k-1}v_{k-1}\right)\end{aligned}$$

Kertomalla yhtälö (6.1) vakiolla λ_k ja vähentämällä saadusta yhtälöstä, saadaan

$$\begin{aligned}0 &= \lambda_k\left(\frac{a_1}{a_k}v_1 + \cdots + \frac{a_{k-1}}{a_k}v_{k-1}\right) - \left(\frac{a_1}{a_k}\lambda_1 v_1 + \cdots + \frac{a_{k-1}}{a_k}\lambda_{k-1}v_{k-1}\right)\\ &= \frac{a_1}{a_k}(\lambda_k - \lambda_1)v_1 + \cdots + \frac{a_{k-1}}{a_k}(\lambda_k - \lambda_{k-1})v_{k-1}.\end{aligned}$$

Koska jono $(v_1, \ldots, v_{k-1})$ on vapaa, niin

$$\frac{a_i}{a_k}(\lambda_k - \lambda_i) = 0$$

jokaisella $i \in \{1, \ldots, k-1\}$. Koska oletuksen nojalla $\lambda_k \neq \lambda_i$ kaikilla $i \in \{1, \ldots, k-1\}$, niin $a_i = 0$ jokaisella $i \in \{1, \ldots, k-1\}$. Tällöin yhtälön (6.1) perusteella $v_k = 0$. Tämä on ristiriita, koska v_k on ominaisvektori. Vastaoletus on siis väärä ja jono $(v_1, \ldots, v_n)$ on lineaarisesti riippumaton. □

Seuraava korollaari jätetään harjoitustehtäväksi.

Korollaari 6.2.2. *Olkoon V äärellisulotteinen vektoriavaruus, $f\colon V \to V$ operaattori, $\lambda_1, \ldots, \lambda_m$ operaattorin f eri ominaisarvoja. Tällöin*

$$E(\lambda_1, f) + \cdots + E(\lambda_m, f) = E(\lambda_1, f) \oplus \cdots E(\lambda_m, f).$$

Erityisesti

$$\dim E(\lambda_1, f) + \cdots + \dim E(\lambda_m, f) \leq \dim V.$$

6.3 Kompleksikertoimisen operaattorin yläkolmioesitys

Kompleksisten ominaisarvojen olemassaolosta seuraa, että jokaiselle operaattorille $f\colon V \to V$ löytyy sellainen kanta, että tässä kannassa operaattorilla on yläkolmiomatriisi. Yhtäpitävästi tämä tarkoittaa, jokainen kompleksinen neliömatriisi voidaan kannanvaihdolla saattaa yläkolmiomatriisin muotoon.

Määritelmä 6.3.1. *Matriisi* $A = (a_{ji}) \in \mathbb{C}^{n\times n}$ *on* yläkolmiomatriisi, *jos matriisin* A *diagonaalin alapuoliset kertoimet ovat nollia, eli* $a_{ji} = 0$ *kaikilla* $j > i$, *eli*

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & & a_{2n} \\ \vdots & & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ 0 & a_{22} & & a_{2n} \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{bmatrix}$$

Lause 6.3.2. *Olkoon* V n*-ulotteinen kompleksinen vektoriavaruus ja* $f\colon V \to V$ *lineaarioperaattori. Tällöin on olemassa sellainen avaruuden* V *kanta* $(v_1, \ldots, v_n)$, *että kuvauksen* f *matriisi tässä kannassa on yläkolmiomatriisi, eli että lineaarikuvaus* $\tilde{f} = \Phi^{-1} \circ f \circ \Phi\colon \mathbb{C}^{n\times 1} \to \mathbb{C}^{n\times 1}$, *missä* $\Phi\colon \mathbb{C}^{n\times 1} \to V$ *on isomorfismi* $e_i \mapsto v_i$, *on lineaarikuvaus* $f_A\colon \mathbb{C}^{n\times 1} \to \mathbb{C}^{n\times 1}$, $z \mapsto Az$, *missä* $A \in \mathbb{C}^{n\times n}$ *on yläkolmiomatriiisi.*

Ennen lauseen 6.3.2 todistamista muotoillaan se uudelleen matriiseille.

Korollaari 6.3.3. *Olkoon* $A \in \mathbb{C}^{n\times n}$ *matriisi. Tällöin on olemassa sellainen kannanvaihtomatriisi* $P \in \mathbb{C}^{n\times n}$ *ja yläkolmiomatriisi* $T \in \mathbb{C}^{n\times n}$, *että* $A = PTP^{-1}$.

Todistus. Tarkastellaan kuvausta $f\colon \mathbb{C}^{n\times 1} \to \mathbb{C}^{n\times 1}$, $z \mapsto Az$. Olkoot nyt $(v_1, \ldots, v_n)$ avaruuden $\mathbb{C}^{n\times 1}$ kanta ja isomorfismi $\Phi\colon \mathbb{C}^{n\times 1} \to \mathbb{C}^{n\times 1}$ kuten lauseessa 6.3.2. Tällöin lineaarikuvaus $\tilde{f} = \Phi^{-1} \circ f \circ \Phi$ on lineaarikuvaus f_U, missä T on yläkolmiomatriisi.

Olkoon nyt P isomorfismin $\Phi\colon \mathbb{C}^{n\times 1} \to \mathbb{C}^{n\times 1}$ matriisi, eli $\Phi(z) = Pz$. Tällöin

$$Tz = P^{-1}APz$$

kaikilla $z \in \mathbb{C}^{n\times 1}$, eli

$$A = PUP^{-1}.$$

□

Lauseen 6.3.2 perustuu yläkolmiomatriisin tulkintaan invarianttien aliavaruuksien avulla.

Määritelmä 6.3.4. *Kompleksisen vektoriavaruuden* V *aliavaruus* W *on* operaattorin $f\colon V \to V$ invariantti aliavaruus, *jos* $fW \subset W$.

Esimerkki 6.3.5. *Triviaalit esimerkkit invarianteista aliavaruuksista ovat koko avaruus* V *ja nolla-avaruus* $\{0\}$. *Muita esimerkkejä operaattoreihin liittyvistä invarianteista aliavaruuksista ovat ydin ja kuva. Tämän luvun kannalta tärkein esimerkki operaattorin* $f\colon V \to V$ *invariantista aliavaruudesta on ominaisarvoon* $\lambda \in \mathbb{C}$ *liittyvä ominaisavaruus* $E(\lambda, f) \subset V$.

Seuraava lemma kertoo kuinka yläkolmiomatriisit ja invariantit aliavaruudet liittyvät toisiinsa. Lemma pätee sekä kompleksisille että reaalisille vektoriavaruuksille. Todistus jätetään harjoitustehtäväksi.

Lemma 6.3.6. *Olkoon* V *vektoriavaruus,* $(v_1, \ldots, v_n)$ *avaruuden* V *kanta ja* $f\colon V \to V$ *operaattori. Tällöin seuraavat ehdot ovat yhtäpitäviä:*

1. *lineaarikuvauksen f matriisi kannassa $(v_1, \ldots, v_n)$ on yläkolmiomatriisi,*

2. *$f(v_j) \in \mathrm{Sp}\{v_1, \ldots, v_j\}$ kaikilla $j \in \{1, \ldots, n\}$ ja*

3. *$\mathrm{Sp}\{v_1, \ldots, v_j\}$ on operaattorin f invariantti aliavaruus jokaisella $j \in \{1, \ldots, n\}$.*

Lauseen 6.3.2 todistus. Todistetaan väite induktiolla avaruuden V dimension suhteen. Selvästi väite pätee tapauksessa $n = 1$, sillä yksiulotteinen kompleksinen vektoriavaruus on isomorfinen avaruuden $\mathbb{C}$ kanssa ja kompleksisella lineaarikuvauksella $\mathbb{C} \to \mathbb{C}$ on 1×1-matriisi.

Olkoon $n \in \mathbb{N}$. Oletetaan, että väite pätee kaikilla kompleksisilla vektoriavaruuksilla, joiden dimensio on alle n. Olkoon V n-ulotteinen kompleksinen vektoriavaruus ja $f \colon V \to V$ operaattori. Lauseen 6.1.1 nojalla operaattorilla f on ominaisarvo $\lambda \in \mathbb{C}$.

Koska $\ker(f - \lambda \mathrm{id}_V) \neq 0$, niin $\mathrm{im}(f - \lambda \mathrm{id}_V) \neq V$. Näin ollen lauseen 5.2.20 perusteella $\dim \mathrm{im}(f - \lambda \mathrm{id}_V) < \dim V$. Merkitään $U = \mathrm{im}(f - \lambda \mathrm{id}_V)$.

Osoitetaan seuraavaksi, että U on operaattorin f invariantti aliavaruus, eli että $fU \subset U$. Olkoon $u \in U$. Tällöin

$$f(u) = (f - \lambda \mathrm{id}_V)(u) + \lambda u$$

Koska $(f - \lambda \mathrm{id}_V)(u) \in \mathrm{im}(f - \lambda \mathrm{id}_V) = U$ ja $\lambda u \in U$, niin $f(u) \in U$. Näin ollen $fU \subset U$.

Koska $fU \subset U$, niin rajoittuma kuvaus $f|_U \colon U \to U$ on hyvin määrilty, eli rajoittuma $f|_U$ on avaruuden U operaattori. Koska $\dim U < \dim V = n$, niin induktio-oletuksen nojalla on olemassa avaruuden U kanta $(u_1, \ldots, u_k)$, missä $k = \dim U$, jossa kuvauksella $f|_U$ on yläkolmiomatriisi esitys. Lemman 6.3.6 nojalla jokaisella $j \in \{1, \ldots, k\}$ pätee

$$f(u_j) = (f|_U)(u_j) \in \mathrm{Sp}\{u_1, \ldots, u_j\}.$$

Olkoon nyt $(v_1, \ldots, v_n)$ sellainen avaruuden V kanta, että $v_j = u_j$ kaikilla $j \leq k$. Osoitetaan, että $f(v_j) \in \mathrm{Sp}\{v_1, \ldots, v_j\}$ kaikilla $1 \leq j \leq n$. Väite pätee jo kaikilla $1 \leq j \leq k$, joten voidaan olettaa, että $k < j \leq n$. Tällöin

$$f(v_j) = (f - \lambda \mathrm{id}_V)(v_j) + \lambda v_j,$$

missä $(f - \lambda \mathrm{id}_V)(v_j) \in U = \mathrm{Sp}\{v_1, \ldots, v_k\}$. Näin ollen

$$f(v_j) \in \mathrm{Sp}\{v_1, \ldots, v_k, v_{k+1}, \ldots, v_j\}.$$

Kuvauksella f on siis yläkolmiomatriisi esitys kannassa $(v_1, \ldots, v_n)$. $\square$

6.3.1 Reaalisten matriisien yläkolmioesitys

Lause 6.3.2 ei päde reaalisten vektoravaruuksien tapauksessa. Syy tähän on se, että yläkolmiomatriisilla $T \in \mathbb{R}^{n \times n}$ on vähintään yksi ominaisvektori. Tämä ominaisuus on vastaavan lineaarikuvauksen $x \mapsto Tx$ ominaisuus ja siten se säilyy kannanvaihdoissa. Näin ollen ollen esimerkiksi kiertomatriiseille ei ole yläkolmioesitystä. Yläkolmioesitys kuitenkin on olemassa, mikäli polynomi $t \mapsto \det(A - tI)$ faktoroituu ensimmäisen asteen tekiöjöihin, eli polynomilla $t \mapsto \det(A - tI)$ on n reaalista juurta moninkerrat huomioiden.

Lause 6.3.7. *Olkoon $A \in \mathbb{R}^{n\times n}$ sellainen matriisi, että polynomi $t \mapsto \det(A-tI)$ faktoroituu ensimmäisen asteen tekijöihin. Tällöin on olemassa sellainen kääntyvä matriisi $P \in \mathbb{R}^{n\times n}$ ja yläkolmiomatriisi $T \in \mathbb{R}^{n\times n}$, että $A = PTP^{-1}$.*

Todistus. Koska $t \mapsto \det(A - tI)$ faktoroituu ensimmäisen asteen tekijöihin, niin on olemassa sellaiset luvut $\lambda_1, \dots, \lambda_n \in \mathbb{R}$, että

$$\det(A - tI) = c(t - \lambda_1)\cdots(t - \lambda_n),$$

missä $c \in \mathbb{R}$. Tällöin luvut $\lambda_1, \dots, \lambda_n$ ovat matriisin A ominaisarvoja. Olkoon nyt $v_1 \in \mathbb{R}^{n\times n}$ matriisin A sellainen ominaisvektori, jonka ominaisarvo on λ_1.

Merkitään $A_0 = A$ ja olkoon nyt $P_1 \in \mathbb{R}^{n\times n}$ sellainen kääntyvä matriiisi, jonka ensimmäinen sarake on v_1. Tällöin $P_1^{-1}A_0P_1$ on sellainen matriisi, että

$$P_1^{-1}A_0P_1e_1 = P_1^{-1}Av_i = \lambda_1P_1^{-1}v_i = \lambda_1e_1.$$

eli matriisin $P_1^{-1}A_0P_1$ ensimmäinen sarake on λ_1e_1.

Olkoon nyt $A_1 \in \mathbb{R}^{(n-1)\times(n-1)}$ matriisi, joka saadaan matriisista $P_1^{-1}A_0P_1$ poistamalla ensimmäinen rivi ja sarake, eli

$$P_1^{-1}A_0P_1 = \begin{bmatrix} \lambda_1 & b_1 \\ 0 & A_1 \end{bmatrix},$$

missä $b_1 \in \mathbb{R}^{1\times(n-1)}$.

Koska

$$\det(P_1^{-1}A_0P_1-tI) = \det(P_1^{-1}(A_0-tI)P_1) = \det(P_1^{-1})\det(A_0-tI)\det(P_1) = \det(A_0-tI)$$

ja

$$\det(P_1^{-1}A_0P_1 - tI) = (\lambda_1 - t)\det(A_1)$$

niin saadaan, että luvut $\lambda_2, \dots, \lambda_n$ ovat polynomin $t \mapsto \det(A_1 - tI)$ juuria (monikerrat huomioiden). Näin ollen väite seuraa induktiolla. □

6.3.2 Schurin lause

Korollaari 6.3.3 esitetään yleensä vahvemmassa muodossa, että matriisi P voidaan valita ortogonaaliseksi. Tätä tulosta kutsutaan Schurin lauseeksi. Schurin lause seuraa lähes suoraan kompleksisesta Gram–Schmidt metetelmästä (eli kompleksisesta QR-hajotelmasta). Kirjataan tarvittavat tulokset ylös.

Lause 6.3.8 (Kompleksinen QR-hajotelma)**.** *Olkoon $A \in \mathbb{C}^{n\times n}$ kääntyvä matriisi. Tällöin on olemassa sellainen unitaarinen matriisi $Q \in \mathbb{C}^{n\times n}$ ja sellainen yläkolmiomatriisi $R \in \mathbb{C}^{n\times n}$, että $A = QR$.*

Todistus. Koska matriisin A sarakkeet muodostavat avaruuden $\mathbb{C}^{n\times 1}$ kannan, niin Gram–Schmidt algoritmi tuottaa sellaisen ortonormaalin kannan $u_1, \dots, u_n$, että

$$Q = [u_1|\cdots|u_n] \in \mathbb{C}^{n\times n}$$

on unitaarimatriisi, jolle pätee, että matriisi $Q^{-1}A$ on yläkolmiomatriisi. (Harjoitustehtävä). □

Lause 6.3.9 (Schurin lause)**.** *Olkoon $A \in \mathbb{C}^{n\times n}$ matriisi. Tällöin on olemassa unitaarimatriisi $U \in \mathbb{C}^{n\times n}$ ja yläkolmiomatriisi $T \in \mathbb{C}^{n\times n}$, että $A = UTU^*$.*

Todistus. Korollaarin 6.3.3 perusteella on olemassa sellainen kääntyvä matriisi $P \in \mathbb{C}^{n\times n}$ ja yläkolmiomatriisi $Y \in \mathbb{C}^{n\times n}$, että $A = PYP^{-1}$. Kompleksisen QR-hajotelman nojalla on olemassa sellainen unitaarimatriisi $Q \in \mathbb{C}^{n\times n}$ ja yläkolmiomatriisi $R \in \mathbb{C}^{n\times n}$, että $P = QR$. Tällöin

$$A = (QR)Y(QR)^{-1} = QRYR^{-1}Q^{-1} = Q(RYR^{-1})Q^*.$$

Koska yläkolmiomatriisien tulo on yläkolmiomatriisi ja yläkolmiomatriisien käänteismatriisit ovat yläkolmiomatriiseja, niin $T = RYR^{-1}$ on yläkolmiomatriisi. □

Huomautus 6.3.10. *Schurin lauseen todistus paljastaa, että mikäli matriisille $A \in \mathbb{C}^{n\times n}$ on yläkolmioesitys jossain kannassa, niin sille on yläkolmioesitys ortonormaalissa kannassa. Vastaava tulos pätee myös reaalisille matriiseille.*

Schurin lauseella on myös reaalinen vastine.

Lause 6.3.11 (Reaalinen Schurin lause)**.** *Olkoon $A \in \mathbb{R}^{n\times n}$ sellainen matriisi, että sillä on n reaalista ominaisarvoa (kertaluvun mukaan laskettuna). Tällöin on olemassa sellaiset ortogonaalimatriisi $O \in \mathbb{R}^{n\times n}$ ja yläkolmiomatriisi $T \in \mathbb{R}^{n\times n}$, että $A = OTO^T$.*

Todistus. Lauseen 6.3.7 nojalla on olemassa sellaiset kääntyvä matriisi $P \in \mathbb{R}^{n\times n}$ ja yläkolmiomatriisi $Y \in \mathbb{R}^{n\times n}$, että $A = PYP^{-1}$. QR-hajotelman nojalla on olemassa sellainen ortogonaalimatriisi $O \in \mathbb{R}^{n\times n}$ ja yläkolmiomatriisi $R \in \mathbb{R}^{n\times n}$, että $P = OR$. Nyt voidaan valita $T = RYR^{-1}$. □

6.4 Operaattorin diagonalisoituvuus

Palautetaan mieleen, että matriisi $A = (a_{ji}) \in \mathbb{C}^{n\times n}$ on *diagonaalimatriisi*, jos $a_{ji} = 0$ kaikilla $j \neq i$, missä $j, i \in \{1, \dots, n\}$.

Esimerkki 6.4.1. *Matriisi*

$$A = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$$

on diagonaalimatriisi. Selvästi matriisi

$$B = \begin{bmatrix} 2 & 1 \\ 0 & 1 \end{bmatrix}$$

on yläkolmiomatriisi, mutta ei diagonaalimatriisi.

Määritelmä 6.4.2. *Äärellisulotteisen vektoriavaruuden V operaattori $f\colon V \to V$ on* diagonalisoituva*, jos on olemassa sellainen avaruuden V kanta $(v_1, \dots, v_n)$, että*

$$f(v_i) \in \mathrm{Sp}\{v_i\}$$

kaikilla $i \in \{1, \dots, n\}$, eli että operaattorin f matriisi kannassa $(v_1, \dots, v_n)$ on diagonaalimatriisi.

Kaikki neliömatriisit (tai operaattorit) eivät ole diagonalisoituvia. Syy tähän on se, että ominaisarvon algebrallisen monikerran ei tarvitse olla vastaavan ominaisavaruuden dimensio. Tarkemmin asian ilmaisee seuraava diagonalisointilause, joka pätee sekä kompleksisille että reaalisille vektoriavaruuksille, vaikka todistus on kirjoitettu käyttäen kompleksilukuja.

Huomaa, että Lineaarialgebra ja matriisilaskenta I kurssin perusteella tiedetään, että riittävä ehto reaalisen vektoriavaruuden operaattorin $f\colon V \to V$ diagonalisoitumiselle on, että $\dim V$ eri ominaisarvoa. Tämän fakta on seuraavan lauseen erikoistapaus.

Lause 6.4.3. *Olkoon V äärellisulotteinen vektoriavaruus ja $f\colon V \to V$ operaattori. Olkoot $\lambda_1, \dots, \lambda_m$ operaattorin f kaikki erisuuret ominaisarvot. Tällöin seuraavat ehdot ovat yhtäpitäviä:*

1. *operaattori f on diagonalisoituva,*
2. *avaruudella V on kanta, joka koostuu operaattorin f ominaisvektoreista,*
3. *on olemassa sellaiset avaruuden V 1-ulotteiset aliavaruudet $U_1, \dots, U_n$, jotka ovat invariantteja operaattorin f suhteen ja joille pätee $V = U_1 \oplus \cdots \oplus U_n$,*
4. *$V = E(\lambda_1, f) \oplus \cdots \oplus E(\lambda_m, f)$ ja*
5. *$\dim V = \dim E(\lambda_1, f) + \cdots + \dim E(\lambda_m, f)$.*

Todistus. Ehdot (1) ja (2) ovat selvästi yhtäpitäviä. (Harjoitustehtävä)

Oletetaan, että ehto (2) pätee ja olkoon $(v_1, \dots, v_n)$ sellainen avaruuden V kanta, jonka alkiot ovat kuvauksen f ominaisvektoreita. Osoitetaan, että ehto (3) on voimassa.

Olkoon $U_j = \mathrm{Sp}\{v_j\}$ jokaisella $j \in \{1, \dots, n\}$. Olkoon $j \in \{1, \dots, n\}$. Koska v_j on ominaisvektori, niin $f(v_j) = \lambda_j v_i$ jollakin $\lambda_i \in \mathbb{C}$. Näin ollen $f(av_j) = a\lambda_j v \in U_j$ kaikilla $a \in \mathbb{C}$, eli $fU_j \subset U_j$. Näin ollen aliavaruudet $U_1, \dots, U_n$ ovat invariantteja operaattorin f suhteen. Koska $(v_1, \dots, v_n)$ on avaruuden V kanta, niin

$$V = \mathrm{Sp}\{v_1\} \oplus \cdots \oplus \mathrm{Sp}\{v_n\} = U_1 \oplus \cdots \oplus U_n.$$

Näin ollen ehdosta (2) seuraa ehto (3).

Se, että ehdosta (3) seuraa ehto (2), jätetään harjoitustehtäväksi. Tämä osoittaa, että ehdot (1), (2) ja (3) ovat yhtäpitäviä.

Riittää siis osoittaa implikaatiot (2) $\Rightarrow$ (4) $\Rightarrow$ (5) $\Rightarrow$ (2).

Oletetaan, että ehto (2) pätee ja olkoon $(v_1, \dots, v_n)$ avaruuden V kanta, joka koostuu operaattorin f ominaisvektoreista. Koska jokainen avaruuden V vektori on näin ollen lineaarikombinaatio operaattorin f ominaisvektoreista, niin saadaan

$$V = E(\lambda_1, f) + \cdots + E(\lambda_m, f).$$

Korollaarin 6.2.2 nojalla tämä summa on suora. Ehto (4) seuraa.

Oletetaan nyt, että ehto (4) pätee, eli että

$$V = E(\lambda_1, f) \oplus \cdots \oplus E(\lambda_m, f).$$

Valitaan jokaiselle $E(\lambda_i, f)$ kanta $(v_{i1}, \dots, v_{ik_i})$. Koska summa on suora, niin

$$(v_{11}, \dots, v_{1k_1}, v_{21}, \dots, v_{mk_m})$$

on avaruuden V kanta. Näin ollen

$$\dim V = \dim E(\lambda_1, f) + \cdots + \dim E(\lambda_m, f),$$

eli ehto (5) pätee.

Osoitetaan lopuksi, että ehdosta (5) seuraa ehto (2). Oletetaan, että

$$\dim V = \dim E(\lambda_1, f) + \cdots + \dim E(\lambda_m, f).$$

Valitaan jokaiselle $E(\lambda_i, f)$ kanta $(v_{i1}, \dots, v_{ik_i})$. Osoitetaan, että

$$(v_1, \dots, v_n) = (v_{11}, \dots, v_{1k_1}, v_{21}, \dots, v_{mk_m})$$

on avaruuden V kanta. Koska $\dim V = n$, niin riittää osoittaa, että jono $(v_1, \dots, v_n)$ on lineaarisesti riippumaton. Tehdään vastaoletus, että on olemassa sellaiset kertoimet $a_1, \dots, a_n \in \mathbb{C}$, jotka eivät kaikki ole nollia, että

$$a_1 v_1 + \cdots + a_n v_n = 0.$$

Tällöin jokaisella $p \in \{1, \dots, m\}$ ja $j \in \{1, \dots, k_p\}$ on olemassa yksikäsitteinen sellainen indeksi $i \in \{1, \dots, n\}$, että $v_{pj} = v_i$. Merkitään $a_{pj} = a_i$ ja $u_p = a_{p1}v_{p1} + \cdots + a_{pk_p}v_{pk_p}$. Nyt

$$0 = a_1 v_1 + \cdots + a_n v_n = u_1 + \cdots + u_m,$$

missä $u_p \in E(\lambda_p, f)$ jokaisella $p \in \{1, \dots, m\}$. Korollaarin 6.2.2 nojalla pätee $u_p = 0$ kaikilla $p \in \{1, \dots, m\}$. Koska $(v_{p1}, \dots, v_{pk_p})$ on avaruuden $E(\lambda_p, f)$ kanta jokaisella $p \in \{1, \dots, m\}$, niin $a_{pj} = 0$ kaikilla $p \in \{1, \dots, m\}$ ja $j \in \{1, \dots, k_p\}$. Näin ollen $a_i = 0$ jokaisella $i \in \{1, \dots, n\}$. Jono $(v_1, \dots, v_n)$ on siis vapaa. Tämä päättää todistuksen. $\square$

Korollaari 6.4.4. *Olkoon V äärellisulottinen vektoriavaruus. Jos operaattorilla $f \colon V \to V$ on $\dim V$ eri ominaisarvoa, niin f on diagonalisoituva.*

Todistus. Olkoot $\lambda_1, \dots, \lambda_n$, missä $n = \dim V$, operaattorin f ominaisarvot. Koska $\dim E(\lambda_i, f) \geq 1$, niin

$$\dim E(\lambda_1, f) + \cdots + \dim E(\lambda_n, f) \geq n = \dim V$$

Koska aliavaruuksien $E(\lambda_1, f), \dots, E(\lambda_n, f)$ summa on suora, niin

$$\dim V = \dim E(\lambda_1, f) + \cdots + \dim E(\lambda_n, f).$$

Operaattori f on siis diagonalisoituva lauseen 6.4.3 nojalla. $\square$

Esimerkki 6.4.5. *Esimerkin 6.4.1 matriisi* B *on diagonalisoituva. Tämä seuraa siitä, että sillä on kaksi ominaisarvoa* 2 *ja* 1*, joiden ominaisavaruudet ovat* $E(2,B) = \mathrm{Sp}\{e_1\}$ *ja* $E(1,B) = \mathrm{Sp}\{e_1 - e_2\}$*. Selvästi* $(e_1, e_1 - e_2)$ *on avaruuden* $\mathbb{R}^{2\times 1}$ *(tai avaruuden* $\mathbb{C}^{2\times 1}$*) kanta. Tässä kannassa lineaarikuvaus* $\varphi_B \colon \mathbb{C}^{2\times 1} \to \mathbb{C}^{2\times 1}$*,* $z \mapsto Bz$*, saa matriisin*

$$\begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}.$$

Esimerkki 6.4.6. *Olkoon*

$$A = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}.$$

Operaattorin $\varphi_A \colon \mathbb{C}^{n\times 1} \to \mathbb{C}^{n\times 1}$*,* $z \mapsto Az$ *ainoa ominaisarvo on* 0 *ja* $E(0, \varphi_A) = \mathrm{Sp}\{e_1\}$*. Näin ollen* φ_A *ie ole diagonalisoituva. Matriisin* A *yläkolmioesitys saadaan, kun havaitaan, että* $Ae_2 = e_2$*. Näin ollen* (v_1, v_2)*, missä* $v_1 = e_2$ *ja* $v_2 = e_1$*, on sellainen avaruuden* $\mathbb{C}^{2\times 1}$ *kanta, että matriisilla* A *on tässä kannassa yläkolmioesitys. Tämä yläkolmioesitys on itseasiassa matriisi*

$$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}.$$

Luku 7

Symmetrisen neliömatriisin diagonalisointi

Tässä luvussa todistetaan tärkeä symmetristen matriisien spektraalilause:

> Matriisi $A \in \mathbb{R}^{n\times n}$ on symmetrinen, jos ja vain jos on olemassa ortogonaalimatriisi P, että PAP^T on diagonaalimatriisi.

Tulosta varten tarkastellaan itse-adjungoituja lineaarikuvauksia, jotka ovat symmetristen matriisien vastine lineaarikuvausten puolella. Tärkein aputulos tulee olemaan, että itse-adjungoitujen lineaarikuvausten (eli symmetristen matriisien) ominaisarvot ovat reaalisia.

7.1 Symmetriset neliömatriisit ja itseadjungoidut kuvaukset

Kuten luvussa 1.4 havaittiin, matriisin transpoosi vastaa sisätuloavaruuksissa lineaarikuvauksen adjungaattia. Symmetriset matriisit puolestaan vastaavat itseadjungoituja kuvauksia.

Määritelmä 7.1.1. *Matriisi $A \in \mathbb{C}^{n\times n}$ on* symmetrinen*, jos $A^T = \overline{A}$.*

Huomautus 7.1.2. *Neliömatriisin $A = (a_{ji}) \in \mathbb{C}^{n\times n}$ tapauksessa, $\overline{A}$ on matriisi $(\overline{a_{ji}})$. Reaalisen matriisin $A \in \mathbb{R}^{n\times n}$ tapauksessa siis $\overline{A} = A$. Reaalinen neliömatriisi A on siis symmetrinen, jos $A^T = A$.*

Määritelmä 7.1.3. *Sisätuloavaruuden $(V, \langle\cdot,\cdot\rangle)$ lineaarioperaattori $f\colon V \to V$ on* itseadjungoitu*, jos $f^* = f$ eli $\langle f(v), w\rangle = \langle v, f(v)\rangle$ kaikilla $v, w \in V$.*

Reaalisessa tapauksessa näiden käsitteiden yhteys seuraa lauseesta 1.4.7. Tarvittavat muutokset reaalisista vektoriavaruuksista kompleksisiin jätetään lukijalle.

Propositio 7.1.4. *Olkoon $(V, \langle\cdot,\cdot\rangle)$ sisätuloavaruus ja $(v_1, \dots, v_n)$ avaruuden V ortonormaali kanta. Tällöin $f\colon V \to V$ operaattori on itseadjungoitu, jos ja vain jos kuvauksen f esitysmatriisi A kannassa $(v_1, \dots, v_n)$ on symmetrinen.*

Todistus. Oletetaan ensin, että f on itseadjungoitu. Lauseen 1.4.7 kompleksisen version nojalla adjungaatin f^* matriisi tässä kannassa $(v_1, \dots, v_n)$ on $\overline{A}^T$. Koska $f = f^*$, niin $A^T = \overline{A}$, eli A on symmetrinen metriisi.

Oletetaan nyt, että operaattorin f esitysmatriisi A on symmetrinen eli $A^T = \overline{A}$. Olkoon $\Phi\colon \mathbb{C}^{n\times 1} \to V$ isomorfismi $e_i \mapsto v_i$. Koska $(f_A)^* = f_{A^T} = f_A$, niin lauseen 1.3.28 nojalla $f = \Phi \circ f_A \circ \Phi^{-1}$ on oma adjungaattinsa. (Yksityiskohdat harjoitustehtävä.) Näin ollen f on itseadjungoitu. □

7.2 Ominaisarvojen reaalisuus

Itseadjungoitujen lineaarikuvausten tärkein ominaisuus on, että niiden ominaisarvot ovat reaalisia.

Lause 7.2.1. *Olkoon $(V, \langle\cdot,\cdot\rangle)$ kompleksinen sisätuloavaruus ja $f\colon V \to V$ itseadjungoitu operaattori. Tällöin operaattorin f ominaisarvot ovat reaalisia.*

Todistus. Olkoon $\lambda \in \mathbb{C}$ operaattorin f ominaisarvo ja $v \in V$ sitä vastaava ominaisvektori. Tällöin

$$\lambda\langle v, v\rangle = \langle \lambda v, v\rangle = \langle f(v), v\rangle = \langle v, f(v)\rangle = \langle v, \lambda v\rangle = \overline{\lambda}\langle v, v\rangle.$$

Koska $v \neq 0$, niin $\lambda = \overline{\lambda}$ eli $\lambda \in \mathbb{R}$. □

Edellisen lauseen seurauksena saadaan, että symmetrisen matriisin ominaisarvot ovat reaalisia. Erityisesti tämä tulos kertoo, että reaalisten neliömatriisien $A \in \mathbb{R}^{n\times n}$ tapauksessa kaikki yhtälön $\det(A - \lambda I) = 0$ juuret ovat reaalisia.

Korollaari 7.2.2. *Symmetrisen matriisin ominaisarvot ovat reaalisia.*

Todistus. Olkoon $A \in \mathbb{C}^{n\times n}$ ja $\varphi_A\colon \mathbb{C}^{n\times 1} \to \mathbb{C}^{n\times 1}$, $z \mapsto Az$. Olkoon λ matriisin A ominaisarvo. Tällöin λ on operaattorin φ_A ominaisarvo ja siten reaalinen. □

On syvällinen tulos, että reaalisella itseadjungoidulla operaattorilla on aina ominaisarvo. Todistetaan tämä tulos hyödyntäen kompleksisten operaattorien teoriaa.

Lause 7.2.3. *Olkoon $(V, \langle\cdot,\cdot\rangle)$ reaalinen sisätuloavaruus ja $f\colon V \to V$ itseadjungoitu operaattori. Tällöin operaattorilla f on ominaisarvo.*

Todistus. Olkoon $v_1, \dots, v_n$ avaruuden V kanta ja olkoon A operaattorin f esitysmatriisi tässä kannassa. Koska f on itseadjungoitu, niin A on symmetrinen, eli $A^T = A$. Halutaan osoittaa, että lineaarikuvauksella $f_A\colon \mathbb{R}^{n\times 1} \to \mathbb{R}^{n\times 1}$, $x \mapsto Ax$, ominaisarvo.

Koska A on reaalikertoiminen, niin $A^T = \overline{A}$ kompleksisena matriisina. Tarkastellaan kompleksista lineaarikuvausta $\varphi_A\colon \mathbb{C}^{n\times 1} \to \mathbb{C}^{n\times 1}$, $z \mapsto Az$. Lauseen 6.1.1 kuvauksella φ_A on ominaisarvo $\lambda \in \mathbb{C}$. Koska A on symmetrinen, niin lauseen 7.2.1 nojalla $\lambda \in \mathbb{R}$. Osoitetaan, että λ on reaalisen matriisin A, eli lineaarikuvauksen f_A ominaisarvo.

Olkoon

$$z = \begin{bmatrix} z_1 \\ \vdots \\ z_n \end{bmatrix} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} + i \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = x + iy$$

lineaarikuvauksen φ_A ominaisarvoa λ vastaava ominaisvektori. Koska φ_A on $\mathbb{C}$-lineaarinen ja A on reaalinen, niin

$$\varphi_A(z) = \varphi_A(x + iy) = \varphi_A(x) + i\varphi_A(y). = f_A(x) + if_A(y).$$

Toisaalta, koska $\lambda \in \mathbb{R}$, niin

$$\lambda x + i\lambda y = \lambda z = \varphi_A(z) = f_A(x) + if_A(y).$$

Näin ollen $\lambda x = f_A(x)$ ja $\lambda y = f_A(y)$. Koska $z \neq 0$, niin ainakin toinen vektoreista x ja y on nollasta poikkeava ja siten operaattorin f_A ominaisvektori ominaisarvolla λ. $\square$

Huomautus 7.2.4. *Lauseen 7.2.3 voi todistaa käyttämättä kompleksisten vektoriavaruuksien teoriaa. Kiinnostunut lukija ohjataan jälleen Axlerin hienon kirjan [1] pariin.*

Koska itseadjungoidun operaattorin määritelmä perustuu sisätuloon, ei ole suuri yllätys, että itseadjungoidun operaattorin invariantin aliavaruuden kohtisuorakomplementti on myös invariantti aliavaruus. Kirjataan tämä tulos lauseeksi.

Lause 7.2.5. *Olkoon $(V, \langle \cdot, \cdot \rangle)$ sisätuloavaruus, $f\colon V \to V$ itseadjungoitu operaattori ja $W \subset V$ operaattorin f invariantti aliavaruus. Tällöin $W^\perp$ on operaattorin f invariantti aliavaruus eli $fW^\perp \subset W^\perp$.*

Todistus. Olkoon $v \in W^\perp$. Halutaan osoittaa, että $f(v) \in W^\perp$, eli että $\langle f(v), w\rangle = 0$ kaikilla $w \in W$. Olkoon $w \in W$. Koska f on itseadjungoitu ja $fW \subset W$, niin

$$\langle f(v), w\rangle = \langle v, f^*(w)\rangle = \langle v, f(w)\rangle = 0.$$

Näin ollen $f(v) \in W^\perp$. $\square$

Kirjataan ylös myös havainto, että itseadjungoidun operaattorin rajoittuma invarianttiin aliavaruuteen on itseadjungoitu. Huomaa, että sisätulon rajoittuma aliavaruuteen on sisätulo siinä aliavaruudessa.

Lemma 7.2.6. *Olkoon $(V, \langle \cdot, \cdot \rangle)$ sisätuloavaruus, $f\colon V \to V$ itseadjungoitu operaattori ja $W \subset V$ operaattorin f invariantti aliavaruus. Tällöin $f|_W\colon W \to W$ on sisätuloavaruuden $(W, \langle \cdot, \cdot \rangle)$ itseadjungoitu operaattori.*

Todistus. Olkoot $w, w' \in W$. Tällöin

$$\langle (f|_W)(w), w'\rangle = \langle f(w), w'\rangle = \langle w, f^*(w')\rangle = \langle w, f(w')\rangle = \langle w, (f|_W)(w')\rangle.$$

$\square$

7.3 Spektraalihajotelma

Tässä luvussa todistamme spektraalihajotelman reaalisille itseadjungoiduille operaattoreille ja reaalisille symmetrisille matriiseille. Kompleksinen tapaus on käsitelty esimerkiksi Axlerin kirjan [1] luvussa 7.B.

Lause 7.3.1 (Reaalinen spektraalilause)**.** *Olkoon* $(V, \langle \cdot, \cdot \rangle)$ *reaalinen sisätuloavaruus ja* $f \colon V \to V$ *operaattori. Tällöin seuraavat ehdot ovat yhtäpitäviä:*

1. f *on itseadjungoitu,*
2. *avaruudella* V *on operaattorin* f *ominaisvektoreista koostuva ortonormaalikanta ja*
3. *operaattorilla* f *on diagonaalimatriisi, jossain avaruuden* V *ortonormaalissa kannassa.*

Kirjataan spektraalilause myös matriisien tapauksessa. Ominaisuuksien johtaminen jätetään harjoitustehtäväksi.

Korollaari 7.3.2. *Olkoon* $A \in \mathbb{R}^{n \times n}$ *neliömatriisi. Tällöin seuraavat ehdot ovat yhtäpitäviä:*

1. A *on symmetrinen,*
2. *avaruudella* $\mathbb{R}^{n \times 1}$ *on matriisin* A *ominaisvektoreista koostuva ortonormaali kanta,*
3. *on olemassa sellainen ortogonaalimatriisi* $P \in \mathbb{R}^{n \times n}$ *ja sellainen diagonaalimatriisi* $D \in \mathbb{R}^{n \times n}$*, että*
$$A = PDP^T.$$

Lisäksi, tässä tapauksessa, matriisin D *diagonaalielementit ovat matriisin* A *ominaisarvoja.*

Spektraalilause on merkittävä tulos: Se antaa täydellisen karakterisoinnin itseadjungoiduille operaattoreille ja symmetrisille matriiseille. Sen seurauksena tiedetään, että reaalisella symmetrisellä $n \times n$-matriisilla on n lineaarisesti riippumatonta ominaisvektoria ja että se on diagonalisoituva.

Huomautus 7.3.3. *Yleensä sanotaan, että itseadjungoidulla operaattorilla* $f \colon V \to V$ *(tai ekvivalensti symmetrisellä matriisilla) on* $\dim V$ *ominaisarvoa, kun ominaisarvot lasketaan multiplisiteetin mukaan. Tällä tarkoitetaan sitä, että ominaisarvoille* $\lambda_1 > \cdots > \lambda_k$ *pätee*
$$\dim E(\lambda_1, f) + \cdots + \dim E(\lambda_k, f) = \dim V,$$
missä $\dim E(\lambda_j, f)$ *on ominaisarvon* λ_j *multiplisiteetti.*

Huomautus 7.3.4. *Symmetrisen matriisin* A *potenssit voidaan helposti laskea diagonaaliesityksestä:*
$$A^2 = AA = PDP^T PDP^T = PD(P^T P)DP^T = PDDP^T = P(D^2)P^T.$$

Vastaavasti kaikilla $k \in \mathbb{N}$ saadaan

$$A^k = P(D^k)P^T,$$

missä matriisi D^k on diagonaalimatriisi, jonka diagonaalielementit ovat matriisin D diagonaalielementtien potensseja. Jos matriisin A ominaisarvot ovat nollasta poikkeavia, niin lisäksi saadaan kaava

$$A^{-1} = PD^{-1}P^T$$

matriisin A käänteismatriisille, missä D^{-1} on diagonaalimatriisi, jonka diagonaalielementit ovat matriisin D diagonaalielementtien käänteislukuja.

Spektraalilauseen todistuksessa tarvittavat tärkeimmät aputulokset ovat ominaisarvojen olemassaolo (Lause 7.2.3) ja invariantin aliavaruuden komplementin invarianttius (Lause 7.2.5).

Lauseen 7.3.1 todistus. Osoitetaan implikaatiot: (3) $\Rightarrow$ (1), (1) $\Rightarrow$ (2) ja (2) $\Rightarrow$ (3). Näistä ensimmäinen ja viimeinen implikaatio ovat helppoja.

Oletetaan, että oletus (3) pätee ja olkoon $(v_1, \ldots, v_n)$ sellainen avaruuden V ortonormaalikanta, että $f = \Phi_V \circ f_D \circ \Phi_V^{-1}$, missä $\Phi_V : \mathbb{R}^{n\times 1} \to V$ on isomorfismi $e_i \mapsto v_i$ ja D on diagonaalimatriisi. Koska $D^T = D$, niin $f^* = f$, eli (1) pätee.

Oletetaan nyt, että (2) pätee ja olkoon $(v_1, \ldots, v_n)$ operaattorin f ominaisvektoreista koostuva ortonormaalikanta ja olkoot $\lambda_1, \ldots, \lambda_n$ vastaavat ominaisvektorit. Olkoon $A \in \mathbb{R}^{n\times n}$ kuvauksen f esitysmatriisi tässä kannassa. Tällöin

$$\Phi_V(Ae_i) = \Phi(f_A(e_i)) = f(\Phi_V(e_i)) = f(v_i) = \lambda_i v_i = \lambda_i \Phi_V(e_i) = \Phi_V(\lambda_i e_i).$$

Näin ollen $Ae_i = \lambda_i e_i$. Matriisi A on siis ominaisarvoista $\lambda_1, \ldots, \lambda_n$ koostuva diagonaalimatriisi, eli (3) pätee.

Oleteaan lopuksi, että (1) on voimassa. Todistetaan väite induktiolla avaruuden V dimension suhteen. Väite pätee selvästi 1-ulotteisille avaruuksille. Olkoon nyt $n \in \mathbb{N}$ sellainen, että väite pätee kaikille avaruuksille W, joiden dimensio on korkeintaan n.

Olkoon V $(n+1)$-ulotteinen avaruus ja olkoon $f : V \to V$ itseadjungoitu operaattori. Lauseen 7.2.3 nojalla operaattorilla f on ominaisarvo $\lambda \in \mathbb{R}$. Olkoon $u \in V$ ominaisarvoa λ vastaava ominaisvektori, jolle pätee $\|u\| = 1$, ja olkoon $U = \mathrm{Sp}\{u\}$. Koska $fU \subset U$, niin lauseen 7.2.5 nojalla $fU^\perp \subset U^\perp$. Koska $\dim U > 0$, niin $\dim U^\perp < \dim V$. Koska $f|_{U^\perp} : U^\perp \to U^\perp$ on itseadjungoitu operaattori sisätuloavaruudessa $(U^\perp, \langle\cdot,\cdot\rangle)$ lemman 7.2.6 nojalla, niin induktio-oletuksesta seuraa, että on olemassa avaruuden $U^\perp$ ortonormaalikanta $(u_1, \ldots, u_k)$, joka koostuu operaattorin $f|_{U^\perp}$ ominaisvektoreista. Tällöin $(u, u_1, \ldots, u_k)$ on avaruuden V ortonormaalikanta, joka koostuu operaattorin f ominaisvektoreista. $\square$

Luku 8

Positiivisesti semidefiniitit neliömatriisit

Tämä luku aloitetaan tarkastelemalla positiivisesti semidefiniittejä matriiseja ja niitä vastaavia positiivisia operaattoreita, joille todistetaan polaarihajotelma. Polaarihajotelman avulla todistetaan singulaariarvohajotelma yleisille neliömatriiseille ja operaattoreille. Sovelluksena todistetaan Cholesky-hajotelma positiivisesti semidefiniiteille matriiseille.

8.1 Symmetrisen neliömatriisin definiittisyys

Yksi tapa luokitella symmetriset matriisit on tarkastella niiden definiittisyyttä.

Määritelmä 8.1.1. *Symmetrinen matriisi $A \in \mathbb{R}^{n\times n}$ on*

- positiivisesti semidefiniitti, *jos $x^T Ax \geq 0$ kaikilla vektoreilla $x \in \mathbb{R}^{n\times 1}$,*
- positiivisesti definiitti, *jos $x^T Ax > 0$ kaikilla nollasta poikkeavilla vektoreilla $x \in \mathbb{R}^{n\times 1}$,*
- negatiivisesti semidefiniitti, *jos $x^T Ax \leq 0$ kaikilla vektoreilla $x \in \mathbb{R}^{n\times 1}$,*
- nagatiivisesti definiitti, *jos $x^T Ax < 0$ kaikilla nollasta poikkeavilla vektoreilla $x \in \mathbb{R}^{n\times 1}$ ja*
- indefiniitti, *jos A ei ole positiivisesti tai negatiivisesti semidefinitiitti.*

Huomautus 8.1.2. *Jos matriisi $A \in \mathbb{R}^{n\times n}$ on negatiivisesti (semi)definiitti, niin tällöin $B = -A$ on on positiviisesti (semi)definiitti. Näin ollen usein tarkastellaan ainoastaan positiivisesti (semi)definiittejä matriiseja.*

Huomautus 8.1.3. *Symmetrinen matriisi A on indefiniitti, jos on olemassa sellaiset $x \in \mathbb{R}^{n\times n}$ ja $y \in \mathbb{R}^{n\times n}$, että $x^T Ax > 0$ ja $y^T Ay < 0$.*

Jatkossa tarkastellaan pääosin positiviisesti semidefiniittejä matriiseja. Tyypillisiä esimerkkejä ovat seuraavat tutut matriisit.

Esimerkki 8.1.4. *Olkoon $E \in \mathbb{R}^{m\times n}$ matriisi. Tällöin matriisi $A = E^T E \in \mathbb{R}^{n\times n}$ on positiivisesti semidefiniitti. Tämä seuraa huomiosta, että kaikilla $x \in \mathbb{R}^{n\times 1}$ pätee*

$$x^T Ax = x^T E^T Ex = (Ex)^T (Ex) = Ex \cdot Ex \geq 0.$$

Esimerkki 8.1.5. *(Jatkoa edelliseen esimerkkiin.) Positiivisesti definiitit matriisit ovat tuttuja sisätulojen teoriasta:* Symmetrinen matriisi $A \in \mathbb{R}^{n\times n}$ on positiivisesti definiitti, jos ja vain jos funktio $\langle\cdot,\cdot\rangle\colon \mathbb{R}^{n\times 1} \times \mathbb{R}^{n\times 1} \to \mathbb{R}$, $\langle x, y\rangle = x^T Ay$ kaikilla $x, y \in \mathbb{R}^{n\times 1}$, on sisätulo. *Asiaa on käsitelty tarkemmin liitteessä A.*

Esimerkki 8.1.6. *Olkoon $A \in \mathbb{R}^{n\times n}$ positiivisesti semidefiniitti matriisi ja olkoon $\tilde{A} \in \mathbb{R}^{k\times k}$ sellainen matriisi, että*

$$A = \begin{bmatrix} \tilde{A} & B \\ B^T & C \end{bmatrix},$$

eli olkoon $\tilde{A}$ matriisin A ns. $(k\times k)$-pääminori. Tällöin $\tilde{A}$ on positiivisesti semidefiniitti. Tämä seuraa havainnosta, että jokaisella $x \in \mathbb{R}^{k\times k}$ pätee

$$x^T \tilde{A}x = \begin{bmatrix} x^T & 0 \end{bmatrix} \begin{bmatrix} \tilde{A} & B \\ B^T & C \end{bmatrix} \begin{bmatrix} x \\ 0 \end{bmatrix} \geq 0.$$

Vastaavasti, jos A on positiivisesti definiitti, niin myös $\tilde{A}$ on positiivisesti definiitti. Samat väitteet pätevät myös matriisille C.

Huomautus 8.1.7. *Yllä olevat esimerkit johdattavat helppoon huomioon, että positiivisesti semidefiniitin matriisin diagonaalialkiot ovat ei-negatiivisa ja että positiivisesti definiitin matriisin diagonaalialkiot ovat positiiviisia. Jälkimmäinen väite ei kuitenkaan päde kääntäen. Matriisi*

$$A = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

on positiivisesti semi-definiitti matriisi, jonka diagonaalialkiot ovat positiivisia.

Positiivisesti semidefiniittit matriisit vastaavat (terminologialtaan harhaanjohtavasti) positiivisia operattoreita.

Määritelmä 8.1.8. *Olkoon $(V, \langle\cdot,\cdot\rangle)$ sisätuloavaruus. Lineaarioperaattori $f\colon V \to V$ on* positiviinen, *jos $\langle f(v), v\rangle \geq 0$ kaikilla $v \in V$.*

Termien vastaavuus perustuu seuraavaan lemmaan. Huomaa, että yhteys ei riipu avaruuteen V valitusta kannasta.

Lemma 8.1.9. *Olkoon $(V, \langle\cdot,\cdot\rangle)$ sisätuloavaruus. Tällöin operaattori $f\colon V \to V$ on positiivinen, jos ja vain jos on olemassa avaruuden V kanta, jossa kuvauksen f matriisi on positiviisesti semidefiniitti.*

Todistus. Olkoon $(v_1, \ldots, v_n)$ avaruuden V kanta ja $\Phi\colon \mathbb{R}^{n\times 1} \to V$ isomorfismi $e_i \to v_i$. Tällöin $f = \Phi \circ f_A \circ \Phi^{-1}$, missä $A \in \mathbb{R}^{n\times n}$ on matriisi. Nyt kaikilla $x \in \mathbb{R}^{n\times 1}$ pätee

$$x^T Ax = x \cdot f_A(x) = \langle \Phi(x), \Phi(f_A(x)) \rangle = \langle \Phi(x), f(\Phi(x)) \rangle.$$

Näin ollen f on positiviinen, jos ja vain jos A on positiivisesti semidefiniitti. □

8.2 Sovellus: Neliömuodot ja toisen asteen käyrät

Symmetristen neliömatriisien luokittelussa esiintyvää funktiota $q_A\colon \mathbb{R}^{n\times 1} \to \mathbb{R}$,

$$x \mapsto x^T Ax,$$

missä $A \in \mathbb{R}^{n\times n}$ on siis symmetrinen neliömatriisi, kutsutaan matriisin A *neliömuodoksi* (engl. *quadratic form*). Näin ollen neliömetriisien luokittulu (semi)definiitteihin ja indefiniitteihin neliömatriiseihin on itseasiassa näiden neliömuotojen luokittelu.

Neliömuodoilla on tärkeä rooli klassisessa geometriassa, esimerkiksi kartioleikkausten teoriassa, mutta myös pääakselien teoriassa. Raapaistaan nyt hieman pintaa tästä asiasta.

Ennen varsinaista esimerkkiä, palautetaan mieleen neliömatriisien yhteys sisätuloihin.

Esimerkki 8.2.1. *Olkoon $A \in \mathbb{R}^{n\times n}$ positiivisesti definiitti matriisi. Tällöin funktio $\langle \cdot, \cdot \rangle\colon \mathbb{R}^{n\times 1} \times \mathbb{R}^{n\times 1} \to \mathbb{R}$, joka on määritelty kaavalla*

$$\langle x, y \rangle = x^T Ay$$

kaikilla $x, y \in \mathbb{R}^{n\times 1}$, on sisätulo avaruudessa $\mathbb{R}^{n\times 1}$. (Harjoitustehtävä) Lisäksi tämän sisätulon normille $\|\cdot\|\colon \mathbb{R}^{n\times 1} \to \mathbb{R}$ pätee

$$\|x\| = \sqrt{x^T Ax} = \sqrt{q_A(x)}.$$

Toisaalta, jos $\langle \cdot, \cdot \rangle\colon \mathbb{R}^{n\times 1} \to \mathbb{R}$ on sisätulo, niin on olemassa sellainen symmetrinen positiivisesti definiitti matriisi $A \in \mathbb{R}^{n\times n}$, että $x^T Ay = \langle x, y \rangle$ kaikilla $x, y \in \mathbb{R}^{n\times 1}$. (Harjoitustehtävä)

Tarkastellaan nyt yhteyttä toisen asteen käyriin ja aloitetaan koordinaattiakselien suuntaisista ellipseistä.

Esimerkki 8.2.2. *.*

Olkoot $a_1, \ldots, a_n > 0$ reaalilukuja ja olkoon

$$E = \{(x_1, \ldots, x_n) \in \mathbb{R}^n \colon \frac{x_1^2}{a_1} + \cdots + \frac{x_n^2}{a_n} \leq 1\}$$

(Piirrä tapaus $n = 2$ ja $a_1 = a_2 = 1$ sekä tapaus $n = 2$ ja $a_1 = 1$ sekä $a_2 = 2$.)

Olkoon $A \in \mathbb{R}^{n\times n}$ diagonaalimatriisi

$$A = \begin{bmatrix} \frac{1}{a_1} & & \\ & \ddots & \\ & & \frac{1}{a_n} \end{bmatrix}.$$

Tällöin jokaisella $x = [x_1 \cdots x_n]^T \in \mathbb{R}^{n\times 1}$ pätee

$$q_A(x) = x^T A x = [x_1 \cdots x_n] \begin{bmatrix} a_1 & & \\ & \ddots & \\ & & a_n \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \frac{x_1^2}{a_1} + \cdots + \frac{x_n^2}{a_n}.$$

Näin ollen

$$E = \{(x_1, \ldots, x_n) \in \mathbb{R}^n \colon q_A([x_1 \cdots x_n]^T) \le 1\}.$$

Huomautus 8.2.3. *Huomaa, että edellisessä esimerkissä, matriisi A määrittelee sisätulon, joten joukko E voidaan kirjoittaa myös tätä sisätuloa vastaavan normin $\|\cdot\|$ avulla muodossa*

$$E = \{(x_1, \ldots, x_n) \in \mathbb{R}^n \colon \|[x_1 \cdots x_n]^T\| = \sqrt{q_A([x_1 \cdots x_n]^T)} \le 1\}.$$

Joukko E on siis kaikkien niiden pisteiden $x \in \mathbb{R}^n$ joukko, joiden etäisyys origosta on alle 1 normin $(x_1, \ldots, x_n) \mapsto \sqrt{x^T A x}$ määräämässä metriikassa. (Tässä vektori $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ on samastettu sarakevektorin $x = [x_1, \ldots, x_n]^T$ kanssa.)

Tarkastellaan nyt samaa ilmiötä toisinpäin.

Esimerkki 8.2.4. *Olkoon $A \in \mathbb{R}^{n\times n}$ symmetrinen positiivisesti definiitti neliömatriisi ja tarkastellaan joukkoa*

$$E_A = \{x \in \mathbb{R}^{n\times 1} \colon q_A(x) \le 1\}.$$

Spektraalilauseen nojalla on olemassa sellainen ortonormaalimatriisi $P \in \mathbb{R}^{n\times n}$ ja diagonaalimatriisi $D \in \mathbb{R}^{n\times n}$, että $PDP^T = A$. Tarkastellaan nyt neliömuotoa q_A ja esitetään se neliömuodon q_D avulla.

Olkoon $x \in \mathbb{R}^{n\times n}$. Tällöin

$$q_A(x) = x^T P D P^T x = (P^T x)^T D (P^T x) = q_D(P^T x).$$

Näin ollen

$$\begin{aligned} E_A &= \{x \in \mathbb{R}^{n\times 1} \colon q_D(P^T x) \le 1\} \\ &= \{x \in \mathbb{R}^{n\times 1} \colon q_D(P^{-1} x) \le 1\} \\ &= \{Py \in \mathbb{R}^{n\times 1} \colon y \in \mathbb{R}^{n\times 1},\ q_D(y) \le 1\}. \end{aligned}$$

Määritellään nyt

$$E_D = \{y \in \mathbb{R}^{n\times 1} \colon q_D(y) \le 1\}.$$

Nyt

$$E_A = PE_D.$$

eli joukko E_A on joukon E_D kuva kuvauksessa $f_P\colon y \mapsto Py$.

Koska D on diagonaalimatriisi, niin saadaan, että

$$q_D([y_1 \cdots y_n]^T) = d_1 y_1^2 + \cdots + d_n y_n^2$$

missä $d_1, \ldots, d_n > 0$ ovat matriisin D diagonaalialkiot, eli

$$D = \begin{bmatrix} d_1 & & \\ & \ddots & \\ & & d_n \end{bmatrix}.$$

Näin ollen E_D on ellipsi avaruudessa $\mathbb{R}^{n\times 1}$, joka edellisen esimerkin merkinnöin vastaa ellipsiä

$$\{(y_1, \ldots, y_n) \in \mathbb{R}^n \colon \frac{y_1^2}{1/d_1} + \cdots + \frac{y_n^2}{1/d_n} \leq 1\}.$$

Koska matriisi P on ortogonaalimatriisi, niin kuvaus f_P on isometria, eli vektoreilla y ja Py on sama pituus pistetulon määräämässä metriikassa kaikilla $y \in \mathbb{R}^{n\times 1}$. Näin ollen myös joukko E_A on ellipsi. Siinä missä E_D on ellipsi, jonka akselit ovat koordinaattiakselien suuntaisia, ellipsin E_A akselit ovat matriisin P sarakevektoreiden suuntaisia. Näitä suuntia kutsutaan ellipsin E_A pääakseleiksi.

Määritelmä 8.2.5. *Olkoon $A \in \mathbb{R}^{n\times n}$ symmetrinen matriisi ja olkoot $P \in \mathbb{R}^{n\times n}$ sellainen ortogonaalimatriisi ja $D \in \mathbb{R}^{n\times n}$ sellainen diagonaalimatriisi, että $A = PDP^{-1}$. Neliömuodon q_A* pääekseliesitys *on neliömuoto $q_D\colon \mathbb{R}^{n\times 1} \to \mathbb{R}$. Lisäksi matriisin P sarakkeita kutsutaan neliömuodon q_A* pääakseleiksi.

Huomaa, että yllä esitetty määritelmä ei rajoitu positiivisesti definiitteihin neliömatriiseihin vaan on yleinen symmetristen matriisien määritelmä. Samat tarkastelut jotka tehtiin yllä positiivisesti definiiteille matriiseille on mahdollista tehdä yleisille symmetrisille matriiseille, sillä erotuksella, että yhteyttä sisätuloon ei tässä tapauksessa enää ole. Tarkastellaan kahta esimerkkiä.

Esimerkki 8.2.6. *Olkoot $a, b > 0$ ja*

$$H = \{(x_1, x_2) \in \mathbb{R}^2 \colon \frac{x_1^2}{a} - \frac{x_2^2}{b} = 1\}$$

Tällöin H on hyperbeli avaruudessa $\mathbb{R}^2$.

Olkoon

$$A = \begin{bmatrix} \frac{1}{a} & 0 \\ 0 & -\frac{1}{b} \end{bmatrix}.$$

Tällöin

$$H = \{(x_1, x_2) \in \mathbb{R}^2 \colon q_A([x_1 x_2]^T) = 1\}.$$

Esimerkki 8.2.7. *Olkoon*

$$M = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}$$

Tällöin $x \mapsto x^T Mx$ on neliömuoto

$$[x_1 \cdots x_4] \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = x_1^2 + x_2^2 + x_3^2 - x_4^2.$$

Matriisi M ei (selvästi) määrittele sisätuloa, mutta määrittelee niin sanotun Minkowskin sisätulon *avaruuteen $\mathbb{R}^4$. Minkowskin sisätuloa käytetään yleisessä suhteellisuusteoriassa. Huomaa, että toisin kuin edellisissä esimerkeissä, tässä tapauksessa*

$$\{x \in \mathbb{R}^{r \times 1} \colon q_M(x) = 0\}$$

ei ole piste, vaan paraboloidi avaruudessa $\mathbb{R}^{4\times 1}$.

Huomautus 8.2.8. *Matriisien neliömuotojen hyödyllisyys seuraa havainnosta, että toisen asteen polynomi yhtälöt voidaan ratkaista neliömuotojen avulla. Tarkemmin asiaa on käsitelty liitteessä D.*

8.3 Semidefiniitin matriisin ominaisarvot

Spektraalilauseen avulla on helppo havaita, että operaattori on positiviinen, jos ja vain jos sen ominaiarvot ovat ei-negatiivisia. Kirjataan tämä tulos matriiseille ja jätetään vastava todistus positiivisille operaattoreille harjoitustehtävksi.

Lause 8.3.1. *Olkoon $A \in \mathbb{R}^{n\times n}$ symmetrinen matriisi. Tällöin A on positiivisesti semidefiniitti, jos ja vain jos matriisin A ominaisarvot ovat ei-negatiivisia. Vastaavasti matriisi A on positiivisesti definiitti, jos ja vain jos matriisin A ominaisrvot ovat positiivisia.*

Todistus. Oletetaan, että A on positiviisesti semidefiniitti. Olkoon $\lambda \in \mathbb{R}$ matriisin A ominaisarvo ja $x \in \mathbb{R}^{n\times 1}$ vastaava ominaisvektori. Tällöin

$$0 \leq x^T Ax = x^T(\lambda x) = \lambda x^T x.$$

Koska $x^T x > 0$, niin $\lambda \geq 0$.

Oletetaan nyt, että A on symmetrinen matriisi, jonka ominaiarvot ovat ei-negatiivisia. Tällöin on olemassa sellainen ortogonaalimatriisi $P \in \mathbb{R}^{n\times n}$ ja sellainen diagonaalimatriisi $D \in \mathbb{R}^{n\times n}$, jonka diagonaalialkiot $\lambda_1, \ldots, \lambda_n$ ovat matriisin A ominaisarvoja, että $A = PDP^T$. Lisäksi matriisin P sarakkeet $(v_1, \ldots, v_n)$ muodostavat avaruuden $\mathbb{R}^{n\times 1}$ kannan. Olkoon nyt $x \in \mathbb{R}^{n\times 1}$ ja olkoot $a_i, \ldots, a_n \in \mathbb{R}$ sellaisia, että

$x = a_1 v_1 + \cdots + a_n v_n$. (Itseasiassa $a_i = x \cdot v_i$ jokaisella $i \in \{1, \ldots, \}$.) Olkoon $a \in \mathbb{R}^{n\times 1}$ sarakevektori $a = [a_1 \cdots a_n]^T$. Tällöin $x = Pa$ ja saadaan

$$x^T A x = (Pa)^T PDP^T (Pa) = a^T P^T PDP^T Pa = a^T D a = \lambda_1 a_1^2 + \cdots + \lambda_n a_n^2 \geq 0.$$

Matriisi A on siis positiivisesti semidefiniitti.

Vastaavat päättelyt antavat väitteen toisen osan, että matriisi A on positiivisesti definiitti, jos ja vain jos matriisin A ominaisrvot ovat positiivisia. $\square$

Edellinen lause yhdessä spektraalilauseen kanssa karakterisoi positiivisesti semidefiniitin matriisin kääntyvyyden.

Olkoon $A \in \mathbb{R}^{n\times n}$ positiivisesti semidefiniitti n matriisi. Tällöin A on kääntyvä, jos ja vain jos A on positiivisesti definiitti.

8.4 Positiivisesti semidefiniitin neliömatriisin neliöjuuri

Positiivisten operaattoreiden mielenkiintoinen ominaisuus on, että niille voidaan määritellä neliöjuuri.

Määritelmä 8.4.1. *Olkoon V vektoriavaruus. Operaattori $g\colon V \to V$ on operaattorin $f\colon V \to V$* neliöjuuri, *jos $g^2 = g \circ g = f$.*

Matriiseille sama määritelmä saa seuraavan muodon.

Määritelmä 8.4.2. *Matriisi $B \in \mathbb{R}^{n\times n}$ on matriisin $A \in \mathbb{R}^{n\times n}$ neliöjuuri, jos $B^2 = A$.*

Matriisien kielelle käännettynä positiivisen operaattorin neliöjuuren olemassaolo tarkoittaa, että jokaiselle positiivisesti semidefiniitille matriisille $A \in \mathbb{R}^{n\times n}$ on olemassa sellainen positiivisesti semidefiniitti matriisi B, että $B^2 = A$. Tämä on helppo seuraus spektraalilauseesta. Kirjataan tämä tulos kahdella tavalla. Aloitetaan teknisestä huomiosta, josta tulokset seuraavat.

Lemma 8.4.3. *Olkoon $P \in \mathbb{R}^{n\times n}$ ortonormaalimatriisi ja*

$$D = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} \in \mathbb{R}^{n\times n}$$

diagonaalimatriisi, jonka diagonaalialkiot ovat ei-negatiivisia. Tällöin matriisille $B = PDP^T$ pätee

$$B^2 = PD^2P^T.$$

Lisäksi matriisin D diagonaalialkiot ovat matriisin B ominaisarvoja, matriisin D^2 diagonaalialkiot ovat matriisin B^2 ominaisarvoja ja matriisin P sarakkeet ovat matriiseiden B ja B^2 ominaisvektoreita.

Todistus. Koska $P^T P = I$, niin

$$B^2 = PDP^T PDP^T = PDDP^T = PD^2 P^T.$$

Osoitetaan nyt, että matriisin P sarakkeet ovat molempien matriisien ominaisvektoreita. Olkoon $P = [v_1 \cdots v_n]$. Tällöin jokaisella $i \in \{1, \dots, n\}$ pätee

$$Bv_i = (PDP^T)v_i = PD(P^T v_i) = PDe_i = P(\lambda_i e_i) = \lambda_i Pe_i = \lambda_i v_i$$

eli matriisin P sarakkeet ovat matriisin B ominaisvektoreita ja matriisin D diagonaalialkiot ovat niitä vastaavat ominaisarvot. Vastavasti

$$B^2 v_i = BBv_i = B(\lambda_i v_i) = \lambda_i Bv_i = \lambda_i^2 v_i.$$

□

Lause 8.4.4. *Olkoon $A \in \mathbb{R}^{n\times n}$ positiivisesti semidefiniitti matriisi. Tällöin on olemassa positiviisesti semidefiniitti matriisi $B \in \mathbb{R}^{n\times n}$, jolle pätee $B^2 = A$.*

Todistus. Spektraalilauseen nojalla matriisilla A on ominaisarvot $\lambda_1, \dots, \lambda_n$ sekä sellainen ortonormaalimatriisi $P \in \mathbb{R}^{n\times n}$, joka koostuu matriisin A ominaivektoreista, ja sellainen diagonaalimatriisi $D \in \mathbb{R}^{n\times n}$, jonka diagonaalialkiot ovat matriisin A ominaisarvoja, että $A = PDP^T$. Merkitään

$$D = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}.$$

Lauseen 8.3.1 perusteella ominaisarvot $\lambda_1, \cdots, \lambda_n$ ovat ei-negatiivisia. Näin ollen matriisi

$$D^{1/2} = \begin{bmatrix} \lambda_1^{1/2} & & \\ & \ddots & \\ & & \lambda_n^{1/2} \end{bmatrix}$$

on hyvin määritelty, jolle matriisitulon määritelmän nojalla pätee

$$D^{1/2} D^{1/2} = D.$$

Näin ollen matriisi $B = PD^{1/2}P^T$ on positiivisesti semidefiniitti matriisi, jolle pätee lemman 8.4.3 perusteella pätee $B^2 = A$. □

Positiivisella operaattorilla on itseasiassa täsmälleen yksi neliöjuuri, joka on positiivinen operaattori. Sama tulos pätee tietysti myös semidefiniiteille matriiseille. Huomaa kuitenkin, että matriisiesitys $PD^{1/2}P^T$ ei ole yksikäsitteinen vaan riippuu sarakevektoreiden järjestyksestä.

Lause 8.4.5. *Olkoon* $(V, \langle\cdot,\cdot\rangle)$ *sisätuloavaruus ja* $f\colon V \to V$ *postiviinen operaattori. Tällöin on olemassa yksikäsitteinen positiivinen operaattori* $g\colon V \to V$*, jolle pätee* $f = g^2 = g \circ g\colon V \to V$*.*

Todistus. Olkoon $(v_1, \ldots, v_n)$ avaruuden V kanta ja $A \in \mathbb{R}^{n\times n}$ kuvauksen f matriisi tässä kannassa. Koska A on positiivisesti semidefiniitti, niin lauseen 8.4.4 perusteella on olemassa positiivisesti semidefiniitti matriisi $B \in \mathbb{R}^{n\times n}$, jolle pätee $B^2 = A$. Olkoon nyt $g\colon V \to V$ lineaarikuvaus, jonka matriisi on B. Tällöin g on positiivisesti definiitti ja $g^2 = g \circ g = f$.

Osoitetaan nyt yksikäsitteisyys. Olkoon $h\colon V \to V$ sellainen positiivinen operaattori, että $h^2 = f$. Olkoon $v_1, \ldots, v_n$ avaruuden V kanta, joka koostuu operaattorin f ominaisvektoreista, ja olkoot $\lambda_1, \ldots, \lambda_n$ vastaavat ominaisarvot, eli $f(v_i) = \lambda_i v_i$ kaikilla $i \in \{1, \ldots, n\}$. Osoitetaan, että $h(v_i) = \sqrt{\lambda_i} v_i$ kaikilla $i \in \{1, \ldots, n\}$. Tämä osoittaa halutun yksikäsitteisyyden, sillä tällöin $h(v_i) = g(v_i)$ kaikilla $i \in \{1, \ldots, n\}$.

Olkoon $v \in V$ operaattorin f ominaisvektori ja $\lambda \in \mathbb{R}$ vastaava ominaisarvo. Osoitetaan, että $h(v) = \sqrt{\lambda} v$.

Koska h on positiivinen operaattori, on olemassa avaruuden V kanta $w_1, \ldots, w_n$, joka koostuu operaattorin h ominaisvektoreista. Olkoot $\mu_1, \ldots, \mu_n$ vastaavat ominaisarvot. Tällöin $h^2(w_i) = h(h(w_i)) = h(\mu_i w_i) = \mu_i^2 w_i$ jokaisella $i \in \{1, \ldots, n\}$. Koska $w_1, \ldots, w_n$ on kanta, niin on olemassa sellaiset $a_1, \ldots, a_n \in \mathbb{R}$, että

$$v = a_1 w_1 + \cdots + a_n w_n.$$

Tällöin

$$\begin{aligned} a_1\lambda w_1 + \cdots + a_n \lambda w_n = \lambda v = f(v) &= f(a_1 w_1 + \cdots + a_n w_n) \\ &= a_1 f(w_1) + \cdots + a_n f(w_n) \\ &= a_1 \mu_1^2 w_1 + \cdots + a_1 \mu_n^2 w_n. \end{aligned}$$

Näin ollen $a_i\lambda = a_i\mu_i^2$ kaikilla $i \in \{1, \ldots, n\}$. Jokaisella $i \in \{1, \ldots, n\}$ pätee siis, joko $a_i = 0$ tai $\mu_i = \sqrt{\lambda}$. Näin ollen

$$h(v) = a_1 h(w_1) + \cdots + a_n h(w_n) = a_1\sqrt{\lambda} w_1 + \cdots + a_n \sqrt{\lambda} w_n = \sqrt{\lambda} v.$$

□

8.5 Sovellus: Choleskyn hajotelma

Cholesky hajotelmalla tarkoitetaan positiivisesti semi-definitiin matriisin $A \in \mathbb{R}^{n\times n}$ esittämistä kahden kolmiomatriisin tulona

$$A = R^T R$$

missä R on yläkolmiomatriisi, jonka diagonaalielementit ovat ei-negatiivisia. Koska Choleskyn hajotelma kirjoiteaan usein yläkolmiomatriisin R sijaan alakolmiomatriisin $T = R^T$ avulla muodossa

$$A = TT^T,$$

niin jatkossa käytetään tätä esitystapaa Choleskyn hajotelmasta. Tarkka määritelmä on siis seuraava.

Määritelmä 8.5.1. *Positiivisesti semi-definiitin matriisin $A \in \mathbb{R}^{n\times n}$ Choleskyn hajotelma on $A = TT^T$, missä $T \in \mathbb{R}^{m\times n}$ on alakolmiomatriisi, jolle pätee $t_{jj} \geq 0$ kaikilla $j = 1, \ldots, n$.*

Huomautus 8.5.2. *Määritelmässä oletetaan, että matriisi A on positiivisesti semidefiniitti. Tämä ei ole lisäoletus matriisille A, sillä kaikki matriisit, jotka voidaan kirjoittaa muodossa TT^T, jollakin alakolmiomatriisilla T, ovat positiivisesti semidefiniittejä. Tämä seuraa suoraan (tutusta) huomioista, että kaikilla $x \in \mathbb{R}^{n\times 1}$ pätee*

$$(TT^T)x \cdot x = T(T^T x) \cdot x = (T^T x) \cdot (T^T x) \geq 0.$$

Choleskyn hajotelman hyödyllisyys liittyy yhtälöryhmien numeeriseen ratkaisemiseen. Tarkastellaan matriisiyhtälöä $Ax = y$, missä $A \in \mathbb{R}^{n\times n}$. Oletetaan, että tunnetaan matriisin A Cholesky hajotelma, eli että $A = TT^T$, missä $T \in \mathbb{R}^{n\times n}$ on alakolmiomatriisi. Tällöin

$$T(T^T x) = Ax = y$$

eli voidaan aloittaa ratkaisemalla yhtälö

$$Tz = y$$

Tämä yhtälöryhmä voidaan ratkaista tehokkaalla eliminointimenetelmällä kuten seuraava esimerkki osoittaa. Tämän jälkeen riittää ratkaista yhtälö $T^T x = z$ samalla menetelmällä.

Esimerkki 8.5.3. *Olkoot*

$$T = \begin{bmatrix} 2 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 1 & 2 \end{bmatrix} \quad \text{ja} \quad y = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

Tällöin yhtälöstä

$$\begin{bmatrix} 2 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 1 & 2 \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

seuraa välittömästi, että $z_1 = 1/2$, $z_2 = 2 - z_1 = 3/2$ ja $z_3 = \frac{1}{2}(3 - z_2) = 3/4$.

Seuraava lause on luvun päätulos.

Lause 8.5.4. *Positiivisesti semidefiniitillä matriisilla $A \in \mathbb{R}^{n\times n}$ on Choleskyn hajotelma.*

Todistus. Lauseen 8.4.4 nojalla matriisilla A on olemassa yksikäsitteinen positiviisesti semidefiniitti neliöjuuri B. Olkoon $k = \mathrm{rank}(A)$.

Olkoon nyt ortogonaalimatriisi $Q \in \mathbb{R}^{n\times k}$ ja yläkolmiomatriisi $R \in \mathbb{R}^{k\times n}$ matriisin B QR-hajotelma, eli $B = QR$. Koska B on symmetrinen matriisin, niin

$$R^T R = R^T Q^T QR = (QR)^T QR = B^T B = B^2 = A.$$

Valitaan nyt $T = R^T$. □

Huomautus 8.5.5. *Huomaa, että mikäli* $\mathrm{rank}(A) < n$, *niin edellisen lauseen todistus antaa alakolmiomatriisin* T, *joka ei ole neliömatriisi vaan* $n \times k$*-matriisi. Huomaa kuitenkin, että lisäämällä matriisiin* Q *nollasarakkeita ja matriisiin* R *nollarivejä, molemmat matriisit voidaan täydentää neliömatriiseiksi* Q' *ja* R', *joille pätee, että* $(R')^T R = A$. *Tällöin* $T = (R')^T$ *on Cholesky hajotelman mukainen alakolmiomatriisi. (Harjoitustehtävä)*

Positiivisesti definiittien matriisien Choleskyn hajotelma on yksikäsitteinen ja alakolmiomatriisi on neliömatriisi. Todistetaan tämä seuraavaksi.

Lause 8.5.6. *Positiivisesti definiitin matriisin* $A \in \mathbb{R}^{n\times n}$ *Choleskyn hajotelma* $A = TT^T$ *on yksikäsitteinen.*

Todistus. Oletetaan, että T ja L ovat sellaisia alakolmiomatriiseja, että $A = TT^T$ ja $A = LL^T$.

Koska A on positiviisesti definiitti, niin se on kääntyvä korollaarin 8.3 perusteella. Näin ollen myös matriisit T ja L ovat kääntyviä.

Koska $TT^T = LL^T$, niin

$$I = T^{-1}LL^T(T^T)^{-1} = (L^{-1}T)(T^T(L^T)^{-1})^{-1} = (L^{-1}T)\left((L^{-1}T)^T\right)^{-1}.$$

Näin ollen

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

Koska matriisi $L^{-1}T$ on alakolmiomatriisi ja $(L^{-1}T)^T$ on yläkolmiomatriisi, niin $L^{-1}T$ on diagonaalimatriisi. Merkitään $D = L^{-1}T$. Tällöin $T = LD$ ja

$$LL^T = A = TT^T = (LD)(LD)^T = LD^2L^T.$$

Koska L on kääntyvä, niin $D^2 = I$. Näin ollen matriisin D jokainen diagonaalialkio on joko 1 tai -1. Koska matriisien T ja L diagonaalialkiot ovat positiviisia, niin $D = I$ ja $T = L$. □

Positiivisesti semidefiniitin matriisin Choleskyn hajotelman ei tarvitse olla yksikäsitteinen, kuten seuraava (helppo) esimerkki osoittaa.

Esimerkki 8.5.7. *Osoitetaan, että matriisilla*

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

on useampi Choleskyn hajotelma ratkaisemalla matriisin T kertoimet suoraan suoraan yhtälöstä $A = TT^T$.

Olkoon

$$T = \begin{bmatrix} a & 0 & 0 \\ b & c & 0 \\ d & e & f \end{bmatrix}$$

Tällöin

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} a & 0 & 0 \\ b & c & 0 \\ d & e & f \end{bmatrix} \begin{bmatrix} a & b & d \\ 0 & c & e \\ 0 & 0 & f \end{bmatrix} = \begin{bmatrix} a^2 & ab & ad \\ ab & b^2 + c^2 & bd + ce \\ ad & bd + ce & d^2 + e^2 + f^2 \end{bmatrix}.$$

Näin ollen matriisi T on yhtälön $A = TT^T$ ratkaisu, jos ja vain jos $a = 1$, $b = 0$, $c = 0$, $d = 0$ ja $e^2 + f^2 = 1$. Näin ollen yhtälöllä $A = TT^T$ on useampia ratkaisuja.

8.5.1 Choleskyn hajotelman iteratiivinen ratkaiseminen

Todistetaan ratkaisualgoritmia varten aputulos.

Lemma 8.5.8. *Olkoon*

$$A = \begin{bmatrix} a_{11} & b^T \\ b & \tilde{A} \end{bmatrix} \in \mathbb{R}^{n \times n},$$

positiivisesti semidefiniitti matriisi, missä $\tilde{A} \in \mathbb{R}^{(n-1)\times(n-1)}$ on symmetrinen matriisi ja $b \in \mathbb{R}^{(n-1)\times 1}$ on sarakevektori. Jos $a_{11} = 0$, niin $b = 0$. Jos $a_{11} \neq 0$, niin matriisi $\tilde{A} - \frac{1}{a_{11}} bb^T$ on positiviisesti semidefiniitti.

Todistus. Tarkastellaan ensin tapausta $a_{11} = 0$. Osoitetaan väite induktiolla dimension n suhteen. Aloitetaan ensimmäisestä epätriviaalista tapauksesta $n = 2$. Huomaa, että tapauksessa $n = 1$, vektori b on ns. tyhjä vektori.

Koska $a_{11} = 0$, niin

$$A = \begin{bmatrix} 0 & b \\ b & d \end{bmatrix},$$

missä $b, d \in \mathbb{R}$. Koska A on positiivisesti semidefiniitti, niin jokaisella $x \in \mathbb{R}$ pätee

$$0 \leq \begin{bmatrix} x & 1 \end{bmatrix} \begin{bmatrix} 0 & b \\ b & d \end{bmatrix} \begin{bmatrix} x \\ 1 \end{bmatrix} = \begin{bmatrix} b & xb + d \end{bmatrix} \begin{bmatrix} x \\ 1 \end{bmatrix} = xb + xb + d = 2xb + d.$$

Näin ollen $b = 0$.

Oletetaan nyt, että väite pätee kaikille positiivisesti semidefiniiteille $(n-1)\times(n-1)$-matriiseille. Olkoon nyt A positiivisesti semidefiniitti $n \times n$-matriisi

$$A = \begin{bmatrix} 0 & v^T & b' \\ v & \tilde{A}' & w^T \\ b' & w & d \end{bmatrix},$$

missä $v \in \mathbb{R}^{(n-2)\times 1}$ ja $w \in \mathbb{R}^{1\times(n-2)}$ ovat vektoreita, $b', d \in \mathbb{R}$ lukuja ja $\tilde{A}' \in \mathbb{R}^{(n-2)\times(n-2)}$ matriisi.

Koska matriisin A $(n-1)$:s pääminori

$$A_{n-1} = \begin{bmatrix} 0 & v^T \\ v & \tilde{A}' \end{bmatrix},$$

on positiivisesti semidefiniitti, niin induktio-oletuksen nojalla $v = 0$. Näin ollen

$$A = \begin{bmatrix} 0 & 0 & b' \\ 0 & \tilde{A}' & w^T \\ b' & w & d \end{bmatrix}.$$

Koska A on positiivisesti semidefiniitti, niin

$$0 \leq \begin{bmatrix} x & 0 & \cdots & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & b' \\ 0 & \tilde{A}' & w^T \\ b' & w & d \end{bmatrix} \begin{bmatrix} x \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} = 2xb' + d.$$

Näin ollen $b' = 0$ ja induktioaskel on todistettu.

Osoitetaan nyt, että tapauksessa $a_{11} \neq 0$ matriisi $\tilde{B} = \tilde{A} - \frac{1}{a_{11}} bb^T$ on positiivisesti semidefiniitti. Olkoon $x \in \mathbb{R}^{(n-1)\times 1}$ ja

$$y = \begin{bmatrix} -\frac{1}{a_{11}} b^T x \\ x \end{bmatrix} \in \mathbb{R}^{n\times 1}.$$

Koska A on positiivisesti semidefiniitti, niin

$$\begin{aligned} 0 \leq y^T A y &= \left[\begin{array}{c|c} -\frac{1}{a_{11}} b^T x & x^T \end{array}\right] \begin{bmatrix} a_{11} & b^T \\ b & \tilde{A} \end{bmatrix} \begin{bmatrix} -\frac{1}{a_{11}} b^T x \\ x \end{bmatrix} \\ &= \left[\begin{array}{c|c} -b^T x + x^T b & -\frac{1}{a_{11}} b^T x b^T + x^T \tilde{A} \end{array}\right] \begin{bmatrix} -\frac{1}{a_{11}} b^T x \\ x \end{bmatrix} \\ &= \frac{1}{a_{11}} (b^T x)^2 - \frac{1}{a_{11}} x^T b b^T x - \frac{1}{a_{11}} b^T x b^T x + x^T \tilde{A} x \\ &= x^T \tilde{A} x - \frac{1}{a_{11}} x^T (bb^T) x = x^T \tilde{B} x. \end{aligned}$$

Väite on näin osoitettu. □

Algoritmi iteratiiviselle ratkaisemiselle

Choleskyn hajotelman iteratiivinen ratkaiseminen perustuu seuraavaan induktio todistukseen. Selvästi Cholesky-hajotelma voidaan löytää jokaiselle positiivisesti semidefiniitille 1×1-matriisille. Oletetaan, että Cholesky-hajotelma osataan löytää jokaiselle positiivisesti semidefiniitille $(n-1) \times (n-1)$-matriisille.

Olkoon
$$A = \begin{bmatrix} a_{11} & b^T \\ b & \tilde{A} \end{bmatrix} \in \mathbb{R}^{n\times n}$$
positiivisesti semidefiniitti matriisi, missä $\tilde{A} \in \mathbb{R}^{(n-1)\times(n-1)}$ on symmetrinen positiivisesti semidefiniitti matriisi ja $b \in \mathbb{R}^{(n-1)\times 1}$ on sarakevektori.

Olkoon nyt
$$T = \begin{bmatrix} a & 0 \\ c & \tilde{T} \end{bmatrix} \in \mathbb{R}^{n\times n}$$
alakolmiomatriisi, missä $\tilde{T} \in \mathbb{R}^{(n-1)\times(n-1)}$ on alakolmiomatriisi, jolla on ei-negatiivinen diagonaali, $c \in \mathbb{R}^{1\times(n-1)}$ on rivivektori ja $a \in \mathbb{R}$ on ei-negatiivinen reaaliluku.

Koska
$$TT^T = \begin{bmatrix} a & 0 \\ c & \tilde{T} \end{bmatrix} \begin{bmatrix} a & c^T \\ 0 & \tilde{T}^T \end{bmatrix} = \begin{bmatrix} a^2 & ac^T \\ ac & cc^T + \tilde{T}\tilde{T}^T \end{bmatrix},$$
niin matriisi T ratkaisee yhtälön
$$A = TT^T,$$
jos ja vain jos matriisi $\tilde{T}$, vektori c ja luku a toteuttavat yhtälöt
$$\begin{cases} a^2 &= a_{11} \\ ac &= b \\ cc^T + \tilde{T}\tilde{T}^T &= \tilde{A} \end{cases}$$
Käsitellään kaksi tapausta.

Tapaus $a_{11} = 0$. Tässä tapauksessa lemman 8.5.8 nojalla pätee $b = 0$. Koska $\tilde{A}$ on positiivisesti definiitti $(n-1)\times(n-1)$-matriisi, niin on olemassa sellainen alakolmiomatriisi $\tilde{T}$, että $\tilde{T}\tilde{T}^T = \tilde{A}$. Näin ollen voidaan valita $a = 0$, $c = 0$ ja yhtälön $\tilde{A} = \tilde{T}\tilde{T}^T$ ratkaisu $\tilde{T}$.

Tapaus $a_{11} > 0$. Valitaan nyt $a = \sqrt{a_{11}}$ ja $c = \frac{1}{a}b$. Koska lemman 8.5.8 perusteella matriisi $\tilde{A} - \frac{1}{a_{11}}bb^T$ on positiivisesti semidefiniitti, niin induktio-oletuksen nojalla on olemassa sellainen alakolmiomatriisi $\tilde{T}$, että $\tilde{A} - \frac{1}{a_{11}}bb^T = \tilde{T}\tilde{T}^T$.

Tämä päättää induktioaskeleen todistuksen.

Huomautus 8.5.9. *Tässä iteratiivisessa ratkaisussa riittää siis ratkaista luku a ja vektori c matriisin A tapauksessa ja siirtyä sen jälkeen ratkaisemaan vastaavat luku ja vektori matriisille $\tilde{A}$ jne.*

Huomautus 8.5.10. *Lukija on saattanut jo huomata, että olemme antaneet konstruktiivisen todistuksen positiivisesti semidefiniitin matriisin Choleskyn hajotelman olemassa ololle.*

8.6 Polaarihajotelma

Polaarihajotelma jakaa yleisen operaattorin $f\colon V \to V$ kahteen osaan $f = S \circ g$, missä $g\colon V \to V$ on positiivinen operaattori ja $S\colon V \to V$ on isometria. Tuloksen syvälllisyys on siinä, että heuristisesti jokainen operaattori on isometriaa vailla positiivinen. Matriisien kielellä tämä tarkoittaa sitä, että jokaisella $A \in \mathbb{R}^{n\times n}$ on olemassa sellainen positiivisesti semidefiniitti matriisi $B \in \mathbb{R}^{n\times n}$ ja sellainen ortogonaalimatriisi $O \in \mathbb{R}^{n\times n}$, että

$$A = OB.$$

Polaarihajotelman positiivinen operaattori g on itseasiassa operaattorin $f^*f = f^* \circ f\colon V \to V$ neliöjuuri. Tätä varten tulee (tietysti) tietää, että $f^* \circ f$ on positiivinen operaattori. Tämän faktan todistus on sama kuin matriisin $A^T A$ positiivi semidefiniittisyyden todistus. Todistetaan se kuitenkin suoraan määritelmistä.

Lemma 8.6.1. *Olkoon $(V, \langle\cdot,\cdot\rangle)$ sisätuloavaruus ja $f\colon V \to V$ operaattori. Tällöin $f^*f\colon V \to V$ on positiivinen operaattori.*

Todistus. Olkoon $v \in V$. Adjungaatin määritelmän nojalla pätee

$$\langle (f^* \circ f)(v), v\rangle = \langle f^*(f(v)), v\rangle = \langle f(v), f(v)\rangle \geq 0.$$

□

Otetaan nyt käyttöön käyttöön merkintä operaattorin f^*f neliöjuurelle.

Määritelmä 8.6.2. *Olkoon $(V, \langle\cdot,\cdot\rangle)$ sisätuloavaruus ja $f\colon V \to V$ operaattori. Merkitään $\sqrt{f^*f}\colon V \to V$ sitä yksikäsitteistä operaattoria, joka on operaattorin $f^* \circ f\colon V \to V$ positiivisesti definiitti neliöjuuri, eli positiivista operaattoria, jolle pätee $\sqrt{f^*f} \circ \sqrt{f^*f} = f^* \circ f$.*

Lause 8.6.3 (Polaarihajotelma)**.** *Olkoon $(V, \langle\cdot,\cdot\rangle)$ sisätuloavaruus ja $f\colon V \to V$ operaattori. Tällöin on olemassa sellainen isometria $S\colon V \to V$, että*

$$f = S \circ \sqrt{f^*f}.$$

Polaarihajotelman ytimessä on havainto, että jokainen vektori $v \in V$ kuvautuu kuvauksissa f ja $\sqrt{f * f}$ saman pituisiksi, eli $\|f(v)\| = \|\sqrt{f^*f}(v)\|$. Tämä on välttämätön ehto isometrian S olemassaololle. Kirjataan tämä havainto lemmaksi.

Lemma 8.6.4. *Olkoon $(V, \langle\cdot,\cdot\rangle)$ sisätuloavaruus ja $f\colon V \to V$ operaattori. Tällöin*

$$\|f(v)\| = \|\sqrt{f^*f}(v)\|$$

kaikilla $v \in V$.

Todistus. Olkoon $v \in V$. Koska $f^* \circ f = \sqrt{f^*f} \circ \sqrt{f^*f}$, niin

$$\begin{aligned}\|f(v)\|^2 &= \langle f(v), f(v)\rangle = \langle v, f^*(f(v))\rangle = \langle v, \sqrt{f^*f}\sqrt{f^*f}(v)\rangle \\ &= \langle \sqrt{f^*f}(v), \sqrt{f^*f}(v)\rangle = \|\sqrt{f^*f}(v)\|^2.\end{aligned}$$

□

Polaarihajotelman todistus. Koska $\sqrt{f^*f}$ on positiivinen operaattori, voidaan valita avaruuden V kanta, joka koostuu operaattorin $\sqrt{f^*f}$ ominaisvektoreista $v_1, \dots, v_n$. Nämä vektorit voidaan järjestää niin, että vastaavat ominaisarvot ovat suuruusjärjestyksessä $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. Olkoon lisäksi $k \in \{1, \dots, n\}$ suurin luku, jolle $\lambda_k \neq 0$.

Olkoon nyt $V_1 \subset V$ operaattorin $\sqrt{f^*f}$ kuva, eli $V_1 = \operatorname{im} \sqrt{f^*f}$. Koska $v_1, \dots, v_n$ on avaruuden V kanta ja $\sqrt{f^*f}(v_i) = \lambda_i v_i$ jokaisella $i \in \{1, \dots, n\}$, niin

$$V_1 = \operatorname{Sp}\{f(v_1), \dots, f(v_n)\} = \operatorname{Sp}\{v_1, \dots, v_k\},$$

missä $v_1, \dots, v_k$ on aliavaruuden V_1 kanta.

Määritellään nyt lineaarikuvaus $S_1 \colon V_1 \to V$ kaavalla $v_i \mapsto \frac{1}{\lambda_i} f(v_i)$ jokaisella $i \in \{1, \dots, k\}$. Osoitetaan ensin, että

$$S_1 \circ \sqrt{f^*f} = f.$$

Havaitaan ensin, että jokaisella $i \in \{1, \dots, k\}$ pätee Tällöin

$$(S_1 \circ \sqrt{f^*f})(v_i) = S\left(\sqrt{f^*f}(v_i)\right) = S(\lambda_i v_i) = \lambda_i S(v_i) = \lambda_i \frac{1}{\lambda_i} f(v_i) = f(v_i).$$

Olkoon nyt $i \in \{k+1, \dots, n\}$. Tällöin lemman 8.6.4 nojalla

$$\|f(v_i)\| = \|\sqrt{f^*f}(v_i)\| = \|\lambda_i v_i\| = 0.$$

Näin ollen

$$(S \circ \sqrt{f^*f})(v_i) = S\left(\sqrt{f^*f}(v_i)\right) = S(\lambda_i v_i) = 0 = f(v_i).$$

Koska $v_1, \dots, v_n$ on avaruuden V kanta, niin

$$S_1 \circ \sqrt{f^*f} = f.$$

Osoitetaan nyt, että S_1 on isometria. Olkoon $v \in V_1$. Tällöin on olemassa sellainen $w \in V$, että $v = \sqrt{f^*f}(w)$. Näin ollen lemman 8.6.4 nojalla pätee

$$\|S_1(v)\| = \|S_1(\sqrt{f^*f}(w))\| = \|f(w)\| = \|\sqrt{f^*f}(w)\| = \|v\|.$$

Kuvaus S_1 on siis isometria $S_1 \colon V_1 \to W_1$, missä $W_1 = \operatorname{im} S_1 = \operatorname{im} f$. Lauseen 3.1.8 nojalla on olemassa sellainen isometria $S \colon V \to V$, että $S|_{V_1} = S_1$. Koska $V_1 = \operatorname{im} \sqrt{f^*f}$, niin

$$S \circ \sqrt{f^*f} = S_1 \circ \sqrt{f^*f} = f.$$

□

Luku 9

Yleisen neliömatriisin diagonalisointi

Luvun tärkein tulos on todistaa lineaarioperaattoreiden singulaariarvohajotelma (luku 9.1):

> Olkoon $A \in \mathbb{R}^{n\times n}$ neliömatriisi ja $s_1, \dots, s_n$ sen singulaariarvot. Tällöin on olemassa sellaiset ortogonaaliset matriisit $P \in \mathbb{R}^{n\times n}$ ja $Q \in \mathbb{R}^{n\times n}$ sekä diagonaalimatriisi $D \in \mathbb{R}^{n\times n}$, että
> $$A = PDQ^T.$$

Matriisin A singulaariarvohajotelma PDQ^T vaikuttaa muotonsa puolesta spektraalilauseen antamalta hajotelmalta $A = QDQ^T$. Näillä kahdella hajotelmalla on kuitenkin kaksi merkittävää eroa. Ensinnäkin siinä missä spektraalilause pätee ainoastaan symmetrisille matriiseille, niin singulaariarvohajotelma pätee kaikille neliömatriiseille. Toiseksi, spektraalilauseessa diagonaalimatriisi D on matriisiin A ominaisarvojen matriisi ja matriisi Q on matriisin A ominaisvektorien matriisi. Yleisen matriisin singulaariarvoajotelmassa näin ei ole, sillä yleisellä matriisilla A ei välttämättä edes ole ominaisarvoja.

Singulaariarvohajotelman olemassaolon todistus antaa tulkinnan singulaariarvohajotelman matriiseille P, Q ja D. Matriisi D koostuu ns. singulaariarvoista, jotka ovat matriisin A^TA ominaisarvojen neliöjuuria. Lisäksi todistuksesta tullaan havaitsemaan, että matriisi Q on matriisin A^TA ominaisvektoreiden matriisi. Todistuksesta ei kuitenkaan välttämättä heti huomaa, että matriisi P on matriisin AA^T ominaisvektoreiden matriisi. Tämän seikan voi kuitenkin havaita helposti singulaariarvohajotelmasta seuraavasti.

Huomautus 9.0.1. *Olkoon $A = PDQ^T$ matriisin $A \in \mathbb{R}^{n\times n}$ singulaariarvohajotelma. Tällöin*

$$AA^T = (PDQ^T)(PDQ^T)^T = PDQ^T(Q^T)^TD^TP^T = PDQ^TQDP^T = PD^2P^T.$$

Näin ollen diagonaalimatriisi D^2 on matriisin AA^T ominaisarvojen matriisi ja matriisin P sarakkeet koostuvat matriisin AA^T ominaisvektoreista. Tämän voi havaita seuraavasti. Olkoot $d_1, \dots, d_n$ matriisin D diagonaalialkiot, eli $D = \begin{bmatrix} d_1e_1 & \cdots & d_ne_n \end{bmatrix}$.

Tällöin $D^2 = \begin{bmatrix} d_1^2 e_1 & \cdots & d_n^2 e_n \end{bmatrix}$. *Koska* $P = \begin{bmatrix} p_1 & \cdots & p_n \end{bmatrix}$ *on ortogonaalimatriisi, niin* $P^T = P^{-1}$ *ja saadaan*

$$(PD^2P^T)p_i = PD^2(P^Tp_i) = PD^2e_i = P(d_i^2e_i) = d_i^2Pe_i = d_i^2p_i.$$

Näin ollen jokainen matriisin P *sarake on matriisin* PD^2P^T *ominaisvektori. Tarkemmin sanottuna, matriisin* P *i:s sarake* p_i *on matriisin* PD^2P^T *ominaisvektori ominaisarvolla* d_i^2.

9.1 Singulaariarvot ja singulaariarvohajotelman olemassaolo

Määritelmä 9.1.1. *Olkoon* $(V, \langle\cdot,\cdot\rangle)$ *sisätuloavaruus. Luku* $s \in \mathbb{R}$ *on lineaarioperaattorin* $f\colon V \to V$ singulaariarvo, *jos* s *on operaattorin* $\sqrt{f^*f}\colon V \to V$ *ominaisarvo.*

Huomautus 9.1.2. *Huomaa, että* $\sqrt{f^*f}$ *on positiivinen operaattori ja siten lauseen 8.3.1 nojalla singulaariarvot ovat aina ei-negatiivisia,*

Huomautus 9.1.3. *Palautetaan mieleen, että spektraalilauseen (lause 7.3.1) perusteella lineaarioperaattorilla* $\sqrt{f^*f}$ *sanotaan olevan* $\dim V$ *ominaisarvoa multiplisiteetti huomioden, eli että ominaisarvoille* $\lambda_1 \geq \cdots \geq \lambda_k \geq 0$ *pätee*

$$\dim E(\lambda_1, \sqrt{f^*f}) + \cdots + \dim E(\lambda_k, \sqrt{f^*f}) = \dim V.$$

Sanotaan, että operaattorilla $f\colon V \to V$ *on* $\dim V$ *singulaariarvoa, kun multiplisiteetti huomioidaan. Dimensiota* $\dim E(\lambda_i, \sqrt{f^*f})$ *kutsutaan operaattorin* f *singulaariarvon* λ_i *moninkerraksi (eli multiplisiteetiksi).*

Lause 9.1.4. *Olkoon* $(V, \langle\cdot,\cdot\rangle)$ n*-ulotteinen sisätuloavaruus ja* $f\colon V \to V$ *lineaarikuvaus, jonka singulaariarvot ovat* $s_1, \ldots, s_n$. *Tällöin on olemassa avaruuden* V *sellaiset ortonormaalit kannat* $(v_1, \ldots, v_n)$ *ja* $(w_1, \ldots, w_n)$, *että*

$$f(v) = s_1\langle v_1, v\rangle w_1 + \cdots + s_n\langle v_n, v\rangle w_n \tag{9.1}$$

kaikilla $v \in V$.

Todistus. Polaarihajotelman (lause 8.6.3) perusteella on olemassa sellainen isometria $S\colon V \to V$, että

$$f = S \circ \sqrt{f^*f}.$$

Olkoon nyt $v_1, \ldots, v_n$ avaruuden V ortonormaalikanta, jonka muodostavat operaattorin $\sqrt{f^*f}$ sellaiset ominaisvektorit, että $\sqrt{f^*f}(v_i) = s_iv_i$ kaikilla $i \in \{1, \ldots, n\}$. Olkoot lisäksi $w_i = S(v_i)$ jokaisella $i \in \{1, \ldots, n\}$. Koska S on isometria, niin $w_1, \ldots, w_n$ on avaruuden V ortonormaalikanta.

Osoitetaan nyt (9.1). Olkoon $v \in V$. Koska $v_1, \ldots, v_n$ on ortonormaalikanta, niin

$$v = \langle v_1, v\rangle v_1 + \cdots + \langle v_n, v\rangle v_n.$$

Tällöin

$$\begin{aligned}\sqrt{f^*f}(v) &= \sqrt{f^*f}\left(\langle v_1, v\rangle v_1 + \cdots + \langle v_n, v\rangle v_n\right) \\ &= \langle v_1, v\rangle \sqrt{f^*f}(v_1) + \cdots + \langle v_n, v\rangle \sqrt{f^*f}(v_n) \\ &= \langle v_1, v\rangle s_1 v_1 + \cdots + \langle v_n, v\rangle s_n v_n.\end{aligned}$$

Näin ollen

$$\begin{aligned}f(v) = S(\sqrt{f^*f}(v)) &= S\left(\langle v_1, v\rangle s_1 v_1 + \cdots + \langle v_n, v\rangle s_n v_n\right) \\ &= \langle v_1, v\rangle s_1 S(v_1) + \cdots + \langle v_n, v\rangle s_n S(v_n) \\ &= s_1 \langle v_1, v\rangle w_1 + \cdots + s_n \langle v_n, v\rangle w_n.\end{aligned}$$

□

Singulaariarvohajotelma on matriisimuodossaan hieman vaikuttavampi.

Korollaari 9.1.5. *Olkoon $A \in \mathbb{R}^{n\times n}$ ja $s_1, \ldots, s_n$ matriisin A singulaariarvot. Tällöin on olemassa sellaiset ortogonaaliset matriisit $P \in \mathbb{R}^{n\times n}$ ja $Q \in \mathbb{R}^{n\times n}$, että*

$$A = PDQ^T,$$

missä D on singulaariarvojen diagonaalimatriisi

$$D = \begin{bmatrix} s_1 & & \\ & \ddots & \\ & & s_n \end{bmatrix}.$$

Todistus. Olkoon $f_A\colon \mathbb{R}^{n\times 1} \to \mathbb{R}^{n\times 1}$, $x \mapsto Ax$. Tällöin lauseen 9.1.4 perusteella on olemassa sellaiset ortormaalit kannat $v_1, \ldots, v_n$ ja $w_1, \ldots, w_n$, että

$$Ax = f_A(x) = s_1(v_1 \cdot x)w_1 + \cdots + s_n(v_n \cdot x)w_n = s_1(v_1^T x)w_1 + \cdots + s_n(v_n^T x)w_n.$$

Olkoot $P = [w_1 \cdots w_n] \in \mathbb{R}^{n\times n}$ ja $Q = [v_1 \cdots v_n] \in \mathbb{R}^{n\times n}$. Tällöin

$$Ax = s_1(v_1^T x)w_1 + \cdots + s_n(v_n^T x)w_n = PDQ^T x.$$

□

Huomautus 9.1.6. *Singulaariarvohajotelman olemassaolon todistus antaa yhden menetelmän singulaariarvohajotelman löytämiseen seuraavasti. Olkoon $A \in \mathbb{R}^{n\times n}$ neliömatriisi. Resepti:*

1. *Laske matriisin $A^T A$ spektraaliarvohajotelma $A^T A = QDQ^T$. Matriisin A singulaariarvohajotelman matriisi Q on tämä matriisi Q ja singulaariarvohajotelman diagonaalimatriisi on matriisi $D^{1/2}$.*

2. *Ratkaise matriisi P yhtälöstä $A = PD^{1/2}Q^T$.*

Huomautus 9.1.7. *Yllä esitetyn reseptin viimeinen vaihe on houkuttelevaa yrittää ratkaista käyttämällä luvun alussa ollutta havaintoa, että P on matriisin AA^T omainaisvektoreiden matriisi, eli tiedosta, että*

$$AA^T = PDP^T.$$

Tässä on kuitenkin omat vaaransa[1]*. Tarkastellaan seuraavaa esimerkkiä. Olkoon*

$$A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}.$$

Tällöin $A^TA = AA^T = I$. Tällöin matriisin A^TA positiivisesti semidefiniitti neliöjuuri on identtinen matriisi ja sille voidaan valita spektraalihajotelma $Q = D = I$. Jäljellä on siis matriisin P valinta.

Tarkastellaan nyt samaa identtistä matriisi matriisia AA^T, mutta toisesta suunnista. Koska jokaiselle ortogonaalimatriisille $P \in \mathbb{R}^{2\times 2}$ pätee, että $PP^T = I$, niin

$$AA^T = I = PP^T = PIP^T$$

antaa matriisin AA^T spektraalihajotelman, millä tahansa ortogonaalimatriisilla $P \in \mathbb{R}^{2\times 2}$. Tässä tilanteessa kuitenkin ainoastan matriisi $P = A$ antaa halutun singulaariarvohajotelman PDQ^T matriisille A.

Tapaus $A^TA = AA^T = I$ ei ole kuitenkaan mikään erikoistapaus, vaan sama ongelma tulee esiin myös muiden matriisien A kohdalla. Matriisin AA^T spektraalihajotelma PDP^T ei nimittäin ole yksikäsitteinen matriisien P ja D suhteen. Ensinnäkin matriisin D diagonaalialkioden järjestystä voi vaihtaa ja toisaalta matriisi P voidaan vaihtaa matriisiin $P\tilde{Q}$, jos ortogonaalimatriisille $\tilde{Q}$ pätee, että $\tilde{Q}D\tilde{Q}^T$ on diagonaalimatriisi. Eräs tälläinen tapaus syntyy tilanteessa, jossa $\tilde{Q}$ kommutoi matriisin D kanssa, eli $\tilde{Q}D = D\tilde{Q}^T$, sillä tällöin $(P\tilde{Q})D(P\tilde{Q})^T = P(\tilde{Q}D\tilde{Q}^T)P^T = P(D\tilde{Q}\tilde{Q}^T)P^T = PDP^T$. Jokaiselle matriisille D voidaan löytää tälläinen kommutoimvamatriisi $\tilde{Q}$ vaihtamalla jokin sarakevektoreista v_i vektoriksi $-v_i$.

9.2 Singulaariarvohajotelman yleistyksiä

Polaarihajotelma ja singulaariarvohajotelma yleistyvät lineaarikuvauksille $f\colon V \to W$. Todistukset ovat oleellisesti samoja ja osa jätetään harjoitustehtäviksi.

Perustavanlaatuinen havainto on, että sisätuloavaruuksien $(V, \langle\cdot,\cdot\rangle_V)$ ja $(W, \langle\cdot,\cdot\rangle_W)$ lineaarikuvauksella $f\colon V \to W$ on adjungaatti $f^*\colon W \to V$. Näin ollen lineaarikuvaus $f\colon V \to W$ määrittelee operaattorin $f^*f\colon V \to V$, joka on positiivinen. Tämä seuraa seuraavasta havainnosta. Olkoon $v \in V$ operaattorin f^*f ominaisvektori, jonka ominaisarvo on λ. Tällöin

$$\lambda\|v\|^2 = \lambda\langle v, v\rangle = \langle v, \lambda v\rangle = \langle v, f^*f(v)\rangle = \langle f(v), f(v)\rangle = \|f(v)\|^2.$$

[1]Kiitokset Joonas Nuutiselle kommentista. Aiempi komentti oli harhaanjohtava, koska kommentti koski ainoastaan matriisia D eikä matriisin P valintaa liittyviä epäselvyyksiäi.

Näin ollen operaattori $\sqrt{f^*f}$ on yksikäsitteisesti määritelty ja kuvauksen f singulaariarvot voidaan määritellä seuraavasti.

Määritelmä 9.2.1. *Olkoot* $(V, \langle\cdot,\cdot\rangle_V)$ *ja* $(W, \langle\cdot,\cdot\rangle_W)$ *sisätuloavaruuksia. Luku* $s \in \mathbb{R}$ *on lineaarikuvauksen* $f\colon V \to W$ singulaariarvo, *jos* s *on operaattorin* $\sqrt{f^*f}$ *ominaisarvo.*

Polaarihajotelma saa nyt seuraavan muodon.

Lause 9.2.2. *Olkoot* $(V, \langle\cdot,\cdot\rangle_V)$ *ja* $(W, \langle\cdot,\cdot\rangle_W)$ *sisätuloavaruuksia, joille pätee* $\dim V \leq \dim W$*, ja olkoon* $f\colon V \to W$ *lineaarikuvaus. Tällöin on olemassa sellainen isometria* $S\colon V \to W$*, että*

$$f = S\sqrt{f^*f}.$$

Todistuksen idea. Kuten tapauksessa $V = W$, havaitaan ensin, että

$$\|f(v)\| = \|\sqrt{f^*f}\|.$$

Olkoon nyt $v_1, \dots, v_n$ avaruuden V kanta, joka koostuu operaattorin $\sqrt{f^*f}$ ominaisvektoreista, joille pätee $\sqrt{f^*f}(v_i) = s_i v_i$ jokaisella $i \in \{1, \dots, n\}$. Olkoot nyt $V_1 = \operatorname{im}\sqrt{f^*f} = \operatorname{Sp}\{v_1, \dots, v_k\}$ sekä $S_1\colon V_1 \to W$ lineaarikuvaus $v_i \mapsto \frac{1}{s_i} f(v_i)$ kuten lauseessa 8.6.3. Nyt sama argumentti kuin lauseen 8.6.3 todistuksessa antaa väitteen. (Harjoitustehtävä) □

Huomautus 9.2.3. *Huomaa, että* $\dim V \leq \dim W$ *välttämätön oletus isometrian* S *olemassaololle.*

Singulaariarvohajotelma saa tässä tilanteessa seuraavan muodon. Väite todistetaan aivan kuten lauseen 9.1.4 tapauksessa, joten todistus sivuutetaan. (Harjoitustehtävä)

Lause 9.2.4. *Olkoot* $(V, \langle\cdot,\cdot\rangle_V)$ *ja* $(W, \langle\cdot,\cdot\rangle_W)$ *sisätuloavaruuksia ja* $f\colon V \to W$ *lineaarikuvaus sekä* $s_1, \dots, s_n$ *sen singulaariarvot. Tällöin on olemassa sellainen avaruuden* V *ortonormaalikanta* $v_1, \dots, v_n$ *ja sellainen avaruuden* W *ortonormallikanta* $w_1, \dots, w_m$*, että*

$$f(v) = s_1\langle v_1, v\rangle w_1 + \cdots + s_n\langle v_n, v\rangle w_n.$$

Matriiseille tätä singulaariarvohajotelman yleisys saa seuraavan muodon. Väitteen todistaminen jätetään jälleen harjoitustehtäväksi.

Lause 9.2.5. *Olkoot* $m \geq n$ *ja olkoon* $A \in \mathbb{R}^{m\times n}$ *matriisi, jonka singulaariarvot ovat* $s_1, \dots, s_n$*. Tällöin on olemassa ortogonaaliset matriisit* $P \in \mathbb{R}^{m\times m}$ *ja* $Q \in \mathbb{R}^{n\times n}$*, joille pätee*

$$A = P\begin{bmatrix} D \\ 0 \end{bmatrix} Q^T,$$

missä $0 \in \mathbb{R}^{(m-n)\times n}$ *on nollamatriisi ja* $D \in \mathbb{R}^{n\times n}$ *on singulaariarvojen diagonaalimatriisi*

$$D = \begin{bmatrix} s_1 & & \\ & \ddots & \\ & & s_n \end{bmatrix}.$$

Huomautus 9.2.6. *Edellisen luvun havainnot singulaariarvohajotelmassa esiintyvistä matriiseista P, Q ja D ovat edelleen voimassa:*

1. *Matriisi D on matriisin $A^T A$ singulaariarvohajotelman diagonaalimatriisin neliöjuuri.*

2. *Matriisi Q on matriisin $A^T A$ ominaisvektoreiden matriisi.*

3. *Matriisi P on matriisin AA^T ominaisvektoreiden matriisi.*

Tämän faktan todistus jätetään harjoistehtäväksi.

9.3 Sovellus: Pääkomponenttianalyysi

Singulaariarvohajotelmaa kutsutaan historiallisista syistä usein (esimerkiksi tilastotieteessä) pääkomponenttianalyysiksi. Määritelmää varten kerrataan singulaariarvohajotelman tulos.

Olkoot $(V, \langle\cdot,\cdot\rangle_V)$ ja $(W, \langle\cdot,\cdot\rangle_W)$ sisätuloavaruuksia ja $f\colon V \to W$ operaattori, jonka singulaariarvot ovat $s_1 \geq \cdots \geq s_n \geq 0$. Tällöin on olemassa sellaiset avaruuden V ja W ortonormaalit kannat $(v_1, \ldots, v_n)$ ja $(w_1, \ldots, w_n)$, että

$$f(v) = \sum_{i=1}^{n} s_i \langle v_i, v\rangle w_i.$$

Määritelmä 9.3.1. *Kannan $(v_1, \ldots, v_n)$ vektoria v_k sanotaan operaattorin f k:nneksi* pääakseliksi.

Seuraava lause on singulaariarvohajotelman suora seuraus.

Lause 9.3.2. *Olkoot $(V, \langle\cdot,\cdot\rangle)$ ja $(W, \langle\cdot,\cdot\rangle_W)$ sisätuloavaruuksia ja $f\colon V \to W$ lineaarikuvaus, jonka singulaariarvot ovat $s_1 \geq \cdots \geq s_n \geq 0$. Tällöin*

$$s_n = \min_{\|v\|=1} \|f(v)\| \leq \max_{\|v\|=1} \|f(v)\| = s_1.$$

Todistus. Olkoon nyt $v \in V$ sellainen, että $\|v\| = 1$. Tällöin

$$\|f(v)\|^2 = \sum_{i=1}^{n} (s_i \langle v_i, v\rangle)^2 = \sum_{i=1}^{n} s_i^2 \langle v_i, v\rangle^2 \leq s_1^2 \sum_{i=1}^{n} \langle v_i, v\rangle^2 = s_1^2 \|v\|^2 = s_1^2.$$

Näin ollen

$$\max_{\|v\|=1} \|f(v)\| \leq s_1.$$

Vastaavasti

$$\|f(v)\|^2 = \sum_{i=1}^{n} (s_i \langle v_i, v\rangle)^2 = \sum_{i=1}^{n} s_i^2 \langle v_i, v\rangle^2 \geq s_n^2 \sum_{i=1}^{n} \langle v_i, v\rangle^2 = s_n^2 \|v\|^2 = s_n^2.$$

Näin ollen

$$\min_{\|v\|=1} \|f(v)\| \geq s_n.$$

Koska $\|f(v_1)\| = s_1$ ja $\|f(v_n)\| = s_n$, väite seuraa. $\square$

Muilla singulaariarvoilla voidaan tehdä vastaava analyysi. Kirjataan se seuraavasti.

Korollaari 9.3.3. *Olkoot* $(V, \langle\cdot,\cdot\rangle_V)$ $(W, \langle\cdot,\cdot\rangle_W)$ *sisätuloavaruuksia ja* $f\colon V \to W$ *positiivinen operaattori, jonka singulaariarvot ovat* $s_1 \geq \cdots \geq s_n \geq 0$ *ja vastaavat pääakselit ovat* $v_1, \ldots, v_n$. *Merkitään* $V_k = \mathrm{Sp}\{v_k, \ldots, v_n\}$ *jokaisella* $1 \leq k \leq n$. *Tällöin*

$$\max_{\|v\|=1, v\in V_k} \|f(v)\| = s_k$$

Todistus. Olkoon $W_k = \mathrm{Sp}\{w_k, \ldots, w_n\}$. Tällöin $f|_{V_k}\colon V_k \to W_k$ on hyvin määritelty ja

$$\max_{\|v\|=1, v\in V_k} \|f(v)\| = \max_{\|v\|=1} \|f|_{V_k}(v)\|.$$

Näin ollen edellisen lauseen perusteella

$$\max_{\|v\|=1} \|f|_{V_k}(v)\| = s_k.$$

$\square$

Luku 10

Determinantti ja jälki

Matriisin determinantti on näissä luentomuistiinpanoissa oletettu tunnetuksi käsitetteeksi. Sitä on tosin käytetty vain kerran: osoittamaan, että yhtälöllä $Ax = \lambda x$ on epätriviaaleja ratkaisuaj, jos ja vain jos $\det(A - \lambda I) = 0$. Kompleksisten matriisien determinantista emme ole puhuneet vielä mitään.

Tässä luvussa tarkastellaan ensin neliömatriisien determinantin perusominaisuuksia ja tämän jälkeen operaattorin determinanttia. Jotkin determinantin perusominaisuuksien todistuksista on jätetty liitteeseen E. Determinantin jälkeen käsitellään lyhyesti toista neliömatriiseihin liittyvää suuretta eli jälkeä.

Determinantilla ja jäljella on molemmilla tärkeä ominaisuus:

> Operaattorin $f\colon V \to V$ jokaisella esitysmatriisilla on sama determinantti ja jokaisella esitysmatriisilla on sama jälki.

Tämä ominaisuus tarkoittaa, että determinantti ja jälki ovat operaattorin niin kutusuttuja *invariantteja*.

10.1 Matriisin determinantti

Suurin osa determinanttien ominaisuuksista on helpointa todistaa matriisien avualla.

Matriisin determinantti määritellään tyypillisesti induktiivisesti matriisin koon suhteen. Palautetaan tämä lähestymistapa ensin mieleen. Aloitetaan kuintekin 2×2-matriisien sijaan 1×1-matriisista.

Matriisin $A = [a] \in \mathbb{C}^{1\times1}$ determinantiksi määritellään $\det A = a$. Olkoon nyt $n \geq 2$ ja oletetaan, että $(n-1) \times (n-1)$-matriisien determinantti on määritelty. Tällöin neliömatriisin $A = [a_{ji}] \in C^{n\times n}$ determinantiksi määritellään

$$\det A = \sum_{j=1}^{n} (-1)^{j+1} a_{j1} \det(\hat{A}_{j1}),$$

missä $\hat{A}_{j1} \in \mathbb{C}^{(n-1)\times(n-1)}$ on se neliömatriisi, joka on saatu matriisista A poistamalla j:s rivi ja ensimmäinen sarake. Tämä määritelmä johtaa välittömästi tuttuun kaavaan 2×2-matriisin determinantille.

Esimerkki 10.1.1. *Olkoon*

$$A = \begin{bmatrix} a & c \\ b & d \end{bmatrix} \in \mathbb{C}^{2\times 2}.$$

Tällöin

$$\hat{A}_{11} = [d] \quad \textit{ja} \quad \hat{A}_{21} = [c]$$

ja

$$\det A = (-1)^{1+1} a \det(\hat{A}_{11}) + (-1)^{2+1} b \det(\hat{A}_{21}) = ad - bc.$$

Tästä indutiivisesta määritelmästä voidaan todistaa helposti induktiolla joitakin matriisien ominaisuuksia kuten, että kaikilla $A \in \mathbb{C}^{n\times n}$ ja $a \in \mathbb{C}$ pätee $\det(aA) = a^n \det(A)$. Monet determinantin perusominaisuudet seuraaavat kuitenkin luonnollisemmin determinantin määritelmästä permutaatioiden avulla. Tämä määritelmä ja perustulosten todistukset on annettu liitteessä E. Luetellaan tässä nyt niistä keskeisimmät. Aloitetaan determinantin tulosäännöstä.

Lause 10.1.2. *Olkoot* $A, B \in \mathbb{C}^{n\times n}$. *Tällöin* $\det(AB) = \det(A)\det(B)$.

Korollaari 10.1.3. *Olkoon* $P \in \mathbb{C}^{n\times n}$ *kääntyvä matriisi. Tällöin* $\det(P^{-1}) = \frac{1}{\det(P)}$.

Korollaari 10.1.4. *Olkoon* $A \in \mathbb{C}^{n\times n}$ *ja* $P \in \mathbb{C}^{n\times n}$ *kääntyvä matriisi. Tällöin*

$$\det(PAP^{-1}) = \det(A).$$

Toinen determinantin tunnettu ominaisuus on, että neliömatriisilla ja sen transpoosilla on sama determinantti. Tämä perustellaan usein niin sanotuilla kehityskaavoilla. Liitteessä tämä tulos todistetaan suoraan determinantin toisesta määritelmästä.

Lause 10.1.5. *Olkoon* $A \in \mathbb{C}^{n\times n}$. *Tällöin* $\det(A^T) = \det(A)$.

Korollaari 10.1.6. *Olkoon* $P \in \mathbb{R}^{n\times n}$ *ortogonaalimatriisi. Tällöin* $|\det(P)| = 1$.

Todistus. Koska $P^T = P^{-1}$, niin

$$\det(P)^2 = \det(P^T)\det(P) = \det(P^T P) = \det(I) = 1.$$

□

10.1.1 Determinentti ominaisarvojen tulona

Aloitetaan kompleksisten matriisien tuloksesta.

Lause 10.1.7. *Olkoon* $A \in \mathbb{C}^{n\times n}$ *neliömatriisi ja olkoot* $\lambda_1, \dots, \lambda_n$ *matriisin* A *ominaisarvot algebrallisen kertaluvun mukaan laskettuna. Tällöin* $\det A = \lambda_1 \cdots \lambda_n$.

Todistus. Olkoon

$$A = PTP^{-1}$$

matriisin A yläkolmioesitys, missä P on kääntyvä matriisi ja T yläkolmiomatriisi

$$T = \begin{bmatrix} t_{11} & t_{12} & \cdots & t_{1n} \\ 0 & t_{22} & & t_{2n} \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & t_{nn} \end{bmatrix}.$$

Tällöin

$$\det(A) = \det(PTP^{-1}) = \det(P)\det(T)\det(P^{-1}) = \det(T) = t_{11}\cdots t_{nn},$$

missä on käytetty kehityskaavasta seuraavaa tietoa, että yläkolmiomatriisin determinantti on diagonaalialkioiden tulo.

Jäljellä on siis havaita, että matriisin T diagonaalialkiot ovat matriisin A ominaisarvoja. Tämä seuraa suoraan seuraavasta havainnosta. Olkoon $\lambda \in \mathbb{C}$. Tällöin

$$A - \lambda I = PTP^{-1} - \lambda I = P(T - \lambda I)P^{-1},$$

joten

$$\det(A - \lambda I) = \det(T - \lambda I) = (t_{11} - \lambda)\cdots(t_{nn} - \lambda).$$

Väite seuraa. □

Reaalisessa tapauksessa sama tulos pätee symmetristen matriisien erikoistapauksessa.

Lause 10.1.8. *Olkoon $A \in \mathbb{R}^{n\times n}$ symmetrinen matriisi ja $\lambda_1, \dots, \lambda_n$ sen ominaisarvot kertaluvun mukaan laskettuna. Tällöin*

$$\det(A) = \lambda_1 \cdots \lambda_n$$

Todistus. Spektraalilauseen nojalla on olemassa ortogonaalimatriisi $P \in \mathbb{R}^{n\times n}$ ja diagonaalimatriisi $D \in \mathbb{R}^{n\times n}$, jonka diagonaalialkiot $\lambda_1, \dots, \lambda_n$ ovat matriisin A ominaiarvot multiplisiteetti huomioiden, joille pätee $A = PDP^T$. Näin ollen

$$\begin{aligned} \det(A) &= \det(PDP^T) = \det(P)\det(D)\det(P^T) = \det(P)\det(D)\det(P^{-1}) \\ &= \det(D) = \lambda_1 \cdots \lambda_n. \end{aligned}$$

□

Yleisesti reaalisessa tilanteessa determinantin itseisarvo on singulaariarvojen tulo.

Lause 10.1.9. *Olkoon $A \in \mathbb{R}^{n\times n}$ neliömatriisi. Tällöin*

$$|\det(A)| = s_1 \cdots s_n$$

missä $s_1, \dots, s_n$ ovat matriisin A singulaariarvot kertaluvun mukaan laskettuna.

Todistus. Singulaariarvohajotelman perusteella on olemassa sellaiset ortogonaaliset matriisit $P \in \mathbb{R}^{n\times n}$ ja $Q \in \mathbb{R}^{n\times n}$ sekä diagonaalimatriisi $D \in \mathbb{R}^{n\times n}$, jonka diagonaalialkiot $s_1, \ldots, s_n$ ovat matriisin A singulaariarvot kertaluku huomioiden, että

$$A = QDP^T.$$

Koska $|\det(Q)| = 1$ ja $|\det(P^T)| = 1$, niin

$$|\det(A)| = |\det(QDP^T)| = |\det(Q)||\det(D)||\det(P^T)| = s_1 \cdots s_n.$$

□

10.2 Sovellus: Ositetun matriisin determinantti

Kuten luvussa 2 havaittiin, ositetun matriisin kääntyvyyttä ei voi lukea suoraan alimatriisien kääntyvyydestä. Determinantin kehityskaavojen (tai luvussa E esitetyn permutaatioihin perustuvan määritelmän) avulla voidaan kuitenkin helposti laskea joidenkin ositettujen matriisien determinantteja. Kootaan tähän joitakin tälläisiä tuloksia.

Aloitetaan havainnosta, että ositetujen ylä- ja alakolmiomatriisien determinantti on diagonaalimatriisien tulo.

Lause 10.2.1. *Olkoon*

$$M = \begin{bmatrix} A & C \\ 0 & D \end{bmatrix} \in \mathbb{R}^{n\times n},$$

missä A ja D ovat neliömatriiseja. Tällöin

$$\det M = (\det A)(\det D).$$

Vastaavasti

$$\det \begin{bmatrix} A & 0 \\ C & D \end{bmatrix} = (\det A)(\det D).$$

Johdetaan nyt kaava ositetun matriisien determinantille erikoistapauksessa, että ositetun matriisin

$$M = \begin{bmatrix} A & C \\ B & D \end{bmatrix}$$

alimatriisi A on kääntyvä. Tällöin

$$M = \begin{bmatrix} A & C \\ B & D \end{bmatrix} = \begin{bmatrix} I & 0 \\ BA^{-1} & D - BA^{-1}C \end{bmatrix} \begin{bmatrix} I & C \\ 0 & I \end{bmatrix} \begin{bmatrix} A & 0 \\ 0 & I \end{bmatrix}$$

Näin ollen determinantin tulokaavan avulla saadaan seuraava tulos.

Lause 10.2.2. *Olkoon*

$$M = \begin{bmatrix} A & C \\ B & D \end{bmatrix}$$

sellainen ositettu neliömatriisi, jonka alimatriisi A on kääntyvä. Tällöin

$$\det(M) = \det(D - BA^{-1}C)\det(A).$$

10.3 Sovellus: Cramerin sääntö ja käänteismatriisin kaava

Yksi determinantin (teoreettisia) sovelluksia on antaa kaava matriisin käänteismatriisille. Tarkka väite on seuraava.

Lause 10.3.1. *Olkoon $A \in \mathbb{C}^{n\times n}$ kääntyvä matriisi. Tällöin*

$$A^{-1} = \frac{1}{\det A}\mathrm{adj}A,$$

missä $\mathrm{adj}A \in \mathbb{C}^{n\times n}$ *on matriisi, jonka alkiot ovat* $(\mathrm{adj}A)_{ji} = (-1)^{j+i}\det(\hat{A}_{ji})$.

Matriisia $\mathrm{adj}A$ kutsutaan matriisin A *(klassiseksi) adjungaatiksi* ja yllä matriisi $\hat{A}_{ji}$ on aiemmin esitelty matriisi, joka saadaan matriisista A poistamalla j:s rivi ja i:s sarake. Valittu terminologia on harmittavasi harhaanjohtavaa, sillä matriisin (klassisella) adjungaatilla ei ole mitään tekemistä lineaarikuvauksen adjungaatin kanssa.

Lauseen 10.3.1 todistuksen ytimessä on niin kutsuttu Cramerin sääntö, joka todistetaan seuraavassa lemmassa. Lemman väitettä varten annetaan seuraava määritelmä. Olkoon $A = [a_1 \cdots a_n] \in \mathbb{C}^{n\times n}$ matriisi, jonka sarakeet ovat $a_1, \ldots, a_n \in \mathbb{C}^{n\times 1}$, ja olkoon $b \in \mathbb{C}^{n\times 1}$ sarake. Tällöin jokaisella $i \in \{1, \ldots, n\}$ merkitään

$$A_i(b) = [a_1 \cdots a_{i-1}\ b\ a_{i+1} \cdots a_n] \in \mathbb{C}^{n\times n},$$

eli $A_i(b)$ on se matriisi, joka saadaan korvaamalla matriisin A i:s sarake sarakkeella b.

Cramerin sääntö sanoo, että yhtälön $Ax = b$ ratkaisu x voidaan laskea matriisien $A_i(b)$ determinanteista. Väite on seuraava.

Lemma 10.3.2. *Olkoon $A \in \mathbb{C}^{n\times n}$ kääntyvä matriisi ja $b \in \mathbb{C}^{n\times 1}$. Tällöin yhtälön $Ax = b$ yksikäsitteiselle ratkaisulle $x = [x_1 \cdots x_n]^T \in \mathbb{C}^{n\times 1}$ pätee*

$$x_i = \frac{\det A_i(b)}{\det A}$$

jokaisella $i \in \{1, \ldots, n\}$.

Todistus. Olkoon $e_1, \ldots, e_n$ avaruuden $\mathbb{C}^{n\times 1}$ standardikanta ja olkoon $i \in \{1, \ldots, n\}$. Koska

$$I_i(x) = [e_1 \cdots e_{i-1}\ x\ e_{i+1} \cdots e_n],$$

niin

$$\begin{aligned} AI_i(x) &= A[e_1 \cdots e_{i-1}\ x\ e_{i+1} \cdots e_n] \\ &= [Ae_1 \cdots Ae_{i+1}\ Ax\ Ae_{i+1} \cdots Ae_n] \\ &= [a_1 \cdots a_{i-1}\ b\ a_{i+1} \cdots a_n] = A_i(b). \end{aligned}$$

Koska

$$\det I_i(x) = x_i,$$

niin

$$\det(A)x_i = \det(A)\det(I_i(x)) = \det(AI_i(x)) = \det(A_i(b)).$$

Koska A on kääntyvä, niin $\det(A) \neq 0$ ja

$$x_i = \frac{\det(A_i(b))}{\det(A)}.$$

□

Todistetaan nyt Cramerin säännön avulla käänteimatriisin kaava.

Lauseen 10.3.1 todistus. Merkitään $A^{-1} = [c_1 \cdots c_n] \in \mathbb{C}^{n\times n}$, missä $c_i = [c_{1i} \cdots c_{ni}]^T \in \mathbb{C}^{n\times n}$ jokaisella $i \in \{1, \dots, n\}$.

Koska $AA^{-1} = I = [e_1 \cdots e_n]$, niin $Ac_i = e_i$ jokaisella $i \in \{1, \dots, n\}$. Cramerin kaavan (lemma 10.3.2) nojalla

$$c_{ji} = \frac{\det A_j(e_i)}{\det A}.$$

jokaisella $j \in \{1, \dots, n\}$.

Kehitetään matriisin $A_j(e_i)$ determinantti sarakkeen j suhteen, jolloin saadaan

$$\det(A_j(e_i)) = (-1)^{j+i}\det(A_{ji}) = \operatorname{adj}(A)_{ji}.$$

Näin ollen

$$A^{-1} = \frac{1}{\det A}\operatorname{adj}(A).$$

□

10.4 Sovellus: Sylvesterin kriteeri

Determinantin avulla voidaan karakterisoida positiivisesti definiitit matriisit. Tätä kutsutaan Sylvesterin kriteeriksi. Lausetta varten esitellään matriisin pääminorin käsite.

Määritelmä 10.4.1. *Neliömatriisin $A = [a_{ji}] \in \mathbb{C}^{n\times n}$ k:s* pääminori *$A_k \in \mathbb{C}^{k\times k}$ on matriisi $A_k = [a_{ji}]_{j,i=1}^k \in \mathbb{C}^{n\times n}$, eli $k \times k$-matriisi A_k, jolle pätee*

$$A = \begin{bmatrix} A_k & C_k \\ B_k & D_k \end{bmatrix},$$

missä $B_k \in \mathbb{C}^{(n-k)\times k}$, $C_k \in \mathbb{C}^{k\times(n-k)}$ ja $D_k \in \mathbb{C}^{(n-k)\times(n-k)}$.

Huomautus 10.4.2. *Matriisin A pääminorilla A_k on tulkinta lineaarikuvauksena. Olkoon $\varphi_A \colon \mathbb{C}^{n\times 1} \to \mathbb{C}^{n\times 1}$ matriisia A vastaava lineaarikuvaus $z \mapsto Az$. Tällöin $\varphi_{A_k} \colon \mathbb{C}^{k\times 1} \to$*

$\mathbb{C}^{k\times 1}$ on yhdistetty lineaarikuvaus $\varphi_{A_k} = \mathrm{pr}_k \circ \varphi_A \circ i_k$, missä $\mathrm{pr}_k \colon \mathbb{C}^{n\times 1} \to \mathbb{C}^{k\times 1}$ on projektio

$$\begin{bmatrix} z_1 \\ \vdots \\ z_k \\ \vdots \\ z_n \end{bmatrix} \mapsto \begin{bmatrix} z_1 \\ \vdots \\ z_k \end{bmatrix}$$

ja $i_k \colon \mathbb{C}^{k\times 1} \to \mathbb{C}^{n\times 1}$ on lineaarikuvaus

$$\begin{bmatrix} z_1 \\ \vdots \\ z_k \end{bmatrix} \mapsto \begin{bmatrix} z_1 \\ \vdots \\ z_k \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

Tämä havaitaan seuraavasti. Olkoon $z = [z_1 \cdots z_k]^T \in \mathbb{C}^{k\times 1}$. Tällöin

$$(\varphi_A \circ i_k)(z) = \begin{bmatrix} A_k & C_k \\ B_k & D_k \end{bmatrix} \begin{bmatrix} z \\ 0 \end{bmatrix} = \begin{bmatrix} A_k z + C_k 0 \\ B_k z + D_k 0 \end{bmatrix} = \begin{bmatrix} A_k z \\ B_k z \end{bmatrix}.$$

Näin ollen

$$(\mathrm{pr}_k \circ \varphi_A \circ i_k)(z) = \begin{bmatrix} I_{k\times k} & 0 \end{bmatrix} \begin{bmatrix} A_k z \\ B_k z \end{bmatrix} = A_k z = \varphi_{A_k}(z).$$

Muotoillaan ja todistetaan nyt Sylvesterin kriteerio.

Lause 10.4.3. *Symmetrinen matriisi $A \in \mathbb{R}^{n\times n}$ on positiivisesti definiitti, jos ja vain jos $\det(A_k) > 0$ jokaisella matriisin A pääminorilla A_k, missä $k \in \{1, \dots, k\}$.*

Todistus. Oletetaan ensin, että A on positiivisesti definiitti. Olkoon $k \in \{1, \dots, n\}$ ja merkitään

$$A = \begin{bmatrix} A_k & C_k \\ B_k & D_k \end{bmatrix}.$$

Tällöin jokaisella $x \in \mathbb{R}^{k\times 1}$ pätee

$$x^T A_k x = \begin{bmatrix} x & 0 \end{bmatrix} \begin{bmatrix} A_k & C_k \\ B_k & D_k \end{bmatrix} \begin{bmatrix} x \\ 0 \end{bmatrix} \geq 0$$

ja epäyhtälö on aito, jos $x \neq 0$. Näin ollen A_k on positiviisesti definiitti. Lauseen 8.3.1 nojalla matriisin A_k ominaisarvot ovat reaalisia ja positiivisia. Koska $\det A_k$ on ominaisarvojen tulo, niin $\det A_k > 0$.

Osoitetaan nyt induktiolla dimension suhteen, että matriisin $A \in \mathbb{R}^{n\times n}$ pääminoreiden A_k ehdosta $\det A_k > 0$ jokaisella $k \in \{1, \dots, n\}$ seuraa matriisin A positiivi definiittiys. Väite selvästi pätee, jos $n = 1$. Oletetaan, että väite pätee kaikilla symmetrisillä $(n-1) \times (n-1)$-matriiseilla. Olkoon nyt A symmetrinen $(n \times n)$-matriisi, jolle pätee $\det A_k \geq 0$ kaikilla $k \in \{1, \dots, n\}$.

Osoitetaan ensin, että matriisilla A on korkeintaan yksi negatiivinen ominaisarvo. Tehdään vastaoletus, että matriisilla A on kaksi negatiivista ominaisarvoa $\lambda \in \mathbb{R}$ ja $\mu \in \mathbb{R}$ ja olkoot $x = [x_1 \cdots x_n]^T \in \mathbb{R}^{n\times 1}$ ja $y = [y_1 \cdots y_n]^T \in \mathbb{R}^{n\times 1}$ näitä ominaisarvoja vastaavat ominaisvektorit. Tällöin $x^T Ax = \lambda x^T x < 0$ ja $y^T Ay = \mu y^T y < 0$. Koska $\lambda \neq \mu$, niin spektraalilauseen (eli lauseen 7.3.1) perusteella vektorit x ja y ovat kohtisuorassa toisiaan vastaan, eli $x^T y = 0$. Olkoon $w = y_n x - x_n y \in \mathbb{R}^{n\times 1}$. Merkitään $w = [w_1 \cdots w_n]^T \in \mathbb{R}^{n\times 1}$ ja $\tilde{w} = [w_1 \cdots w_{n-1}] \in \mathbb{R}^{(n-1)\times 1}$.

Koska A_{n-1} on symmetrinen $(n-1) \times (n-1)$-matriisi, jolle pätee $\det(A_{n-1})_k = \det A_k > 0$ kaikilla $k \in \{1, \dots, n-1\}$, niin A_k on positiivisesti semidefiniitti. Näin ollen $\tilde{w}^T A_k \tilde{w} \geq 0$. Koska $w_n = y_n x_n - x_n y_n = 0$, niin $\tilde{w}^T A_k \tilde{w} = w^T Aw$. Koska x ja y ovat matriisin A ominaisvektoreita, niin saadaan

$$\begin{aligned}
w^T Aw &= (y_n x - x_n y)^T A(y_n x - x_n y) \\
&= (y_n x^T - x_n y^T)A(y_n x - x_n y) \\
&= y_n^2 x^T Ax - y_n x_n \left(x^T Ay + y^T Ax\right) + x_n^2 y^T Ay \\
&= y_n^2 x^T Ax - y_n x_n \left(\mu x^T y + \lambda y^T x\right) + x_n^2 y^T Ay \\
&= y_n^2 x^T Ax + x_n^2 y^T Ay < 0.
\end{aligned}$$

Tämä on ristiriita. Näin ollen matriisilla A on korkeintaan yksi negatiivinen ominaisarvo. Myös tämä on ristiriita, koska tällöin $\det A < 0$ vastoin oletustsa. Näin ollen matriisin A ominaisarvot ovat positiivisia. Matriisi A on siis positiivisesti definiitti lauseen 8.3.1 nojalla. □

10.5 Operaattorin determinantti

Matriisin determinantin vastine lineaarikuvauksille on operaattorin determinantti. Määritelmä perustuu seuraavaan tulokseen.

Lemma 10.5.1. *Olkoon $f : V \to V$ äärellisulotteisen vektoriavaruuden operaattori sekä olkoot $(v_1, \dots, v_n)$ ja $(w_1, \dots, w_n)$ avaruuden V kantoja. Tällöin operaattorin f esitysmatriisille $A \in \mathbb{R}^{n\times n}$ kannassa $(v_1, \dots, v_n)$ ja esitysmatriisille $B \in \mathbb{R}^{n\times n}$ kannassa $(w_1, \dots, w_n)$ pätee*

$$\det(A) = \det(B).$$

Todistus. Olkoon $P \in \mathbb{R}^{n\times n}$ kannanvaihtomatriisi avaruuden V kannasta $(v_1, \dots, v_n)$ kantaan $(w_1, \dots, w_n)$. Tarkemmin sanottuna, olkoon $\Phi : \mathbb{R}^{n\times 1} \to V$ isomorfismi $e_i \mapsto v_i$ ja olkoon $\Psi : \mathbb{R}^{n\times 1} \to V$ isomorfismi $e_i \mapsto w_i$. Tällöin $A \in \mathbb{R}^{n\times n}$ on määritelmän nojalla lineaarikuvauksen $\Phi^{-1} \circ f \circ \Phi \colon \mathbb{R}^{n\times 1} \to \mathbb{R}^{n\times 1}$ esitysmatriisi ja $B \in \mathbb{R}^{n\times n}$ on lineaarikuvauksen $\Psi^{-1} \circ f \circ \Psi \colon \mathbb{R}^{n\times 1} \to \mathbb{R}^{n\times 1}$ esitysmatriisi. Lisäksi

$$B = PAP^{-1},$$

missä $P \in \mathbb{R}^{n\times n}$ on lineaarikuvauksen $\Psi^{-1} \circ \Phi \colon \mathbb{R}^{n\times 1} \to \mathbb{R}^{n\times 1}$ matriisi.

Näin ollen determinantin tulosäännön nojalla (lause 10.1.2) pätee

$$\det(B) = \det(PAP^{-1}) = \det(P)\det(A)\det(P^{-1}) = \det(P)\det(A)\det(P)^{-1} = \det(A).$$

□

Määritelmä 10.5.2. *Operaattorin* $f \colon V \to V$ determinantti $\det(f)$ *on sen esitysmatriisin* A *determinantti* $\det(A)$.

Suoraan matriisitulon määritelmästä saadaan, että yhdistetyn kuvauksen determinantti on determinanttien tulo. Todistus jätetään harjoitustehtäväksi.

Lemma 10.5.3. *Olkoot* $f, g \colon V \to V$ *operaattoreita. Tällöin* $\det(g \circ f) = \det(g)\det(f)$.

Samoin suoraan määritelmästä ja vastaavasta tulokseksesta matriiseille seuraa, että operaattorilla ja sen adjungaatilla on sama determinantti.

Korollaari 10.5.4. *Olkoon* $(V, \langle \cdots, \cdots \rangle)$ *sisätuloavaruus ja* $f \colon V \to V$ *operaattori. Tällöin* $\det(f^*) = \det(f)$.

Todistus. Olkoon $(v_1, \dots, v_n)$ avaruuden V kanta ja $A \in \mathbb{R}^{n\times n}$ operaattorin f esitysmatriisi tässä kannassa. Koska A^T on operaattorin f^* esitysmatriisi samassa kannassa, niin

$$\det(f^*) = \det(A^T) = \det(A) = \det(f).$$

□

Vastavasti voidaan osoittaa, että kompleksisen operaattorin determinantti on ominaisarvojen tulo. Tulos seuraa suoraan vastaavasta tuloksesta matriiseille.

10.6 Matriisin ja operaattorin jälki

Jälki on determinantin ohella toinen operaattoreiden ja neliömatriisien tärkeä invariantti. Matriisin jäljen määritelmä on erittäin helppo.

Määritelmä 10.6.1. *Neliömatriisin* $A \in \mathbb{C}^{n\times n}$ jälki *(engl.* trace*) on matriisin* A *diagonaalielemettien summa, eli*

$$\operatorname{tr}(A) = a_{11} + \cdots + a_{nn},$$

missä $A = [a_{ji}]$.

Kuten determinantille myös matriisin jäljelle pätee tulokaava. Jäljen tapauksessa tulokaava sanoo, että tulomatriisin jälki ei riipu tulontekijöiden järjestyksestä.

Lause 10.6.2. *Olkoot* $A, B \in \mathbb{C}^{n\times n}$. *Tällöin* $\operatorname{tr}(AB) = \operatorname{tr}(BA)$.

Todistus. Koska

$$\operatorname{tr}(AB) = \sum_{i=1}^{n}(AB)_{ii} = \sum_{i=1}^{n}\sum_{k=1}^{n} A_{ik}B_{ki} = \sum_{k=1}^{n}\sum_{i=1}^{n} B_{ki}A_{ik} = \sum_{k=1}^{n}(BA)_{kk} = \operatorname{tr}(BA),$$

niin väite pätee. □

Jäljen tulokaavasta seuraa, että operaattorin esitysmatriiseilla on sama jälki. Merkitään tämä tulos lemmaksi. Todistus jätetään harjoitustehtäväksi.

Lemma 10.6.3. *Olkoon* $f\colon V \to V$ *äärellisulotteisen vektoriavaruuden operaattori sekä olkoot* $(v_1, \dots, v_n)$ *ja* $(w_1, \dots, w_n)$ *avaruuden* V *kantoja. Tällöin operaattorin* f *esitysmatriisille* $A \in \mathbb{R}^{n\times n}$ *kannassa* $(v_1, \dots, v_n)$ *ja esitysmatriisille* $B \in \mathbb{R}^{n\times n}$ *kannassa* $(w_1, \dots, w_n)$ *pätee*

$$\operatorname{tr}(A) = \operatorname{tr}(B).$$

Määritelmä 10.6.4. *Operaattorin* $f\colon V \to V$ jälki *on sen esitysmatriisin* A *jälki* $\operatorname{tr}(A)$.

Lauseen suorana sovelluksena saadaan, että matriisin jäljelle triviaalisti pätevät tulokset pätevät myös lineaarikuvauksille.

Kirjataan nyt muutama matriisin ja operaattorin jälkeen liittyvä perustulos lauseiksi. Toisin kuin determinantin tapauksessa nämä tulokset ovat hyvin suoraviivaisia todistaa ja ne jätetään harjoitustehtäviksi.

Lause 10.6.5. *Olkoot* $A, B \in \mathbb{C}^{n\times n}$ *neliömatriiseja ja* $a \in \mathbb{C}$. *Tällöin*

$$\operatorname{tr}(A + aB) = \operatorname{tr}(A) + a\operatorname{tr}(B).$$

Lisäksi $\operatorname{tr}(A^T) = \operatorname{tr}(A)$.

Korollaari 10.6.6. *Olkoon äärellisulotteinen vektoriavaruus, olkoot* $f, g\colon V \to V$ *operaattorieta ja olkoon* $a \in \mathbb{C}$. *Tällöin*

$$\operatorname{tr}(f + ag) = \operatorname{tr}(f) + a\operatorname{tr}(g).$$

Lisäksi $\operatorname{tr}(f^*) = \operatorname{tr}(f)$.

Kuten determinantin tapauksessa, operaattorin jälki voidaan määritellä yleisesti ilman esitysmatriisia: jälki on operaattorin ominaisarvojen summa, kun ominaisarvot lasketaan algebrallisen multiplisiteetin mukaan.

Lause 10.6.7. *Olkoon* $A \in \mathbb{C}^{n\times n}$ *neliömatriisi ja olkoot* $\lambda_1, \dots, \lambda_n \in \mathbb{C}$ *sen ominaisarvot kertaluvun mukaan laskettuna. Tällöin*

$$\operatorname{tr}(A) = \lambda_1 + \dots + \lambda_n.$$

Todistus. Olkoon $A = PTP^{-1}$ matriisin A yläkolmioesitys, missä matriisin T diagonaalialkiot ovat $t_{11}, \dots, t_{nn} \in \mathbb{C}$. Tällöin

$$\operatorname{tr}(A) = \operatorname{tr}(PTP^{-1}) = \operatorname{tr}(TPP^{-1}) = \operatorname{tr}(T) = t_{11} + \cdots + t_{nn}.$$

Koska matriisin T diagonaalialkiot ovat matriisin A ominaisarvoja samoilla kertaluvuilla, niin väite seuraa. □

Jälleen vastaava symmetrisiä matriiseja koskeva tulos seuraa spektraalilauseesta.

Lause 10.6.8. *Olkoon $A \in \mathbb{R}^{n\times n}$ symmetrinen matriisi. Tällöin*

$$\operatorname{tr}(A) = \lambda_1 + \cdots + \lambda_n$$

missä $\lambda_1 \geq \cdots \geq \lambda_n$ ovat matriisin A ominaisarvot kertaluku huomioiden.

Todistus. Olkoot $P \in \mathbb{R}^n n \times n$ ortogonaalimatriisi ja $D \in \mathbb{R}^{n\times n}$ diagonaalimatriisi, jonka diagonaalialkiot $\lambda_1, \dots, \lambda_n$ matriisin A ominaisarvot multiplisiteetti huomioiden, joille pätee $A = PDP^T$. Tällöin

$$\operatorname{tr}(A) = \operatorname{tr}(PDP^T) = \operatorname{tr}(DPP^T) = \operatorname{tr}(D) = \lambda_1 + \cdots + \lambda_n.$$

□

Tätä ominaisarvoihin liittyvää tulosta voidaan soveltaa monin eri tavoin. Annetaan tästä esimerkki.

Esimerkki 10.6.9. *Ei ole olemassa sellaisia matriiseja $A, B \in \mathbb{C}^{n\times n}$, että $AB - BA = I$. Jos tälläiset matriisit A ja B olisivat olemassa, niin*

$$\operatorname{tr}(I) = \operatorname{tr}(AB - BA) = \operatorname{tr}(AB) - \operatorname{tr}(BA) = \operatorname{tr}(AB) - \operatorname{tr}(BA) = 0.$$

Toisaalta tiedetään, että $\operatorname{tr}(I) = n \neq 0$.

Liite A

Sisätulo extra

A.1 Normista sisätuloon

Koska normi kertoo ainoastaan yksittäisen vektorin pituuden, on helppo jäädä vaikutelmaan, että sisätulon määrittelemä normifunktio sisältää vähemmän informaatiota kuin alkuperäinen sisätulo, eli että normifunktiosta ei voi päätellä vektoreiden välisiä sisätuloja. Näin ei kuitenkaan ole siitä yksinkertaisesta syystä, että kolmion sivujen pituuksien suhteet määräävät vektoreiden väliset kulmat (ks. kosinilause). Tarkemmin voidaan sanoa, että mikäli normi on sisätulon määräämä, niin sisätulo voidaan rekonstruoida normifunktiosta. Tämä havainto voidaan formalisoida seuraavasti.

Lemma A.1.1. *Olkoon* $(V, \langle\cdot,\cdot\rangle)$ *sisätuloavaruus ja* $\|\cdot\|$ *sisätulon* $\langle\cdot,\cdot\rangle$ *määräämä normi. Tällöin kaikilla* $v, w \in V$ *pätee*

$$\langle v, w\rangle = \frac{1}{2}\left(\|v+w\|^2 - \|v\|^2 - \|w\|^2\right) \tag{A.1}$$

Todistus. Todistus on suora lasku, joka jätetään harjoitustehtäväksi. (Aloita yhtälöstä $\|v+w\|^2 = \langle v+w, v+w\rangle$.) □

Toinen asiaan liittyvä kysymys on, että milloin vektoriavaruuden V normi $\|\cdot\|\colon V \to [0,\infty[$ on jonkin sisätulon $\langle\cdot,\cdot\rangle\colon V\times V \to \mathbb{R}$ määräämä. Yllättäen vastaus on hyvin elegantti.

Sisätuloavaruuksissa $(V, \langle\cdot,\cdot\rangle)$ pätee niin sanottu *suunnikassääntö*, eli sisätulon määräämälle normille $\|\cdot\|$ pätee kaikilla $v, w \in V$ yhtälö

$$\|v+w\|^2 + \|v-w\|^2 = 2\left(\|v\|^2 + \|w\|^2\right). \tag{A.2}$$

Suunnikassäännön todistus on jälleen suora lasku ja se jätetään harjoitustehtäväksi.

Eleganttia on se, että sunnikassääntö karakterisoi ne normifunktiot, jotka ovat sisätulon määräämiä. Todistus ei kuitenkaan ole aivan helppo.

Lause A.1.2. *Olkoon* $(V, \|\cdot\|)$ *normiavaruus. Tällöin* $\|\cdot\|$ *on sisätulon* $\langle\cdot,\cdot\rangle\colon V\times V \to \mathbb{R}$ *määräämä, jos ja vain jos normi* $\|\cdot\|$ *toteuttaa suunnikassäännön* (A.2) *kaikilla* $v, w \in V$*. Lisäksi sisätulolle* $\langle\cdot,\cdot\rangle$ *pätee* (A.1) *kaikilla* $v, w \in V$*.*

Todistus. Yllä todettiin, että sisätulon määräämä normi toteuttaa suunnikassäännön, eli että suunnikassääntö on välttämätön ehto. Osoitetaan nyt, että suunnikassääntö on riittävä ehto. Oletetaan siis, että normi $\|\cdot\|$ toteuttaa suunnikassäännön (A.2). Määritellään nyt $\langle\cdot,\cdot\rangle\colon V\times V\to\mathbb{R}$ kaavalla

$$\langle v,w\rangle=\frac{1}{2}\left(\|v+w\|^2-\|v\|^2-\|w\|^2\right)$$

kaikilla $v,w\in V$ ja havaitaan, että

$$\langle v,v\rangle=\frac{1}{2}\left(\|2v\|^2-\|v\|^2-\|v\|^2\right)=\frac{1}{2}\left(4\|v\|^2-2\|v\|^2\right)=\|v\|^2.$$

Normi $\|\cdot\|$ on siis sisätulon $\langle\cdot,\cdot\rangle$ määräämä, jos $\langle\cdot,\cdot\rangle$ on sisätulo. Osoitetaan siis, että $\langle\cdot,\cdot\rangle$ on sisätulo. Havaitaan kuitenkin ensin, että suunnikassäännön perusteella

$$\begin{aligned}\langle v,w\rangle&=\frac{1}{2}\left(\|v+w\|^2-\|v\|^2-\|w\|^2\right)=\frac{1}{2}\left(\|v+w\|^2-(\|v\|^2+\|w\|^2)\right)\\&=\frac{1}{2}\left(\|v+w\|^2-\frac{\|v+w\|^2+\|v-w\|^2}{2}\right)=\frac{\|v+w\|^2-\|v-w\|^2}{4}\end{aligned}$$

kaikilla $v,w\in V$.

Koska $\|\cdot\|$ on normi ja kaikilla $v\in V$ pätee $\langle v,v\rangle=\|v\|^2$, niin $\langle\cdot,\cdot\rangle$ on positiividefiniitti. Lisäksi, $\langle\cdot,\cdot\rangle$ on symmetrinen, sillä kaikilla $v,w\in V$ pätee

$$\langle v,w\rangle=\frac{\|v+w\|^2-\|v-w\|^2}{4}=\frac{\|w+v\|^2-\|-(v+w)\|^2}{4}=\frac{\|w+v\|^2-\|v+w\|^2}{4}=\langle w,v\rangle.$$

Näin ollen jäljellä on osoittaa, että

$$\langle v+v',w\rangle=\langle v,w\rangle+\langle v',w\rangle \tag{A.3}$$

ja että

$$\langle av,w\rangle=a\langle v,w\rangle \tag{A.4}$$

kaikilla $v,v',w\in V$ ja $a\in\mathbb{R}$.

Aloitetaan yhtälön (A.3) todistuksella. Olkoot $v,v',w\in V$. Havaitaan ensin, että suunnikassäännön perusteella pätee

$$\begin{aligned}\langle 2v,w\rangle&=\frac{\|2v+w\|^2-\|2v-w\|^2}{4}\\&=\frac{\|v+(v+w)\|^2-\|v+(v-w)\|^2}{4}\\&=\frac{2\|v\|^2+2\|v+w\|^2-\|v-(v+w)\|^2-2\|v\|^2-2\|v-w\|^2-\|v-(v-w)\|^2}{4}\\&=\frac{2\|v+w\|^2-\|-w\|^2-2\|v-w\|^2+\|w\|^2}{4}\\&=\frac{2\|v+w\|^2-2\|v-w\|^2}{4}=2\langle v,w\rangle.\end{aligned}$$

Vastaavasti, tai käyttämällä symmetrisyyttä, päätellään, että

$$\langle v, 2w\rangle = 2\langle v, w\rangle.$$

Näin ollen suunnikassäännön nojalla saadaan

$$\begin{aligned}
\langle v+v', w\rangle &= \frac{1}{4}\left(\|v+v'+w\|^2 - \|v+v'-w\|^2\right)\\
&= \frac{1}{4}\left(\|(v+\frac{1}{2}w)+(v'+\frac{1}{2}w)\|^2 - \|(v-\frac{1}{2}w)+(v'-\frac{1}{2}w)\|^2\right)\\
&= \frac{1}{4}\left(2\|(v+\frac{1}{2}w)\|^2 + 2\|(v'+\frac{1}{2}w)\|^2 - \|(v+\frac{1}{2}w)-(v'+\frac{1}{2}w)\|^2\right)\\
&\quad - \frac{1}{4}\left(2\|(v-\frac{1}{2}w)\|^2 + 2\|(v'-\frac{1}{2}w)\|^2 - \|(v-\frac{1}{2}w)-(v'-\frac{1}{2}w)\|^2\right)\\
&= \frac{1}{4}\left(2\|(v+\frac{1}{2}w)\|^2 + 2\|(v'+\frac{1}{2}w)\|^2 - \|v-v'\|^2\right)\\
&\quad - \frac{1}{4}\left(2\|(v-\frac{1}{2}w)\|^2 + 2\|(v'-\frac{1}{2}w)\|^2 - \|v-v'\|^2\right)\\
&= \frac{1}{4}\left(2\|v+\frac{1}{2}w\|^2 - 2\|v-\frac{1}{2}w\|^2\right) + \frac{1}{4}\left(2\|v'+\frac{1}{2}w\|^2 - 2\|v'-\frac{1}{2}w\|^2\right)\\
&= 2\langle v, \frac{1}{2}w\rangle + 2\langle v', \frac{1}{2}w\rangle\\
&= \langle v, w\rangle + \langle v', w\rangle.
\end{aligned}$$

Yhtälö (A.3) on näin todistettu.

Yhtälön (A.4) todistus on hankalampi. Se tehdään vaiheittain: ensin luvulle -1, sitten luonnollisille- ja kokonaisluvuille, sitten rationaaliluvuille ja lopuksi reaaliluvuille Schwartzin epäyhtälön avulla.

Ensimmäinen havainto on siis, että kaikilla $v, w \in V$ pätee

$$\langle -v, w\rangle = \frac{1}{4}\left(\|-v+w\|^2 - \|-v-w\|^2\right) = \frac{1}{4}\left(\|v-w\|^2 - \|v+w\|^2\right) = -\langle v, w\rangle.$$

Toinen havainto on, että induktiolla saadaan, että kaikilla $v, w \in V$ ja $m \in \mathbb{N}$ pätee

$$\langle mv, w\rangle = m\langle v, w\rangle.$$

Koska väite pätee nyt kaikilla luonnollisilla luvuilla $m \in \mathbb{N}$, niin ensimmäisen havainnon perusteella se pätee myös kaikilla $m \in \mathbb{Z}$. Tästä saadaan, että väite pätee kaikilla rationaaliluvuilla $p/q \in \mathbb{Q}$. Tämä havaitaan seuraavasti. Olkoot $q \in \mathbb{N}$ ja $p \in \mathbb{Z}$. Tällöi

$$q\langle (p/q)v, w\rangle = \langle q(p/q)v, w\rangle = \langle pv, w\rangle = p\langle v, w\rangle$$

eli

$$\langle (p/q)v, w\rangle = \frac{p}{q}\langle v, w\rangle.$$

Ennenkuin osoitetaan, että väite pätee kaikilla reaaliluvuilla, osoitetaan, että funktio $\langle \cdot, \cdot \rangle$ toteuttaa Schwarzin epäyhtälön, eli että kaikilla $v, w \in V$ pätee

$$|\langle v, w\rangle| \leq \|v\|\|w\|.$$

Olkoot $v, w \in V$. Jos $w = 0$, niin väite pätee, joten voidaan olettaa, että $w \neq 0$. Valitaan

$$t_0 = \frac{\langle v, w\rangle}{\|w\|^2}.$$

Tällöin

$$\begin{aligned} 0 \leq \|v - t_0 w\|^2 &= \langle v - t_0 w, v - t_0 w\rangle \\ &= \|v\|^2 - 2t_0\langle v, w\rangle + t_0^2\|w\|^2 \\ &= \|v\|^2 - 2\frac{\langle v, w\rangle^2}{\|w\|^2} + \frac{\langle v, w\rangle^2}{\|w\|^2} = \|v\|^2 - \frac{\langle v, w\rangle^2}{\|w\|^2}. \end{aligned}$$

Schwarzin epäyhtälö seuraa.

Osoitetaan nyt Schwarzin epäyhtälön avulla, että kaikilla $a \in \mathbb{R}$ pätee

$$\langle av, w\rangle = a\langle v, w\rangle.$$

Olkoot $v, w \in V$ ja olkoon $a \in \mathbb{R}$. Olkoon lisäksi (r_i) jono rationaalilukuja, joka suppenee lukuun a, eli $r_i \to a$ kun $i \to \infty$. Tällöin

$$\begin{aligned} |\langle av, w\rangle - r_i\langle v, w\rangle| &= |\langle av, w\rangle - \langle r_i v, w\rangle| = |\langle av - r_i v, w\rangle| = |\langle (a - r_i)v, w\rangle| \\ &\leq \|(a - r_i)v\|\|w\| = |a - r_i|\|v\|\|w\| \to 0, \end{aligned}$$

kun $i \to \infty$, eli

$$\lim_{i\to\infty} r_i\langle v, w\rangle = \langle av, w\rangle.$$

Näin ollen

$$a\langle v, w\rangle = \lim_{i\to\infty} r_i\langle v, w\rangle = \langle av, w\rangle.$$

Väite on todistettu. □

A.2 Jokainen vektoriavaruus on sisätuloavaruus

Kantalause sanoo, että jokaisella äärellisulotteisella avaruudella V on kanta $(v_1, \ldots, v_n)$, missä $n = \dim V$. Kanta puolestaan samastaa avaruuden V sarakeavaruuden $\mathbb{R}^{n\times 1}$ kanssa; isomorfismiksi voidaan ottaa $\Phi_V \colon \mathbb{R}^{n\times 1} \to V$, joka kuvaa standardikannan $(e_1, \ldots, e_n)$ kannaksi $(v_1, \ldots, v_n)$. Avaruus $\mathbb{R}^{n\times 1}$ on myös sisätuloavaruus; standardi valinta sisätuloksi on pistetulo $\cdot \colon \mathbb{R}^{n\times 1} \times \mathbb{R}^{n\times 1} \to \mathbb{R}$. Siinä missä Φ_v siirtää lineaarisen rakenteen, siirtää se myös sisätulon ja tekee avaruudesta V sisätuloavaruuden. Formaalisti tämä osoitetaan seuraavasti.

Lause A.2.1. *Olkoon V äärellisulotteinen vektoriavaruus ja $(v_1, \ldots, v_n)$ avaruuden V kanta. Olkoon lisäksi $\Phi_V \colon \mathbb{R}^{n\times 1} \to V$ lineaarinen isomorfismi, jolle pätee $\Phi_V(e_i) = v_i$ kaikilla $i \in \{1, \ldots, n\}$. Tällöin funktio $\langle \cdot, \cdot \rangle \colon V \times V \to \mathbb{R}$, joka on määritelty kaavalla*

$$\langle v, w \rangle = \Phi_V^{-1}(v) \cdot \Phi_V^{-1}(w)$$

kaikilla $v, w \in V$, on sisätulo avaruudessa V.

Todistus. Pitää tarkastaa sisätulon ehdot niiden tietojen avulla, että Φ_V on isomorfismi ja että pistetulo on sisätulo. Tämä jätetään harjoitustehtäväksi. □

Korollaari A.2.2. . *Jokainen äärellisulotteinen vektoriavaruus on sisätuloavaruus: Olkoon V äärellisulotteinen vektoriavaruus. Tällöin on olemassa sisätulo $\langle \cdot, \cdot \rangle \colon V \times V \to \mathbb{R}$.*

Todistus. Jokaisella äärellisulotteisella vektoriavarudella on kanta. Näin ollen sillä on myös sisätulo lauseen A.2.1. □

Huomautus A.2.3. *Tarkkaavainen lukija on jo varmaan huomannut, että luvun otsikossa puhutaan ainoastaan vektoriavaruuksista eikä äärellisulotteisista vektoriavaruuksissa kuten lauseessa A.2.1 ja korollaarissa A.2.2. Tämä ei ole vahinko. Itseasiassa jokaisella vektoriavaruudella on kanta ja jokaisella vektoriavaruudella on sisätulo. Tämä ei kuitenkaan ole tämän kurssin keskeistä asiaa ja siksi näitä asioita käsitellään ainoastaan liitteessä G.*

Tarkastellaan vielä millainen on lauseessa A.2.1 määritelty sisätulo $\langle \cdot, \cdot \rangle \colon V \times V \to \mathbb{R}$, sillä se selvästi riippuu kannan $(v_1, \ldots, v_n)$ valinnasta.

Tätä sisätuloa on luonnollisinta tarkastella tapauksessa, jossa V on sarakeavaruus $\mathbb{R}^{n\times 1}$ ja $(v_1, \ldots, v_n)$ on jokin avaruuden $\mathbb{R}^{n\times 1}$ standardikannasta $(e_1, \ldots, e_n)$ poikkeava kanta. Tällöin isomorfismi $\Phi_V \colon \mathbb{R}^{n\times 1} \to V$ on lineaarikuvaus $x \mapsto Ax$, missä A on sarakevektoreiden $(v_1, \ldots, v_n)$ muodostama matriisi, eli

$$A = \Big[v_1 | \cdots | v_n \Big].$$

Näillä merkinnöillä käänteiskuvaus Φ_V^{-1} on lineaarikuvaus $x \mapsto A^{-1}x$.

Näillä merkinnöillä saadaan, että kaikilla $v, w \in \mathbb{R}^{n\times 1}$ pätee

$$\langle v, w \rangle = \Phi_V^{-1}(v) \cdot \Phi_V^{-1}(w) = (A^{-1}v) \cdot (A^{-1}w).$$

Koska $x \cdot y = x^T y$ kaikilla $x, y \in \mathbb{R}^{n\times 1}$, niin näin ollen saadaan

$$\langle v, w \rangle = (A^{-1}v) \cdot (A^{-1}w) = (A^{-1}x)^T (A^{-1}w = x^T (A^{-1})^T A^{-1} y = x^T \left((A^{-1})^T A^{-1} \right) y.$$

Muutama huomio on paikallaan:

- Lukuunottamatta matriisia $(A^{-1})^T A^{-1}$, tämän sisätulon matriisiesitys on kuin pistetulon.

- Matriisissa $(A^{-1})^T A^{-1}$ on kirjoitettu kannanvaihtomatriisin A^{-1} avulla. Matriisi A^{-1} vie kannan $(v_1, \ldots, v_n)$ standardikannalle $(e_1, \ldots, e_n)$. Jos isomorfisimin Φ_V sijaan aloitettaisiin isomorfismista $\Psi_V \colon \Phi_V^{-1} \colon V \to \mathbb{R}^{n \times 1}$, joka vei kannan $(v_1, \ldots, v_n)$ standardikannalle, kirjoitettaisiin sisätulon kaava matriisin A^{-1} sijasta kuvauksenn Ψ_V matriisin $B = A^{-1}$ avulla, eli $\langle v, w \rangle = v^T (B^T B) w$ kaikilla $v, w \in V$.

A.3 Sisätulon matriisiesitys ja symmetriset matriisit

Ensinäkemältä sisätulon määritelmästä voi jää vaikutelma, että kyseessä on hyvin abstrakti käsite. Näin onkin ääretönulotteisten vektoriavaruuksien tapauksessa. Tässä luvussa osoitetaan, että äärellisulotteisissa vektoriavaruuksissa sisätulot kuitenkin vastaavat symmetrisiä positiivideﬁniittejä neliömatriiseja. Tässä mielessä kaikki äärellisulotteisen vektoriavaruuden sisätulot ovat samankaltaisia. Aloitetaan määritelmistä.

Määritelmä A.3.1. *Neliömatriisi* $A = (A_{ij})_{ij} \in \mathbb{R}^{n \times n}$ *on* symmetrinen*, jos* $A_{ij} = A_{ji}$ *kaikilla* $i, j \in \{1, \ldots, n\}$*, eli* $A^T = A$*.*

Määritelmä A.3.2. *Neliömatriisi* $A = (A_{ij})_{ij} \in \mathbb{R}^{n \times n}$ *on* positiividefiniitti*, jos* $x^T A x > 0$ *kaikilla* $x \in \mathbb{R}^{n \times 1} \setminus \{0\}$*.*

Jokainen symmetrinen ja positiividefiniitti matriisi $B \in \mathbb{R}^{n \times n}$ määrittelee sisätulon n-ulotteisessa vektoriavaruudessa V. Tätä varten riittää valita vektoriavaruudelle V kanta.

Lause A.3.3. *Olkoon* V *äärellisulotteinen vektoriavaruus,* $(v_1, \ldots, v_n)$ *avaruuden* V *kanta ja* $B \in \mathbb{R}^{n \times n}$ *symmetrinen ja positiividefiniitti matriisi. Määritellään funktio* $\langle \cdot, \cdot \rangle \colon V \times V \to \mathbb{R}$ *kaavalla*

$$\langle v, w \rangle = x^T B y$$

kaikilla $v, w \in V$*, missä* x *ja* y *ovat vektoreita* v *ja* w *vastaavat sarakevektorit kannassa* $(v_1, \ldots, v_n)$*. Tällöin* $\langle \cdot, \cdot \rangle$ *on sisätulo avaruudessa* V*.*

Todistus. Osoitetaan sisätulon ominaisuudet. Aloitetaan lineaarisuudesta. Olkoot $v, v', w \in V$ ja $a, b \in \mathbb{R}$. Olkoot lisäksi $x, x', y \in \mathbb{R}^{n \times 1}$ vektoreita v, v' ja w vastaavat sarakevektorit kannassa $(v_1, \ldots, v_n)$. Tällöin $ax + bx'$ on vektoria $av + bv'$ vastaava sarakevektori, sillä luvun 1 merkinnöillä pätee $\Phi_V(ax + bx') = \Phi_V(ax) + \Phi_V(bx') = a\Phi_V(x) + b\Phi_V(x') = av + bv'$.

Näin ollen

$$\begin{aligned}\langle av + bv', w \rangle &= (ax + bx')^T B y = (ax^T + b(x')^T) B y = ax^T B y + b(x')^T B y \\ &= a\langle v, w \rangle + b\langle v', w \rangle.\end{aligned}$$

Osoitetaan nyt symmetrisyys. Olkoot jälleen $v, w \in V$ ja olkoot $x \in \mathbb{R}^{n \times 1}$ ja $y \in \mathbb{R}^{n \times 1}$ vektoreita v ja w vastaavat sarakevektorit kannassa $(v_1, \ldots, v_n)$.

Koska oletuksen mukaan $B^T = B$ ja yleisesti tiedetään, että $(CD)^T = D^T C^T$ kaikilla matriiseilla $C \in \mathbb{R}^{n\times m}$ ja $D \in \mathbb{R}^{m\times p}$, niin

$$\langle v, w\rangle = x^T B y = x^T B^T (y^T)^T = x^T (y^T B)^T = (y^T B x)^T = y^T B x = \langle w, v\rangle.$$

Toiseksi viimeisessä yhtäsuuruudessa käytettiin trivialiteettia, että $x^T = x$ kaikilla $x \in \mathbb{R}^{1\times 1}$.

Osoitetaan vielä sisätulon positiividefiniittisyys ehdon (ST3) mukaisesti. Olkoon $v \in V \setminus \{0\}$ ja $x \in \mathbb{R}^{n\times 1}$ vastaava sarakevektori. Tällöin $x \neq 0$. Näin ollen

$$\langle v, v\rangle = x^T B x > 0,$$

koska B on positiividefiniitti matriisi. □

Edellisen lauseen mukaan, jokainen symmetrinen ja positiividefiniitti neliömatriisi määrittelee sisätulon. Seuraava lause osoittaa, että kääntäen jokainen sisätulo voidaan esittää tälläisenä matriisina.

Lause A.3.4. *Olkoon $(V, \langle\cdot,\cdot\rangle)$ äärellisulotteinen sisätuloavaruus ja $(v_1, \ldots, v_n)$ avaruuden V kanta. Tällöin matriisi*

$$B = \begin{bmatrix} b_{11} & \cdots & b_{n1} \\ \vdots & & \vdots \\ b_{1n} & \cdots & b_{nn} \end{bmatrix} = \begin{bmatrix} \langle v_1, v_1\rangle & \cdots & \langle v_n, v_1\rangle \\ \vdots & & \vdots \\ \langle v_1, v_n\rangle & \cdots & \langle v_n, v_n\rangle \end{bmatrix}.$$

on sisätulon $\langle\cdot,\cdot\rangle$ matriisiesitys*, eli symmetrinen ja positiividefiniitti matriisi, joka toteuttaa ehdon*

$$\langle v, w\rangle = x^T B y \tag{A.5}$$

kaikilla $v, w \in V$, missä $x \in \mathbb{R}^{n\times 1}$ ja $y \in \mathbb{R}^{n\times 1}$ ovat vastaavat sarakevektorit.

Todistuksen päättely on kaksiosainen. Ensin havaitaan, että koska sisätulo on bilinaarinen, niin se määräytyy täysin kanta-alkioiden sisätuloista. Tämä havainto on analoginen sen kanssa, että kanta-alkioden kuvat määräävät lineaarikuvauksen. Matriisin B olemassaolo seuraa tästä havainnosta. Tämän jälkeen on helppo tarkastaa, että matriisin B ominaisuudet seuraavat vastaavista sisätulon ominaisuuksista.

Lauseen A.3.4 todistus. Osoitetaan ensin (A.5). Olkoot $v = x_1 v_1 + \cdots + x_n v_n \in V$ ja $w = y_1 v_1 + \cdots + y_n v_n \in V$ sekä olkoot $x, y \in \mathbb{R}^{n\times 1}$ vastaavat sarakevektorit.

Sisätulon bilineaarisuuden ja matriisitulon perusteella

$$\begin{aligned}
\langle v, w\rangle &= \langle x_1 v_1 + \cdots + x_n v_n, y_1 v_1 + \cdots + y_n v_n\rangle \\
&= \left\langle \sum_{i=1}^n x_i v_i, \sum_{j=1}^n y_j v_j \right\rangle = \sum_{i=1}^n \left\langle x_i v_i, \sum_{j=1}^n y_j v_j \right\rangle = \sum_{i=1}^n x_i \left\langle v_i, \sum_{j=1}^n y_j v_j \right\rangle \\
&= \sum_{i=1}^n x_i \sum_{j=1}^n \langle v_i, y_j v_j\rangle = \sum_{i=1}^n x_i \sum_{j=1}^n y_j \langle v_i, v_j\rangle = \sum_{i=1}^n x_i \sum_{j=1}^n y_j B_{ij} = \sum_{i=1}^n x_i \sum_{j=1}^n B_{ij} y_j \\
&= \sum_{i=1}^n \sum_{j=1}^n x_i B_{ij} y_j = x^T B y.
\end{aligned}$$

Ehto (A.5) on näin osoitettu.

Sisätulon symmetrisyyden nojalla

$$B_{ij} = \langle v_i, v_j \rangle = \langle v_j, v_i \rangle = B_{ij}$$

kaikilla $i, j \in \{1, \ldots, n\}$. Näin ollen $B^T = B$.

Osoitetaan vielä, että matriisi B on positiividefiniittinen. Olkoon

$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^{n \times 1}.$$

sarakevektori, joka ei ole nollavektori. Olkoon $v = x_1 e_1 + \cdots + x_n e_n \in V$. Koska $x \neq 0$, niin $v \neq 0$. Näin ollen ehdon (A.5) perusteella pätee

$$x^T B x = \langle x_1 e_1 + \cdots + x_n e_n, x_1 e_1 + \cdots + x_n e_n \rangle = \langle v, v \rangle > 0.$$

$\square$

Liite B

Ositetut matriisit ja suorat summat

Ositettu matriisi liittyy vektoriavaruuksien suoriin summiin seuraavalla tavalla. Seuraavassa lauseessa oletetaan, että äärellisulotteiset avaruudet V ja W on esitetty aliavaruuksiensa suorana summana, eli $V = V' \oplus V''$ ja $W = W' \oplus W''$, joillain aliavaruuksilla $V', V'' \subset V$ ja $W', W'' \subset W$.

Oletetaan, että avaruuden V aliavaruuksille V' ja V'' on valittu kannat $(v'_1, \cdots, v'_\ell)$ ja $(v''_1, \cdots, v''_{n-\ell})$ ja että avaruuden W aliavaruuksille W' ja W'' on valittu kannat $(w'_1, \ldots, w'_k)$ ja $(w''_1, \ldots, w''_{m-k})$. Lisäksi oletetaan, että avaruudet V ja W on varustettu kannoilla $(v'_1, \ldots, v'_\ell, v''_1, \ldots, v''_{n-\ell})$ ja $(w'_1, \ldots, w'_k, w''_1, \ldots, w''_{m-k})$.

Lause B.0.1. *Olkoot $f' \colon V' \to W'$, $f'' \colon V' \to W''$, $g' \colon V'' \to W'$ ja $g'' \colon V'' \to W''$ lineaarikuvauksia. Tällöin kuvaus $F \colon V \to W$, joka on määritelty kaavalla*

$$\begin{aligned} F(x'_1 v'_1 + \cdots + x'_\ell v'_\ell + x''_1 v''_1 + \cdots + x''_{n-\ell} v''_{n-\ell}) &= f'(x'_1 v'_1 + \cdots + x'_\ell v'_\ell) \\ &\quad + f''(x'_1 v'_1 + \cdots + x'_\ell v'_\ell) \\ &\quad + g'(x''_1 v''_1 + \cdots + x''_{n-\ell} v''_{n-\ell}) \\ &\quad + g''(x''_1 v''_1 + \cdots + x''_{n-\ell} v''_{n-\ell}) \end{aligned}$$

kaikilla $v = x'_1 v'_1 + \cdots + x'_\ell v'_\ell + x''_1 v''_1 + \cdots + x''_{n-\ell} v''_{n-\ell} \in V$, on hyvin määritelty lineaarikuvaus, jonka matriisi on

$$\begin{bmatrix} A & C \\ B & D \end{bmatrix},$$

missä

- *$A = (a_{ij}) \in \mathbb{R}^{k \times \ell}$ on kuvauksen f' matriisi,*
- *$B = (b_{ij}) \in \mathbb{R}^{k \times (n-\ell)}$ on kuvauksen f'' matriisi,*
- *$C = (c_{ij}) \in \mathbb{R}^{(m-k) \times \ell}$ on kuvauksen g' matriisi ja*
- *$D = (d_{ij}) \in \mathbb{R}^{k \times (n-\ell)}$ on kuvauksen g'' matriisi.*

Huomautus B.0.2. *Kommentoidaan jo ennen todistusta, mitä tarkoittaa, että F on hyvin määritelty. Huomaa, että kuvausten f' ja g' maaliavaruus on W' ja että kuvausten f'' ja g'' maaliavaruus on W''. Koska W' ja W'' ovat avaruuden W aliavaruuksia, niin voidaan jokaisen kuvauksen maaliavaruudeksi vaihtaa W. Näin ollen kuvauken F määritelmässä oleva yhteenlasku on siis avaruuden W vektoreiden yhteenlasku ja kuvaus on hyvin määritelty.*

Lauseen B.0.1 todistus. Osoitetaan ensin, että F on lineaarikuvaus. Olkoot $v = x'_1 v'_1 + \cdots + x'_\ell v'_\ell + x''_1 v''_1 + \cdots + x''_{n-\ell} v''_{n-\ell} \in V$ ja $w = y'_1 v'_1 + \cdots + y'_\ell v'_\ell + y''_1 v''_1 + \cdots + y''_{n-\ell} v''_{n-\ell} \in V$ sekä $a \in \mathbb{R}$. Merkitään $v' = x'_1 v'_1 + \cdots + x'_\ell v'_\ell \in V'$, $v'' = x''_1 v''_1 + \cdots + x''_{n-\ell} v''_{n-\ell} \in V''$ sekä $w' = y'_1 v'_1 + \cdots + y'_\ell v'_\ell \in V'$ ja $w'' = y''_1 v''_1 + \cdots + y''_{n-\ell} v''_{n-\ell} \in W''$. Tällöin $v = v' + v''$ ja $w = w' + w''$. Nyt

$$\begin{aligned}
F(av + w) &= F((av' + w') + (av'' + w'')) \\
&= f'(av' + w') + f''(av' + w') + g'(av'' + w'') + g''(av'' + w'') \\
&= \big(af'(v') + f'(w')\big) + \big(af''(v') + f''(w')\big) \\
&+ \big(ag'(v'') + g'(w'')\big) + \big(ag''(v'') + g''(w'')\big) \\
&= aF(v) + F(w).
\end{aligned}$$

Näin ollen F on lineaarinen.

Osoitetaan vielä, että kuvauksella F on haluttu matriisi, eli lasketaan kuvavektoreiden $F(v'_1), \ldots, F(v'_\ell), F(v''_1), \ldots, F(v''_{n-\ell})$ esitykset kannassa $(w'_1, \ldots, w'_k, w''_1, \ldots, w''_{m-k})$. Olkoon $M \in \mathbb{R}^{m \times n}$ kuvauksen F matriisi näissä avaruuksien V ja W kannoissa.

Olkoon ensin $i \in \{1, \ldots, \ell\}$. Tällöin

$$\begin{aligned}
F(v'_i) = f'(v'_i) + f''(v'_i) &= \big(a_{1i} w'_1 + \cdots + a_{ki} w'_k\big) + \big(b_{1i} w''_1 + \cdots + b_{(n-k)i} w''_{n-k}\big) \\
&= a_{1i} w'_1 + \cdots + a_{ki} w'_k + b_{1i} w''_1 + \cdots + b_{(n-k)i} w''_{n-k}.
\end{aligned}$$

Koska vektori v'_i on kannan $(v'_1, \ldots, v'_\ell, v''_1, \ldots, v''_{n-\ell})$ i:s vektori, niin saadaan, että matriisin M i:s sarake on matriisin $\begin{bmatrix} A \\ B \end{bmatrix}$ i:s sarake.

Olkoon nyt $i \in \{1, \ldots, n - \ell\}$. Tällöin

$$\begin{aligned}
F(v''_i) = g'(v''_i) + g''(v''_i) &= \big(c_{1i} w'_1 + \cdots + c_{ki} w'_k\big) + \big(d_{1i} w''_1 + \cdots + d_{(n-k)i} w''_{n-k}\big) \\
&= c_{1i} w'_1 + \cdots + c_{ki} w'_k + d_{1i} w''_1 + \cdots + d_{(n-k)i} w''_{n-k}
\end{aligned}$$

Koska vektori v''_i on kannan $(v'_1, \ldots, v'_\ell, v''_1, \ldots, v''_{n-\ell})$ $\ell + i$:s vektori, niin saadaan, että matriisin M $\ell + i$:s sarake on matriisin $\begin{bmatrix} C \\ D \end{bmatrix}$ i:s sarake. Näin ollen

$$M = \begin{bmatrix} A & C \\ B & D \end{bmatrix}.$$

□

Liite C

Duaaliavaruudet

Tässä luvussa tarkastellaan uudestaan transpoosiin ja adjungaattiin liittyviä käsitteitä ja havaitaan kuinka ne johdattavat vektoriavaruuden duaalin käsitteeseen. Aloitetaan sisätuloavaruudesta $(V, \langle\cdot,\cdot\rangle)$.

Määritellään jokaisella $v \in V$ kuvaus $v^*\colon V \to \mathbb{R}$ kaavalla $w \mapsto \langle v, w\rangle$. Koska sisätulo on bilineaarinen, niin v^* on lineaarikuvaus. Lisäksi, jos $v_1, v_2 \in V$ ja $a \in \mathbb{R}$, niin vektorin $v = v_1 + av_2$ määrittelemä lineaarikuvaus $v^* = (v_1 + av_2)^*$ on itseasiassa lineaarikuvuas $v_1^* + av_2^*$, eli kaikilla $w \in V$ pätee

$$v^*(w) = \langle v, w\rangle = \langle v_1 + av_2, w\rangle = \langle v_1, w\rangle + a\langle v_2, w\rangle = v_1^*(w) + av_2^*(w) = (v_1^* + av_2^*)(w),$$

missä viimeisessä yhtäsuuruudessa käytettiin kuvausten yhteenlaskun määritelmää.

Itseasiassa jokainen lineaarikuvaus avaruudelta reaaliluvuille voidaan kirjoittaa kuvausten v^* lineaarikombinaatioina. Tarkka lause on seuraava

Lause C.0.1. *Olkoon $(V, \langle\cdot,\cdot\rangle)$ sisätuloavaruus ja $v_1, \ldots, v_n$ avaruuden V ortonormaalikanta. Tällöin $v_1^*, \ldots, v_n^*$ on vektoriavaruuden*

$$V^* = \{f\colon V \to \mathbb{R}\colon f \text{ on lineaarinen}\}$$

kanta.

Todistus. Osoitetaan ensin, että $\mathrm{Sp}\{v_1^*, \ldots, v_n^*\} = V^*$. Olkoon $f \in V^*$ eli $f\colon V \to \mathbb{R}$ on lineaarikuvaus. Tälllöin jokaisella $w = a_1v_1 + \cdots + a_nv_n \in V$ pätee

$$\begin{aligned} f(w) = f(a_1v_1 + \cdots + a_nv_n) &= a_1f(v_1) + \cdots + a_nf(v_n) \\ &= f(v_1)\langle v_1, w\rangle + \cdots + f(v_n)\langle v_n, w\rangle \\ &= f(v_1)v_1^*(w) + f(v_n)v_n^*(w) \\ &= (f(v_1)v_1^* + \cdots + f(v_n)v_n^*)(w). \end{aligned}$$

Näin ollen

$$f = f(v_1)v_1^* + \cdots + f(v_n)v_n^* \in \mathrm{Sp}\{v_1^*, \ldots, v_n^*\}.$$

Osoitetaan nyt, että jono $v_1^*, \ldots, v_n^*$ on vapaa. Oletetaan, että on olemassa sellaiset $a_1, \ldots, a_n \in \mathbb{R}$, että

$$a_1 v_1^* + \cdots + a_n v_n^* = 0.$$

Olkoon $v = a_1 v_1 + \ldots + a_n v_n$. Tällöin aiemman laskun perusteella

$$v^* = a_1 v_1^* + \cdots a_n v_n^* = 0,$$

eli $v^*(w) = 0$ kaikilla $w \in V$. Koska

$$0 = v^*(v) = \langle v, v \rangle = \|v\|,$$

niin $v = 0$. Näin ollen $a_i = 0$ kaikilla $i \in \{1, \ldots, n\}$. □

Määritelmä C.0.2. *Avaruus $V^* = \{f \colon V \to \mathbb{R} \colon f$ on lineaarinen$\}$ on vektoriavaruuden V* duaaliavaruus *ja sen alkiota kutsutaan* funktionaaleiksi.

Huomautus C.0.3. *Duaaliavaruuden V^* määritelmässä ei vaadita, että V on sisätuloavaruus, vaan ainoastaan, että V on vektoriavaruus. Sisätuloa ei itseasiassa tarvita myöskään edellisessä lauseessa vaan seuraava tulos pätee: Olkoon V äärellisulotteinen vektoriavaruus ja $v_1, \ldots, v_n$ sen kanta. Tällöin funktionaalit $v_j^\# \colon V \to \mathbb{R}$, jotka on määritelty kaavalla $v_j^\#(v_i) = \delta_{ij}$ kaikilla $i, j \in \{1, \ldots, n\}$, on avaruuden V^* kanta. (Harjoitustehtävä)*

Lauseesta C.0.1 seuraa, että äärellisulotteinen avaruus V on isomorfinen duaaliavaruuden V^* kanssa. Sisätuloavaruuksille isomorfisuuden osoittaminen ei vaadi kannan valintaa ja siksi todistus tehdään sisätuloavaruuksien tapauksissa. Huomaa, että kaikki äärellisulotteiset vektoriavaruudet voidaan varustaa sisätulolla.

Korollaari C.0.4. *Olkoon $(V, \langle \cdot, \cdot \rangle)$ äärellisulotteinen sisätuloavaruus. Tällöin kuvaus $\iota \colon V \to V^*$, $v \mapsto v^*$, on isomorfismi.*

Todistus. Kuvaus ι^* on jo osoitettu lineaarikuvaukseksi. Koska $\dim V = \dim V^*$ riittää osoittaa, että $\ker \iota = \{0\}$. Olkoon $v \in V$ sellainen, että $\iota(v) = 0$. Tällöin $v^* = 0$. Koska $0 = v^*(v) = \langle v, v \rangle$, niin $v = 0$. Kuvaus ι on siis injektio. □

Huomautus C.0.5. *Tottakai väite seuraa myös tiedosta, että $\iota(v_1), \ldots, \iota(v_n)$ on avaruuden V^* kanta, jos $v_1, \ldots, v_n$ on avaruuden V kanta.*

Tarkastellaan nyt vektoriavaruuksien V ja W välistä lineaarikuvausta $L \colon V \to W$. Määritellään uusi kuvaus $L^* \colon W^* \to V^*$ kaavalla $f \mapsto f \circ L$. Huomaa, että tässä $f \in W^*$ eli $f \colon W \to \mathbb{R}$ on lineaarikuvaus ja $L^*(f) = f \circ L \colon V \to \mathbb{R}$ on lineaarikuvaus. Kuvaus L^* on itseasiassa lineaarikuvaus.

Lemma C.0.6. *Olkoot V ja W sisätuloavaruuksia ja $L \colon V \to W$ lineaarikuvaus. Tällöin $L^* \colon W^* \to V^*$, $f \mapsto f \circ L$, on lineaarikuvaus.*

Todistus. Olkoot $f_1, f_2 \in W^*$ ja $a \in \mathbb{R}$. Tällöin

$$L^*(f_1 + af_2) = (f_1 + af_2) \circ L = f_1 \circ L + af_2 \circ L = L^*(f_1) + aL^*(f_2).$$

□

Sisätuloavaruuksien $(V, \langle\cdot,\cdot\rangle_V)$ ja $(W, \langle\cdot,\cdot\rangle_W)$ välinen lineaarikuvaus $L^*\colon W^* \to V^*$ vastaa lineaarikuvauksen $L\colon V \to W$ adjungaattia $L^*\colon W \to V$. Merkitään hetkellisesti adjungaattia symbolilla $\mathrm{adj}(L)\colon W \to V$.

Lause C.0.7. *Olkoot $(V, \langle\cdot,\cdot\rangle_V)$ ja $(W, \langle\cdot,\cdot\rangle_W)$ sisätuloavaruuksia ja $L\colon V \to W$ lineaarikuvaus. Tällöin jokaisella $w \in W$ pätee*

$$L^*(w^*) = (\mathrm{adj}(L)w)^*$$

Todistus. Olkoon $v \in V$. Tällöin

$$(L^*(w^*))(v) = w^*(L(v)) = \langle w, L(v)\rangle_W = \langle \mathrm{adj}(L)w, v\rangle_V = (\mathrm{adj}(L)w)^*(v).$$

□

Huomautus C.0.8. *Formaalisti tämä yhteys voidaan esittää isomorfismien $\iota_V\colon V \to V^*$ ja $\iota_W\colon W \to W^*$ avulla käyttäen kommutoivaa kaavioita*

$$\begin{tikzcd}
W \arrow[r, "\mathrm{adj}(L)"] \arrow[d, "\iota_W"'] & V \arrow[d, "\iota_V"] \\
W^* \arrow[r, "L^*"] & V^*
\end{tikzcd}$$

Koska adjungaatin esitysmatriisi on alkuperäisen kuvauksen esitysmatriisin transpoosi, saadaan sama tulos duaaleiden välille indusoidulle kuvaukselle. Todistus jätetään harjoitustehtäväksi.

Korollaari C.0.9. *Olkoot $(V, \langle\cdot,\cdot\rangle_V)$ ja $(W, \langle\cdot,\cdot\rangle_W)$ sisätuloavaruuksia ja $L\colon V \to W$ lineaarikuvaus. Mikäli $v_1, \ldots, v_n$ ja $w_1, \ldots, w_n$ ovat avaruuksien V ja W kantoja ja $A \in \mathbb{R}^{m\times n}$ kuvauksen L esitysmatriisi näissä kannoissa, niin kuvauksen $f^*\colon W^* \to V^*$ esitysmatriisi duaalikannoissa $v_1^*, \ldots, v_n^*$ ja $w_1^*, \ldots, w_m^*$ on A^T.*

Liite D

Neliömuodot ja geometria

D.1 Neliömuodosta matriisiin

Matriisien neliömuotojen hyödyllisyys seuraa havainnosta, että toisen asteen polynomi yhtälöt voidaan ratkaista neliömuotojen avulla.

Aloitetaan yleisen neliömuodon määritelmästä

Määritelmä D.1.1. *Funktio* $q\colon \mathbb{R}^n \to \mathbb{R}$ *on* neliömuoto*, jos se on homogeeninen toisen asteen polynomi, eli muotoa*

$$q(x_1, \ldots, x_n) = \sum_{1 \le i \le j \le n} c_{ij} x_i x_j$$

kaikilla $(x_1, \ldots, x_n) \in \mathbb{R}^n$*, missä* $c_{ij} \in \mathbb{R}$ *kaikilla* $1 \le i \le j \le n$*.*

Huomautus D.1.2. *Yleisesti asteen* $m \in \mathbb{N}$ *polynomi* $p\colon \mathbb{R}^n \to \mathbb{R}$ *on homogeeninen, jos* $p(\lambda x) = \lambda^m p(x)$ *kaikilla* $x \in \mathbb{R}^n$*.*

Jokainen neliömuoto on jonkin symmetrisen neliömatriisin neliömuoto.

Lemma D.1.3. *Olkoon* $q\colon \mathbb{R}^n \to \mathbb{R}$ *neliömuoto. Tällöin on olemassa sellainen symmetrinen matriisi* $A \in \mathbb{R}^{n\times n}$*, että* $q = q_A$*.*

Todistus. Oletaan, että

$$q(x_1, \ldots, x_n) = \sum_{1 \le i \le j \le n} c_{ij} x_i x_j$$

kaikilla $(x_1, \ldots, x_n) \in \mathbb{R}^n$, missä $c_{ij} \in \mathbb{R}$ kaikilla $1 \le i \le j \le n$.

Määritellään

$$a_{ij} = \begin{cases} c_{ij}/2, & i < j \\ c_{ij}, & i = j \\ c_{ji}/2, & i > j \end{cases}$$

Olkoon $A = [a_{ji}] \in \mathbb{R}^{n\times n}$. Tällöin jokaisella $x = [x_1 \cdots x_n]^T \in \mathbb{R}^{n\times 1}$ pätee

$$q_A(x) = x^T A x = q(x).$$

(Harjoitustehtävä) $\square$

D.2 Toisen asteen polynomiyhtälön ratkaiseminen neliömuotojen avulla

Tarkastellaan nyt pääakseleiden sovelluksena yleistä toisen asteen polynomia $p\colon \mathbb{R}^n \to \mathbb{R}$ ja yhtälön

$$p(x) = 0 \tag{D.1}$$

ratkaisemiseta, eli ratkaisujoukkoa

$$R_p = \{x \in \mathbb{R}^n \colon p(x) = 0\}.$$

Huomautus D.2.1. *Tason $\mathbb{R}^2$ erikoistapauksessa tämä vastaa toisen asteen yhtälön*

$$c_{11}x_1^2 + c_{12}x_1x_2 + c_{22}x_2^2 + b_1x_1 + b_2x_2 + a = 0,$$

missä $c_{11}, c_{12}, c_{22}, b_1, b_2, a \in \mathbb{R}$, ratkaisujen etsimistä.

Yleinen tulos jakautuu tapauksiin, jotka riippuvat polynomista p.
Tarkastellaan tämän vuoksi ongelmaa neljässä vaiheessa:

- Ensimmäisessä yhtälö $p(x) = 0$ kirjoitetaan neliömuodon ja sisätulon avulla.
- Toisessa neliömuoto kirjoitetaan pääakselimuodossa.
- Kolmannessa vaiheessa sisätuloon liittyvät termit yhdistetään neliömuodon termeihin, mikäli neliömuodon pääakseliesitys sen mahdollistaa. Tämä on neliöksitäydentämisen korkeampiulotteinen vastine.
- Jäljelle jäävät tapaukset käsitellään esimerkin muodossa ja jätetään harjoitustehtäviksi.

Ensimmäinen vaihe on siis polynomin tulkitseminen neliömuodon ja sisätulon avulla.

Lemma D.2.2. *Olkoon $p\colon \mathbb{R}^{n\times 1} \to \mathbb{R}$ toisen asteen polynomi. Tällöin on olemassa sellainen neliömatriisi $A \in \mathbb{R}^{n\times n}$, sellainen vektori $\mathbb{R}^{n\times 1}$ ja sellainen luku $a \in \mathbb{R}$, että*

$$p(x) = q_A(x) + b \cdot x + a.$$

Todistus. Olkoon

$$p(x_1, \ldots, x_n) = \sum_{1 \le i \le j} c_{ij}x_ix_j + \sum_{i=1}^{n} b_ix_i + a$$

tarkasteltava polynomi. Olkoon $q\colon \mathbb{R}^{n\times 1} \to \mathbb{R}$ neliömuoto

$$(x_1, \ldots, x_n) = \sum_{1 \le i \le j} c_{ij}x_ix_j.$$

Olkoon nyt $A \in \mathbb{R}^{n\times n}$ se symmetrinen matriisi, jolle pätee $q = q_A$. Olkoon lisäksi $b = [b_1 \cdots b_n]^T \in \mathbb{R}^{n\times 1}$. Tällöin

$$p(x) = q(x) + b \cdot x + a = q_A(x) + b \cdot x + a$$

jokaisella $x \in \mathbb{R}^{n\times 1}$. $\square$

Toinen vaihe on yleisen neliömuodon q_A pääakseliesitykseen siirtyminen. Tämä antaa meille matriisin P. Huomaa, että seuraavan lemman todistus on oleellisesti tehty esimerkissä 8.2.2.

Lemma D.2.3. *Olkoon $p\colon \mathbb{R}^{n\times 1} \to \mathbb{R}$ toisen asteen polynomi, joka on muotoa*

$$p(x) = q_A(x) + b \cdot x + a$$

jollakin symmetrisellä matriisilla $A \in \mathbb{R}^{n\times n}$, jollakin vektorilla $b \in \mathbb{R}^{n\times 1}$ ja jollaikin $a \in \mathbb{R}$. Tällöin on olemassa sellainen ortogonaalimatriisi $P \in \mathbb{R}^{n\times n}$ ja diagonaalimatriisi $D \in \mathbb{R}^{n\times n}$, että

$$p(Py) = q_D(y) + (P^T b) \cdot y + a.$$

kaikilla $y \in \mathbb{R}^{n\times 1}$.

Todistus. Olkoon $y \in \mathbb{R}^{n\times 1}$. Tällöin

$$\begin{aligned} p(Py) &= q_A(Py) + b \cdot Py + a \\ &= (Py)^T A(Py) + b^T Py + a \\ &= y^T P^T PDP^T Py + (P^T b)^T y + a \\ &= y^T Dy + (P^T b) \cdot y + a = q_D(y) + (P^T b) \cdot y + a. \end{aligned}$$

□

Edellisen lemman varsinainen hyöty paljastuu seuraavasta lemmasta, joka vastaa neliöksi täydentämistä. Huomaa, että neliöksi ei voi täydentää sellaisten muuttujjien y_i suhteen, joiden toisen potenssin kerroin on nolla.

Lemma D.2.4. *Olkoon $D \in \mathbb{R}^{n\times n}$ diagonaalimatriisi, jonka diagonaalialkiot ovat $d_1, \ldots, d_n \in \mathbb{R}$ ovat kaikki nollasta poikkeavia ja olkoon $b = [b_1 \cdots b_n]^T \in \mathbb{R}^{n\times 1}$. Tällöin on olemassa sellainen vektori $\tilde{b} \in \mathbb{R}$ ja vakio $\tilde{a} \in \mathbb{R}$, että*

$$q_D(y) + b \cdot y = q_D(y - \tilde{b}) + \tilde{a}$$

kaikilla $y \in \mathbb{R}^{n\times 1}$.

Todistus. Olkoon $y = [y_1 \cdots y_n]^T \in \mathbb{R}^{n\times 1}$. Koska $d_i \neq 0$ kaikilla $i \in \{1, \ldots, n\}$, niin

$$\begin{aligned} q_D(y) + b \cdot y &= d_1 y_1^2 + \cdots + d_n y_n^2 + b_1 y_1 + \cdots + b_n y_n \\ &= (d_1 y_1^2 + b_1 y_1) + \cdots + (d_n y_n^2 + b_n y_n) \\ &= d_1 \left(y_1^2 + \frac{b_1}{d_1} y_1 \right) + \cdots + d_n \left(y_n^2 + \frac{b_n}{d_n} y_n \right) \\ &= d_1 \left(\left(y_1 + \frac{b_1}{2d_1} \right)^2 - \left(\frac{b_1}{2d_1} \right)^2 \right) + \cdots + d_n \left(\left(y_n + \frac{b_n}{2d_n} \right)^2 - \left(\frac{b_n}{2d_n} \right)^2 \right) \\ &= d_1 \left(y_1 + \frac{b_1}{2d_1} \right)^2 + \cdots + d_n \left(y_n + \frac{b_n}{2d_n} \right)^2 - \left(d_1 \left(\frac{b_1}{2d_1} \right)^2 + \cdots + d_n \left(\frac{b_n}{2d_n} \right)^2 \right) \end{aligned}$$

Merkitään $\tilde{b}_i = b_i/(2d_i)$ jokaisella $i \in \{1, \ldots, n\}$ ja $\tilde{b} = [\tilde{b}_1 \cdots \tilde{b}_n]^T \in \mathbb{R}^{n\times 1}$. Olkoon myös

$$\tilde{a} = -\left(d_1\left(\frac{b_1}{2d_1}\right)^2 + \cdots + d_n\left(\frac{b_n}{2d_n}\right)^2\right).$$

Tällöin

$$q_D(y) + b \cdot y = q_D(y + \tilde{b}) + \tilde{a}.$$

□

Kirjataan nyt yleinen tulos, jonka nämä lemmat yhdessä todistavat.

Lause D.2.5. *Olkoon $p\colon \mathbb{R}^{n\times 1} \to \mathbb{R}$ toisen asteen polynomi. Mikäli*

$$p(x) = q_A(x) + b \cdot y + a$$

jollain neliömatriisilla $A \in \mathbb{R}^{n\times n}$, vektorilla $b \in \mathbb{R}^{n\times 1}$ ja luvulla $a \in \mathbb{R}$ ja matriisin A spektraalihajotelma $A = PDP^T$ on sellainen, että matriisin D diagonaalialkiot $d_1, \ldots, d_n$ ovat kaikki nollasta poikkeavia, niin

$$p(x) = 0$$

jos ja vain jos vektorille $y = [y_1 \cdots y_n]^T = Px$ pätee

$$d_1\left(y_1 - \frac{b_1}{2d_1}\right)^2 + \cdots + d_n\left(y_n - \frac{b_1}{2d_1}\right)^2 + a = \sum_{j=1}^{n}\left(\frac{b_i}{2d_i}\right)^2.$$

Huomautus D.2.6. *Lauseen sanoma on, että yleinen toisen asteen polynomiyhtälö $p(x) = 0$ avaruudessa $\mathbb{R}^n$ voidaan tässä tapauksessa saattaa avaruuden $\mathbb{R}^n$ ortogonaalikuvauksella, jota edustaa matriisi P, yhtälöksi, jossa ei ole toisen kertaluvun ristitermejä, eli muotoon*

$$d_1(y_1 - c_1)^2 + \cdots + d_n(y_n - c_n)^2 + a = e.$$

Yhden muuttujan tapauksessa tämä vastaa yhtälöä

$$d(y - c)^2 + a = 0 \quad \text{eli} \quad (y - c)^2 = \frac{e - a}{d},$$

jonka ratkaisut ovat

$$y = c \mp \sqrt{\frac{e - a}{d}},$$

jos $e - a/d \geq 0$.

Lukijaa voi tässä vaiheessa alkaa vaivaamaan, että mitä tapahtuu, jos spektraalihajotelmassa matriisi D sisältää nollia, eli että nolla on matriisin A ominaisrvo. Tällöin polynomin p ratkaisujoukko on paraabeli (tai sen yleistys). Käsitellään tämä yhden esimerkin valossa.

Esimerkki D.2.7. *Olkoon* $p\colon \mathbb{R}^{3\times 1} \to \mathbb{R}$ *neliömuoto*

$$p([x_1 x_2 x_3]^T) = x_1^2 - x_2^2 - x_3$$

kaikilla $\begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix}^T \in \mathbb{R}^{3\times 1}$.

Olkoot

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

ja $b = \begin{bmatrix} 0 & 0 & -1 \end{bmatrix}^T \in \mathbb{R}^{3\times 1}$. *Tällöin*

$$p(x) = q_A(x) + b \cdot x.$$

Koska A *on jo diagonaalinen, niin spektraalihajotelmassa* $P = I$ *ja* $D = A$.

Nyt yhtälö

$$p(x) = 0$$

vastaa yhtälöä

$$x_3 = x_1^2 - x_2^2.$$

avaruudessa $\mathbb{R}^3$. *(Piirrä kuva.)*

Liite E

Determinantin määritelmä

Tässä luvussa osoittetaan, että determinantin induktiivinen määritelmä voidaan purkaa myös auki kaavaksi, joka antaa deteminantin ilman induktiivista määritelmää. Määritellään tätä varten permutaation ja transposition käsitteet ja permutaation merkki.

E.1 Permutaatiot

Määritelmä E.1.1. *Joukon* $\{1,\dots,n\}$ *bijektiota* $\sigma\colon \{1,\dots,n\} \to \{1,\dots,n\}$ *kutsutaan* permutaatioksi. *Permutaatio* $\sigma\colon \{1,\dots,n\} \to \{1,\dots n\}$ *on* transpositio, *jos on olemassa sellaiset luvut* $i, j \in \{1,\dots,n\}$, $i \neq j$, *että* $\sigma(i) = j$, $\sigma(j) = i$ *ja* $\sigma(k) = k$ *kaikilla* $k \in \{1,\dots,n\} \setminus \{i, j\}$.

Joukon $\{1,\dots,n\}$ permutaatioiden joukkoa

$$\mathrm{Sym}(n) = \{\sigma\colon \{1,\dots,n\} \to \{1,\dots,n\} : \sigma \text{ on permutaatio}\}$$

kutsutaan joukon $\{1,\dots,n\}$ *symmetriaryhmäksi*, koska kahden permutaation yhdiste on permutaatio ja permutaation käänteiskuvaus on permutaatio. Näin ollen joukon $\{1,\dots,n\}$ permutaatiot muodostavat ryhmän algebrallisessa mielessä.

E.1.1 Permutaation merkki

Permutaation merkki on permutaatioon liittyvä luku 1 tai -1, joka kuvastaa samanaikaisesti kahta permutaation $\sigma\colon \{1,\dots,n\} \to \{1,\dots,n\}$ ominaisuutta:

- sitä vaihtaako σ lukujen $1 \leq i < j \leq n$ järjestyksen parillisen vai parittoman monta kertaa ja
- sitä tarvitaanko permutaation σ esittämiseen parillinen vai pariton määrä transpositioita.

Yleensä määritelmäksi valitaan jälkimmäinen ominaisuus. Tässä esityksessä valitaan kuitenkin ensimmäinen, koska se on sekä määritelmällisesti helpompi että teoreettisesti mielenkiintoisempi.

Määritelmä E.1.2. *Permutaation* $\sigma \in \mathrm{Sym}(n)$ *merkki* $\mathrm{sign}(\sigma) \in \{-1, +1\}$ *on*

$$\mathrm{sign}(\sigma) = (-1)^k,$$

missä $k \in \mathbb{N}$ *on niiden lukuparien* (i, j), *missä* $1 \leq i < j \leq n$, *lukumäärä, joille pätee* $\sigma(j) < \sigma(i)$.

Huomautus E.1.3. *Suoraan määritelmästä havaitaan, että transposition merkki on aina* -1 *ja että identtisen permutaation merkki on* 1.

Osoitetaan nyt, että permutaatioiden yhdisteen merkki on merkkien tulo.

Lause E.1.4. *Olkoot* $\sigma, \tau \in \mathrm{Sym}(n)$. *Tällöin* $\mathrm{sign}(\sigma \circ \tau) = \mathrm{sign}(\sigma)\mathrm{sign}(\tau)$.

Todistus. Ennen varsinaisen väitteen todistamista tehdään havainto. Olkoon $n \in \mathbb{N}$ ja $P \colon \mathbb{R}^n \to \mathbb{R}$ polynomi

$$P(x_1, \ldots, x_n) = \prod_{1 \leq i < j \leq n} (x_j - x_i)$$

kaikilla $(x_1, \ldots, x_n) \in \mathbb{R}^n$.

Olkoon $\sigma \in \mathrm{Sym}(n)$ permutaatio ja määritellään polynomi $P_\sigma \colon \mathbb{R}^n \to \mathbb{R}$ kaavalla

$$P_\sigma(x_1, \ldots, x_n) = P(x_{\sigma(1)}, \ldots, x_{\sigma(n)})$$

kaikilla $(x_1, \ldots, x_n) \in \mathbb{R}^n$. Tällöin

$$P_\sigma(x_1, \ldots, x_n) = \prod_{1 \leq i < j \leq n} (x_{\sigma(j)} - x_{\sigma(i)})$$

kaikilla $(x_1, \ldots, x_n) \in \mathbb{R}^n$.

Olkoot $1 \leq i < j \leq n$. Tällöin joko $\sigma(i) < \sigma(j)$ tai $\sigma(i) > \sigma(j)$. Jos $\sigma(i) < \sigma(j)$, niin monomi $(x_{\sigma(j)} - x_{\sigma(i)})$ on polynomin P tulontekijä. Jos puolestaan $\sigma(i) > \sigma(j)$, niin monomi $-(x_{\sigma(i)} - x_{\sigma(j)})$ on polynomin P tulontekijä. Havainnon perusteella

$$P_\sigma(x_1, \ldots, x_n) = \mathrm{sign}(\sigma) P(x_1, \ldots, x_n),$$

kaikilla $x_1, \ldots, x_n$, eli polynomeina

$$P_\sigma = \mathrm{sign}(\sigma) P.$$

Olkoot nyt $\sigma, \tau \in \mathrm{Sym}(n)$ permutaatioita. Koska

$$\begin{aligned}
\mathrm{sign}(\sigma \circ \tau) P(x_1, \ldots, x_n) &= P(x_{\sigma(\tau(1))}, \ldots, x_{\sigma(\tau(n))}) \\
&= P_\tau(x_{\sigma(1)}, \ldots, x_{\sigma(n)}) \\
&= \mathrm{sign}(\tau) P(x_{\sigma(1)}, \ldots, x_{\sigma(n)}) \\
&= \mathrm{sign}(\tau) P_\sigma(x_1, \ldots, x_n) \\
&= \mathrm{sign}(\tau)\mathrm{sign}(\sigma) P(x_1, \ldots, x_n)
\end{aligned}$$

kaikilla $(x_1, \ldots, x_n) \in \mathbb{R}^n$, niin

$$\mathrm{sign}(\sigma \circ \tau) = \mathrm{sign}(\sigma)\mathrm{sign}(\tau).$$

□

Edellisestä lauseesta seuraa suoraan permutaation merkin toinen tulkinta.

Korollaari E.1.5. *Olkoon $\sigma \in \mathrm{Sym}(n)$ permutaatio ja $\alpha_1, \ldots, \alpha_\ell$ sellaisia transpositioita, että $\sigma =_1 \circ \cdots \circ \alpha_\ell$. Tällöin*

$$\mathrm{sign}(\sigma) = (-1)^\ell.$$

Huomautus E.1.6. *Lause E.1.4 osoittaa, että permutaation merkki* $\mathrm{sign}(\cdot)$ *määrittelee ryhmähomomorfismin* $\mathrm{sign}\colon \mathrm{Sym}(n) \to \mathbb{F}_2$*, missä $\mathbb{F}_2 = \{-1, 1\}$ on kahden alkion multipilikatiivinen ryhmä.*

Huomautus E.1.7. *Lauseen E.1.4 todistus saattaa vaikuttaa täysin hihasta vedetyltä. (Se ei ole kirjoittajan keksimä!) Sillä on kuitenkin syvällinen tulkinta.*

Joukon $\mathrm{Sym}(n)$ *permutaatiot permutoivat luonnolisella tavalla avaruuden $\mathbb{R}^n$ koorodinaatteja. (Itseasiassa jokaisella permutatiolla σ kuvaus $(x_1, \ldots, x_n) \mapsto (x_{\sigma(1)}, \ldots, x_{\sigma(n)})$ on avaruuden $\mathbb{R}^n$ lineaarinen isomorfismi.) Näin ollen jokainen permutaatio $\sigma \in \mathrm{Sym}(n)$ itseasiassa määrittelee kuvauksen avaruuden $\mathbb{R}^n$ polynomien avaruudelta itseensä kaavalla $Q \mapsto Q_\sigma$, missä Q on avaruuden $\mathbb{R}^n$ polynomi ja Q_σ on määritelty vastaavalla tavalla kuin todistuksessa polynomi P_σ.*

Koska joukko $\mathrm{Sym}(n)$ *on kuvausten yhdistämisen suhteen ryhmä, niin edellä olevat kuvaukset $Q \mapsto Q_\sigma$ määrittelevät ryhmän* $\mathrm{Sym}(n)$ *toiminnan polynomien avaruuteen. Koska polynomille P pätee, että joko $P_\sigma = P$ tai $P_\sigma = -P$, niin paljastuu, että ryhmä* $\mathrm{Sym}(n)$ *toimii aliavaruuteen* $\mathrm{Sp}\{P\}$ *kuten ryhmä $\mathbb{Z}_2$ ja toiminta määräytyy permutaation merkistä. Aiheesta lisää maisteriopinnoissa tai hakusanalla* group action *internetistä.*

E.1.2 Permutaatiot transpositioiden yhdisteenä

Osoitetaan vielä täydellisyyden vuoksi, että jokainen permutaatio on transpositioiden yhdiste. Yhden alkion joukon $\{1\}$ tapaus jätetään lukijalle.

Lemma E.1.8. *Olkoon $n \geq 2$. Jokainen joukon $\{1, \ldots, n\}$ permutaatio $\sigma\colon \{1, \ldots, n\} \to \{1, \ldots, n\}$ voidaan kirjoittaa yhdistettynä kuvauksena transpositioista.*

Todistus. Todistetaan väite induktiolla. Oletetaan ensin, että $n = 2$ ja että $\sigma\colon \{1, 2\} \to \{1, 2\}$ on permutaatio. Tällöin, joko $\sigma = \mathrm{id}$ tai σ on transpoosi $\tau\colon \{1, 2\} \to \{1, 2\}$, jolle pätee $\tau(1) = 2$ ja $\tau(2) = 1$. Koska $\mathrm{id} = \tau \circ \tau$, niin väite pätee.

Oletetaan nyt, että väite pätee luvulla $n \in \mathbb{N}$ ja olkoon $\sigma\colon \{1, \ldots, n+1\} \to \{1, \ldots n+1\}$ permutaatio. Olkoon nyt $i = \sigma(n+1)$ ja olkoon $\tau\colon \{1, \ldots, n+1\} \to \{1, \ldots, n+1\}$ transpoosi $\tau(i) = n + 10$. Tällöin $\tau \circ \sigma$ on permutaatio, jolle pätee $(\tau \circ \sigma)(n+1) = \tau(\sigma(n+1)) = \tau(i) = n + 1$. Näin ollen $\rho\colon \{1, \ldots, n\} \to \{1, \ldots, n\}$, $j \mapsto (\tau \circ \sigma)(j)$, on joukon $\{1, \ldots, n\}$ permutaatio.

Induktio-oletuksen mukaan on olemassa sellaiset transpositiot $\tau_1, \ldots, \tau_k$, että $\rho = \tau_1 \circ \cdots \circ \tau_k$. Laajennetaan transposititiot $\tau_1, \ldots, \tau_k$ joukon $\{1, \ldots, n+1\}$ transpositioiksi asettamalla $\tau_j(n+1) = n+1$ jokaisella $j \in \{1, \ldots, k\}$. Näin ollen näille uusille transpositioille $\tau_1, \ldots, \tau_k$ (joista käytetään samaa merkintää) pätee

$$\tau \circ \tau_1 \circ \tau_k = \sigma.$$

(Harjoitustehtävä) $\square$

E.2 Determinentin määritelmä permutaatiolla

Määritellään nyt neliömatriisin determinantti toisella tavalla käyttäen permutaatioita.

Määritelmä E.2.1. *Matriisin* $A = [a_{ji}] \in \mathbb{C}^{n\times n}$ determinantti $\widehat{\det}(A)$ *on*

$$\widehat{\det}(A) = \sum_{\sigma\in\mathrm{Sym}(n)} \mathrm{sign}(\sigma) a_{\sigma(1)1}a_{\sigma(2)2}\cdots a_{\sigma(n)n}.$$

Determinantin $\widehat{\det}(\cdot)$ määritelmä on suunnattoman hyödyllinen. Todistetaan sen avulla determinantin perusominaisuudet. Osoitetaan ensin, että determinantti ei muutu transpoosissa. Tämä perustuu havaintoon, että määritelmä voidaan antaa myös toisella tavalla.

Lemma E.2.2. *Olkoon* $A \in [a_{ji}] \in \mathbb{C}^{n\times n}$ *neliömatriisi. Tällöin*

$$\widehat{\det}(A) = \sum_{\sigma\in\mathrm{Sym}(n)} \mathrm{sign}(\sigma) a_{1\sigma(1)}a_{2\sigma(2)}\cdots a_{n\sigma(n)}.$$

Todistus. Olkoon $\sigma \in \mathrm{Sym}(n)$ permutaatio. Koska myös σ^{-1} on permutaatio, niin luvut $a_{\sigma(\sigma^{-1}(1))\sigma^{-1}(1)}, \ldots, a_{\sigma(\sigma^{-1}(n))\sigma^{-1}(n)}$ ovat luvut $a_{\sigma(1)1}, \ldots, a_{\sigma(n)n}$ uudelleen järjestettynä. Näin ollen

$$a_{\sigma(1)1}a_{\sigma(2)2}\cdots a_{\sigma(n)n} = a_{\sigma(\sigma^{-1}(1))\sigma^{-1}(1)}\cdots, a_{\sigma(\sigma^{-1}(n))\sigma^{-1}(n)} = a_{1\sigma^{-1}(1)}\cdots a_{n\sigma^{-1}(n)}.$$

Koska jokaisella $\sigma \in \mathrm{Sym}(n)$ pätee $\mathrm{sign}(\sigma^{-1}) = \mathrm{sign}(\sigma)$, niin

$$\begin{aligned}
\widehat{\det}(A) &= \sum_{\sigma\in\mathrm{Sym}(n)} \mathrm{sign}(\sigma) a_{1\sigma(1)}a_{2\sigma(2)}\cdots a_{n\sigma(n)} \\
&= \sum_{\sigma\in\mathrm{Sym}(n)} \mathrm{sign}(\sigma) a_{1\sigma^{-1}(1)}\cdots a_{n\sigma^{-1}(n)} \\
&= \sum_{\sigma\in\mathrm{Sym}(n)} \mathrm{sign}(\sigma^{-1}) a_{1\sigma^{-1}(1)}\cdots a_{n\sigma^{-1}(n)}
\end{aligned}$$

Näin ollen, koska summa yli joukon $\mathrm{Sym}(n)$ summausindeksillä $\sigma^{-1} \in \mathrm{Sym}(n)$ vastaa summausjärjesyksen muuttamista, niin

$$\begin{aligned}
\widehat{\det}(A) &= \sum_{\sigma\in\mathrm{Sym}(n)} \mathrm{sign}(\sigma^{-1}) a_{1\sigma^{-1}(1)}\cdots a_{n\sigma^{-1}(n)} \\
&= \sum_{\sigma^{-1}\in\mathrm{Sym}(n)} \mathrm{sign}(\sigma^{-1}) a_{1\sigma^{-1}(1)}\cdots a_{n\sigma^{-1}(n)} \\
&= \sum_{\tau\in\mathrm{Sym}(n)} \mathrm{sign}(\tau) a_{1\tau(1)}\cdots a_{n\tau(n)}.
\end{aligned}$$

$\square$

Seuraava lause on oikeasti vain edellisen lemman korollaari, mutta kirjataan kuitenkin lauseeksi sen tärkeyden vuoksi.

Lause E.2.3. *Olkoon $A \in \mathbb{C}^{n\times n}$ neliömatriisi. Tällöin $\widehat{\det}(A) = \widehat{\det}(A^T)$.*

Todistus. Lemman E.2.2 perusteella

$$\begin{aligned}\widehat{\det}(A^T) &= \sum_{\sigma\in\mathrm{Sym}(n)} \mathrm{sign}(\sigma)(A^T)_{\sigma(1)1}\cdots(A^T)_{\sigma(n)n}\\ &= \sum_{\sigma\in\mathrm{Sym}(n)} \mathrm{sign}(\sigma)A_{1\sigma(1)}\cdots A_{n\sigma(n)} = \widehat{\det}(A).\end{aligned}$$

□

Osoitetaan seuraavaksi, että matriisin rivejä tai sarakkeita permutoitaessa matriisin determinantti muuttuu permuaation merkillä. Todistetaan väite sarakkeille . Riveille väite seuraa transponoimalla.

Lause E.2.4. *Olkoon $A = [a_{ji}] = [a_1 \cdots a_n] \in \mathbb{C}^{n\times n}$ neliömatriisi, missä vektorit $a_1, \ldots, a_n \in \mathbb{C}^{n\times 1}$ ovat matriisin A sarakkeet. Olkoon $\xi \in \mathrm{Sym}(n)$ permutaatio. Tällöin*

$$\widehat{\det}\left[a_{\xi(1)}\cdots a_{\xi(n)}\right] = \mathrm{sign}(\xi)\widehat{\det}A.$$

Todistus. Merkitään

$$B = [b_1 \cdots b_n] = [a_{\xi(1)} \cdots a_{\xi(n)}] \in \mathbb{C}^{n\times n}.$$

Tällöin

$$\begin{aligned}\widehat{\det}(B) &= \sum_{\sigma\in\mathrm{Sym}(n)} \mathrm{sign}(\sigma)b_{\sigma(1)1}\cdots b_{\sigma(n)n}\\ &= \sum_{\sigma\in\mathrm{Sym}(n)} \mathrm{sign}(\sigma)a_{\sigma(\xi(1))1}\cdots b_{\sigma(\xi(n))n}\\ &= \sum_{\sigma\in\mathrm{Sym}(n)} \mathrm{sign}(\xi)\mathrm{sign}(\sigma\circ\xi)a_{(\sigma\circ\xi)(1))1}\cdots a_{(\sigma\circ\xi)(n))n}\\ &= \mathrm{sign}(\xi)\sum_{\sigma\in\mathrm{Sym}(n)} \mathrm{sign}(\sigma\circ\xi)a_{(\sigma\circ\xi)(1))1}\cdots a_{(\sigma\circ\xi)(n))n}\end{aligned}$$

Havaitaan nyt, että kuvaus $P\colon \mathrm{Sym}(n) \to \mathrm{Sym}(n)$, joka on määritelty kaavalla $\sigma \mapsto \sigma\circ\xi$, on bijektio, jonka käänteiskuvaus P^{-1} on määritelty kaavalla $\sigma \mapsto \sigma\circ\xi^{-1}$. Näin ollen summa yli kaikkien permutaatioiden $\sigma \in \mathrm{Sym}(n)$ sama kuin summa yli kaikkien permutaatioiden $\sigma \circ \xi \in \mathrm{Sym}(n)$. Näin ollen

$$\begin{aligned}
\widehat{\det}(B) &= \operatorname{sign}(\xi) \sum_{\sigma \in \mathrm{Sym}(n)} \operatorname{sign}(\sigma \circ \xi) a_{(\sigma\circ\xi)(1))1} \cdots a_{(\sigma\circ\xi)(n))n} \\
&= \operatorname{sign}(\xi) \sum_{\sigma\circ\xi \in \mathrm{Sym}(n)} \operatorname{sign}(\sigma \circ \xi) a_{(\sigma\circ\xi)(1))1} \cdots a_{(\sigma\circ\xi)(n))n} \\
&= \operatorname{sign}(\xi) \sum_{\tau \in \mathrm{Sym}(n)} \operatorname{sign}(\tau) a_{\tau(1))1} \cdots a_{\tau(n))n} = \operatorname{sign}(\xi)\widehat{\det}(A).
\end{aligned}$$

□

Äskeisellä lauseella on tärkeä korollaari: jos neliömatriisissa on kaksi samaa sareketta tai riviä, niin determinantti on silloin nolla.

Korollaari E.2.5. *Olkoon $A = [a_1 \cdots a_n] \in \mathbb{C}^{n\times n}$ sellainen matriisi, että $a_i = a_j$ jollain $i \neq j$. Tällöin $\widehat{\det}(A) = 0$.*

Todistus. Olkoon $\sigma \in \mathrm{Sym}(n)$ transpositio, jolle $\sigma(i) = j$. Koska $a_{\sigma(i)} = a_j = a_i$ ja $a_{\sigma(j)} = a_i = a_j$, niin

$$\widehat{\det}(A) = \widehat{\det}\,[a_1 \cdots a_n] = \widehat{\det}\left[a_{\sigma(1)} \cdots a_{\sigma(n)}\right] = \operatorname{sign}(\sigma)\widehat{\det}\,[a_1 \cdots a_n] = -\widehat{\det}(A).$$

Näin ollen $\widehat{\det}(A) = 0$. □

Osoitetaan myös determinentin lineaarisuus sarakkeiden suhteen. Otetaan tätä varten käyttöön merkintä. Olkoon $A = [a_1 \cdots a_n] \in \mathbb{C}^{n\times n}$ matriisi, jonka sarakkeet ovat $a_1, \ldots, a_n \in \mathbb{C}^{n\times 1}$. Olkoot $i \in \{1, \ldots, n\}$ ja $v \in \mathbb{C}^{n\times 1}$. Merkitään

$$A_i(v) = [a_1 \cdots a_{i+1}\ v\ a_{i+1} \cdots a_n] \in \mathbb{C}^{n\times n}.$$

Lause E.2.6. *) Olkoot $v, w \in \mathbb{C}^{n\times 1}$, $a_1, \ldots, a_n \in \mathbb{C}^{n\times 1}$ ja $a \in \mathbb{C}$. Tällöin jokaisella $i \in \{1, \ldots, n\}$ pätee*

$$\widehat{\det}(A_i(v + aw)) = \widehat{\det}(A_i(v)) + a\widehat{\det}(A_i(w)).$$

Todistus. Todistus on suora lasku. Koska

$$\begin{aligned}
\widehat{\det}(A_i(v+aw)) &= \sum_{\sigma\in\mathrm{Sym}(n)} \mathrm{sign}(\sigma)(A_i(v+aw))_{\sigma(1)1}\cdots(A_i(v+aw))_{\sigma(i)i}\cdots(A_i(v+aw))_{\sigma(n)n} \\
&= \sum_{\sigma\in\mathrm{Sym}(n)} \mathrm{sign}(\sigma)a_{\sigma(1)1}\cdots\left(v_{\sigma(i)}+aw_{\sigma(i)}\right)\cdots a_{\sigma(n)n} \\
&= \sum_{\sigma\in\mathrm{Sym}(n)} \mathrm{sign}(\sigma)a_{\sigma(1)1}\cdots v_{\sigma(i)}\cdots a_{\sigma(n)n} \\
&\quad + a\sum_{\sigma\in\mathrm{Sym}(n)} \mathrm{sign}(\sigma)a_{\sigma(1)1}\cdots w_{\sigma(i)}\cdots a_{\sigma(n)n} \\
&= \sum_{\sigma\in\mathrm{Sym}(n)} \mathrm{sign}(\sigma)(A_i(v))_{\sigma(1)1}\cdots(A_i(v))_{\sigma(i)i}\cdots(A_i(v))_{\sigma(n)n} \\
&\quad + a\sum_{\sigma\in\mathrm{Sym}(n)} \mathrm{sign}(\sigma)(A_i(w))_{\sigma(1)1}\cdots(A_i(w))_{\sigma(i)i}\cdots(A_i(w))_{\sigma(n)n} \\
&= \widehat{\det}(A_i(v)) + a\widehat{\det}(A_i(w)),
\end{aligned}$$

niin väite pätee. □

Todistetaan äskeisen tuloksen avulla tärkeä determinanttien tulokaava.

Lause E.2.7. *Olkoot* $A = [a_1\cdots a_n], B = [b_1\cdots b_n] \in \mathbb{C}^{n\times n}$ *neliömatriiseja, missä* $a_1,\ldots,a_n,b_1,\ldots,b_n \in \mathbb{C}^{n\times 1}$. *Tällöin*

$$\widehat{\det}(AB) = \widehat{\det}(A)\widehat{\det}(B).$$

Todistus. Koska

$$AB = [Ab_1\cdots Ab_n] = [A(Be_1)\cdots A(Be_n)]$$

ja

$$Be_i = \sum_{m_i=1}^{n} b_{m_i i}e_{m_i}$$

jokaisella $i \in \{1,\ldots,m\}$, niin matriisitulon lineaarisuuden ja determinantin sarakelineaarisuuden (lause E.2.6) nojalla

$$\begin{aligned}
\widehat{\det}(AB) &= \widehat{\det}\left[A\left(\sum_{m_1=1}^{n} b_{m_1 1} e_{m_1}\right)\cdots A\left(\sum_{m_n=n}^{n} b_{m_n n} e_{m_n}\right)\right] \\
&= \widehat{\det}\left[\left(\sum_{m_1=1}^{n} b_{m_1 1} A e_{m_1}\right)\cdots\left(\sum_{m_n=n}^{n} b_{m_n n} A e_{m_n}\right)\right] \\
&= \sum_{m_1=1}^{n}\cdots\sum_{m_n=1}^{n} b_{m_1 1}\cdots b_{m_n n}\widehat{\det}\left[Ae_{m_1}\cdots Ae_{m_n}\right] \\
&= \sum_{m_1=1}^{n}\cdots\sum_{m_n=1}^{n} b_{m_1 1}\cdots b_{m_n n}\widehat{\det}\left[a_{m_1}\cdots a_{m_n}\right]
\end{aligned}$$

Korollaarin E.2.5 nojalla determinantti $\det\left[a_{m_1}\cdots a_{m_n}\right]$ on nolla, jos indeksit $m_1,\ldots,m_n$ saavat saman arvon. Toisaalta, jos indeksit $m_1,\ldots,m_n$ saavat kaikki eri arvoja, niin silloin kyseessa on joukon $\{1,\ldots,n\}$ permutaatio. Näin ollen

$$\begin{aligned}
\widehat{\det}(AB) &= \sum_{m_1=1}^{n}\cdots\sum_{m_n=1}^{n} b_{m_1 1}\cdots b_{m_n n}\widehat{\det}\left[a_{m_1}\cdots a_{m_n}\right] \\
&= \sum_{\sigma\in\mathrm{Sym}(n)} b_{\sigma(1)1}\cdots b_{\sigma(n)n}\widehat{\det}\left[a_{\sigma(1)}\cdots a_{\sigma(n)}\right] \\
&= \sum_{\sigma\in\mathrm{Sym}(n)} b_{\sigma(1)1}\cdots b_{\sigma(n)n}\mathrm{sign}(\sigma)\widehat{\det}\left[a_1\cdots a_n\right] \\
&= \widehat{\det}(A)\sum_{\sigma\in\mathrm{Sym}(n)} \mathrm{sign}(\sigma) b_{\sigma(1)1}\cdots b_{\sigma(n)n} = \widehat{\det}(A)\widehat{\det}(B).
\end{aligned}$$

□

E.3 Determinantin kaksi määritelmää yhtyvät

Osoitetaan nyt, että nämä kahdella eri määritelmää yhtyvät.

Lause E.3.1. *Olkoon $A\in\mathbb{C}^{n\times n}$ neliömatriisi. Tällöin $\widehat{\det}(A)=\det(A)$.*

Todistus. Käsitellään ensin tapaus $n=1$. Koska $\mathrm{Sym}(1)=\{\mathrm{id}\colon\{1\}\to\{1\}\}$, niin matriisille $A=[a]\in\mathbb{C}^{1\times 1}$ pätee

$$\widehat{\det}(A)=\mathrm{sign}(\mathrm{id})a=a=\det(A).$$

Oletetaan nyt, että $\widehat{\det}(B)=\det(B)$ pätee kaikilla $(n-1)\times(n-1)$-matriiseilla B, missä $n\in\mathbb{N}$. Olkoon $A\in\mathbb{C}^{n\times n}$. Merkitään

$$\mathrm{Sym}(n,i)=\{\sigma\in\mathrm{Sym}(n)\colon \sigma(1)=i\}$$

jokaisella $i \in \{1, \ldots, n\}$. Tällöin joukot $\mathrm{Sym}(n, 1), \ldots, \mathrm{Sym}(n, n)$ ovat pistevieraita ja niiden yhdiste on $\mathrm{Sym}(n)$.

Määritellään jokaisella $i \in \{1, \ldots, n\}$ permutaatio $\tau_i \colon \{1, \ldots, n\} \to \{1, \ldots, n\}$ kaavalla

$$\tau_i(j) = \begin{cases} j+1, & j < i \\ 1, & j = i \\ j, & j > i. \end{cases}$$

Määritellään lisäksi kuvaus $T_i \colon \mathrm{Sym}(n) \to \mathrm{Sym}(n)$ kaavalla $\sigma \mapsto \tau \circ \sigma$. Tällöin jokaiselle $\sigma \in \mathrm{Sym}(n, i)$ pätee, että $T_i(\sigma)$ on permutaatio, jolle pätee

$$(T_i(\sigma))(1) = (\tau_i \circ \sigma)(1) = \tau_i(\sigma(1)) = \tau_i(i) = 1,$$

eli $T_i(\sigma) \in \mathrm{Sym}(n, 1)$. Itseasiassa T_i on joukon $\mathrm{Sym}(n)$ bijektio itselleen, jonka käänteiskuvaus on $\sigma \mapsto \tau_i^{-1} \circ \sigma$ ja jolle pätee $T_i(\mathrm{Sym}(n, i)) = \mathrm{Sym}(n, 1)$. (Harjoitustehtävä) Koska $\mathrm{sign}(\tau_i) = (-1)^{i-1}$, niin myös $\mathrm{sign}(\tau_i^{-1}) = (-1)^{i-1}$. (Harjoitustehtävä)

Määritellään nyt, että A_i on se matriisi, joka saadaan permutoimalla matriisin A rivit permutaatiolla τ_i, eli matriisin A_i rivi j on matriisin A rivi $\tau_i^{-1}(j)$ jokaisella $j \in \{1, \ldots, n\}$. Tällöin

$$\begin{aligned}
\widehat{\det}(A) &= \sum_{\sigma \in \mathrm{Sym}(n)} \mathrm{sign}(\sigma) a_{\sigma(1)1} a_{\sigma(2)2} \cdots a_{\sigma(n)n} \\
&= \sum_{i=1}^{n} \sum_{\sigma \in \mathrm{Sym}(n,i)} \mathrm{sign}(\sigma) a_{\sigma(1)1} a_{\sigma(2)2} \cdots a_{\sigma(n)n} \\
&= \sum_{i=1}^{n} \sum_{\xi \in \mathrm{Sym}(n,1)} \mathrm{sign}(\tau_i^{-1} \circ \xi) a_{(\tau_i^{-1} \circ \xi)(1)1} a_{(\tau_i^{-1} \circ \xi)(2)2} \cdots a_{(\tau_i^{-1} \circ \xi)(n)n} \\
&= \sum_{i=1}^{n} \mathrm{sign}(\tau_i^{-1}) \sum_{\xi \in \mathrm{Sym}(n,1)} \mathrm{sign}(\xi) (A_i)_{\xi(1)1} (A_i)_{\xi(2)2} \cdots (A_i)_{\xi(n)n} \\
&= \sum_{i=1}^{n} (-1)^{i-1} a_{i1} \sum_{\xi \in \mathrm{Sym}(n,1)} \mathrm{sign}(\xi) (A_i)_{\xi(2)2} \cdots (A_i)_{\xi(n)n}
\end{aligned}$$

Koska matriisi A_{i1} saadaan matriisista A_i poistamalla sen ensimmäinen rivi ja bijektio $s \colon \{1, \ldots, n-1\} \to \{2, \ldots, n\}$, $j \mapsto j+1$, samaistaa permutaatiot $\mathrm{Sym}(n, 1)$ ja permutaatioita $\mathrm{Sym}(n-1)$ luonnollisella tavalla (harjoitustehtävä), niin

$$\begin{aligned}
&\sum_{i=1}^{n}(-1)^{i-1}a_{i1}\sum_{\xi\in\mathrm{Sym}(n,1)}\mathrm{sign}(\xi)(A_i)_{\xi(2)2}\cdots(A_i)_{\xi(n)n}\\
&=\sum_{i=1}^{n}(-1)^{i-1}a_{i1}\sum_{\zeta\in\mathrm{Sym}(n-1)}\mathrm{sign}(\zeta)(A_{i1})_{\zeta(1)1}\cdots(A_{i1})_{\zeta(n-1),n-1}\\
&=\sum_{i=1}^{n}(-1)^{i-1}a_{i1}\widehat{\det}(A_{i1}).
\end{aligned}$$

Näin ollen induktio-oletuksen perusteella pätee

$$\widehat{\det}(A)=\sum_{i=1}^{n}(-1)^{i-1}a_{i1}\widehat{\det}(A_{i1})=\sum_{i=1}^{n}(-1)^{i-1}a_{i1}\det(A_{i1})=\det(A).$$

□

E.4 Sovellus: Determinantin kehityskaava ja matriisin kääntyvyys

Todistetaan kehitetyn teorian sovelluksen determinantin kehityskaava. Palautetaan tätä varten mieleen seuraava merkintä. Olkoon $A=[a_{ji}]\in\mathbb{C}^{n\times n}$ matriisi ja olkoot $i,j\in\{1,\ldots,n\}$. Tällöin $A_{ji}\in\mathbb{C}^{(n-1)\times(n-1)}$ on se matriisi, joka saadaan matriisista A poistamalla sen j:s rivi ja i:s sarake.

Lause E.4.1. *Olkoon $A=[a_{ji}]=[a_1\cdots a_n]\in\mathbb{C}^{n\times n}$, missä $a_1,\ldots,a_n\in\mathbb{C}^{n\times 1}$ ovat matriisin A sarakkeet. Tällöin jokaisella $i\in\{1,\ldots,n\}$ pätee*

$$\det(A)=\sum_{j=1}^{n}(-1)^{j+i}a_{ji}\det(A_{ji}).$$

Todistus. Olkoon $\xi\in\mathrm{Sym}(n)$ permutaatio, jolle pätee $\xi(i)=1$, $\xi(k)=k+1$ kaikilla $k<i$ ja $\xi(k)=k$ kaikilla $k>i$. Tällöin $\mathrm{sign}(\xi)=(-1)^{i-1}$.

Merkitään $B=[b_{kl}]=\left[a_{\xi(1)}\cdots a_{\xi(n)}\right]$. Tällöin $B_{j1}=A_{ji}$. Näin ollen determinantin induktiivisen määritelmän perusteella pätee

$$\begin{aligned}
\det(A)=\widehat{\det}(A)&=\mathrm{sign}(\xi)\widehat{\det}(B)=\mathrm{sign}(\xi)\det(B)\\
&=\mathrm{sign}(\xi)\sum_{j=1}^{n}(-1)^{j+1}b_{j1}\det(B_{j1})\\
&=(-1)^{i-1}\sum_{j=1}^{n}(-1)^{j+1}a_{ji}\det(A_{ji})\\
&=\sum_{j=1}^{n}(-1)^{j+i}a_{ji}\det(A_{ji})
\end{aligned}$$

□

Osoitetaan seuraavaksi matriisin kääntyvyyttä ja determinantteja koskeava tulos. Tulos perustuu havaintoon, että kääntyvän matriisin sarakkeet ovat lineaarisesti riippumaattomia.

Lause E.4.2. *Olkoon* $A = [a_1 \cdots a_n] \in \mathbb{C}^{n\times n}$*, jolle pätee* $\dim \mathrm{Col}(A) < n$*, eli jonka sarekkeet eivät ole riippumattomia. Tällöin* $\det(A) = 0$*.*

Todistus. Koska $\dim \mathrm{Col}(A) < n$, niin on olemassa sellainen $i \in \{1, \dots, n\}$, että sarake a_i on muiden sarakkeiden lineaarikombinaatio, eli

$$a_i = z_1 a_1 + \cdots + z_n a_n,$$

missä $z_i = 0$. Näin ollen determinantin sarakelineaarisuuden ja permutaatio-ominaisuuden nojalla pätee

$$\begin{aligned} \det(A) &= \widehat{\det}(A) = \widehat{\det}\,[a_1 \cdots a_{i-1}\ a_i\ a_{i+1} \cdots a_n] \\ &= \widehat{\det}\,[a_1 \cdots a_{i-1}\ (z_1 a_1 + \cdots + z_n a_n)\ a_{i+1} \cdots a_n] \\ &= \sum_{k=1}^{n} z_k \widehat{\det}\,[a_1 \cdots a_{i-1}\ a_k\ a_{i+1} \cdots a_n] = 0. \end{aligned}$$

$\square$

Korollaari E.4.3. *Neliömatriisi* $A \in \mathbb{C}^{n\times n}$ *on kääntyvä, jos ja vain jos* $\det(A) \neq 0$*.*

Todistus. Oletetaan ensin, että neliömatriisi $A \in \mathbb{C}^{n\times n}$ on kääntyvä. Tällöin tulokaavan perusteella $\det(A)\det(A^{-1}) = \widehat{\det}(A)\widehat{\det}(A^{-1}) = \widehat{\det}(AA^{-1}) = \widehat{\det}(I) = 1$. Näin ollen $\det(A) \neq 0$.

Toisaalta, jos $\det(A) \neq 0$, niin $\dim \mathrm{Col}(A) = n$. Näin ollen matriisin A sarakkeet muodostavat avaruuden $\mathbb{C}^{n\times 1}$ kannan. Näin ollen kuvaus $\varphi_A \colon \mathbb{C}^{n\times 1} \to \mathbb{C}^{n\times 1}$, $z \mapsto Az$, on isomorfismi. Käänteiskuvauksen φ_A^{-1} matriisi on siis matriisin A käänteismatriisi. $\square$

Liite F

Toinen todistus kompleksisen ominaisarvon olemassaololle

.

Tässä liitteessa annetaan toinen todistus kompleksisen ominaisarvon olemassaololle, eli lauseelle 6.1.1. Todistus perustuu seuraavaan määritelmään.

Määritelmä F.0.1. *Olkoon $f\colon V \to V$ lineaarioperaattori ja $P\colon \mathbb{C} \to \mathbb{C}$ polynomi $z \mapsto a_n z^n + \cdots + a_1 z + a_0$. Tällöin $P(f)\colon V \to V$ on kaavalla*

$$v \mapsto a_n f^n(v) + \cdots + a_1 f(v) + a_0,$$

määritelty lineaarioperaattori, missä $f^n(v) = f(f^{n-1}(v))$ kaikilla $n \geq 1$ ja $f^0 = \mathrm{id}_V$.

Tarvitsemme myös pienen lemman.

Lemma F.0.2. *Olkoon $f\colon V \to V$ lineaarioperaattori ja olkoot $P\colon \mathbb{C} \to \mathbb{C}$ ja $Q\colon \mathbb{C} \to \mathbb{C}$ polynomeja. Merkitään myöt $PQ\colon \mathbb{C} \to \mathbb{C}$ näiden polynomien tuloa, eli funktiota $z \mapsto P(z)Q(z)$. Tällöin*

$$PQ(f) = P(f) \circ Q(f) = Q(f) \circ P(f).$$

Todistus. Olkoon $P\colon \mathbb{C} \to \mathbb{C}$ polynomi $z \mapsto \sum_{j=0}^m a_j z^j$ ja $Q\colon \mathbb{C} \to \mathbb{C}$ polynomi $z \mapsto \sum_{k=0}^n b_k z^k$. Tällöin

$$PQ(z) = \left(\sum_{j=0}^m a_j z^j\right)\left(\sum_{k=0}^n b_k z^k\right) = \sum_{j=0}^m \left(a_j z^j \sum_{k=0} n b_k z^k\right) = \sum_{j=0}^m \sum_{k=0}^n a_j b_k z^{j+k}.$$

Koska f on lineaarikuvaus, niin kaikilla $v \in V$ pätee

$$\begin{aligned} PQ(f)(v) &= \sum_{j=0}^m \sum_{k=0}^n a_j b_k f^{j+k}(v) = \sum_{j=0}^m \sum_{k=0}^n a_j b_k f^j(f^k(v)) \\ &= \sum_{j=0}^m a_j f^j \left(\sum_{k=0}^n b_k f^{j+k}(v)\right) = \sum_{j=0}^m a_j f^j(Q(f)(v)) \\ &= P(f)\left(Q(f)(v)\right) = (P(f) \circ Q(f))(v). \end{aligned}$$

$\square$

Edellisen lemman merkitys on seuraava. Olkoon $P\colon \mathbb{C} \to \mathbb{C}$ polynomi $z \mapsto a_0 + a_1 z + \cdots + a_n z^n$. Tekijöihin jaon perusteella P voidaan kirjoittaa muodossa $P(z) = a(z-\lambda_1)\cdots(z-\lambda_n) = aP_1(z)\cdots P_n(z)$, missä $P_i\colon \mathbb{C} \to \mathbb{C}$ on polynomi $z \mapsto z - \lambda_i$. Lemman F.0.2 nojalla kaikilla $f\colon V \to V$ pätee

$$P(f)(v) = a_0 + a_1 f(v) + \cdots + a_n f^n(v) = a\left(P_1(f) \circ \cdots \circ P_n(f)\right)(v),$$

missä $P_i(f)\colon V \to V$ on operaattori $P_i(f) = f - \lambda_i \mathrm{id}_V$.

Lauseen 6.1.1 vaihtoehtoinen todistus. Olkoon $n = \dim V$ ja olkoon $v \in V \setminus \{0\}$ (jokin) vektori. Koska jonossa

$$(v, f(v), f^2(v), \ldots f^n(v))$$

on $n+1$ jäsentä, niin se ei ole lineaarisesti riippumaton. Näin ollen on olemassa sellaiset kertoimet $a_0, \ldots, a_n \in \mathbb{C}$, joista jokin on nollasta poikkeava, että

$$a_0 v + a_1 f(v) + \cdots + a_n f^n(v) = 0.$$

Havaitaan ensin, että jokin kertoimista $a_1, \ldots, a_n$ ei ole nolla. Jos näin pätisi, niin silloin $0 = a_0 v$. Tällöin $a_0 = 0$, sillä $v \neq 0$. Tämä on ristiriita, koska jokin kertoimista $a_0, \ldots, a_n$ ei ole nolla. Olemme siis päätelleet, että jokin kertoimista $a_1, \ldots, a_n$ ei ole nolla.

Olkoon $P\colon \mathbb{C} \to \mathbb{C}$ polynomi $z \mapsto a_0 + a_1 z + \cdots + a_n z^n$. Kompleksisten polynomien tekijöihin jaon (Korollaari 6.1.3) nojalla on olemassa $a \in \mathbb{C}$ ja sellaiset $\lambda_1, \ldots, \lambda_n \in \mathbb{C}$, että

$$P(z) = a(z-\lambda_1)\cdots(z-\lambda_n).$$

Lemman F.0.2 nojalla

$$0 = a_0 + a_1 f(v) + \cdots + a_n f^n(v) = a(f - \lambda_1 \mathrm{id}_V) \circ \cdots \circ (f - \lambda_n \mathrm{id}_V)(v).$$

Koska $v \neq 0$, niin päätellään, että on olemassa sellainen $i \in \{1, \ldots, n\}$, että $w = \left((f - \lambda_{i+1}\mathrm{id}_V) \cdots \circ \cdots (f - \lambda_n \mathrm{id}_V)\right)(v) \neq 0$, mutta $(f - \lambda_i \mathrm{id}_V)(w) = 0$. Näin ollen jokin operaattoreista $f - \lambda_i \mathrm{id}_V$ ei ole injektio ja on olemassa sellainen vektori $\tilde{v} \in V$, että $f(v) = \lambda_i \tilde{v}$. Operaattorilla f on siis ominaisarvo λ_i. $\square$

Liite G

Ääretönulotteiset vektoriavaruudet

.

To be written.