TAMPEREEN
YLIOPISTO

PRO GRADU -TUTKIELMA

Matematiikan, tilastotieteen ja filosofian laitos
Tilastotiede
Syyskuu 2005

ERNO MÄNTYSALO

Antiominaisarvot ja -vektorit tilastotieteessä

Tampereen yliopisto

Matematiikan, tilastotieteen ja filosofian laitos

MÄNTYSALO, ERNO: Antiominaisarvot ja -vektorit tilastotieteessä

Pro gradu -tutkielma, 60 s., 6 liites.

Tilastotiede

Syyskuu 2005

Tiivistelmä

Tässä tutkielmassa tarkastellaan reaalisen symmetrisen ja ei-negatiivisesti definiitin matriisin antiominaisarvojen ja antiominaisvektoreiden sovellutuksia tilastotieteessä. Tutkielman alussa esitellään ne ominaisarvoja koskevat tulokset, joita antiominaisarvoja ja -vektoreita tarkasteltaessa tarvitaan. Tutkielmassa johdetaan yhtälöt antiominaisarvoille ja -vektoreille ja tarkastellaan niiden geometrista merkitystä. Lisäksi johdetaan antiominaisarvon tulkinta OLSE:n ja BLUE:n välisenä Watsonin tehokkuuden alarajana kun mallimatriisi on vektori. Esimerkkinä tarkastellaan tapauksia joissa jäännöstermillä on AR(1)-, MA(1) ja tasakorrelaatiorakenne.

Tutkielmassa tarkastellaan pienimmän antiominaisarvon tulkintaa maksimaalisena kulmana kahden vektorin välillä, ja esitellään sovellutus, jossa pienin antiominaisarvo tulkitaan pienimmäksi mahdolliseksi korrelaatiokertoimeksi kahden havaintovektorin välillä, kun kovarianssirakenne tunnetaan. Lisäksi tutkielmassa esitellään kanoniset korrelaatiot, joitakin niiden ominaisuuksia ja johdetaan kanoniset korrelaatiot antiominaisarvojen funktiona. Tutkielmaan on myös kerätty kirjallisuudesta joitakin erilaisia esitysmuotoja antiominaisarvolle, sekä yhteyksiä joihinkin muihin epäyhtälöihin.

Antiominaisarvojen lisäksi tutkielmassa määritellään yleistetty antiominaisarvo ja sitä vastaava antiominaismatriisi. Yleistetylle antiominaisarvolle ja antiominaismatriisille johdetaan yhtälöt ominaisarvojen ja ominaisvektoreiden funktiona. Osoitetaan, että Watsonin tehokkuuden alaraja voidaan esittää pienimmän yleistetyn antiominaisarvon funktiona. Tutkielmassa tarkastellaan lyhyesti kuntoisuusluvun 'condition number' ja antiominaisarvon välistä suhdetta matriisin kuntoisuuden arvioinnissa. Lopuksi tutkielmassa johdetaan myös uskottavuussuhdetesti kovarianssirakenteen testaamiseen ja osoitetaan, että uskottavuussuhdetestin testisuure on yleistetyn antiominaisarvon neliö.

Asiasanat: antiominaisarvo, antiominaisvektori, yleistetty antiominaisarvo, OLSE:n tehokkuus, kollineaarisuus, kanoninen korrelaatio

Sisältö

1 Johdanto

Antiominaisarvot ja antiominaisvektorit ovat suhteellisen tuoretta käsitteistöä matematiikassa ja tilastotieteessä. Termi antiominaisarvo nousi esille vasta Gustafsonin (1972) artikkelissa, vaikka asiaa on käsitelty aikaisemmin matematiikassa ja tilastotieteessä muilla nimillä. Gustafsonin varhaisimmat artikkelit ovat suuntautuneet matematiikan puolelle, mutta myöhemmin hän on kirjoittanut artikkelin antiominaisarvojen ja -vektoreiden sovelluksista tilastotieteessä (Gustafson 2002, 2005). Antiominaisarvo-termin lanseerauksen myötä antiominaisarvoihin ja -vektoreihin liittyviä artikkeleita ovat kirjoittaneet Gustafsonin lisäksi ainakin Davis (1980), Rao (2005) ja Khattree (2001, 2002, 2003).

Tilastotieteen terminologiassa antiominaisarvo on vielä nuorempi käsite kuin matematiikassa. Ensimmäiset artikkelit tilastotieteen sovellutuksista tai tulkinnoista ovat tehneet Gustafson (2002) sekä Khattree (2003). Tätä varhaisemmassa kirjallisuudessa antiominaisarvoa ei olla käytetty terminä, mutta varsinainen sisältö on ollut käytössä ainakin Kantorovichin epäyhtälön (1948) julkaisusta lähtien. Gustafson (2002) käyttää artikkelissa antiominaisarvoja pienimmän neliösumman estimaattorin (OLSE:n) tehokkuuden sekä kanonisten korrelaatioiden tarkasteluun. Tätä aikaisemmin Khattree (2001, 2002) on kirjoittanut artikkeleita antiominaisarvojen laskemisesta, mutta vuonna 2003 Khattree julkaisi kokoavan artikkelin antiominaisarvoista tilastotieteessä (Khattree 2003). Tämä artikkeli onkin toiminut merkittävänä tiedonlähteenä tätä tutkielmaa kirjoitettaessa.

Tässä tutkielmassa rajoitutaan tarkastelemaan ainoastaan reaalisten, symmetristen ja ei-negatiivisesti definiittien matriisien antiominaisarvoja. Antiominaisarvojen teoria, kuten myös jotkut muut tutkielmassa esitetyt epäyhtälöt, on käsitelty kirjallisuudessa yleisemmässä muodossa. Tässä tutkielmassa ne kuitenkin esitetään ainoastaan reaalisten symmetristen ja ei-negatiivisesti definiittien matriisien erikoistapauksessa, sillä tutkielman valmistumiseen mennessä löytämäni tilastotieteen sovellutukset eivät vaadi tätä yleisempiä tuloksia.

Tutkimus on jaettu kuuteen lukuun. Luvussa 2 tarkastellaan ominaisarvoja ja ominaisvektoreita ja johdetaan graafinen esitys, jonka tarkoitus on tukea antiominaisarvojen graafista esitystä. Kappaleessa 2 otetaan myös todistamatta käyttöön ominaisarvohajotelma, jota tarvitaan myöhemmin tulevissa todistuksissa. Kappaleessa 3.1 tarkastellaan hieman antiominaisarvojen historiaa, sekä antiominaisarvojen kehityksen keskeisiä henkilöitä. Kappaleessa 3.2 määritellään antiominaisarvot ja antiominaisvektorit, sekä todistetaan että pienin antiominaisarvo saadaan suoraan Kantorovichin epäyhtälöstä. Tässä kappaleessa myös todistetaan keskeisin antiominaisarvoja koskeva lause, joka kertoo että antiominaisarvot ja -vektorit voidaan kirjoittaa ominaisarvojen ja ominaisasvektoreiden funktiona. Kappaleessa 3.3 johdetaan tehokkuuden minimi lineaarisessa mallissa jossa mallimatriisina on pelkkä vektori. Esimerkkeinä käsitellään tilannetta jossa kovarianssimatriisilla on AR(1), MA(1) sekä tasakorrelaatiorakenne. Luvussa 4 tarkastellaan pienimmän antiominaisarvon tulkintaa minimaalisena korrelaationa kahden havaintovektorin välillä, kun kovarianssirakenne on tunnettu.

Kappaleessa 5.1 esitellään kanoniset korrelaatiot, sekä joitakin kanonisten korrelaatioiden jatkossa tarvittavia tuloksia. Kappaleessa 5.2 johdetaan kanonisten korrelaatioiden ja antiominaisarvojen yhteys. Kappaleessa 6.1 esitellään joitakin kirjallisuudesta löytyviä antiominaisarvojen esitysmuotoja. Kappaleessa 6.2 esitellään lyhyesti muita epäyhtälöitä jotka ovat suoraan yhteydessä Kantorovichin epäyhtälöön ja samalla pienimpään antiominaisarvoon. Kappaleessa 6.3 esitellään ja otetaan todistamatta käyttöön Wieladtin epäyhtälö sekä esitellään kolmen esimerkin avulla sen sovellutuksia tilastotieteessä.

Kappaleessa 7.1 määritellään yleistetty antiominaisarvo ja osoitetaan, että yleistetty antiominaisarvo voidaan esittää antiominaisarvojen tulona. Lisäksi esitellään kaksi yksinkertaista esimerkkiä yleistetystä antiominaisarvosta. Kappaleessa 7.2 johdetaan, että Watsonin tehokkuus yhtyy ongelmaan yleistetyistä antiominaisarvoista. Tästä esitellään myös esimerkki, Watsonin tehokkuuden alaraja on piirrettynä autokorrelaation funktiona AR(1)- rakenteen tapauksessa. Kappaleessa yleistetään yleistetty antiominaisarvo, ja samalla siis Watsonin tehokkuus, koskemaan ei-negatiivisesti definiittiä kovarianssirakennetta. Tällöin voidaan tarkastella vakiotermin ulkopuolisen osavektorin tehokkuutta. Esimerkkeinä esitellään kyseisen osavektorin Watsonin tehokkuuden alarajat korrelaation funktiona malleissa AR(1)-, MA(1)- ja tasakorrelaatiorakenne. Kappaleessa 7.3 esitellään lyhyesti mitä tarkoitetaan matriisin kuntoisuudella, ja kuinka pienintä antiominaisarvoa voidaan käyttää hyödyksi aineiston kuntoisuuden arvioinnissa. Lisäksi esitellään eräs kuntoisuuluvun 'condition number' heikkous aineiston kuntoisuuden mittana sekä herätellään keskustelua yleistetyn antiominaisarvon käyttämisestä kuntoisuusmittana. Lopuksi kappaleessa 7.4 johdetaan uskottavuussuhdetesti otoskovarianssimatriisin testaamiseen, ja todetaan että se on yleistetyn antiominaisarvon neliö.

2 Ominaisarvot ja -vektorit

Ominaisarvoilla ovat keskeisiä työvälineitä tilastotieteessä sekä matematiikassa. Yksi erityinen sovellutus ominaisarvoille ja ominaisvektoreille on pääkomponenttianalyysi, jossa pyritään etsimään sellaista satunnaismuuttujien lineaarikombinaatio, jonka varianssi on mahdollisimman suuri. Esitellään tämä ominaisarvojen sovellutus, sillä se tukee jatkossa olevia graafisia esityksiä.

Olkoon $\boldsymbol{x}_{p\times 1}$ satunnaisvektori siten, että $E(\boldsymbol{x}) = \mathbf{0}$. Koska $\boldsymbol{x}$:n odotusarvo on nolla, $\boldsymbol{x}$:n kovarianssimatriisi $\boldsymbol{\Sigma} = E(\boldsymbol{x}\boldsymbol{x}')$. Olkoon $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \ldots, \gamma_p)'$ kerroinvektori, jonka avulla keskistetystä satunnaisvektorista $\boldsymbol{x}$ muodostetaan uusi muuttuja $\xi = \boldsymbol{\gamma}'\boldsymbol{x}$. Tämän uuden muuttujan varianssi on

$$\mathrm{var}(\xi) = E(\boldsymbol{\gamma}'\boldsymbol{x}\boldsymbol{x}'\boldsymbol{\gamma}) = \boldsymbol{\gamma}'E(\boldsymbol{x}\boldsymbol{x}')\boldsymbol{\gamma} = \boldsymbol{\gamma}'\boldsymbol{\Sigma}\boldsymbol{\gamma}.$$

Nyt ongelmana on muodostaa sellainen kerroinvektori $\boldsymbol{\gamma}$, jotta uuden muuttujan ξ varianssi $\boldsymbol{\gamma}'\boldsymbol{\Sigma}\boldsymbol{\gamma}$ on mahdollisimman suuri ehdon $\boldsymbol{\gamma}'\boldsymbol{\gamma} = 1$ vallitessa. Tämä ehto tarkoittaa, että kerroinvektorin $\boldsymbol{\gamma}$ alkioiden neliöiden summa $\gamma_1^2 + \gamma_2^2 + \cdots + \gamma_p^2 = 1$. Ehto on tarpeen, sillä mielivaltaisen suurilla $\boldsymbol{\gamma}$:n alkion arvoilla $\mathrm{var}(\xi) = \boldsymbol{\gamma}'\boldsymbol{\Sigma}\boldsymbol{\gamma}$ olisi ääretön ja mitään ääriarvoa ei olisi olemassakaan.

Maksimointitehtävä voidaan ratkaista esimerkiksi Lagrangen kertojamenettelyn avulla. Muodostetaan Lagrangen funktio

$$Lag(\boldsymbol{\gamma}, \boldsymbol{\Sigma}) = \boldsymbol{\gamma}'\boldsymbol{\Sigma}\boldsymbol{\gamma} - \lambda(\boldsymbol{\gamma}'\boldsymbol{\gamma} - 1),$$

jossa $\lambda \in \mathbb{R}$ on Lagrangen kerroin. Derivoidaan lauseke $\boldsymbol{\gamma}$:n suhteen, jolloin stationaaripisteen välttämättömäksi ehdoksi saadaan, että

$$\frac{\partial Lag(\boldsymbol{\gamma}, \boldsymbol{\Sigma})}{\partial \boldsymbol{\gamma}} = 2\boldsymbol{\Sigma}\boldsymbol{\gamma} - 2\lambda\boldsymbol{\gamma} = \mathbf{0}.$$

Jakamalla yhtälö kakkosella, ja ottamalla $\boldsymbol{\gamma}$ yhteiseksi tekijäksi, saadaan ehto muotoon:

$$(\boldsymbol{\Sigma} - \lambda\boldsymbol{I})\boldsymbol{\gamma} = \mathbf{0}.$$

Jotta tällä homogeenisella yhtälöryhmällä olisi olemassa ei-triviaali ratkaisu, determinantin $|\boldsymbol{\Sigma} - \lambda \boldsymbol{I}|$ on oltava nolla, eli

$$|\boldsymbol{\Sigma} - \lambda \boldsymbol{I}| = 0. \tag{2.1}$$

Ominaisyhtälö (2.1) on λ:n p:nnen asteen polynomi, jonka juuria kutsutaan matriisin $\boldsymbol{\Sigma}$ ominaisarvoiksi $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$. Koska jokainen λ_i on ominaisyhtälön ratkaisu, ne toteuttavat yhtälöryhmän

$$\begin{aligned} (\boldsymbol{\Sigma} - \lambda \boldsymbol{I})\boldsymbol{\gamma} &= \boldsymbol{0} \\ \boldsymbol{\gamma}'\boldsymbol{\gamma} &= 1. \end{aligned}$$

Tällöin siis luonnollisesti myös ensimmäinen ominaisvektori $\boldsymbol{\gamma}_1$ ja sitä vastaava ominaisarvo λ_1 toteuttavat yhtälöryhmän

$$\begin{aligned} (\boldsymbol{\Sigma} - \lambda_1 \boldsymbol{I})\boldsymbol{\gamma}_1 &= \boldsymbol{0} \\ \boldsymbol{\gamma}_1'\boldsymbol{\gamma}_1 &= 1. \end{aligned} \tag{2.2}$$

Nyt kertomalla yhtälö (2.2) vasemmalta $\boldsymbol{\gamma}_1'$:lla saadaan

$$\boldsymbol{\gamma}_1'(\boldsymbol{\Sigma} - \lambda_1 \boldsymbol{I})\boldsymbol{\gamma}_1 = 0. \tag{2.3}$$

Jäsjestelemällä yhtälön (2.3) termit uudestaan ja sijoittamalla siihen yhtälöryhmän (2.2) viimeinen yhtälö voidaan yhtälö (2.3) kirjoittaa muodossa:

$$\boldsymbol{\gamma}_1'\boldsymbol{\Sigma}\boldsymbol{\gamma}_1 = \lambda_1.$$

Nyt suurimman ominaisarvon λ_1 tiedetään olevan uuden muuttujan ξ_1 varianssi. Toiseksi suurinta ominaisarvoa λ_2 vastaavan lineaarikombinaation kerroinvektori $\boldsymbol{\gamma}_2$ on siis 2. ominaisarvoa vastaava ominaisvektori. Toiseksi suurin ominaisarvo ja sitä vastaava ominaisvektori saadaan samasta maksimointitehtävästä kuin ensimmäisetkin ominaisarvot ja -vektorit, kuitenkin lisärajoituksilla $\boldsymbol{\gamma}_1'\boldsymbol{\gamma}_2 = 0$ ja $\boldsymbol{\gamma}_2'\boldsymbol{\gamma}_2 = 1$.

Jatkossa tullaan tarvitsemaan ominaisarvojen ja ominaisvektoreiden lisäksi ominaisarvohajotelmaa.

Lause 2.1. *Olkoon matriisi $\boldsymbol{A}_{n\times n}$ symmetrinen neliömatriisi. Tällöin*

$$\boldsymbol{A} = \boldsymbol{T}\boldsymbol{\Lambda}\boldsymbol{T}', \qquad \boldsymbol{T}'\boldsymbol{T} = \boldsymbol{T}\boldsymbol{T}' = \boldsymbol{I}_n, \tag{2.4}$$

missä $\boldsymbol{T}_{n\times n}$:n sarakkeina ovat matriisin $\boldsymbol{A}$ ortogonaaliset ja normeeratut ominaisvektorit, sekä matriisi $\boldsymbol{\Lambda} = diag(\lambda_1, \lambda_2, \ldots, \lambda_n)$ kun λ_i on matriisin $\boldsymbol{A}$ i. ominaisarvo.

Esimerkki 2.1. Olkoon $\boldsymbol{X} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3)$ keskistetty havaintoaineisto, jonka kovarianssimatriisi

$$\boldsymbol{\Sigma} = \begin{pmatrix} 6 & 2 & 1 \\ 2 & 8 & 2 \\ 1 & 2 & 12 \end{pmatrix}.$$

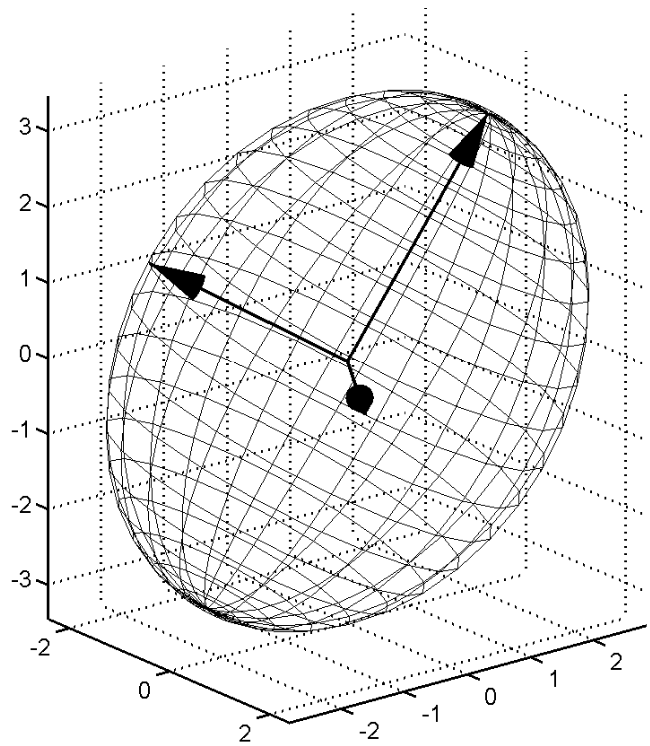
Tällöin ominaisarvohajotelman perusteella on olemassa $\boldsymbol{\Lambda}$ jonka diagonaalilla ovat ominaisarvot ja muualla nollaa, sekä ominaisvektoreiden muodostama matriisi $\boldsymbol{T}$:

$$\boldsymbol{\Lambda} = \begin{pmatrix} 4.7574 & 0 & 0 \\ 0 & 8 & 0 \\ 0 & 0 & 13.2426 \end{pmatrix}, \qquad \boldsymbol{T} = \begin{pmatrix} 0.8415 & 0.4851 & 0.2378 \\ -0.5392 & 0.7276 & 0.4241 \\ 0.0327 & -0.4851 & 0.8739 \end{pmatrix}.$$

Tarkastellaan kovarianssimatriisin $\boldsymbol{\Sigma}$ määrittelemää hyperellipsoidia

$$\boldsymbol{x}'\boldsymbol{\Sigma}^{-1}\boldsymbol{x} = \alpha, \tag{2.5}$$

missä $\alpha = 1$ ja $\boldsymbol{x} = (x, y, z)'$ on muuttujavektori. Kuviossa 2.1 on piirrettynä yhtälön (2.5) määrittelemä hyperellipsoidi sekä ominaisvektorit. Ominaisvektorit voidaan aina valita ykkösen mittaisiksi, mutta kuviossa ominaisvektoreiden pituudet ovat ominaisarvojen neliöjuuret, eli hyperellipsoidin pääakselien pituudet. Jos $\alpha \neq 1$, pääakselien pituudet ovat $\sqrt{\alpha\lambda_i}$, missä λ_i on matriisin $\boldsymbol{\Sigma}$ i:s ominaisarvo.



Kuvio 2.1. Kovarianssimatriisin määrittelemä hyperellipsoidi, johon on lisätty ominaisarvojen neliöjuurien pituiset ominaisvektorit.

3 Antiominaisarvot ja -vektorit

3.1 Antiominaisarvojen historia

Termin antiominaisarvo otti käyttöön Gustafson (1972) artikkelissaan ”Antieigenvalue inequalities in operator theory”. Gustafson (1972) tarkastelee ’disjoint’ -operaattoreita, ja Davis (1980) johtaa tulokset koskemaan myös ’normal’ -operaattoreita. Koska tässä tutkielmassa tarkastellaan vain reaalisia, symmetrisiä, positiivisesti definiittejä matriiseja, vektori $\boldsymbol{x} \in \mathbb{R}^n$ ja matriisi $\boldsymbol{A}$ on reaalinen, symmetrinen ja positiivisesti definiitti.

Gustafson tarkastelee (1972) artikkelissaan vektoreiden $\boldsymbol{x}$ ja $\boldsymbol{Ax}$ välistä kulmaa. Kahden tunnetun vektorin $\boldsymbol{x}$ ja $\boldsymbol{Ax}$ välisen kulman α kosini voidaan tunnetusti kirjoittaa muodossa

$$\cos\alpha = \frac{\boldsymbol{x}'\boldsymbol{Ax}}{\sqrt{\boldsymbol{x}'\boldsymbol{A}^2\boldsymbol{x} \cdot \boldsymbol{x}'\boldsymbol{x}}}. \tag{3.1}$$

Gustafson osoitti, että yhtälön (3.1) oikean puolen ääriarvot saadaan valitsemalla $\boldsymbol{x}$:ksi joko jokin matriisin $\boldsymbol{A}$ ominaisvektori tai antiominaisvektori $\boldsymbol{x}_1$. Kun vektoriksi $\boldsymbol{x}$ valitaan jokin matriisin $\boldsymbol{A}$ ominaisvektori, kulmaksi tulee nolla. Itse asiassa ominaisvektorilla $\boldsymbol{x}$ kertominen ainoastaan venyttää vektoria $\boldsymbol{Ax}$ vastaavan ominaisarvon suhteessa. Kun valitaan $\boldsymbol{x} = \boldsymbol{x}_1$, saavutetaan suurin mahdollinen kulma vektoreiden $\boldsymbol{x}$ ja $\boldsymbol{Ax}$ välille. Koska vektorilla $\boldsymbol{x}_1$ on kulman α suhteen juuri vastakkainen ominaisuus kuin ominaisvektorilla, Gustafson (1972) nimesi tämän suurimman mahdollisen kulman kosinin pienimmäksi antiominaisarvoksi ja vektorin $\boldsymbol{x}_1$ pienintä antiominaisarvoa vastaavaksi antiominaisvektoriksi.

Gustafson (1972) osoitti myös, että pienin antiominaisarvo yhtyy Kantorovichin epäyhtälön alarajaan. Nobel-palkinnon vuonna 1975 voittanut Leonid Vital´evich Kantorovich (1948) osoitti, että

$$\frac{\boldsymbol{y}'\boldsymbol{Ay} \cdot \boldsymbol{y}'\boldsymbol{A}^{-1}\boldsymbol{y}}{(\boldsymbol{y}'\boldsymbol{y})^2} \leq \frac{(\lambda_1 + \lambda_n)^2}{4\lambda_1\lambda_n}, \tag{3.2}$$

missä $\boldsymbol{A}_{n\times n}$ on reaalinen symmetrinen positiivisesti definiitti matriisi, $\boldsymbol{y} \in \mathbb{R}^n$ ja $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n > 0$ ovat matriisin $\boldsymbol{A}$ ominaisarvot.

Kantorovich julkaisi epäyhtälön venäjäksi vuonna 1948, mutta Watson, Alpargu ja Styan (1997) tekivät havainnon, että graafiteoreetikko Roberto Wertheimer Frucht (1943) oli julkaissut epäyhtälön espanjaksi viittä vuotta aikaisemmin. Tämän vuoksi jatkossa tulemmekin käyttämään yhtälöstä (3.2) nimeä Fruchtin–Kantorovichin epäyhtälö.

Kantorovichin (1948) ja Fruchtin (1943) erillään tekemien julkaisuiden tulokset olivat siis käytössä jo vuosista 1943 ja 1948 lähtien. Tämän vuoksi Gustafsonin (1972) reaalisia symmetrisiä positiivisesti definiittejä matriiseja koskevat tulokset olivat käytössä jo vuosikymmeniä aikaisemmin ennen antiominaisarvonimen lanseeraamista. Tilastotieteeseen ehtikin tänä aikana tulla huomattavasti sovellutuksia Fruchtin–Kantorovichin epäyhtälölle. Ensimmäisen varsinaisen antiominaisarvojen ja -vektoreiden tilastotieteen sovellutuksista kertovan artikkelin kirjoitti Gustafson (2002). Artikkelissaan hän lähestyy "operator trigonometry":n näkökulmasta Bloomfieldin, Watsonin (1975) ja Knottin (1975) esittelemää OLSE:n ja BLUE:n suhteellista tehokkuutta sekä kanonisia korrelaatioita.

3.2 Antiominaisarvot ja -vektorit

Antiominaisarvot ovat saaneet nimensä matemaattisten ominaisuuksiensa perusteella, kuten edellisessä luvussa 3.1 lyhyesti todettiin. Antiominaisarvoja tai antiominaisvektoreita ei ole kuitenkaan määritelty vielä sen tarkemmin.

Tarkastellaan Cauchyn–Schwarzin epäyhtälöä (Lause L.1), eli

$$(\boldsymbol{u}'\boldsymbol{v})^2 \leq \boldsymbol{u}'\boldsymbol{u} \cdot \boldsymbol{v}'\boldsymbol{v}, \qquad \boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^p.$$

Tehdään valinta $\boldsymbol{v} = \boldsymbol{x}$ ja $\boldsymbol{u} = \boldsymbol{A}\boldsymbol{x}$, missä $\boldsymbol{A}$ on reaalinen symmetrinen positiivisesti definiitti matriisi, ja vektori $\boldsymbol{x} \neq \boldsymbol{0}$. Tällöin Cauchyn–Schwarzin epäyhtälöstä seuraavat keskenään yhtäpitävät epäyhtälöt

$$\begin{aligned} (\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x})^2 &\leq \boldsymbol{x}'\boldsymbol{A}^2\boldsymbol{x} \cdot \boldsymbol{x}'\boldsymbol{x}, \\ \boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} &\leq \sqrt{\boldsymbol{x}'\boldsymbol{A}^2\boldsymbol{x} \cdot \boldsymbol{x}'\boldsymbol{x}} \end{aligned}$$

ja

$$\frac{\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}}{\sqrt{\boldsymbol{x}'\boldsymbol{A}^2\boldsymbol{x} \cdot \boldsymbol{x}'\boldsymbol{x}}} \leq 1. \tag{3.3}$$

Yhtälön (3.3) vasen puoli saavuttaa maksiminsa 1 kun $\boldsymbol{x}$ ja $\boldsymbol{A}\boldsymbol{x}$ ovat lineaarisesti riippuvia, kuten Cauchyn–Schwarzin epäyhtälön todistuksessa (Lause L.1) todetaan. Tällöin optimointitehtävän (3.5) ratkaisuksi $\boldsymbol{x}$ sopii esimerkiksi mikä tahansa matriisin $\boldsymbol{A}$ ominaisvektori. Tämä voidaan osoittaa helposti matriisin $\boldsymbol{A}$ ominaisarvohajotelman avulla.

Koska $\boldsymbol{A}$ on symmetrinen, sillä on ominaisarvohajotelma

$$\boldsymbol{A} = \boldsymbol{T\Lambda T}', \qquad \boldsymbol{T}'\boldsymbol{T} = \boldsymbol{T}\boldsymbol{T}' = \boldsymbol{I},$$

jossa $\boldsymbol{\Lambda}$ on diagonaalimatriisi jonka diagonaalilla ovat matriisin $\boldsymbol{A}$ ominaisarvot $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n > 0$ sekä $\boldsymbol{T} = (\boldsymbol{t}_1, \boldsymbol{t}_2, \cdots, \boldsymbol{t}_n)$, jossa $\boldsymbol{t}_i$ on λ_i:tä vastaava ominaisvektori.

Sijoittamalla ominaisarvohajotelma yhtälöön (3.3) saadaan

$$\text{(3.4)} \qquad \frac{\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}}{\sqrt{\boldsymbol{x}'\boldsymbol{A}^2\boldsymbol{x} \cdot \boldsymbol{x}'\boldsymbol{x}}} = \frac{\boldsymbol{x}'\boldsymbol{T\Lambda T}'\boldsymbol{x}}{\sqrt{\boldsymbol{x}'\boldsymbol{T\Lambda T}'\boldsymbol{T\Lambda T}'\boldsymbol{x} \cdot \boldsymbol{x}'\boldsymbol{x}}} = \frac{\boldsymbol{x}'\boldsymbol{T\Lambda T}'\boldsymbol{x}}{\sqrt{\boldsymbol{x}'\boldsymbol{T\Lambda}^2\boldsymbol{T}'\boldsymbol{x} \cdot \boldsymbol{x}'\boldsymbol{x}}},$$

jossa viimeinen yhtäsuuruus seuraa ominaisvektoreista koostuneen matriisin $\boldsymbol{T}$:n ortogonaalisuudesta $\boldsymbol{T}'\boldsymbol{T} = \boldsymbol{T}\boldsymbol{T}' = \boldsymbol{I}$.

Olkoon vektori $\boldsymbol{x}$ matriisin $\boldsymbol{A}$ k:nnetta ominaisarvoa vastaava ominaisvektori. Tällöin vektori $\boldsymbol{T}'\boldsymbol{x} = \boldsymbol{e}_k = (0, \cdots, 0, 1, 0, \cdots, 0)'$ on k:s kantavektori, ja yhtälö (3.4) saadaan muotoon

$$\frac{\boldsymbol{x}'\boldsymbol{T\Lambda T}'\boldsymbol{x}}{\sqrt{\boldsymbol{x}'\boldsymbol{T\Lambda}^2\boldsymbol{T}'\boldsymbol{x} \cdot \boldsymbol{x}'\boldsymbol{x}}} = \frac{\boldsymbol{e}_k'\boldsymbol{\Lambda}\boldsymbol{e}_k}{\sqrt{\boldsymbol{e}_k'\boldsymbol{\Lambda}^2\boldsymbol{e}_k}} = \frac{\lambda_k}{\sqrt{\lambda_k^2}} = 1,$$

missä λ_k on matriisin $\boldsymbol{\Lambda}$ k:s diagonaalialkio, eli matriisin $\boldsymbol{A}$ k:s ominaisarvo.

Cauchyn–Schwarzin epäyhtälön perusteella tiedetään, että

$$\text{(3.5)} \qquad \max_{\boldsymbol{x} \neq \boldsymbol{0}} \frac{\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}}{\sqrt{\boldsymbol{x}'\boldsymbol{A}^2\boldsymbol{x} \cdot \boldsymbol{x}'\boldsymbol{x}}} = 1$$

ja maksimi saavutetaan esimerkiksi matriisin $\boldsymbol{A}$ jollakin ominaisvektorilla. Tarkastellaan seuraavana maksimointitehtävää (3.5) vastaavaan minimointitehtävään:

$$\text{(3.6)} \qquad \min_{\boldsymbol{x} \neq \boldsymbol{0}} \frac{\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}}{\sqrt{\boldsymbol{x}'\boldsymbol{A}^2\boldsymbol{x} \cdot \boldsymbol{x}'\boldsymbol{x}}} = \mu_1.$$

Yhtälön (3.6) minimiarvoa μ_1 kutsutaan pienimmäksi antiominaisarvoksi ja Minimin tuottavaa vektoria $\boldsymbol{x}_1$ kutsutaan pienintä antiominaisarvoa vastaavaksi antiominaisvektoriksi.

Tehdään muunnos $\boldsymbol{y} = \boldsymbol{A}^{\frac{1}{2}}\boldsymbol{x}$, missä $\boldsymbol{A}^{\frac{1}{2}}$ on matriisin $\boldsymbol{A}$ symmetrinen ja positiivisesti definiitti neliöjuuri. Matriisin $\boldsymbol{A}$ neliöjuurella tarkoitetaan matriisia joka toteuttaa yhtälön $\boldsymbol{A} = \boldsymbol{A}^{\frac{1}{2}}\boldsymbol{A}^{\frac{1}{2}}$.

Sijoittamalla tämä yhtälöön (3.6) saadaan, että

$$\begin{aligned}
\mu_1 &= \min_{\boldsymbol{x}\neq\boldsymbol{0}} \frac{\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}}{\sqrt{\boldsymbol{x}'\boldsymbol{A}^2\boldsymbol{x}\cdot\boldsymbol{x}'\boldsymbol{x}}} \\
&= \min_{\boldsymbol{x}\neq\boldsymbol{0}} \frac{\boldsymbol{x}'\boldsymbol{A}^{\frac{1}{2}}\boldsymbol{A}^{\frac{1}{2}}\boldsymbol{x}}{\sqrt{\boldsymbol{x}'\boldsymbol{A}^{\frac{1}{2}}\boldsymbol{A}\boldsymbol{A}^{\frac{1}{2}}\boldsymbol{x}\cdot\boldsymbol{x}'\boldsymbol{A}^{\frac{1}{2}}\boldsymbol{A}^{-1}\boldsymbol{A}^{\frac{1}{2}}\boldsymbol{x}}} \\
&= \min_{\boldsymbol{y}\neq\boldsymbol{0}} \frac{\boldsymbol{y}'\boldsymbol{y}}{\sqrt{\boldsymbol{y}'\boldsymbol{A}\boldsymbol{y}\cdot\boldsymbol{y}'\boldsymbol{A}^{-1}\boldsymbol{y}}} \qquad (3.7) \\
&= \frac{2\sqrt{\lambda_1\lambda_p}}{\lambda_1+\lambda_p}, \qquad (3.8)
\end{aligned}$$

missä $\lambda_1 \geq \lambda_2 \geq \ldots \lambda_n > 0$ ovat matriisin $\boldsymbol{A}$ ominaisarvot. Yhtälöiden (3.7) ja (3.8) välinen yhtäsuuruus seuraa Fruchtin–Kantorovichin epäyhtälöstä (Lause L.3).

Määritelmä 3.1. *Olkoon $\boldsymbol{A}_{p\times p}$ symmetrinen ja positiivisesti definiitti matriisi, ja $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p > 0$ sen ominaisarvot. Matriisin $\boldsymbol{A}$ pienin antiominaisarvo määrittellään yhtälön*

$$\mu_1 = \min_{\boldsymbol{x}\neq\boldsymbol{0}} \frac{\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}}{\sqrt{\boldsymbol{x}'\boldsymbol{A}^2\boldsymbol{x}\cdot\boldsymbol{x}'\boldsymbol{x}}} \qquad (3.9)$$

ratkaisuna ja ratkaisun tuottavaa vektoria $\boldsymbol{x}_1$ kutsutaan pienintä antiominaisarvoa vastaavaksi antiominaisvektoriksi. Matriisin $\boldsymbol{A}$ yleinen antiominaisarvo määritellään yhtälöllä

$$\mu_i = \min_{\boldsymbol{x}_i\perp\boldsymbol{x}_j} \frac{\boldsymbol{x}_i'\boldsymbol{A}\boldsymbol{x}_i}{\sqrt{\boldsymbol{x}_i'\boldsymbol{A}^2\boldsymbol{x}_i\cdot\boldsymbol{x}_i'\boldsymbol{x}_i}}, \quad \textit{kun } j < i. \qquad (3.10)$$

Yhtälön (3.10) *ratkaisuja μ_i kutsutaan matriisin $\boldsymbol{A}$ i:nneksi antiominaisarvoiksi, ja ratkaisun tuottavaa vektoria $\boldsymbol{x}_i$ kutsutaan i:nnettä antiominaisarvoa μ_i vastaavaksi antiominaisvektoriksi.*

Edellä ollaan jo nähty, että pienin antiominaisarvo saadaan suoraan Fruchtin–Kantorovichin epäyhtälön raja-arvona. Vastaava tulos voidaan yleistää koskemaan myös muita antiominaisarvoja.

Lause 3.1. *Olkoon $\boldsymbol{A}_{p\times p}$ symmetrinen ja positiivisesti definiitti matriisi, jonka ominaisarvot ovat $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p > 0$. Tällöin matriisin $\boldsymbol{A}$ pienin antiominaisarvo on*

$$\mu_1 = \min_{\boldsymbol{x}\neq\boldsymbol{0}} \frac{\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}}{\sqrt{\boldsymbol{x}'\boldsymbol{A}^2\boldsymbol{x}\cdot\boldsymbol{x}'\boldsymbol{x}}} = \frac{2\sqrt{\lambda_1\lambda_p}}{\lambda_1+\lambda_p}, \qquad (3.11)$$

ja sitä vastaava antiominaisvektori $\boldsymbol{x}_1 = \frac{\sqrt{\lambda_p}}{\sqrt{\lambda_1+\lambda_p}}\boldsymbol{q}_1 \pm \frac{\sqrt{\lambda_1}}{\sqrt{\lambda_1+\lambda_p}}\boldsymbol{q}_p$, missä $\boldsymbol{q}_i$ on i:nnettä ominaisarvoa vastaava ominaisvektori.

Muut antiominaisarvot saadaan yhtälöstä:

$$\mu_i = \min_{\boldsymbol{x}_i \perp \boldsymbol{x}_j} \frac{\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}}{\sqrt{\boldsymbol{x}'\boldsymbol{A}^2\boldsymbol{x} \cdot \boldsymbol{x}'\boldsymbol{x}}} = \frac{2\sqrt{\lambda_i \lambda_{p-i+1}}}{\lambda_1 + \lambda_{p-1+1}}, \tag{3.12}$$

ja niitä vastaavat antiominaisvektorit $\boldsymbol{x}_i = \frac{\sqrt{\lambda_{p-i+1}}}{\sqrt{\lambda_i+\lambda_{p-i+1}}}\boldsymbol{q}_i \pm \frac{\sqrt{\lambda_i}}{\sqrt{\lambda_i+\lambda_{p-i+1}}}\boldsymbol{q}_{p-i+1}$, *missä* $\boldsymbol{q}_k$ *on k:nnetta ominaisarvoa vastaava ominaisvektori.*

Todistus. Todistuksen on esittänyt Khattree (2001), mutta se on keskeisiltä osin sama kuin Raon ja Raon (1987) esittämä todistus. Tarkastellaan funktiota

$$h(\boldsymbol{x}) = \frac{\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}}{\sqrt{\boldsymbol{x}'\boldsymbol{A}^2\boldsymbol{x} \cdot \boldsymbol{x}'\boldsymbol{x}}}. \tag{3.13}$$

Derivoidaan $h(\boldsymbol{x})$ vektorin $\boldsymbol{x}$ suhteen (Rao, 1973, s.72), jolloin saadaan välttämätön ehto

$$\frac{\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}'\boldsymbol{A}^2\boldsymbol{x}}\boldsymbol{A}^2\boldsymbol{x} + \frac{\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}'\boldsymbol{x}} \cdot \boldsymbol{x} = 2\boldsymbol{A}\boldsymbol{x}. \tag{3.14}$$

Ratkaisua koskevan välttämättömän ehdon (3.14) perusteella voidaan päätellä, että

$$\mathcal{C}[\boldsymbol{A}(\boldsymbol{x} : \boldsymbol{A}\boldsymbol{x})] = \mathcal{C}(\boldsymbol{A}\boldsymbol{x} : \boldsymbol{A}^2\boldsymbol{x}) \subset \mathcal{C}(\boldsymbol{x} : \boldsymbol{A}\boldsymbol{x}), \tag{3.15}$$

missä $\mathcal{C}(\boldsymbol{A})$:lla tarkoitetaan matriisin $\boldsymbol{A}$ sarakeavaruutta. Tällöin $\boldsymbol{x}$:n ja $\boldsymbol{A}\boldsymbol{x}$:n virittämä avaruus on suljettu $\boldsymbol{A}$:n suhteen. Tällöin maksimissaan kaksi $\boldsymbol{A}$:n ominaisvektoria riittää virittämään sarakeavaruuden $\mathcal{C}(\boldsymbol{x} : \boldsymbol{A}\boldsymbol{x})$. Tällöin mahdolliset ratkaisut saadaan siis yhden tai kahden $\boldsymbol{A}$:n ominaisvektorin lineaarikombinaationa (katso s. 41).

Koska matriisi $\boldsymbol{A}$ on symmetrinen, niin sillä on ominaisarvohajotelma

$$\boldsymbol{A} = \boldsymbol{T}\boldsymbol{\Lambda}\boldsymbol{T}', \tag{3.16}$$

missä $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_p)$, ja matriisi $\boldsymbol{T}$ muodostuu $\boldsymbol{A}$:n ominaisvektoreista. Kun tehdään sijoitus $\boldsymbol{y} = \boldsymbol{T}'\boldsymbol{x}$, saadaan minimoitava funktio $h(\boldsymbol{y})$ muotoon

$$h(\boldsymbol{y}) = \frac{\boldsymbol{y}'\boldsymbol{\Lambda}\boldsymbol{y}}{\sqrt{\boldsymbol{y}'\boldsymbol{\Lambda}^2\boldsymbol{y} \cdot \boldsymbol{y}'\boldsymbol{y}}}. \tag{3.17}$$

Tehdään sijoitukset $\boldsymbol{y} = \boldsymbol{T}'\boldsymbol{x}$, $a_1 = \frac{\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}'\boldsymbol{A}^2\boldsymbol{x}}$ ja $a_2 = \frac{\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}'\boldsymbol{x}}$ yhtälöön (3.14), jolloin

$$a_1\boldsymbol{A}^2\boldsymbol{T}\boldsymbol{y} + a_2\boldsymbol{T}\boldsymbol{y} = 2\boldsymbol{A}\boldsymbol{T}\boldsymbol{y}. \tag{3.18}$$

Ominaisarvohajotelman perusteella yhtälö (3.18) voidaan kirjoittaa muodossa

$$2\boldsymbol{T}\boldsymbol{\Lambda}\boldsymbol{y} = a_1\boldsymbol{\Lambda}^2\boldsymbol{T}\boldsymbol{y} + a_2\boldsymbol{T}\boldsymbol{y}. \tag{3.19}$$

Kertomalla yhtälö (3.19) vasemmalta matriisilla $\boldsymbol{T}'$, saadaan

$$2\boldsymbol{\Lambda}\boldsymbol{y} = a_1\boldsymbol{\Lambda}^2\boldsymbol{y} + a_2\boldsymbol{y}. \tag{3.20}$$

Kerrotaan yhtälö (3.14) vasemmalta vektorilla $\boldsymbol{y}'\boldsymbol{T}'$, jolloin saadaan

$$a_1\boldsymbol{y}'\boldsymbol{T}'\boldsymbol{A}^2\boldsymbol{T}\boldsymbol{y} = \boldsymbol{y}'\boldsymbol{T}'\boldsymbol{A}\boldsymbol{T}\boldsymbol{y}. \tag{3.21}$$

Ominaisarvohajotelman perusteella yhtälö (3.21) voidaan kirjoittaa muodossa

$$\boldsymbol{y}'\boldsymbol{\Lambda}\boldsymbol{y} = a_1\boldsymbol{y}'\boldsymbol{\Lambda}^2\boldsymbol{y}. \tag{3.22}$$

Stationaaritilassa yhtälöiden (3.20) ja (3.22) on pädettävä, jolloin ne voidaankin kirjoittaa yhtälöryhmänä

$$\begin{cases} 2\boldsymbol{\Lambda}\boldsymbol{y} = a_1\boldsymbol{\Lambda}^2\boldsymbol{y} + a_2\boldsymbol{y} \\ \boldsymbol{y}'\boldsymbol{\Lambda}\boldsymbol{y} = a_1\boldsymbol{y}'\boldsymbol{\Lambda}^2\boldsymbol{y} \end{cases}. \tag{3.23}$$

Yhtälöryhmän (3.23) ylempi matriisiyhtälö voidaan kirjoittaa p:n yhtälön yhtälöryhmänä, eli

$$2\lambda_i y_i = (a_1\lambda_i^2 + a_2)y_i, \qquad i = 1, \ldots, p. \tag{3.24}$$

Edellä on todettu, että yhtälön (3.14) ratkaisu $\boldsymbol{x}$ voidaan esittää yhden tai kahden matriisin $\boldsymbol{A}$ ominaisvektorin lineaarikombinaationa. Koska $\boldsymbol{y} = \boldsymbol{T}'\boldsymbol{x}$, missä $\boldsymbol{T}$ on ominaisvektoreista muodostunut matriisi, niin $\boldsymbol{y} = \boldsymbol{e}_i$ tai $\boldsymbol{y} = d_i\boldsymbol{e}_i + d_j\boldsymbol{e}_i$. Tällöin yhtälön (3.23) ratkaisut löytyvät kahdesta vaihtoehdosta:

Tapaus 1. Olkoon ehdokasvektori $\boldsymbol{y} = \boldsymbol{e}_i$ eli i:s kantavektori. Sijoitetaan tämä yhtälöryhmään (3.23), eli

$$\begin{aligned} 2\boldsymbol{\Lambda}\boldsymbol{e}_i &= a_1\boldsymbol{\Lambda}^2\boldsymbol{e}_i + a_2\boldsymbol{e}_i \\ \boldsymbol{e}_i'\boldsymbol{\Lambda}\boldsymbol{e}_i &= a_1\boldsymbol{e}_i'\boldsymbol{\Lambda}^2\boldsymbol{e}_i, \end{aligned}$$

josta voidaan ratkaista, että $a_1 = \lambda_i^{-1}$ ja $a_2 = \lambda_i$. Tässä stationaaritilassa kuitenkin funktio $h(\boldsymbol{x})$ saavuttaa maksiminsa 1, kuten jo aikaisemmin on Cauchyn–Schwarzin epäyhtälön perusteella todettu.

Tapaus 2. Olkoon ehdokasvektorit muotoa $\boldsymbol{y} = d_i\boldsymbol{e}_i + d_j\boldsymbol{e}_j$, missä $d_i, d_j \neq 0$. Sijoitetaan tämä yhtälöryhmään (3.24) ja oletetaan, että $\lambda_i \neq \lambda_j$. Tällöin yhtälöryhmän (3.23) ylempi yhtälö voidaan esittää muodossa:

$$2\lambda_i = a_1\lambda_i^2 + a_2 \tag{3.25}$$

$$2\lambda_j = a_1\lambda_j^2 + a_2, \tag{3.26}$$

joista a_1 ja a_2 voidaan ratkaista:

$$a_1 = \frac{2}{\lambda_i + \lambda_j} \tag{3.27}$$

ja

$$a_2 = \frac{2\lambda_i\lambda_j}{\lambda_i + \lambda_j}. \tag{3.28}$$

Muotoa $\boldsymbol{y} = d_1\boldsymbol{e}_i + d_2\boldsymbol{e}_j$ oleva ratkaisu löytyy vain jos toisen asteen yhtälöillä (3.25) ja (3.26) on olemassa ratkaisu. Yhtälöiden (3.25) ja (3.26) diskriminantti on

$$D = 4 - 4(-a_1)(-a_2)$$

ja on oltava ei-negatiivinen, jotta yhtälöillä (3.25) ja (3.26) olisi olemassa ratkaisu. Sijoitetaan yhtälöt (3.27) ja (3.28), jolloin diskriminantin perusteella saadaan ehto:

$$\frac{4\lambda_i\lambda_j}{(\lambda_i + \lambda_j)^2} \leq 1. \tag{3.29}$$

Kerrotaan ehto (3.29) puolittain $(\lambda_i + \lambda_j)^2$:lla ja kerrotaan auki sulkulauseke, jolloin ehto voidaan kirjoittaa muodossa

$$4\lambda_i\lambda_j \leq \lambda_i^2 + \lambda_j^2 + 2\lambda_i\lambda_j$$

tai

$$(\lambda_i + \lambda_j)^2 \geq 0. \tag{3.30}$$

Epäyhtälöstä (3.30) nähdään, että diskriminantti-ehto pätee aina ja tällöin ratkaisu on aina olemassa. Nyt siis yhtälöparin (3.23) jälkimmäisestä yhtälöstä seuraa, että

$$\left(\frac{d_i}{d_j}\right)^2 = \frac{\lambda_i}{\lambda_j}, \tag{3.31}$$

Koska $a_1 = \frac{\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}'\boldsymbol{A}^2\boldsymbol{x}}$ ja $a_2 = \frac{\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}'\boldsymbol{x}}$, niin voidaan helposti huomata, että

$$h(\boldsymbol{x}) = \sqrt{a_1 a_2} = \frac{2\sqrt{\lambda_i\lambda_j}}{\lambda_i + \lambda_j}.$$

Yhtälön (3.31) avulla nähdään, että $\boldsymbol{y} = \sqrt{\lambda_j}\boldsymbol{e}_i \pm \sqrt{\lambda_i}\boldsymbol{e}_j$, missä $\lambda_i \neq \lambda_j$ on ratkaisu yhtälöparille (3.23). Tällöin siis kaikki yhtälön (3.23) ratkaisut sisältyvät tapauksiin 1 ja 2, joista tapaus 1 tuottaa funktion $h(\boldsymbol{y})$ maksimin ja tapaus 2 minimin. Nyt koska yhtälöllä (3.14) ja $h(\boldsymbol{y})$ on samat ääriarvot, niin apulauseen (L.2) perusteella voidaan sanoa, että pienin antiominaisarvo μ_1 on:

$$\mu_1 = \frac{2\sqrt{\lambda_1 \lambda_p}}{\lambda_1 + \lambda_p},$$

jota vastaava antiominaisvektori on $\boldsymbol{x}_1 = \boldsymbol{T}\boldsymbol{y}_1 = \frac{\sqrt{\lambda_p}}{\sqrt{\lambda_1+\lambda_p}}\boldsymbol{q}_1 \pm \frac{\sqrt{\lambda_1}}{\sqrt{\lambda_1+\lambda_p}}\boldsymbol{q}_p$, missä $\boldsymbol{q}_i$ on i:nnettä ominaisarvoa vastaava ominaisvektori. Muut antiominaisarvot saadaan siis luonnollisesti yhtälöstä:

$$\mu_i = \frac{2\sqrt{\lambda_i \lambda_{p-i+1}}}{\lambda_i + \lambda_{p-i+1}},$$

ja sitä vastaava antiominaisvektori $\boldsymbol{x}_i = \boldsymbol{T}\boldsymbol{y}_i = \frac{\sqrt{\lambda_{p-i+1}}}{\sqrt{\lambda_i+\lambda_{p-i+1}}}\boldsymbol{q}_i \pm \frac{\sqrt{\lambda_i}}{\sqrt{\lambda_i+\lambda_{p-i+1}}}\boldsymbol{q}_{p-i+1}$, missä $\boldsymbol{q}_k$ on k:nnetta ominaisarvoa vastaava ominaisvektori.

□

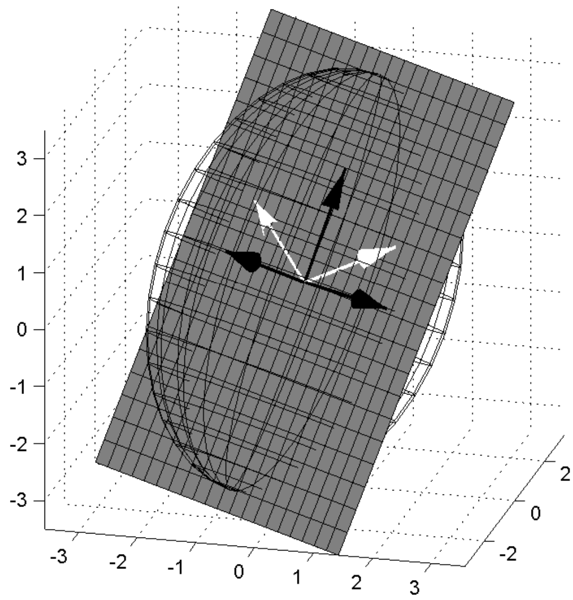
Edellä on jo todettu, että antiominaisvektorit ovat ortogonaalisia ja ne ilmoitetaan yleensä ykkösen mittaisina. Ortogonaalisuudesta seuraa, että matriisilla $\boldsymbol{A}$ on maksimissaan $[p/2]$ kappaletta antiominaisarvoja ja antiominaisvektoreita. Merkinnällä $[p/2]$ tarkoitetaan osamäärän $p/2$ kokonaislukuosaa, eli esimerkiksi $[5/2] = 2$.

Antiominaisarvojen ja vektoreiden lukumäärä on maksimissaan $[p/2]$, sillä matriisilla $\boldsymbol{A}$ saattaa olla useampikertaisia antiominaisarvoja, ja niihin liittyviä antiominaisvektoreita. Useampikertaisiin antiominaisarvoihin liittyvät antiominaisvektorit eivät kuitenkaan ole välttämättä ortogonaalisia, joten ne eivät ole yhtälön (3.10) ratkaisuja. Tämän vuoksi antiominaisarvoja ja antiominaisvektoreita on maksimissaan $[p/2]$ kappaletta.

Edellä olevasta todistuksesta huomattiin että antiominaisvektorit voidaan aina esittää kahden ominaisvektorin lineaarikombinaationa. Kuviossa 3.1 on esimerkin 2.1 kovarianssimatriisin määrittelemä hyperellipsoidi, ominaisvektorit, antiominaisvektorit ja ensimmäistä ja kolmatta ominaisarvoa vastaavien ominaisvektorien virittämä taso. Pienintä antiominaisarvoa vastaava antiominaisvektori ei ole yksikäsitteinen vaan $\boldsymbol{x}_1 = \frac{\sqrt{\lambda_3}}{\sqrt{\lambda_1+\lambda_3}}\boldsymbol{q}_1 \pm \frac{\sqrt{\lambda_1}}{\sqrt{\lambda_1+\lambda_3}}\boldsymbol{q}_3$, ja se virittää saman tason kuin ensimmäistä ja kolmatta ominaisarvoa vastaavat ominaisvektorit. On kuitenkin syytä huomata, että tämän tason muut vektorit eivät ole yhtälön (3.11) ratkaisuja.

3.3 Pienin antiominaisarvo ja tehokkuus

Ehkä tunnetuin sovellutus antiominaisarvolle on OLSE:n ja BLUE:n välinen tehokkuus. Tarkastellaan erikoistapausta, jossa mallimatriisi $\boldsymbol{x}$ on vektori (Puntanen 1999, s. 450).



Kuvio 3.1. Esimerkin 2.1 kovarianssimatriisin määrittelemä hyperellipsoidi, johon on mustalla lisätty ominaisvektorit sekä valkoisella pienintä antiominaisarvoa vastaavat antiominaisvektorit. Lisäksi pienimmän ja suurimman ominaisvektorin virittämä taso, jossa pienintä antiominaisarvoa vastaavat antiominaisvektorit sijaitsevat.

Olkoon malli $\{\boldsymbol{y},\, \boldsymbol{x}\beta,\, \sigma^2\boldsymbol{V}\}$, jossa mallimatriisi $\boldsymbol{x}$ on siis vektori ja tällöin β on reaaliluku. Tarkastellaan pienimmän neliösumman estimaattorin OLSE:n tehokkuutta parhaaseen lineaariseen estimaattoriin BLUE verrattuna. Merkitään pienimmän neliösumman estimaatoria $\hat{\beta}$:lla ja parasta lineaarista estimaattoria $\tilde{\beta}$:llä. Oletetaan tunnetuksi, että

$$\begin{aligned}
\tilde{\beta} &= (\boldsymbol{x}'\boldsymbol{V}^{-1}\boldsymbol{x})^{-1}\boldsymbol{x}'\boldsymbol{V}^{-1}\boldsymbol{y} = \frac{\boldsymbol{x}'\boldsymbol{V}^{-1}\boldsymbol{y}}{\boldsymbol{x}'\boldsymbol{V}^{-1}\boldsymbol{x}},\\
\hat{\beta} &= (\boldsymbol{x}'\boldsymbol{x})^{-1}\boldsymbol{x}'\boldsymbol{y} = \frac{\boldsymbol{x}'\boldsymbol{y}}{\boldsymbol{x}'\boldsymbol{x}}.
\end{aligned}$$

Tällöin

$$\begin{aligned}
\operatorname{var}(\tilde{\beta}) &= (\boldsymbol{x}'\boldsymbol{V}^{-1}\boldsymbol{x})^{-1}\boldsymbol{x}'\boldsymbol{V}^{-1}\operatorname{cov}(\boldsymbol{y})\boldsymbol{V}^{-1}\boldsymbol{x}(\boldsymbol{x}'\boldsymbol{V}^{-1}\boldsymbol{x})^{-1}\\
&= (\boldsymbol{x}'\boldsymbol{V}^{-1}\boldsymbol{x})^{-1}\boldsymbol{x}'\boldsymbol{V}^{-1}\boldsymbol{V}\boldsymbol{V}^{-1}\boldsymbol{x}(\boldsymbol{x}'\boldsymbol{V}^{-1}\boldsymbol{x})^{-1}\\
&= (\boldsymbol{x}'\boldsymbol{V}^{-1}\boldsymbol{x})^{-1},\\
\operatorname{var}(\hat{\beta}) &= (\boldsymbol{x}'\boldsymbol{x})^{-1}\boldsymbol{x}'\operatorname{cov}(\boldsymbol{y})\boldsymbol{x}(\boldsymbol{x}'\boldsymbol{x})^{-1}\\
&= (\boldsymbol{x}'\boldsymbol{x})^{-1}\boldsymbol{x}'\boldsymbol{V}\boldsymbol{x}(\boldsymbol{x}'\boldsymbol{x})^{-1} = \frac{\boldsymbol{x}'\boldsymbol{V}\boldsymbol{x}}{(\boldsymbol{x}'\boldsymbol{x})^2}.
\end{aligned}$$

Määritellään tehokkuus ϕ BLUE:n ja OLSE:n varianssien suhteeksi (Watsonin tehokkuus)

$$\phi = \frac{\operatorname{var}(\tilde{\beta})}{\operatorname{var}(\hat{\beta})} = \frac{(\boldsymbol{x}'\boldsymbol{V}^{-1}\boldsymbol{x})^{-1}}{\frac{\boldsymbol{x}'\boldsymbol{V}\boldsymbol{x}}{(\boldsymbol{x}'\boldsymbol{x})^2}} = \frac{(\boldsymbol{x}'\boldsymbol{x})^2}{\boldsymbol{x}'\boldsymbol{V}\boldsymbol{x}\cdot\boldsymbol{x}'\boldsymbol{V}^{-1}\boldsymbol{x}}. \tag{3.32}$$

Nyt siis nähdään suoraan, että tehokkuus saavuttaa maksiminsa 1 aina kun $\boldsymbol{V}\boldsymbol{x} = \lambda\boldsymbol{x}, \quad \lambda \in \mathbb{R}$, eli kun $\boldsymbol{x}$ on jonkin $\boldsymbol{V}$:n ominaisvektorin suuntainen. Maksimi ei kuitenkaan ole kovin mielenkiintoinen, sillä satunnaisvektorin $\boldsymbol{x}$ valinnassa voi aina käydä hyvä tuuri. Mielenkiintoisempi kuitenkin on tehokkuuden minimi, jota ei voida huonoimmallakaan tuurilla alittaa. Tehokkuuden minimi on pienimmän antiominaisarvon neliö, joka nähdään helposti yhtälöistä (3.7) ja (3.8). Samalla nähdään, että yhtäsuuruus saavutetaan, kun $\boldsymbol{x} = \boldsymbol{V}^{\frac{1}{2}}\boldsymbol{x}_1 = \frac{1}{\sqrt{2}}\boldsymbol{q}_1 \pm \frac{1}{\sqrt{2}}\boldsymbol{q}_n$, missä $\boldsymbol{x}_1$ on kovarianssimatriisin $\boldsymbol{V}$ pienintä antiominaisarvoa vastaava antiominaisvektori ja $\boldsymbol{q}_i$ on kovarianssimatriisin $\boldsymbol{V}$ i:nnettä ominaisarvoa vastaava ominaisvektori.

Esimerkki 3.1. Tarkastellaan tilannetta jossa virhetermi noudattaa astetta 1 olevaa autoregressiivistä prosessia

$$\epsilon_i = a\epsilon_{i-1} + u_i, \qquad i = \ldots, -2, -1, 0, 1, 2, \ldots, \quad |a| < 1,$$

missä u_i:t ovat riippumattomia satunnaismuuttujia joiden odotusarvo $E(u_i) = 0$ ja varianssi $\operatorname{var}(u_i) = \sigma_u^2$. Tästä prosessista käytetään yleisesti lyhennettyä merkintää AR(1). Tällöin voidaan näyttää, että

$$\mathrm{cov}(\boldsymbol{\epsilon}) = \mathrm{cov}(\boldsymbol{y}) = \sigma^2 \boldsymbol{V} = \frac{\sigma_u^2}{1-a^2} \begin{pmatrix} 1 & a^2 & a^3 & \cdots & a^{n-1} \\ a^2 & 1 & a^2 & \cdots & a^{n-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a^{n-1} & a^{n-2} & a^{n-3} & \cdots & 1 \end{pmatrix},$$

missä n on havaintojen lukumäärä. OLSE:n ja BLUE:n välisen Watsonin tehokkuuden minimi ϕ_{min} voidaan siis yhtälön (3.32) perusteella kirjoittaa muodossa

$$\phi_{min} = \frac{(\boldsymbol{x}'\boldsymbol{x})^2}{\boldsymbol{x}'\boldsymbol{V}\boldsymbol{x} \cdot \boldsymbol{x}'\boldsymbol{V}^{-1}\boldsymbol{x}} = \frac{4\lambda_1\lambda_n}{(\lambda_1+\lambda_n)^2} = \mu_1^2, \tag{3.33}$$

missä λ_i on matriisin $\boldsymbol{V}$ i. ominaisarvo. Kuten yhtälöstä (3.33) nähdään, termi $\frac{\sigma_u^2}{1-a^2}$ supistuu nimittäjässä ja voidaan siksi jättää huomiotta.

Piirretään kuvio 3.2, jossa tehokkuuden minimiä tarkastellaan parametrin a suhteen. AR(1)-mallissa parametri a on itse asiassa autokorrelaatio viiveellä 1. Tällöin kuviosta 3.2 nähdään, että pienillä autokorrelaatioilla, eli parametrin a arvoilla OLSE on vielä pahimmillaankin suhteellisen tehokas AR(1)-kovarianssirakenteen tapauksessa. Tehokkuuden minimi kuitenkin pienenee mitä suuremman arvon parametri a saa. Suurilla havaintomäärillä kuvaaja tulee kuitenkin huipukkaammaksi, eli parametrin a arvo vaikuttaa voimakkaammin tehokkuuden alarajaan. Yksityiskohtaisempaa tietoa AR(1)-mallin vaikutuksesta regressioanalyysiin löytää esimerkiksi Puntasen (1999) kirjasta.

Esimerkki 3.2. Kuten edellisessä esimerkissä (3.1) tarkasteltiin OLSE:n tehokkuuden minimiä AR(1)-mallin tapauksessa, tarkastellaan nyt liukuvan keskiarvon mallia viiveellä 1, eli MA(1)-mallia. Tällöin oletetaan, että:

$$\epsilon_i = bu_{i-1} + u_i \qquad i = \ldots, -2, -1, 0, 1, 2, \ldots, \qquad |b| < 1.$$

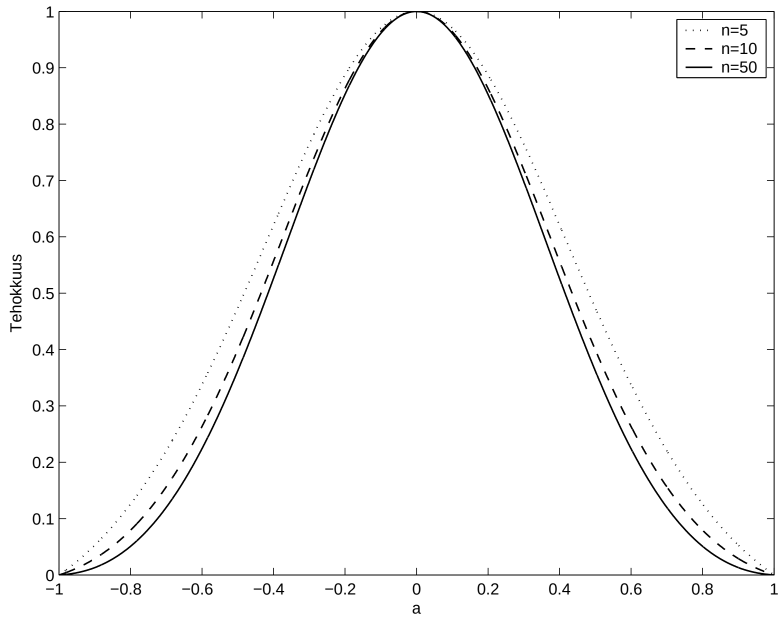
missä u_i:t ovat riippumattomia satunnaismuuttujia joiden odotusarvo $E(u_i) = 0$ ja varianssi $\mathrm{var}(u_i) = \sigma_u^2$.

Kuten edellä, voidaan myös nyt helposti osoittaa että

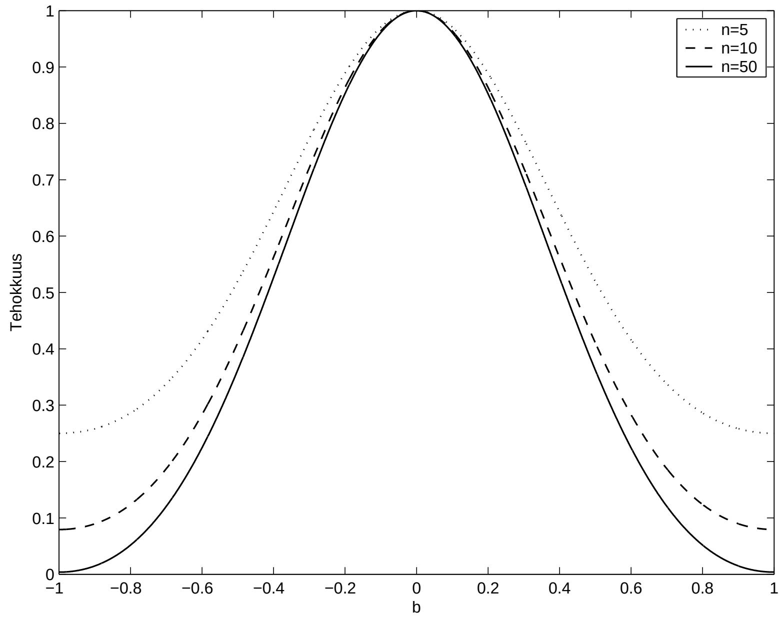
$$\mathrm{cov}(\boldsymbol{y}) = \sigma^2 \boldsymbol{V} = \sigma^2 \begin{pmatrix} 1+b^2 & b & 0 & \cdots & 0 \\ b & 1+b^2 & b & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1+b^2 \end{pmatrix},$$

josta σ^2 voidaan taas jättää huomiotta.

Kuviosta 3.3 nähdään että parametrin b vaikutus tehokkuuden minimiin MA(1)-kovarianssirakenteen tapauksessa on hyvin samanlaista kuin parametrin a vaikutus AR(1)-kovarianssirakenteen tapauksessa. Pienillä havaintomäärillä kuitenkin MA(1)-kovarianssirakenteen tapauksessa tehokkuuden minimi



Kuvio 3.2. Kuvion pystyakselilla on OLSE:n tehokkuuden minimi ϕ_{min} AR(1)-kovarianssirakenteen tapauksessa, kun vaaka-akselilla on parametri a.



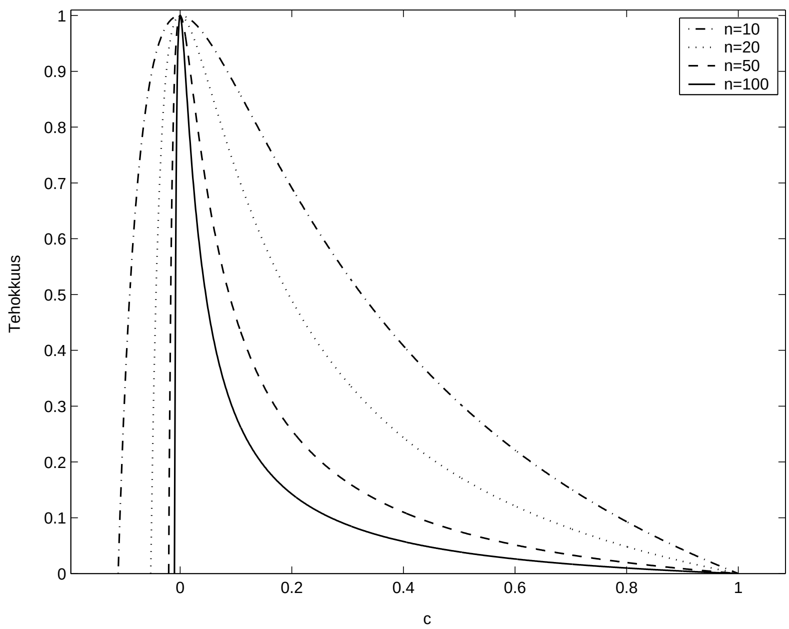
Kuvio 3.3. Kuvion pystyakselilla on OLSE:n tehokkuuden minimi ϕ_{min} MA(1)-kovarianssirakenteen tapauksessa, kun vaaka-akselilla on parametri b.

ei mene nollaan parametrin b lähestyessä ykköstä, kuin vastaavassa AR(1)-kovarianssirakenteen tapauksessa.

Esimerkki 3.3. Tarkastellaan vielä kolmantena esimerkkinä OLSE:n tehokkuuden minimiä tasakorrelaation tapauksessa, jolloin

$$\operatorname{cov}(\boldsymbol{y}) = \sigma^2 \boldsymbol{V} = \sigma^2 \begin{pmatrix} 1 & c & c & \cdots & c \\ c & 1 & c & \cdots & c \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c & c & c & \cdots & 1 \end{pmatrix}, \qquad -\frac{1}{n-1} < c < 1.$$

Kuten edellä, niin myös nytkin σ^2 supistuu tehokkuustarkastelussa, ja voidaan tarkastella ainoastaan matriisin $\boldsymbol{V}$ pienintä antiominaisarvoa. Piirretään kuviota 3.2 vastaava kuvio, jossa OLSE:n tehokkuuden minimi on pystyakselilla ja vaaka-akselilla on parametri c.



Kuvio 3.4. Kuvion pystyakselilla on OLSE:n tehokkuuden minimi ϕ_{min} tasakorrelaation tapauksessa, kun vaaka-akselilla on parametri c.

Kuvioista 3.2, 3.4 ja 3.3 nähdään, että AR(1)- ja MA(1)-kovarianssirakenne vaikuttaa tehokkuuden minimiin hyvin samalla tavalla. Kummassakin tapauksessa parametrin ollessa lähellä nollaa, tehokkuuden minimi on lähellä ykköstä. Parametrien vaikutus tehokkuuden alarajaan on sekä AR(1)- että MA(1)-kovarianssirakenteiden tapauksessa symmetristä nollan suhteen. AR(1)- kovarianssirakenteen tapauksessa (kuvio 3.2) pienilläkin otoskoilla a:n ollessa lähellä ykköstä, tehokkuuden alaraja lähenee nollaa. Näin ei kuitenkaan tapahdu MA(1)-kovarianssirakenteen tapauksessa (Kuvio 3.3).

Tasakorrelaation tapauksessa (Kuvio 3.4) parametrin c vaikutus tehokkuuden alarajaan ei ole symmetristä nollan suhteen. Negatiiviset parametrin c arvot vaikuttavat huomattavasti voimakkaammin tehokkuuden alarajaan kuin positiiviset c:n arvot. Lisäksi parametrin c vaikutus tehokkuuden alarajaan riippuu voimakkaasti havaintojen lukumäärästä, mikä näkyy kuviossa 3.4 kuvaajan huipukkuutena suurilla n:n arvoilla.

Huomautus. Voidaan osoittaa, että mallissa $\{\boldsymbol{y}, \boldsymbol{X\beta}, \sigma^2\boldsymbol{V}\}$, missä kovarianssimatriisilla $\boldsymbol{V}$ on tasakovarianssirakenne, OLSE on BLUE (esim. Puntanen 1999 s. 168). Esimerkissä 3.3 mallimatriisi $\boldsymbol{x}$ on vektori, eli mallimatriisista puuttuu vakiotermiä vastaava vektori $\mathbf{1} = (1, 1, \ldots, 1)'$. Tämän vuoksi OLSE:n tehokkuus menee alle ykkösen esimerkin 3.3 mallissa.

4 Antiominaisarvot ja korrelaatio

Gustafson (1972) tarkasteli matriisin $\boldsymbol{A}$ kosinia $\cos \boldsymbol{A}$, joka reaalisen symmetrisen ja positiivisesti definiitin matriisin $\boldsymbol{A}_{p \times p}$ tapauksessa on

$$\cos \boldsymbol{A} = \min_{\boldsymbol{x} \neq \boldsymbol{0}} \frac{\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}}{\sqrt{\boldsymbol{x}'\boldsymbol{A}^2\boldsymbol{x} \cdot \boldsymbol{x}'\boldsymbol{x}}} = \mu_1, \tag{4.1}$$

missä $\boldsymbol{x}$ on reaalinen vektori.

Matriisin $\boldsymbol{A}$ kosini on siis vektoreiden $\boldsymbol{x}$ ja $\boldsymbol{A}\boldsymbol{x}$ välisen kulman maksimi. Tilastotieteessä tällä kulman kosinilla on tulkinta korrelaatiokertoimena.

Antiominaisarvojen etsiminen voidaan tulkita myös minimaalisen korrelaation etsimiseksi lineaarikombinaatioiden $\boldsymbol{z}'\boldsymbol{x}$ ja $\boldsymbol{z}'\boldsymbol{A}\boldsymbol{x}$ välillä

$$\min_{\boldsymbol{x} \neq \boldsymbol{0}} \operatorname{cor}(\boldsymbol{z}'\boldsymbol{x}, \boldsymbol{z}'\boldsymbol{A}\boldsymbol{x}) = \min_{\boldsymbol{x} \neq \boldsymbol{0}} \frac{\operatorname{cov}(\boldsymbol{z}'\boldsymbol{x}, \boldsymbol{z}'\boldsymbol{A}\boldsymbol{x})}{\sqrt{\operatorname{var}(\boldsymbol{z}'\boldsymbol{A}\boldsymbol{x}) \cdot \operatorname{var}(\boldsymbol{z}'\boldsymbol{x})}}, \tag{4.2}$$

missä vektori $\boldsymbol{z}$ on satunnaisvektori jonka kovarianssimatriisi on identiteettimatriisi. Yhtälö (4.2) saadaan muotoon

$$\begin{aligned}
\min_{\boldsymbol{x} \neq \boldsymbol{0}} \operatorname{cor}(\boldsymbol{z}'\boldsymbol{x}, \boldsymbol{z}'\boldsymbol{A}\boldsymbol{x}) &= \min_{\boldsymbol{x} \neq \boldsymbol{0}} \frac{\operatorname{cov}(\boldsymbol{z}'\boldsymbol{x}, \boldsymbol{z}'\boldsymbol{A}\boldsymbol{x})}{\sqrt{\operatorname{var}(\boldsymbol{z}'\boldsymbol{A}\boldsymbol{x}) \cdot \operatorname{var}(\boldsymbol{z}'\boldsymbol{x})}} \\
&= \min_{\boldsymbol{x} \neq \boldsymbol{0}} \frac{\operatorname{cov}(\boldsymbol{x}'\boldsymbol{z}, \boldsymbol{x}'\boldsymbol{A}\boldsymbol{z})}{\sqrt{\operatorname{var}(\boldsymbol{x}'\boldsymbol{A}\boldsymbol{z}) \cdot \operatorname{var}(\boldsymbol{x}'\boldsymbol{z})}} \\
&= \min_{\boldsymbol{x} \neq \boldsymbol{0}} \frac{\boldsymbol{x}' \operatorname{cov}(\boldsymbol{z})\boldsymbol{A}\boldsymbol{x}}{\sqrt{\boldsymbol{x}'\boldsymbol{A} \operatorname{cov}(\boldsymbol{z})\boldsymbol{A}\boldsymbol{x} \cdot \boldsymbol{x}' \operatorname{cov}(\boldsymbol{z})\boldsymbol{x}}}. \qquad (4.3)
\end{aligned}$$

Koska $\boldsymbol{z}$:n kovarianssimatriisi on identiteettimatriisi $\operatorname{cov}(\boldsymbol{z}) = \boldsymbol{I}_p$, yhtälöstä (4.3) seuraa, että

$$\min_{\boldsymbol{x} \neq \boldsymbol{0}} \operatorname{cor}(\boldsymbol{z}'\boldsymbol{x}, \boldsymbol{z}'\boldsymbol{A}\boldsymbol{x}) = \min_{\boldsymbol{x} \neq \boldsymbol{0}} \frac{\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}}{\sqrt{\boldsymbol{x}'\boldsymbol{A}^2\boldsymbol{x} \cdot \boldsymbol{x}'\boldsymbol{x}}} = \mu_1, \tag{4.4}$$

mikä on matriisin $\boldsymbol{A}$ pienin antiominaisarvo.

Esimerkki 4.1. Antiominaisarvon tulkintaa korrelaatiokertoimena kahden satunnaismuuttujan välillä voidaan soveltaa mielekkäästi isien ja poikien välisen korrelaation tarkastelussa (Rao ja Rao 1987). Olkoon vektoreissa $\boldsymbol{f}$ ja $\boldsymbol{c}$ p samoin kerättyä mittaustulosta isästä ja pojasta. Olkoon mittausten välinen kovarianssimatriisi

$$\text{cov}\begin{pmatrix}\boldsymbol{f}\\ \boldsymbol{c}\end{pmatrix} = \begin{pmatrix}\boldsymbol{I} & \boldsymbol{\Sigma}\\ \boldsymbol{\Sigma} & \boldsymbol{\Sigma}^2\end{pmatrix}, \tag{4.5}$$

missä $\boldsymbol{\Sigma}$:n oletetaan olevan symmetrinen ja positiivisesti definiitti matriisi. Tarkastellaan nyt kahden lineaarisen funktion $\boldsymbol{y}'\boldsymbol{f}$ ja $\boldsymbol{y}'\boldsymbol{c}$ välistä korrelaatiota, jossa $\boldsymbol{y}_{p\times 1}$ on reaalinen kerroinvektori. Tällöin korrelaatio voidaan yhtälön (4.4) perusteella kirjoittaa muodossa

$$\begin{aligned}\text{cor}(\boldsymbol{y}'\boldsymbol{f},\boldsymbol{y}'\boldsymbol{c}) &= \frac{\text{cov}(\boldsymbol{y}'\boldsymbol{f},\boldsymbol{y}'\boldsymbol{c})}{\sqrt{\text{cov}(\boldsymbol{y}'\boldsymbol{c})\cdot\text{cov}(\boldsymbol{y}'\boldsymbol{f})}} = \frac{\boldsymbol{y}'\,\text{cov}(\boldsymbol{f},\boldsymbol{c})\boldsymbol{y}}{\sqrt{\boldsymbol{y}'\,\text{cov}(\boldsymbol{c})\boldsymbol{y}\cdot\boldsymbol{y}'\,\text{cov}(\boldsymbol{f})\boldsymbol{y}}}\\ &= \frac{\boldsymbol{y}'\boldsymbol{\Sigma}\boldsymbol{y}}{\sqrt{\boldsymbol{y}'\boldsymbol{\Sigma}^2\boldsymbol{y}\cdot\boldsymbol{y}'\boldsymbol{y}}}.\end{aligned} \tag{4.6}$$

Kuten edellä todettiin, yhtälön (4.6) maksimi saavutetaan valitsemalla kerroinvektoriksi $\boldsymbol{y}$ jokin $\boldsymbol{\Sigma}$:n ominaisvektori. Käyttämällä kerroinvektorina jotain matriisin $\boldsymbol{\Sigma}$ ominaisvektoria isät ja pojat ovat korrelaation mielessä mahdollisimman samanlaisia.

Miniminsä yhtälö (4.6) saavuttaa kun kerroinvektoriksi $\boldsymbol{y}$ valitaan $\boldsymbol{\Sigma}$:n pienintä antiominaisarvoa μ_1 vastaava antiominaisvektori. Tällöin minimiarvoksi tulee matriisin $\boldsymbol{\Sigma}$ pienin antiominaisarvo. Jos siis kerroinvektorina $\boldsymbol{y}$ käytetään pienintä antiominaisarvoa vastaavaa antiominaisvektoria, isät ja pojat muistuttavat mahdollisimman vähän toisiansa. Voidaan siis sanoa, että isät ja pojat muistuttavat toisiaan vähintään sen verran, että korrelaatioksi jää pienin antiominaisarvo.

Edellisen esimerkin tulosta voidaan yleistää (Rao ja Rao 1987) koskemaan kovarianssimatriisia

$$\text{cov}\begin{pmatrix}\boldsymbol{f}\\ \boldsymbol{c}\end{pmatrix} = \begin{pmatrix}\boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12}\\ \boldsymbol{\Sigma}_{12} & \boldsymbol{\Sigma}_{22}\end{pmatrix}, \tag{4.7}$$

missä matriisit $\boldsymbol{\Sigma}_{11}$ ja $\boldsymbol{\Sigma}_{22}$ ovat positiivisti definiittejä sekä $\boldsymbol{\Sigma}_{12}$ symmetrinen. Tällöin päädytään samaan lopputulokseen kuin edellisessä yksinkertaistetussa esimerkissä.

5 Antiominaisarvot ja kanoninen korrelaatio

5.1 Kanoniset korrelaatiot

Tarkastellaan vektoria $\boldsymbol{x}_{(p+q)\times 1}$ ja olkoon $E(\boldsymbol{x}) = \mathbf{0}$. Ositetaan vektori $\boldsymbol{x}$ kahteen osavektoriin $\boldsymbol{x}^{(1)}_{q\times 1}$ ja $\boldsymbol{x}^{(2)}_{p\times 1}$ siten, että $q \leq p$. Vektorin $\boldsymbol{x}$:n kovarianssimatriisin oletetaan olevan

$$\text{cov}(\boldsymbol{x}) = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}. \tag{5.1}$$

Muodostetaan uudet muuttujat $w = \boldsymbol{\alpha}'\boldsymbol{x}^{(1)}$ ja $v = \boldsymbol{\gamma}'\boldsymbol{x}^{(2)}$, jossa vektorit $\boldsymbol{\alpha}$ ja $\boldsymbol{\gamma}$ on valittu siten että $\text{var}(w) = \text{var}(v) = 1$. Täsmällisesti tämä siis tarkoittaa, että

$$\text{var}(w) = E(w^2) = \text{var}(\boldsymbol{\alpha}'\boldsymbol{x}^{(1)}) = \boldsymbol{\alpha}'\boldsymbol{\Sigma}_{11}\boldsymbol{\alpha} = 1, \tag{5.2}$$

$$\text{var}(v) = E(v^2) = \text{var}(\boldsymbol{\gamma}'\boldsymbol{x}^{(2)}) = \boldsymbol{\gamma}'\boldsymbol{\Sigma}_{22}\boldsymbol{\gamma} = 1, \tag{5.3}$$

ja tällöin w:n ja v:n välinen kovarianssi, samoin kuin korrelaatio on

$$\text{cov}(w, v) = E(wv) = \boldsymbol{\alpha}' \,\text{cov}(\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)})\boldsymbol{\gamma} = \boldsymbol{\alpha}'\boldsymbol{\Sigma}_{12}\boldsymbol{\gamma}. \tag{5.4}$$

Kanonisen korrelaatioanalyysin ongelma on valita kerroinvektorit $\boldsymbol{\alpha}$ ja $\boldsymbol{\gamma}$ siten että w:n ja v:n välinen korrelaatio saavuttaa maksiminsa. Eli maksimoidaan kaavan (5.4) korrelaatiota (kovarianssia) siten että ehdot (5.2) ja (5.3) ovat voimassa. Maksimointiongelma voidaan ratkaista käyttämällä Lagrangen kertojamenettelyä, jolloin Lagrangen funktioksi tulee

$$\psi = \boldsymbol{\alpha}'\boldsymbol{\Sigma}_{12}\boldsymbol{\gamma} - \frac{1}{2}\lambda(\boldsymbol{\alpha}'\boldsymbol{\Sigma}_{11}\boldsymbol{\alpha} - 1) - \frac{1}{2}\mu(\boldsymbol{\gamma}'\boldsymbol{\Sigma}_{22}\boldsymbol{\gamma} - 1), \tag{5.5}$$

missä λ ja μ ovat Lagrangen kertojia. Derivoidaan funktio ψ muuttujavektoreiden $\boldsymbol{\alpha}$ ja $\boldsymbol{\gamma}$ suhteen, sekä asetetaan osittaisderivaatat nolliksi:

$$\frac{\partial\psi}{\partial\boldsymbol{\alpha}} = \boldsymbol{\Sigma}_{12}\boldsymbol{\gamma} - \lambda\boldsymbol{\Sigma}_{11}\boldsymbol{\alpha} = \mathbf{0}, \tag{5.6}$$

$$\frac{\partial\psi}{\partial\boldsymbol{\gamma}} = \boldsymbol{\Sigma}'_{12}\boldsymbol{\alpha} - \mu\boldsymbol{\Sigma}_{22}\boldsymbol{\gamma} = \mathbf{0}. \tag{5.7}$$

Kertomalla yhtälö (5.6) vasemmalta $\boldsymbol{\alpha}'$:lla ja (5.7) $\boldsymbol{\gamma}'$:lla saadaan

$$\boldsymbol{\alpha}'\boldsymbol{\Sigma}_{12}\boldsymbol{\gamma} - \lambda\boldsymbol{\alpha}'\boldsymbol{\Sigma}_{11}\boldsymbol{\alpha} = 0, \tag{5.8}$$

$$\boldsymbol{\gamma}'\boldsymbol{\Sigma}_{12}'\boldsymbol{\alpha} - \mu\boldsymbol{\gamma}'\boldsymbol{\Sigma}_{22}\boldsymbol{\gamma} = 0. \tag{5.9}$$

Koska $\boldsymbol{\alpha}'\boldsymbol{\Sigma}_{11}\boldsymbol{\alpha} = 1$ ja $\boldsymbol{\gamma}'\boldsymbol{\Sigma}_{22}\boldsymbol{\gamma} = 1$, niin nähdään että $\lambda = \mu = \boldsymbol{\alpha}'\boldsymbol{\Sigma}_{12}\boldsymbol{\gamma}$. Tällöin yhtälöt (5.6) ja (5.7) voidaan kirjoittaa muodossa

$$-\lambda\boldsymbol{\Sigma}_{11}\boldsymbol{\alpha} + \boldsymbol{\Sigma}_{12}\boldsymbol{\gamma} = \mathbf{0}, \tag{5.10}$$

$$\boldsymbol{\Sigma}_{21}\boldsymbol{\alpha} - \lambda\boldsymbol{\Sigma}_{22}\boldsymbol{\gamma} = \mathbf{0}. \tag{5.11}$$

Yhtälöryhmä voidaan esittää myös matriisimuodossa

$$\begin{pmatrix} -\lambda\boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & -\lambda\boldsymbol{\Sigma}_{22} \end{pmatrix} \begin{pmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\gamma} \end{pmatrix} = \mathbf{0}, \tag{5.12}$$

joka taas voidaan esittää matriisin $\boldsymbol{\Sigma}$ ominaisarvoyhtälön muodossa

$$\left[\begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix} - \lambda\boldsymbol{I}\right] \begin{pmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\gamma} \end{pmatrix} = \mathbf{0}. \tag{5.13}$$

Tällöin maksimiarvon etsiminen palautuu matriisin $\boldsymbol{\Sigma}$ suurimman ominaisarvon λ_1 etsimiseksi. Kaikki ominaisarvot ovat ei-negatiivisia sekä reaalisia, sillä kovarianssimatriisi $\boldsymbol{\Sigma}$ on ei-negatiivisesti definiitti sekä symmetrinen. Triviaaliratkaisut $\boldsymbol{\alpha} = \mathbf{0}$ ja $\boldsymbol{\gamma} = \mathbf{0}$ eivät käy ratkaisuiksi ehtojen (5.2) ja (5.3) perusteella.

Kaavasta (5.8) nähdään että $\lambda = \boldsymbol{\alpha}'\boldsymbol{\Sigma}_{12}\boldsymbol{\gamma}$ on w:n ja v:n välinen korrelaatio stationaaritilassa. Jos halutaan valita maksimikorrelaatio, niin tulee valita että $\lambda = \lambda_1$, mikä on suurin ominaisarvo, jota vastaava ominaisvektori on $\begin{pmatrix} \boldsymbol{\alpha}^{(1)} \\ \boldsymbol{\gamma}^{(1)} \end{pmatrix}$. Tällöin w_1 ja v_1 ovat $\boldsymbol{x}^{(1)}$:n ja $\boldsymbol{x}^{(2)}$:n lineaarikombinaatiot, joilla on maksimikorrelaatio.

Toinen kanoninen korrelaatio voidaan etsiä kuten edellä, mutta lisätään ehtojen (5.2) ja (5.3) lisäksi rajoitus, että ensimmäisen ja toisen kanonisen muuttujan välillä ei ole korrelaatiota. Tällöin toiseksi suurin kanoninen korrelaatio on $\rho_2 = \lambda_2$, eli kovarianssimatriisin $\boldsymbol{\Sigma}$ toiseksi suurin ominaisarvo. Täsmällinen todistus esitetään esimerkiksi kirjassa (Anderson 2003, s. 488).

Kanoninen korrelaatio voidaan kirjoittaa myös ainoastaan joko kerroinvektorin $\boldsymbol{\alpha}$:n tai $\boldsymbol{\gamma}$:n suhteen. Kertomalla yhtälö (5.10) λ:lla ja (5.11) vasemmalta $\boldsymbol{\Sigma}_{22}^{-1}$:lla, saadaan että

$$\lambda\boldsymbol{\Sigma}_{12}\boldsymbol{\gamma} = \lambda^2\boldsymbol{\Sigma}_{11}\boldsymbol{\alpha}, \tag{5.14}$$

$$\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\boldsymbol{\alpha} = \lambda\boldsymbol{\gamma}. \tag{5.15}$$

Sijoitetaan yhtälö (5.15) yhtälöön (5.14), jolloin

$$\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21}\boldsymbol{\alpha} = \lambda^2\mathbf{\Sigma}_{11}\boldsymbol{\alpha}. \tag{5.16}$$

Järjestelemällä yhtälön (5.16) termit uudestaan, saadaan

$$(\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21} - \lambda^2\mathbf{\Sigma}_{11})\boldsymbol{\alpha} = \mathbf{0}. \tag{5.17}$$

Kertomalla yhtälö (5.17) vasemmalta $\mathbf{\Sigma}_{11}^{-1}$:llä, päädytään tarkastelemaan matriisin

$$\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21}$$

ominaisarvoja. Ominaisarvot toteuttavat siis yhtälön:

$$(\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21} - \lambda^2\boldsymbol{I})\boldsymbol{\alpha} = \mathbf{0}.$$

Vastaava yhtälö kerroinvektorille $\boldsymbol{\gamma}$ saadaan kertomalla yhtälö (5.11) λ:lla ja (5.10) vasemmalta $\mathbf{\Sigma}_{11}^{-1}$:lla:

$$(\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12} - \lambda^2\boldsymbol{I})\boldsymbol{\gamma} = \mathbf{0}.$$

Matriisit $\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}$ ja $\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21}$ ovat kokoa $p \times q$ ja $q \times p$. Tällöin matriisitulon nollasta poikkeavia ominaisarvoja koskevan lauseen (Lause L.4) nojalla tiedetään, että

$$\left\{\mathrm{nzch}(\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21(p\times p)})\right\} = \left\{\mathrm{nzch}(\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12(q\times q)})\right\},$$

eli matriiseilla ovat samat nollasta eroavat ominaisarvot. Koska $p < q$ niin kyseisillä matriiseilla on p kappaletta nollasta eroavia ominaisarvoja $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p > 0$.

Koska kovarianssimatriisi $\mathbf{\Sigma}_{11}$ oletetaan positiivisesti definiitiksi, niin on olemassa symmetrinen ja positiivisesti definiitti matriisi $\mathbf{\Sigma}_{11}^{-\frac{1}{2}}$ siten, että $\mathbf{\Sigma}_{11}^{-\frac{1}{2}}\mathbf{\Sigma}_{11}^{-\frac{1}{2}} = \mathbf{\Sigma}_{11}^{-1}$. Tällöin edellä mainitun lauseen L.4 nojalla

$$\left\{\mathrm{nzch}(\mathbf{\Sigma}_{11}^{-\frac{1}{2}}\mathbf{\Sigma}_{11}^{-\frac{1}{2}}\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21})\right\} = \left\{\mathrm{nzch}(\mathbf{\Sigma}_{11}^{-\frac{1}{2}}\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-\frac{1}{2}})\right\}. \tag{5.18}$$

Koska matriisit $\mathbf{\Sigma}_{11}^{-\frac{1}{2}}$ ja $\mathbf{\Sigma}_{11}^{-\frac{1}{2}}\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21}$ ovat samankokoisia neliömatriiseja, niin yhtälö (5.18) voidaan kirjoittaa myös muodossa

$$\left\{\mathrm{ch}(\mathbf{\Sigma}_{11}^{-\frac{1}{2}}\mathbf{\Sigma}_{11}^{-\frac{1}{2}}\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21})\right\} = \left\{\mathrm{ch}(\mathbf{\Sigma}_{11}^{-\frac{1}{2}}\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-\frac{1}{2}})\right\},$$

eli kaikki matriisien ominaisarvot ovat samat.

Nyt siis kanoniset korrelaatiot on johdettu useammalla eri tavalla, liittyen matriisien ominaisarvoihin. Suurin kanoninen korrelaatio ρ_1 on siis kovarianssimatriisin $\mathbf{\Sigma}$ suurin ominaisarvo λ_1, sekä $\rho_2 = \lambda_2, \rho_3 = \lambda_3, \cdots, \rho_p = \lambda_p$. Toisaalta suurin kanoninen korrelaatio ρ_1 on matriisin $\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}$ suurin ominaisarvo, joka on λ_1^2. Lisäksi tiedetään, että

$$\{\operatorname{nzch}(\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12})\} = \{\operatorname{nzch}(\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21})\} \tag{5.19}$$

$$= \left\{\operatorname{nzch}(\mathbf{\Sigma}_{11}^{-\frac{1}{2}}\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-\frac{1}{2}})\right\}. \tag{5.20}$$

5.2 Kanoniset korrelaatiot ja pienin antiominaisarvo

Luvussa 4 tarkasteltiin antiominaisarvojen tulkintaa korrelaatiokertoimena. Tämän vuoksi onkin luonnollista että antiominaisarvoilla on sovellutuksia myös kanonisten korrelaatioiden yhteydessä.

Ensimmäinen kanoninen korrelaatio voidaan esittää pienimmän antiominaisarvon funktiona kuten Eaton (1976) on esittänyt. Maksimoidaan satunnaismuuttujien $w = \boldsymbol{\alpha}'\boldsymbol{x}^{(1)}$ ja $v = \boldsymbol{\gamma}'\boldsymbol{x}^{(2)}$ välistä korrelaatiota, kun

$$\operatorname{cov}\begin{pmatrix}\boldsymbol{x}^{(1)}\\ \boldsymbol{x}^{(2)}\end{pmatrix} = \mathbf{\Sigma} = \begin{pmatrix}\mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12}\\ \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22}\end{pmatrix}. \tag{5.21}$$

Ensimmäisen kanonisen korrelaation neliö voidaan silloin luonnollisesti kirjoittaa muodossa

$$\max \frac{\operatorname{cov}(w,v)^2}{\operatorname{cov}(w)\operatorname{cov}(v)} = \max_{\boldsymbol{\alpha},\boldsymbol{\gamma}} \frac{(\boldsymbol{\alpha}'\mathbf{\Sigma}_{12}\boldsymbol{\gamma})^2}{\boldsymbol{\alpha}'\mathbf{\Sigma}_{11}\boldsymbol{\alpha}\cdot\boldsymbol{\gamma}'\mathbf{\Sigma}_{22}\boldsymbol{\gamma}}. \tag{5.22}$$

Nyt kun teemme sijoitukset

$$\boldsymbol{a} = \begin{pmatrix}\boldsymbol{\alpha}\\ \mathbf{0}\end{pmatrix} \in \mathbb{R}^n, \qquad \boldsymbol{g} = \begin{pmatrix}\mathbf{0}\\ \boldsymbol{\gamma}\end{pmatrix} \in \mathbb{R}^n, \tag{5.23}$$

niin saadaan epäyhtälö

$$\max_{\boldsymbol{\alpha},\boldsymbol{\gamma}} \frac{(\boldsymbol{\alpha}'\mathbf{\Sigma}_{12}\boldsymbol{\gamma})^2}{\boldsymbol{\alpha}'\mathbf{\Sigma}_{11}\boldsymbol{\alpha}\cdot\boldsymbol{\gamma}'\mathbf{\Sigma}_{22}\boldsymbol{\gamma}} \le \max_{\boldsymbol{a},\boldsymbol{g}} \frac{(\boldsymbol{a}'\mathbf{\Sigma}\boldsymbol{g})^2}{\boldsymbol{a}'\mathbf{\Sigma}\boldsymbol{a}\cdot\boldsymbol{g}'\mathbf{\Sigma}\boldsymbol{g}}, \tag{5.24}$$

sillä epäyhtälön vasen puoli on oikean puolen osajoukko, ja siksi pienempi tai yhtäsuuri kuin oikea puoli.

Epäyhtälön (5.24) perusteella Eaton (1976) osoittaa ensimmäisen kanonisen korrelaation ja pienimmän antiominaisarvon välisen yhteyden. Nyt tämä yhteys kuitenkin osoitetaan toisella tavalla.

Lause 5.1. *Olkoon* $\boldsymbol{\Sigma}$ $p\times p$ *symmetrinen ja positiivisesti definiitti matriisi, jolla on ominaisarvot* $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p > 0$, *ja niitä vastaavat ominaisvektorit* $\boldsymbol{q}_1, \boldsymbol{q}_2, \ldots, \boldsymbol{q}_p$, $||\boldsymbol{q}_i|| = 1$, $i = 1, \ldots, p$, $\boldsymbol{q}_i'\boldsymbol{q}_j = 0$ *kun* $i \neq j$.

Olkoon vektorit $\boldsymbol{y} \neq \boldsymbol{0} \in \mathbb{R}^p$ *ja* $\boldsymbol{x} \neq \boldsymbol{0} \in \mathbb{R}^p$. *Tällöin kanonisen korrelaation neliö voidaan kirjoittaa muodossa*

$$\max_{\substack{\boldsymbol{x}\neq\boldsymbol{0},\boldsymbol{y}\neq\boldsymbol{0},\\ \boldsymbol{x}'\boldsymbol{y}=\boldsymbol{0}}} \left(\frac{\boldsymbol{x}'\boldsymbol{\Sigma}\boldsymbol{y}}{\sqrt{\boldsymbol{x}'\boldsymbol{\Sigma}\boldsymbol{x}\cdot\boldsymbol{y}'\boldsymbol{\Sigma}\boldsymbol{y}}} \right)^2 = 1 - \frac{(\boldsymbol{x}'\boldsymbol{x})^2}{\boldsymbol{x}'\boldsymbol{\Sigma}\boldsymbol{x}\cdot\boldsymbol{x}'\boldsymbol{\Sigma}^{-1}\boldsymbol{x}} = 1 - \mu_1^2. \tag{5.25}$$

Yhtäsuuruus yhtälössä (5.25) *toteutuu kun* $\boldsymbol{x} = \boldsymbol{q}_1 + \boldsymbol{q}_p$ *ja* $\boldsymbol{y} = \boldsymbol{q}_1 - \boldsymbol{q}_p$.

Todistus. Koska vektorit $\boldsymbol{x}$ ja $\boldsymbol{y}$ ovat ortogonaalisia, niin voidaan sanoa että $\exists\boldsymbol{a} \in \mathbb{R}^p : \boldsymbol{y} = (\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{x}})\boldsymbol{a}$. Tällöin rajoite $\boldsymbol{x}'\boldsymbol{y} = \boldsymbol{0}$ voidaan sisällyttää maksimointitehtävään siten, että rajoitutaan etsimään korrelaation maksimoivaa kerroinvektoria $\boldsymbol{a} \in \mathbb{R}^p$ ainoastaan $\mathcal{C}(\boldsymbol{x})^{\perp}$:lta. Merkitään vielä että $(\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{x}}) = \boldsymbol{Q}$, jolloin voidaan kirjoittaa, että

$$\begin{aligned} \max_{\substack{\boldsymbol{x}\neq\boldsymbol{0},\boldsymbol{y}\neq\boldsymbol{0},\\ \boldsymbol{x}'\boldsymbol{y}=\boldsymbol{0}}} \operatorname{cor}^2(\boldsymbol{x}'\boldsymbol{z}, \boldsymbol{y}'\boldsymbol{z}) &= \max_{\substack{\boldsymbol{x}\neq\boldsymbol{0},\boldsymbol{y}\neq\boldsymbol{0},\\ \boldsymbol{x}'\boldsymbol{y}=\boldsymbol{0}}} \frac{(\boldsymbol{x}'\boldsymbol{\Sigma}\boldsymbol{y})^2}{\boldsymbol{x}'\boldsymbol{\Sigma}\boldsymbol{x}\cdot\boldsymbol{y}'\boldsymbol{\Sigma}\boldsymbol{y}} \\ &= \max_{\boldsymbol{x}\neq\boldsymbol{0}} \frac{(\boldsymbol{x}'\boldsymbol{\Sigma}\boldsymbol{Q}\boldsymbol{a})^2}{\boldsymbol{x}'\boldsymbol{\Sigma}\boldsymbol{x}\cdot\boldsymbol{a}'\boldsymbol{Q}\boldsymbol{\Sigma}\boldsymbol{Q}\boldsymbol{a}}. \end{aligned} \tag{5.26}$$

Tehdään valinta

$$\boldsymbol{u} = \boldsymbol{Q}\boldsymbol{\Sigma}\boldsymbol{x}, \quad \boldsymbol{v} = \boldsymbol{Q}\boldsymbol{a},$$

niin voidaan soveltaan lauseen L.1, eli Cauchyn–Schwarzin epäyhtälön toista muotoa (K.2):

$$(\boldsymbol{u}'\boldsymbol{v})^2 \leq \boldsymbol{u}'\boldsymbol{A}\boldsymbol{u}\cdot\boldsymbol{v}'\boldsymbol{A}^-\boldsymbol{v}, \qquad \boldsymbol{v} \in \mathcal{C}(\boldsymbol{A}).$$

Valitaan $\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{\Sigma}\boldsymbol{Q}$, joilloin ehto $\boldsymbol{v} \in \mathcal{C}(\boldsymbol{A})$ toteutuu, sillä

$$\mathcal{C}(\boldsymbol{Q}\boldsymbol{a}) \subset \mathcal{C}(\boldsymbol{Q}\boldsymbol{\Sigma}\boldsymbol{Q}).$$

Cauchyn–Schwarzin epäyhtälön perusteella yhtälön (5.26) maksimoitavalle funktiolle pätee epäyhtälö:

$$\begin{aligned} \frac{(\boldsymbol{x}'\boldsymbol{\Sigma}\boldsymbol{Q}\boldsymbol{a})^2}{\boldsymbol{x}'\boldsymbol{\Sigma}\boldsymbol{x}\cdot\boldsymbol{a}'\boldsymbol{Q}\boldsymbol{\Sigma}\boldsymbol{Q}\boldsymbol{a}} &\leq \frac{\boldsymbol{x}'\boldsymbol{\Sigma}\boldsymbol{Q}(\boldsymbol{Q}\boldsymbol{\Sigma}\boldsymbol{Q})^-\boldsymbol{Q}\boldsymbol{\Sigma}\boldsymbol{x}\cdot\boldsymbol{a}'\boldsymbol{Q}(\boldsymbol{Q}\boldsymbol{\Sigma}\boldsymbol{Q})\boldsymbol{Q}\boldsymbol{a}}{\boldsymbol{x}'\boldsymbol{\Sigma}\boldsymbol{x}\cdot\boldsymbol{a}'\boldsymbol{Q}\boldsymbol{\Sigma}\boldsymbol{Q}\boldsymbol{a}} \\ &= \frac{\boldsymbol{x}'\boldsymbol{\Sigma}\boldsymbol{Q}(\boldsymbol{Q}\boldsymbol{\Sigma}\boldsymbol{Q})^-\boldsymbol{Q}\boldsymbol{\Sigma}\boldsymbol{x}\cdot\boldsymbol{a}'(\boldsymbol{Q}\boldsymbol{\Sigma}\boldsymbol{Q})\boldsymbol{a}}{\boldsymbol{x}'\boldsymbol{\Sigma}\boldsymbol{x}\cdot\boldsymbol{a}'\boldsymbol{Q}\boldsymbol{\Sigma}\boldsymbol{Q}\boldsymbol{a}} \\ &= \frac{\boldsymbol{x}'\boldsymbol{\Sigma}\boldsymbol{Q}(\boldsymbol{Q}\boldsymbol{\Sigma}\boldsymbol{Q})^-\boldsymbol{Q}\boldsymbol{\Sigma}\boldsymbol{x}}{\boldsymbol{x}'\boldsymbol{\Sigma}\boldsymbol{x}}. \end{aligned} \tag{5.27}$$

Tällöin voidaan kirjoittaa, että

$$(5.28) \qquad \max_{\substack{\boldsymbol{x}\neq\boldsymbol{0},\boldsymbol{y}\neq\boldsymbol{0},\\ \boldsymbol{x}'\boldsymbol{y}=\boldsymbol{0}}} \frac{(\boldsymbol{x}'\boldsymbol{\Sigma}\boldsymbol{y})^2}{\boldsymbol{x}'\boldsymbol{\Sigma}\boldsymbol{x}\cdot\boldsymbol{y}'\boldsymbol{\Sigma}\boldsymbol{y}} \leq \max_{\boldsymbol{x}\neq\boldsymbol{0}} \frac{\boldsymbol{x}'\boldsymbol{\Sigma}\boldsymbol{Q}(\boldsymbol{Q}\boldsymbol{\Sigma}\boldsymbol{Q})^{-}\boldsymbol{Q}\boldsymbol{\Sigma}\boldsymbol{x}}{\boldsymbol{x}'\boldsymbol{\Sigma}\boldsymbol{x}}.$$

Oletetaan tunnetuksi, että $\boldsymbol{Q}(\boldsymbol{Q}\boldsymbol{\Sigma}\boldsymbol{Q})^{-}\boldsymbol{Q} = \dot{\boldsymbol{M}}$ ja sille ekvivalentti esitystapa on $\boldsymbol{\Sigma}^{-1}+\boldsymbol{\Sigma}^{-1}\boldsymbol{x}(\boldsymbol{x}'\boldsymbol{\Sigma}^{-1}\boldsymbol{x})^{-1}\boldsymbol{x}'\boldsymbol{\Sigma}^{-1}$. Sijoittamalla tämä epäyhtälöön (5.28), saadaan

$$\begin{aligned}
\max_{\substack{\boldsymbol{x}\neq\boldsymbol{0},\boldsymbol{y}\neq\boldsymbol{0},\\ \boldsymbol{x}'\boldsymbol{y}=\boldsymbol{0}}} \frac{(\boldsymbol{x}'\boldsymbol{\Sigma}\boldsymbol{y})^2}{\boldsymbol{x}'\boldsymbol{\Sigma}\boldsymbol{x}\cdot\boldsymbol{y}'\boldsymbol{\Sigma}\boldsymbol{y}} &\leq \max_{\boldsymbol{x}\neq\boldsymbol{0}} \frac{\boldsymbol{x}'\boldsymbol{\Sigma}\left[\boldsymbol{\Sigma}^{-1}-\boldsymbol{\Sigma}^{-1}\boldsymbol{x}(\boldsymbol{x}'\boldsymbol{\Sigma}^{-1}\boldsymbol{x})^{-1}\boldsymbol{x}'\boldsymbol{\Sigma}^{-1}\right]\boldsymbol{\Sigma}\boldsymbol{x}}{\boldsymbol{x}'\boldsymbol{\Sigma}\boldsymbol{x}} \\
&= \max_{\boldsymbol{x}\neq\boldsymbol{0}} \left[\frac{\boldsymbol{x}'\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}\boldsymbol{x}}{\boldsymbol{x}'\boldsymbol{\Sigma}\boldsymbol{x}} - \frac{\boldsymbol{x}'\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{-1}\boldsymbol{x}(\boldsymbol{x}'\boldsymbol{\Sigma}^{-1}\boldsymbol{x})^{-1}\boldsymbol{x}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}\boldsymbol{x}}{\boldsymbol{x}'\boldsymbol{\Sigma}\boldsymbol{x}}\right] \\
(5.29) \qquad &= \max_{\boldsymbol{x}\neq\boldsymbol{0}} \left[\frac{\boldsymbol{x}'\boldsymbol{\Sigma}\boldsymbol{x}}{\boldsymbol{x}'\boldsymbol{\Sigma}\boldsymbol{x}} - \frac{\boldsymbol{x}'\boldsymbol{x}(\boldsymbol{x}'\boldsymbol{\Sigma}^{-1}\boldsymbol{x})^{-1}\boldsymbol{x}'\boldsymbol{x}}{\boldsymbol{x}'\boldsymbol{\Sigma}\boldsymbol{x}}\right] \\
(5.30) \qquad &= \max_{\boldsymbol{x}\neq\boldsymbol{0}} \left[1 - \frac{(\boldsymbol{x}'\boldsymbol{x})^2}{\boldsymbol{x}'\boldsymbol{\Sigma}\boldsymbol{x}\cdot\boldsymbol{x}'\boldsymbol{\Sigma}^{-1}\boldsymbol{x}}\right] \\
(5.31) \qquad &= 1-\mu_1^2,
\end{aligned}$$

jossa yhtälöiden (5.29) ja (5.30) välinen yhtäsuuruus pitää paikkansa sillä sekä $\boldsymbol{x}'\boldsymbol{x}$ ja $\boldsymbol{x}'\boldsymbol{\Sigma}^{-1}\boldsymbol{x}$ ovat reaalilukuja.

□

Yhtälöä (5.25) vastaavaa epäyhtälöä kutsutaan joissakin yhteyksissä Wielandin epäyhtälöksi (esim. Drury, Liu, Lu, Puntanen, Styan 2000). Tämä on hyvä huomata, sillä tämä vektoriyhtälö tullaan kappaleessa 6.3 yleistämään matriisimuotoiseksi.

Nyt on siis todettu että ensimmäinen kanoninen korrelaatio voidaan kirjoittaa kovarianssimatriisin $\boldsymbol{\Sigma}$ pienimmän antiominaisarvon μ_1 funktiona:

$$\rho_1^2 = 1-\mu_1^2.$$

Tämä tulos kuitenkin yleistyy koskemaan muitakin kanonisia korrelaatioita:

$$\rho_i^2 = 1-\mu_i^2, \qquad i=1,2,\ldots$$

Edellä ollaan todettu, että matriisin $\boldsymbol{\Sigma}$ pienimmästä antiominaisarvosta μ_1 käytetään varsinkin matemaattisessa kirjallisuudessa termiä matriisin $\boldsymbol{\Sigma}$ kosini, '*the cosine of matrix* $\boldsymbol{\Sigma}$' eli $\cos\boldsymbol{\Sigma}$. Tällä merkintätavalla ensimmäinen kanoninen korrelaatio voidaan esittää muodossa $\sin\boldsymbol{\Sigma}$ (Gustafson ja Rao 1997).

Tarkastellaan kahden satunnaisvektorin $\boldsymbol{v}_1$ ja $\boldsymbol{v}_2$ välistä kanonista korrelaatiota, kun $\boldsymbol{v}_1 = \boldsymbol{P}_1'\boldsymbol{u}$ ja $\boldsymbol{v}_2 = \boldsymbol{P}_2'\boldsymbol{u}$, missä $\boldsymbol{P}_{p\times p} = (\boldsymbol{P}_1 : \boldsymbol{P}_2)$ on ortogonaalinen. Olkoon $\boldsymbol{v}_1$ $r\times 1$ ja $\boldsymbol{v}_2$ $(p-r)\times 1$ vektori siten, että $1\leq r\leq \frac{p}{2}$.

Muodostetaan vektoreiden $\boldsymbol{v}_1$ ja $\boldsymbol{v}_2$ kovarianssimatriisi:

$$\begin{aligned}\operatorname{cov}\begin{pmatrix}\boldsymbol{v}_1\\ \boldsymbol{v}_2\end{pmatrix} &= \begin{pmatrix}\operatorname{cov}(\boldsymbol{v}_1) & \operatorname{cov}(\boldsymbol{v}_1,\boldsymbol{v}_2)\\ \operatorname{cov}(\boldsymbol{v}_2,\boldsymbol{v}_1) & \operatorname{cov}(\boldsymbol{v}_2)\end{pmatrix}\\ &= \begin{pmatrix}\boldsymbol{P}_1'\operatorname{cov}(\boldsymbol{u})\boldsymbol{P}_1 & \boldsymbol{P}_1'\operatorname{cov}(\boldsymbol{u})\boldsymbol{P}_2\\ \boldsymbol{P}_2'\operatorname{cov}(\boldsymbol{u})\boldsymbol{P}_1 & \boldsymbol{P}_2'\operatorname{cov}(\boldsymbol{u})\boldsymbol{P}_2\end{pmatrix}\\ &= \begin{pmatrix}\boldsymbol{P}_1'\boldsymbol{\Sigma}\boldsymbol{P}_1 & \boldsymbol{P}_1'\boldsymbol{\Sigma}\boldsymbol{P}_2\\ \boldsymbol{P}_2'\boldsymbol{\Sigma}\boldsymbol{P}_1 & \boldsymbol{P}_2'\boldsymbol{\Sigma}\boldsymbol{P}_2\end{pmatrix}.\end{aligned}$$

Tällöin satunnaisvektoreiden $\boldsymbol{v}_1$ ja $\boldsymbol{v}_2$ suurin kanoninen korrelaatio voidaan esittää edellä johdettujen matriisien (5.19) ja (5.20) ominaisarvo-ongelmina. Suurimman kanonisen korrelaation neliö voidaan esimerkiksi esittää matriisin

$$(\boldsymbol{P}_1'\boldsymbol{\Sigma}\boldsymbol{P}_1)^{-1}\boldsymbol{P}_1'\boldsymbol{\Sigma}\boldsymbol{P}_2(\boldsymbol{P}_2'\boldsymbol{\Sigma}\boldsymbol{P}_2)^{-1}\boldsymbol{P}_2'\boldsymbol{\Sigma}\boldsymbol{P}_1$$

suurimpana ominaisarvona, eli

$$\rho_1^2 = \max_{\boldsymbol{y}} \frac{\boldsymbol{y}'(\boldsymbol{P}_1'\boldsymbol{\Sigma}\boldsymbol{P}_1)^{-1}\boldsymbol{P}_1'\boldsymbol{\Sigma}\boldsymbol{P}_2(\boldsymbol{P}_2'\boldsymbol{\Sigma}\boldsymbol{P}_2)^{-1}\boldsymbol{P}_2'\boldsymbol{\Sigma}\boldsymbol{P}_1\boldsymbol{y}}{\boldsymbol{y}'\boldsymbol{y}} = \lambda_1^2,$$

missä $\boldsymbol{y} \neq \boldsymbol{0} \in \mathbb{R}^r$. Tässä tarkastelussa matriisit $\boldsymbol{P}_1$ ja $\boldsymbol{P}_2$ olivat siis tunnettuja ja vakioita.

Esimerkki 5.1. Olkoon malli $\{\boldsymbol{y},\, \boldsymbol{X\beta},\, \boldsymbol{V}\}$, jossa $\boldsymbol{X}$ on $n \times k$ –mallimatriisi. Pienimmän neliösumman ennuste $\hat{\boldsymbol{y}}$ ja jäännöstermi $\boldsymbol{\epsilon}$ ovat:

$$\begin{aligned}&\hat{\boldsymbol{y}} = \boldsymbol{Hy},\\ &\boldsymbol{\epsilon} = \boldsymbol{y} - \hat{\boldsymbol{y}} = (\boldsymbol{I} - \boldsymbol{H})\boldsymbol{y} = \boldsymbol{My}.\end{aligned}$$

Tarkastellaan pienimmän neliösumman ennusteen $\hat{\boldsymbol{y}}$ ja jäännöstermin $\boldsymbol{\epsilon}$ välistä suurinta kanonista korrelaatiota, eli

$$\begin{aligned}\operatorname{cor}^2(\hat{\boldsymbol{y}},\boldsymbol{\epsilon}) &= \operatorname{cc}^2(\hat{\boldsymbol{y}},\boldsymbol{\epsilon})\\ &= \operatorname{cc}^2(\boldsymbol{Hy},\boldsymbol{My}).\end{aligned}$$

Tällöin sovitteen $\boldsymbol{Hy}$ ja jäännöstermin $\boldsymbol{My}$ välinen yhteisjakauman kovarianssimatriisi on

$$\operatorname{cov}\begin{pmatrix}\boldsymbol{Hy}\\ \boldsymbol{My}\end{pmatrix} = \begin{pmatrix}\boldsymbol{HVH} & \boldsymbol{HVM}\\ \boldsymbol{MVH} & \boldsymbol{MVM}\end{pmatrix}. \tag{5.32}$$

Edellä ollaan todettu, että ensimmäinen kanonisen korrelaation saadaan esimerkiksi yhteisjakauman kovarianssimatriisin (5.32) suurimpana ominaisarvona. Edellä esitellyn tarkastelun perusteella kuitenkin tiedetään, että ensimmäinen kanoninen korrelaatio voidaan määritellä myös kovarianssimatriisin $\boldsymbol{V}$ pienimmän antiominaisarvon funktiona, eli $\rho_1^2 = 1 - \mu_1^2$.

Edellä käsiteltiin tilannetta jossa ortogonaalinen matriisi $\boldsymbol{P} = (\boldsymbol{P}_1 : \boldsymbol{P}_2)$ oli tunnettu, ja haluttiin ratkaista ongelma

$$\max_{\boldsymbol{\alpha},\boldsymbol{\gamma}} \operatorname{cor}^2(\boldsymbol{\alpha}'\boldsymbol{P}_1'\boldsymbol{u}, \boldsymbol{\gamma}'\boldsymbol{P}_2'\boldsymbol{u}). \tag{5.33}$$

Joskus kuitenkin voidaan halutaan valita sarakkeittain ortogonaaliset matriisit $\boldsymbol{P}_1$ ja $\boldsymbol{P}_2$ siten, että ρ_1^2 maksimoituu:

$$\max_{\boldsymbol{P}_1} \rho_1^2. \tag{5.34}$$

Venables (1976) osoitti, että tällöin tehtävän (5.34) kanssa ekvivalentti tehtävä on etsiä kaksi lineaarista funktiota $\boldsymbol{k}_1'\boldsymbol{u}$ ja $\boldsymbol{k}_2'\boldsymbol{u}$, jotka ratkaisevat ongelman:

$$\begin{aligned} &\max && (\boldsymbol{k}_1'\boldsymbol{\Sigma}\boldsymbol{k}_2)^2 && (5.35)\\ &\text{ehdolla} && \boldsymbol{k}_1'\boldsymbol{k}_2 = 0, && (5.36)\\ & && \boldsymbol{k}_1'\boldsymbol{\Sigma}\boldsymbol{k}_1 = 1, && (5.37)\\ & && \boldsymbol{k}_2'\boldsymbol{\Sigma}\boldsymbol{k}_2 = 1. && (5.38)\end{aligned}$$

Ensimmäiset kanoniset muuttujat v_1 ja w_1 voidaan siis kirjoittaa muodossa

$$\begin{aligned} v_1 &= \boldsymbol{\alpha}'\boldsymbol{v}_1 = \boldsymbol{\alpha}'\boldsymbol{P}_1'\boldsymbol{u} = \boldsymbol{k}_1'\boldsymbol{u},\\ w_1 &= \boldsymbol{\gamma}'\boldsymbol{v}_2 = \boldsymbol{\gamma}'\boldsymbol{P}_2'\boldsymbol{u} = \boldsymbol{k}_2'\boldsymbol{u}.\end{aligned}$$

Koska matriisit $\boldsymbol{P}_1$ ja $\boldsymbol{P}_2$ ovat sarakkeittain ortogonaalisia, ja vektorit $\boldsymbol{\alpha}$ sekä $\boldsymbol{\gamma}$ käyvät maksimoinnissa läpi koko $\mathbb{R}^r$:n, niin vektorit $\boldsymbol{k}_1$ ja $\boldsymbol{k}_2$ käyvät yhdessä läpi koko $\mathbb{R}^p$:n.

Yhtälöstä (5.35) on helppoa huomata, että vastaavan minimointiongelman toteuttaisivat mitkä tahansa $\boldsymbol{\Sigma}$:n ominaisvektorit. Ehdot (5.36)-(5.38) täyttyvät triviaalisti, sillä ominaisvektorit voidaan aina valita siten että ne ovat ortogonaalisia ja ykkösen mittaisia.

Koska symmetrisellä matriisilla $\boldsymbol{\Sigma}$ on ominaisarvohajotelma, niin minimoitava funktion (5.35) neliöjuuri voidaan kirjoittaa muodossa

$$\boldsymbol{k}_1'\boldsymbol{\Sigma}\boldsymbol{k}_2 = \boldsymbol{k}_1'\boldsymbol{T}\boldsymbol{\Lambda}\boldsymbol{T}'\boldsymbol{k}_2 = \boldsymbol{e}_i'\boldsymbol{\Lambda}\boldsymbol{e}_j = 0,$$

missä $\boldsymbol{e}_i = (0, 0, \cdots, 0, 1, 0, \cdots, 0)'$ on i. kantavektori, ja $\boldsymbol{\Lambda}$ on ominaisarvomatriisi. Koska $\boldsymbol{k}_1$:ksi oli valittu matriisin $\boldsymbol{\Sigma}$ i. ominaisvektori, niin se on ortogonaalinen kaikkiin muihin $\boldsymbol{\Sigma}$:n ominaisvektoreihin nähden ja sen normi on voitu valita ykköseksi. Tällöin $\boldsymbol{T}'\boldsymbol{k}$ on i. kantavektori, ja alkuperäisen minimoitavan fuktion (5.35) minimiarvoksi tulee myös nolla.

Lisäksi voidaan huomata, että optimointitehtävä (5.35) palautuu siis yhtälön (5.25) tarkasteluksi.

6 Muista epäyhtälöistä ja muodoista

6.1 Antiominaisarvon monet muodot

Tähänastisissa tarkasteluissa ollaan aina palautettu antiominaisarvon tunnetuinpaan muotoon, eli pienimmän ja suurimman ominaisarvon geometrisen ja aritmeettisen keskiarvon suhteeseen:

$$\mu_1 = \frac{\sqrt{\lambda_1 \lambda_n}}{\frac{1}{2}(\lambda_1 + \lambda_n)} = \frac{2\sqrt{\lambda_1 \lambda_n}}{\lambda_1 + \lambda_n}.$$

Antiominaisarvolla on kuitenkin asiayhteydestä riippuen monia esitysmuotoja. Esimerkiksi artikkeleista Drury et al. (2000) ja Wang ja Ip (1999) kerättyjä ekvivalentteja esitysmuotoja pienimmälle antiominaisarvolle ovat:

$$\begin{aligned}\mu_1 = \frac{2\sqrt{\lambda_1 \lambda_n}}{\lambda_1 + \lambda_n} &= \frac{2}{\sqrt{\frac{\lambda_1}{\lambda_n}} + \sqrt{\frac{\lambda_n}{\lambda_1}}} = \frac{2}{\sqrt{\kappa} + \sqrt{\frac{1}{\kappa}}} = \frac{2}{\sqrt{(\lambda_1 + \lambda_n)\left(\frac{1}{\lambda_1} + \frac{1}{\lambda_n}\right)}} \\ &= 1 - \left(\frac{\lambda_1 - \lambda_n}{\lambda_1 + \lambda_n}\right) = 1 - \left(\frac{\kappa - 1}{\kappa + 1}\right) = \frac{2\sqrt{\kappa}}{(\kappa + 1)},\end{aligned}$$

missä $\kappa = \lambda_1/\lambda_n$ on matriisin $\boldsymbol{A}$ kuntoisuusluku '*condition number*'.

6.2 Muista epäyhtälöistä

Edellä ollaan esitetty, että Fruchtin–Kantorovichin epäyhtälön yläraja pystytään määrittelemään helposti sijoittamalla $\boldsymbol{Ax} = \boldsymbol{y}$, jolloin päädytään tarkastelemaan Caychyn–Schwarzin epäyhtälöä. Fruchtin–Kantorovichin epäyhtälön alaraja on yleensä ylärajaa kiinnostavampi. Kirjallisuudesta löytyykin useita epäyhtälöitä, jotka ovat läheisessä yhteydessä Fruchtin–Kantorovichin epäyhtälöön. Householder (1974) osoittikin, että Wielandtin epäyhtälöstä

$$\frac{(\boldsymbol{x}'\boldsymbol{Ay})^2}{\boldsymbol{x}'\boldsymbol{Ax} \cdot \boldsymbol{y}'\boldsymbol{Ay}} \leq \left(\frac{\lambda_1 - \lambda_n}{\lambda_1 + \lambda_n}\right)^2, \tag{6.1}$$

saadaan Fruchtin–Kantorovichin epäyhtälö sijoituksella

$$\boldsymbol{y} = \boldsymbol{A}^{-1}\boldsymbol{x} - \left(\frac{\boldsymbol{x}'\boldsymbol{A}^{-1}\boldsymbol{x}}{\boldsymbol{x}'\boldsymbol{x}}\right)\boldsymbol{x} = \left(\boldsymbol{I} - \frac{\boldsymbol{xx}'}{\boldsymbol{x}'\boldsymbol{x}}\right)\boldsymbol{A}^{-1}\boldsymbol{x}.$$

Toiseen suuntaan, eli Fruchtin–Kantorovichin epäyhtälöstä saadaan Wielandtin epäyhtälö, ovat todistaneet esimerkiksi Bhatia (1999), Bhatia ja Davis (2000) ja Zang (1999, 2001). Voidaankin sanoa että nämä kaksi epäyhtälöä ovat olennaisilta osin ekvivalentit. Watson, Albargu ja Styan (1997) osoittivat samalla tavalla ekvivalenteiksi Fruchtin–Kantorovichin epäyhtälön kanssa Schweitzer, Pólya–Szegö, Greub–Rheinboldt, Cassels ja Krasnosel'skiĭ–Kreĭn epäyhtälöt. Lisää tietoa Kantorivichin tyyppisistä epäyhtälöistä löytyy esimerkiksi artikkelista Drury et al. (2000).

6.3 Matriisimuotoinen Wielandtin epäyhtälö

Wang ja Ip (1999) ovat osoittaneet, että edellä esitetystä Wielandtin epäyhtälöstä voidaan esittää myös matriisiversio:

Lause 6.1 (Matriisimuotoinen Wielandintin epäyhtälö)**.** *Olkoon $\boldsymbol{A}_{n\times n}$ positiivisesti definiitti matriisi, ja λ_1 ja λ_n sen suurin ja pienin ominaisarvo, sekä niitä vastaavat ortonormaalit ominaisvektorit ovat $\boldsymbol{q}_1$ ja $\boldsymbol{q}_n$. Olkoot $\boldsymbol{X}_{n\times p}$ ja $\boldsymbol{Y}_{n\times q}$ reaalisia matriiseja jotka toteuttavat ehdon $\boldsymbol{X}'\boldsymbol{Y} = \boldsymbol{0}$. Tällöin*

$$\boldsymbol{X}'\boldsymbol{A}\boldsymbol{Y}(\boldsymbol{Y}'\boldsymbol{A}\boldsymbol{Y})^{-}\boldsymbol{Y}'\boldsymbol{A}\boldsymbol{X} \leq_L \left(\frac{\lambda_1 - \lambda_n}{\lambda_1 + \lambda_n}\right)^2 \boldsymbol{X}'\boldsymbol{A}\boldsymbol{X}, \tag{6.2}$$

missä "$\leq_L$" viittaa Löwnerin järjestykseen. Yhtäsuuruus (6.2)*:ssa pätee kun*

$$\frac{1}{\lambda_1}\boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{q}_1 + \frac{1}{\lambda_n}\boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{q}_n = \gamma(\boldsymbol{q}_1 + \boldsymbol{q}_n),$$

missä γ on reaaliluku.

Todistus. Todistus sivuutetaan, mutta löytyy artikkelista Wang ja Ip (1999). □

Tilastotieteen sovellutuksia epäyhtälölle (6.2) löytyy ainakin kun oletetaan, että $n = p + q$ ja $\boldsymbol{X} = \begin{pmatrix} \boldsymbol{I}_p \\ \boldsymbol{0} \end{pmatrix}$ ja $\boldsymbol{Y} = \begin{pmatrix} \boldsymbol{0} \\ \boldsymbol{I}_{n-p} \end{pmatrix}$. Nyt jos matriisi $\boldsymbol{A}$ ositetaan siten, että

$$\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X} = \boldsymbol{A}_{11}, \quad \boldsymbol{X}'\boldsymbol{A}\boldsymbol{Y} = \boldsymbol{A}_{12}, \quad \boldsymbol{Y}'\boldsymbol{A}\boldsymbol{X} = \boldsymbol{A}_{21}, \quad \boldsymbol{Y}'\boldsymbol{A}\boldsymbol{Y} = \boldsymbol{A}_{22},$$

niin epäyhtälö (6.2) saadaan muotoon

$$\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21} \leq_L \left(\frac{\lambda_1 - \lambda_n}{\lambda_1 + \lambda_n}\right)^2 \boldsymbol{A}_{11}. \tag{6.3}$$

Esimerkki 6.1. Muodostetaan satunnaisvektoreiden $\boldsymbol{u}$ ja $\boldsymbol{v}$ ensimmäinen kanoninen korrelaatio Wielandtin epäyhtälön matriisimuotoa käyttäen (Wang ja Ip 1999). Olkoon vektorit $\boldsymbol{u}_{p\times 1}$ ja $\boldsymbol{v}_{q\times 1}$ satunnaisvektoreita, ja

$$\operatorname{cov}\begin{pmatrix}\boldsymbol{u}\\ \boldsymbol{v}\end{pmatrix} = \boldsymbol{\Sigma} = \begin{pmatrix}\boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12}\\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22}\end{pmatrix}.$$

Wiedlandtin epäyhtälön matriisimuodon perusteella

$$\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} \leq_L \left(\frac{\lambda_1-\lambda_n}{\lambda_1+\lambda_n}\right)^2 \boldsymbol{\Sigma}_{11}, \tag{6.4}$$

jossa λ_i on matriisin $\boldsymbol{\Sigma}$ i. ominaisarvo. Kertomalla yhtälö (6.4) vasemmalta ja oikealta matriisilla $\boldsymbol{\Sigma}_{11}^{-\frac{1}{2}}$ saadaan

$$\boldsymbol{\Sigma}_{11}^{-\frac{1}{2}}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-\frac{1}{2}} \leq_L \left(\frac{\lambda_1-\lambda_n}{\lambda_1+\lambda_n}\right)^2 \boldsymbol{I}_p.$$

Koska yleisyyttä menettämättä voidaan olettaa, että $p \leq q$ ja $r(\boldsymbol{\Sigma}_{12}) = p$, niin kanonisille korrelaatioille $\rho_1 \geq \rho_2 \geq \cdots \geq \rho_p$ pätee, että

$$\rho_i < \frac{\lambda_1-\lambda_n}{\lambda_1+\lambda_n}, \qquad i = 1, 2, \ldots, p.$$

Esimerkki 6.2. Olkoon matriisi

$$\boldsymbol{S} = \begin{pmatrix}\boldsymbol{S}_{11} & \boldsymbol{S}_{12}\\ \boldsymbol{S}_{21} & \boldsymbol{S}_{22}\end{pmatrix}$$

esimerkissä (6.1) olevan matriisin $\boldsymbol{\Sigma}$ estimaatti, missä $\boldsymbol{S}_{11}$ on $p \times p$-matriisi. Wiedlandtin epäyhtälön matriisimuodon perusteella

$$\boldsymbol{S}_{12}\boldsymbol{S}_{22}^{-1}\boldsymbol{S}_{21} \leq_L \left(\frac{\lambda_1-\lambda_n}{\lambda_1+\lambda_n}\right)^2 \boldsymbol{S}_{11},$$

jollon matriisin $\boldsymbol{S}_{22}$ Schurin komplementti $\boldsymbol{S}$:n suhteen

$$\begin{aligned}\boldsymbol{S}_{11\cdot 2} &= \boldsymbol{S}_{11} - \boldsymbol{S}_{12}\boldsymbol{S}_{22}^{-1}\boldsymbol{S}_{21}\\ &\geq_L \boldsymbol{S}_{11} - \left(\frac{\lambda_1-\lambda_n}{\lambda_1+\lambda_n}\right)^2 \boldsymbol{S}_{11}\\ &= \left(1-(1-\mu_1^2)\right)\boldsymbol{S}_{11}\\ &= \mu_1^2\boldsymbol{S}_{11}, \end{aligned} \tag{6.5}$$

missä μ_1 on matriisin $\boldsymbol{S}$ pienin antiominaisarvo.

Tarkastellaan uuden havainnon ennustamista, kun satunnaisvektorin $\boldsymbol{y}$ oletetaan muodostuvan lineaarisesta malllista

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

missä $\boldsymbol{X}$ on $n \times k$ mallimatriisi. Uudet havainnot sisältävä satunnaisvektori $\boldsymbol{y}_f$ oletetaan muodostuvan lineaarisesta mallista

$$\boldsymbol{y}_f = \boldsymbol{X}_f\boldsymbol{\beta} + \boldsymbol{\epsilon}_f,$$

missä $\boldsymbol{X}_f$ on $m \times k$ mallimatriisi. Olkoon satunnaisvektoreiden $\boldsymbol{y}_f$ ja $\boldsymbol{y}$ yhteisjakauman kovarianssimatriisi

$$\operatorname{cov}\begin{pmatrix} \boldsymbol{y}_f \\ \boldsymbol{y} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{12} & \boldsymbol{\Sigma}_{22} \end{pmatrix},$$

ja oletetaan se positiivisesti definiitiksi. Tarkastellaan tilannetta, jossa $\boldsymbol{y}$:n $\boldsymbol{y}_f$:n yhteisjakaumaksi oletetaan normaalijakauma

$$\begin{pmatrix} \boldsymbol{y}_f \\ \boldsymbol{y} \end{pmatrix} \sim N\left(\begin{pmatrix} \boldsymbol{y}_f \\ \boldsymbol{y} \end{pmatrix}, \sigma^2 \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{12} & \boldsymbol{\Sigma}_{22} \end{pmatrix}\right).$$

Keskineliövirheen mielessä paras lineaarinen ennuste uudelle havainnolle $\boldsymbol{y}_f$ on ehdollinen odotusarvo.

Vektorin $\boldsymbol{y}_f$:n ehdollinen odotusarvo ehdolla $\boldsymbol{y}$ on

$$E(\boldsymbol{y}_f|\boldsymbol{y}) = E(\boldsymbol{y}_f) - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{y} - E(\boldsymbol{y})),$$

ja tällöin siis satunnaisvektorin $\boldsymbol{y}_f$ ehdollinen varianssi on

$$\operatorname{cov}(\boldsymbol{y}_f|\boldsymbol{y}) = \boldsymbol{\Sigma}_{11\cdot 2} = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{12}. \tag{6.6}$$

Yhtälön (6.6) ja (6.5) perusteella voidaan päätellä, että

$$\mu_1^2\boldsymbol{\Sigma}_{11} \leq_L \boldsymbol{\Sigma}_{11\cdot 2} \leq_L \boldsymbol{\Sigma}_{11}. \tag{6.7}$$

Esimerkki 6.3. Olkoon tarkasteltava aineisto keskistetty, ja olkoon sen ositettu kovarianssimatriisi

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix},$$

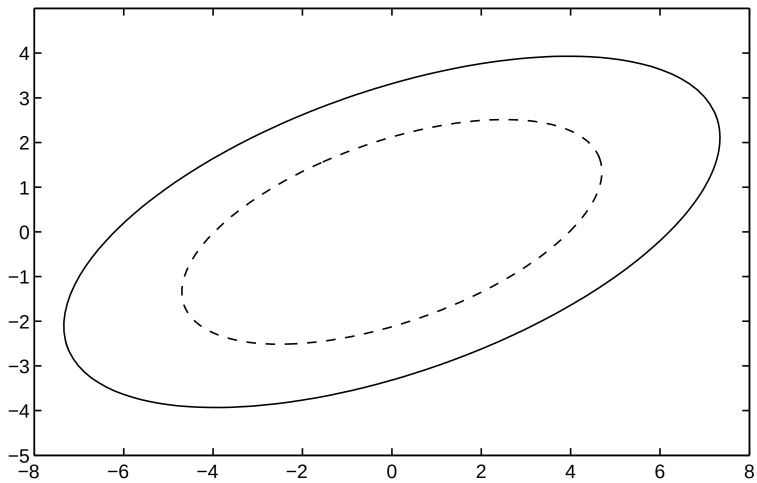
jonka pienin antiominaisarvo $\mu_p = 0.8$. Olkoon vielä vasemman yläkulman osite

$$\boldsymbol{\Sigma}_{11} = \begin{pmatrix} 2 & 2 \\ 2 & 7 \end{pmatrix}.$$

Edellisen epäyhtälön (6.7) perusteella pätee, että

$$\mu_1^2 \boldsymbol{\Sigma}_{11} = \begin{pmatrix} 1.64 & 1.64 \\ 1.64 & 5.74 \end{pmatrix} \leq_L \boldsymbol{\Sigma}_{11 \cdot 2} \leq_L \boldsymbol{\Sigma}_{11} = \begin{pmatrix} 2 & 2 \\ 2 & 7 \end{pmatrix}.$$

Kuviossa (6.1) ovat kovarianssimatriisien $\mu_1^2 \boldsymbol{\Sigma}_{11}$ ja $\boldsymbol{\Sigma}_{11}$ määrittelemät konfidenssiellipsit. Katkoviivalla piirretty $\mu_1^2 \boldsymbol{\Sigma}_{11}$:n konfidenssiellipsi on $\boldsymbol{\Sigma}_{11}$:n konfidenssiellipsoidin sisäpuolla, ja niiden etäisyyden toisistaan määrää pienin antiominaisarvo μ_1. Mitä lähempänä ykköstä pienin antiominaisarvo on, niin sitä lähempänä konfidenssiellipsit ovat toisiansa, ja sitä vähemmän hyötyä kovarianssirakenteen mallintamisesta on mahdollista saada.



Kuvio 6.1. Matriisin $\mu_1^2 \boldsymbol{\Sigma}_{11}$ määrittelemä konfidenssiellipsi on piirretty katkoviivalla ja $\boldsymbol{\Sigma}_{11}$ määrittelemä konfidenssiellipsi yhtenäisellä viivalla.

7 Yleistetty antiominaisarvo

7.1 Yleistetty antiominaisarvo

Khattreen (2002) julkaisussa esitellään antiominaisarvo-ongelman yleistys, jossa antiominaisvektorin sijasta etsitään antiominaismatriisia. Yleistetyt antiominaisarvot voidaan muodostaa helposti antiominaisarvojen avulla, jolloin suurin yleistetty antiominaisarvo $\mu_{[1]}$ on suoraan pienin antiominaisarvo μ_1.

Määritelmä 7.1. *Olkoon matriisi $\boldsymbol{A}_{p\times p}$ reaalinen symmetrinen ja positiivisesti definiitti matriisi. Tarkastellaan optimointitehtävää*

$$\mu_{[r]} = \min_{\boldsymbol{X}} \frac{|\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X}|}{\sqrt{|\boldsymbol{X}'\boldsymbol{A}^2\boldsymbol{X}||\boldsymbol{X}'\boldsymbol{X}|}}, \tag{7.1}$$

missä $\boldsymbol{X}_{p\times r}$ on reaalinen matriisi astetta r siten, että $1 \leq r \leq [p/2]$. Optimointitehtävän (7.1) *ratkaisua $\mu_{[r]}$ kutsutaan matriisin $\boldsymbol{A}$ r:nneksi yleistetyksi antiominaisarvoksi ja ratkaisun tuottavaa matriisia $\boldsymbol{X}^*$ r:ttä yleistettyä antiominaisarvoa vastaavaksi antiominaismatriisiksi.*

Olkoon matriisi $\boldsymbol{A}_{p\times p}$ reaalinen, symmetrinen ja positiivisesti definiitti matriisi ja minimoidaan funktiota

$$h(\boldsymbol{X}) = \frac{|\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X}|}{\sqrt{|\boldsymbol{X}'\boldsymbol{A}^2\boldsymbol{X}|}}, \qquad \text{ehdolla } \boldsymbol{X}'\boldsymbol{X} = \boldsymbol{I}.$$

Koska funktion ja sen logaritmin ääriarvot ovat samat, minimoinnissa voidaan käyttää funktiota $\log h(\boldsymbol{X})$. Funktion $h(\boldsymbol{X})$ osoittaja ja nimittäjä ovat polynomeja matriisin $\boldsymbol{X}$ elementtien suhteen, joten ne molemmat ovat jatkuvasti differentoituvia ja tällöin siis myös funktio $h(\boldsymbol{X})$ on jatkuvasti differentoituva. Tällöin on riittävää tarkastella funktion $\sum_{i=1}^{p} x_{ti}x_{tj}$ gradienttien lineaarista riippumattomuutta, kun $i = 1, 2, \cdots, j$ ja $j = 1, 2, \cdots, r$. Merkitään, että

$$\frac{\partial}{\partial x_{lm}} \sum_{t=1}^{p} x_{ti}x_{tj} = \delta_{im}x_{lj} + \delta_{jm}x_{li} = 0,$$

missä

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

Tehdään vastaoletus, että gradientit ovat lineaarisesti riippuvia. Jotta gradientit voivat olla lineaarisesti riippuvia, täytyy olla olemassa sellainen yläkolmiomatriisi $\boldsymbol{B} \neq \boldsymbol{0}$, siten että

$$\sum_{i=1}^{p}\sum_{j=1}^{p} b_{ij}(\delta_{im}x_{lj} + \delta_{jm}x_{lj}) = 0,$$

mikä pätee jokaisella parilla (l, m). Matriisimerkinnöin tämä voidaan kirjoittaa, että $(\boldsymbol{B} + \boldsymbol{B}')\boldsymbol{X}' = \boldsymbol{0}$. Nyt jos molemmat puolet kerrotaan oikealta matriisilla $\boldsymbol{X}$, niin

$$\begin{aligned}(\boldsymbol{B} + \boldsymbol{B}')\boldsymbol{X}'\boldsymbol{X} &= \boldsymbol{0}\\ \boldsymbol{B} + \boldsymbol{B}' &= \boldsymbol{0},\end{aligned}$$

ja koska $\boldsymbol{B}$ on yläkolmiomatriisi, $\boldsymbol{B} = \boldsymbol{0}$. Tämä on kuitenkin ristiriidassa vastaoletuksen kanssa, ja tällöin alkuperäisestä oletuksesta seuraa, että gradientit ovat lineaarisesti riippumattomia. Tämän perusteella me siis tiedämme, että funktion $h(\boldsymbol{X})$ minimi lötyy stationaariarvojen joukosta.

Nyt siis stationaaritila voidaan löytää käyttämällä Lagrangen kertojamenettelyä. Muodostetaan Lagrangen funktio jonka avulla minimoidaan funktiota $\log h(\boldsymbol{X})$, ehdolla että $\boldsymbol{X}'\boldsymbol{X} = \boldsymbol{I}$, eli

$$Lag(\boldsymbol{X}, \boldsymbol{T}) = \log|\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X}| - \frac{1}{2}\log|\boldsymbol{X}'\boldsymbol{A}^2\boldsymbol{X}| - \mathrm{tr}(\boldsymbol{X}'\boldsymbol{X}\boldsymbol{T}), \tag{7.2}$$

missä Lagrangen kertojamatriisi $\boldsymbol{T}$ on yläkolmiomatriisi. Derivoimalla yhtälö (7.2) $\boldsymbol{X}$:n suhteen saadaan

$$\begin{aligned}\frac{\partial Lag(\boldsymbol{X}, \boldsymbol{T})}{\partial \boldsymbol{X}} &= \boldsymbol{A}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X})^{-1} - \frac{1}{2}\boldsymbol{A}^2\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{A}^2\boldsymbol{X})^{-1}\\ &\quad - \frac{1}{2}\boldsymbol{X}(\boldsymbol{T} + \boldsymbol{T}') = \boldsymbol{0},\end{aligned} \tag{7.3}$$

ja derivoimalla matriisin $\boldsymbol{T}$ suhteen saadaan

$$\frac{\partial Lag(\boldsymbol{X}, \boldsymbol{T})}{\partial \boldsymbol{T}} = \boldsymbol{X}'\boldsymbol{X} = \boldsymbol{I}_r. \tag{7.4}$$

Kertomalla vasemmalta yhtälö (7.3) matriisilla $\boldsymbol{X}'$ ja sijoittamalla (7.4) saadaan

$$\begin{aligned}\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X})^{-1} - \frac{1}{2}\boldsymbol{X}'\boldsymbol{A}^2\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{A}^2\boldsymbol{X})^{-1} - \frac{1}{2}\boldsymbol{X}'\boldsymbol{X}(\boldsymbol{T} + \boldsymbol{T}') &= \boldsymbol{0}\\ \boldsymbol{I} - \frac{1}{2}\boldsymbol{I} - \frac{1}{2}\boldsymbol{I}(\boldsymbol{T} + \boldsymbol{T}') &= \boldsymbol{0}\\ \boldsymbol{T} + \boldsymbol{T}' &= \boldsymbol{I}_r.\end{aligned}$$

Tällöin yhtälö (7.3) sieventyy muotoon

$$AX(X'AX)^{-1} - \frac{1}{2}A^2X(X'A^2X)^{-1} - \frac{1}{2}X = \mathbf{0}. \tag{7.5}$$

Nyt kertomalla yhtälö (7.5) vasemmalta matriisilla $X'A^{-1}$ saadaan että

$$\begin{aligned}
X'A^{-1}AX(X'AX)^{-1} - \frac{1}{2}X'A^{-1}A^2X(X'A^2X)^{-1} - \frac{1}{2}X'A^{-1}X &= \mathbf{0}\\
(X'AX)^{-1} - \frac{1}{2}X'AX(X'A^2X)^{-1} - \frac{1}{2}X'A^{-1}X &= \mathbf{0}\\
(X'AX)^{-1} - \frac{1}{2}X'A^{-1}X &= \frac{1}{2}X'AX(X'A^2X)^{-1}. \qquad (7.6)
\end{aligned}$$

Yhtälön (7.6) vasen puoli on kahden symmetrisen matriisin summa, joten koko vasen puoli on symmetrinen. Tällöin myös oikea puoli on symmetrinen, ja silloin matriisit $X'AX$ ja $(X'A^2X)^{-1}$ kommutoivat. Lauseen L.5 perusteella on olemassa molemmat matriisit $X'AX$ ja $(X'A^2X)^{-1}$ diagonalisoiva ortogonaalinen matriisi P, eli

$$P'(X'A^2X)^{-1}P = D_1 \tag{7.7}$$

$$P'(X'AX)P = D_2, \tag{7.8}$$

missä D_1 ja D_2 ovat diagonaalimatriiseja. Kertomalla yhtälöt (7.7) ja (7.8) vasemmalta matriisilla P ja oikealta P' saadaan

$$(X'A^2X)^{-1} = PD_1P' \tag{7.9}$$

$$(X'AX) = PD_2P', \tag{7.10}$$

Sijoittamalla yhtälöt (7.9) ja (7.10) yhtälöön (7.6) saadaan

$$\begin{aligned}
(PD_2P')^{-1} - \frac{1}{2}X'A^{-1}X &= \frac{1}{2}PD_2P'PD_1P'\\
P'^{-1}D_2^{-1}P^{-1} - \frac{1}{2}X'A^{-1}X &= \frac{1}{2}PD_2D_1P'. \qquad (7.11)
\end{aligned}$$

Koska matriisi P on ortogonaalinen $P'P = PP' = I_r$, $P^{-1} = P'$ ja koska $P'^{-1} = (P^{-1})'$, yhtälö (7.11) voidaan kirjoittaa muodossa

$$\begin{aligned}
PD_2^{-1}P' - \frac{1}{2}X'A^{-1}X &= \frac{1}{2}PD_2D_1P'\\
\frac{1}{2}X'A^{-1}X &= PD_2^{-1}P' - \frac{1}{2}PD_2D_1P'\\
X'A^{-1}X &= 2P(D_2^{-1} - \frac{1}{2}D_2D_1)P'. \qquad (7.12)
\end{aligned}$$

Nyt jos yhtälö (7.12) kerrotaan vasemmalta matriisilla P' ja oikealta matriisilla P, saadaan että

$$P'X'A^{-1}XP = 2(D_2^{-1} - \frac{1}{2}D_2D_1). \tag{7.13}$$

Yhtälön (7.13) oikea puoli on diagonaalimatriisi, sillä diagonaalimatriisien käänteismatriisi ja diagonaalimatriisien tulo ovat diagonaalimatriiseja. Tällöin siis sama matriisi $\boldsymbol{P}$ joka diagonalisoi matriisit $\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X}$ ja $(\boldsymbol{X}'\boldsymbol{A}^2\boldsymbol{X})^{-1}$, diagonalisoi myös matriisin $\boldsymbol{X}'\boldsymbol{A}^{-1}\boldsymbol{X}$.

Olkoon $\boldsymbol{X}^*$ matriisi joka minimoi funktion $h(\boldsymbol{X})$, tällöin myös $\boldsymbol{X} = \boldsymbol{X}^*\boldsymbol{Q}$ minimoi yhtälön, kun $\boldsymbol{Q}$ on ortogonaalinen matriisi. Tämä nähdää helposti sijoittamalla $\boldsymbol{X} = \boldsymbol{X}^*\boldsymbol{Q}$ yhtälöön (7.1). Lisäksi tiedetään että

$$\boldsymbol{X}'\boldsymbol{X} = \boldsymbol{Q}'(\boldsymbol{X}^*)'\boldsymbol{X}^*\boldsymbol{Q} = \boldsymbol{Q}'\boldsymbol{Q} = \boldsymbol{I}_r.$$

Koska minimiarvo saavutetaan millä tahansa ortogonaalisella matriisilla $\boldsymbol{Q}$, niin $\boldsymbol{Q}$:ksi voidaan valita matriisit $\boldsymbol{X}'\boldsymbol{A}^2\boldsymbol{X}$ ja $\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X}$ ortogonalisoiva matriisi $\boldsymbol{P}$. Tällöin yleisyyttä menettämättä voidaan olettaa, että matriisit $\boldsymbol{X}'\boldsymbol{A}^2\boldsymbol{X}$ ja $\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X}$ ovat diagonaalisia. Koska matriisin $\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X}$ oletetaan olevan diagonaalinen, myös $(\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X})^{-1}$ on diagonaalinen.

Koska diagonaalimatriisin determinantti on diagonaalialkioiden tulo, minimoitava funktio $h(\boldsymbol{X})$ voidaan kirjoittaa muodossa

$$h(\boldsymbol{X}) = \frac{\prod_{i=1}^{r} \boldsymbol{x}_i'\boldsymbol{A}\boldsymbol{x}_i}{\sqrt{\prod_{i=1}^{r} \boldsymbol{x}_i'\boldsymbol{A}^2\boldsymbol{x}_i}}. \tag{7.14}$$

Vastaavasti jos ehto (7.5) kirjoitetaan muodossa, jossa otetaan huomioon matriisien diagonaalisuus, saadaan

$$\begin{aligned}
&(\boldsymbol{A}\boldsymbol{x}_1 : \cdots : \boldsymbol{A}\boldsymbol{x}_r)\begin{pmatrix} \frac{1}{\boldsymbol{x}_1'\boldsymbol{A}\boldsymbol{x}_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{\boldsymbol{x}_2'\boldsymbol{A}\boldsymbol{x}_2} & \cdots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ 0 & \cdots & 0 & \frac{1}{\boldsymbol{x}_r'\boldsymbol{A}\boldsymbol{x}_r} \end{pmatrix} \\
&- \frac{1}{2}(\boldsymbol{A}^2\boldsymbol{x}_1 : \cdots : \boldsymbol{A}^2\boldsymbol{x}_r)\begin{pmatrix} \frac{1}{\boldsymbol{x}_1'\boldsymbol{A}^2\boldsymbol{x}_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{\boldsymbol{x}_2'\boldsymbol{A}^2\boldsymbol{x}_2} & \cdots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ 0 & \cdots & 0 & \frac{1}{\boldsymbol{x}_r'\boldsymbol{A}^2\boldsymbol{x}_r} \end{pmatrix} \\
&- \frac{1}{2}(\boldsymbol{x}_1 : \cdots : \boldsymbol{x}_r) = \boldsymbol{0}.
\end{aligned} \tag{7.15}$$

Tällöin nähdään että ehto (7.5) voidaan kirjoittaa myös ositettuna, muodossa

$$\frac{\boldsymbol{A}\boldsymbol{x}_i}{\boldsymbol{x}_i'\boldsymbol{A}\boldsymbol{x}_i} - \frac{1}{2}\frac{\boldsymbol{A}^2\boldsymbol{x}_i}{\boldsymbol{x}_i'\boldsymbol{A}^2\boldsymbol{x}_i} - \frac{1}{2}\boldsymbol{x}_i = \boldsymbol{0}, \qquad i = 1, 2, \cdots, r. \tag{7.16}$$

Yhtälön (7.16) perusteella nähdään, että vektori $\boldsymbol{A}^2\boldsymbol{x}_i$ on vektoreiden $\boldsymbol{x}_i$ ja $\boldsymbol{A}\boldsymbol{x}_i$ lineaarikombinaatio. Tällöin

$$\mathcal{C}[\boldsymbol{A}(\boldsymbol{x}_i : \boldsymbol{A}\boldsymbol{x}_i)] = \mathcal{C}(\boldsymbol{A}\boldsymbol{x}_i : \boldsymbol{A}^2\boldsymbol{x}_i) \subset \mathcal{C}(\boldsymbol{x}_i : \boldsymbol{A}\boldsymbol{x}_i),$$

mikä tarkoittaa, että vektoreiden $\boldsymbol{x}_i$ ja $\boldsymbol{A}\boldsymbol{x}_i$ virittämä avaruus on suljettu $\boldsymbol{A}$:n suhteen. Tällöin $\boldsymbol{x}_i$ ja $\boldsymbol{A}\boldsymbol{x}_i$ virittämä avaruuden virittämiseen riittää maksimissaan kaksi matriisin $\boldsymbol{A}$ ominaisvektoria, ja nämä ominaisvektorit löydetään ratkaisemalla yhtälö

$$\frac{u}{\boldsymbol{x}_i'\boldsymbol{A}\boldsymbol{x}_i} - \frac{1}{2}\frac{u^2}{\boldsymbol{x}_i'\boldsymbol{A}^2\boldsymbol{x}_i} - \frac{1}{2} = 0$$

tai

$$u^2 - 2\frac{\boldsymbol{x}_i'\boldsymbol{A}^2\boldsymbol{x}_i}{\boldsymbol{x}_i'\boldsymbol{A}\boldsymbol{x}_i}u + \boldsymbol{x}_i'\boldsymbol{A}^2\boldsymbol{x}_i = 0. \tag{7.17}$$

Oletetaan nyt, että nämä ominaisarvot ovat λ_i ja λ_i^*. Tällöin yhtälöstä (7.17) nähdään, että

$$\lambda_i + \lambda_i^* = 2\frac{\boldsymbol{x}_i'\boldsymbol{A}^2\boldsymbol{x}_i}{\boldsymbol{x}_i'\boldsymbol{A}\boldsymbol{x}_i} \quad \text{ja} \quad \lambda_i\lambda_i^* = \boldsymbol{x}_i'\boldsymbol{A}^2\boldsymbol{x}_i.$$

Tällöin siis

$$\boldsymbol{x}_i'\boldsymbol{A}\boldsymbol{x}_i = 2\frac{\lambda_i\lambda_i^*}{\lambda_i + \lambda_i^*} \quad \text{ja} \quad \lambda_i\lambda_i^* = \boldsymbol{x}_i'\boldsymbol{A}^2\boldsymbol{x}_i.$$

Nyt siis funktio $h(\boldsymbol{X})$ voidaan kirjoittaa muodossa

$$\begin{aligned} h(\boldsymbol{X}) &= \prod_{i=1}^{r} \frac{\boldsymbol{x}_i'\boldsymbol{A}\boldsymbol{x}_i}{\sqrt{\boldsymbol{x}_i'\boldsymbol{A}^2\boldsymbol{x}_i}} = \prod_{i=1}^{r} \frac{2\frac{\lambda_i\lambda_i^*}{\lambda_i+\lambda_i^*}}{\sqrt{\lambda_i\lambda_i^*}} \\ &= \prod_{i=1}^{r} \frac{2\sqrt{\lambda_i\lambda_i^*}}{\lambda_i + \lambda_i^*}. \end{aligned}$$

Bloomfield ja Watson (1975) ja Knott (1975) osoittivat kombinatorisesti, että minimi saavutetaan valinnalla $\lambda_i^* = \lambda_{p-i+1}$. Sama seuraa myös suoraan apulauseesta (L.2). Tällöin

$$h(\boldsymbol{X}) = \prod_{i=1}^{r} \frac{2\sqrt{\lambda_i\lambda_{p-i+1}}}{\lambda_i + \lambda_{p-i+1}}.$$

Tällöin $\boldsymbol{x}_i$ voidaan esittää ominaisarvoja λ_i ja λ_{p-i+1} vastaavien kantavektoreiden summana

$$\boldsymbol{x}_i = \frac{\sqrt{\lambda_{p-i+1}}}{\sqrt{\lambda_i + \lambda_{p-i+1}}}\boldsymbol{e}_i \pm \frac{\sqrt{\lambda_i}}{\sqrt{\lambda_i + \lambda_{p-i+1}}}\boldsymbol{e}_{p-i+1}, \quad i = 1, \ldots r, \tag{7.18}$$

missä $\boldsymbol{e}_k$ on k. kantavektori.

Tilannetta $\lambda_i = \lambda_{p-i+1}$ vastaava tekijä yhtälössä $h(\boldsymbol{x})$ on siis

$$\frac{\boldsymbol{x}_i'\boldsymbol{A}\boldsymbol{x}_i}{\sqrt{\boldsymbol{x}_i'\boldsymbol{A}^2\boldsymbol{x}_i}} = \frac{2\sqrt{\lambda_i\lambda_{p-i+1}}}{\lambda_i + \lambda_{p-i+1}} = \frac{2\lambda_i}{\lambda_i + \lambda_i} = 1,$$

eli $(\boldsymbol{x}_i'\boldsymbol{A}\boldsymbol{x}_i)^2 = \boldsymbol{x}_i'\boldsymbol{A}^2\boldsymbol{x}_i$. Matriisin $\boldsymbol{A}$ ominaisarvohajotelman avulla kirjoitettuna nähdään

$$\begin{aligned}
(\boldsymbol{x}_i'\boldsymbol{A}\boldsymbol{x}_i)^2 &= \boldsymbol{x}_i'\boldsymbol{A}^2\boldsymbol{x}_i \\
(\boldsymbol{x}_i'\boldsymbol{T}\boldsymbol{\Lambda}\boldsymbol{T}'\boldsymbol{x}_i)^2 &= \boldsymbol{x}_i'(\boldsymbol{T}\boldsymbol{\Lambda}\boldsymbol{T}')^2\boldsymbol{x}_i \\
\boldsymbol{x}_i'\boldsymbol{T}\boldsymbol{\Lambda}\boldsymbol{T}'\boldsymbol{x}_i\boldsymbol{x}_i'\boldsymbol{T}\boldsymbol{\Lambda}\boldsymbol{T}'\boldsymbol{x}_i &= \boldsymbol{x}_i'\boldsymbol{T}\boldsymbol{\Lambda}\boldsymbol{T}'\boldsymbol{T}\boldsymbol{\Lambda}\boldsymbol{T}'\boldsymbol{x}_i \\
\boldsymbol{x}_i'\boldsymbol{T}\boldsymbol{\Lambda}\boldsymbol{T}'\boldsymbol{x}_i\boldsymbol{x}_i'\boldsymbol{T}\boldsymbol{\Lambda}\boldsymbol{T}'\boldsymbol{x}_i &= \boldsymbol{x}_i'\boldsymbol{T}\boldsymbol{\Lambda}^2\boldsymbol{T}'\boldsymbol{x}_i,
\end{aligned}$$

jolloin $\boldsymbol{T}'\boldsymbol{x}_i\boldsymbol{x}_i'\boldsymbol{T} = \boldsymbol{I}$. Näin tapahtuu ainakin silloin kun $\boldsymbol{x}_i$ on matriisin $\boldsymbol{A}$ jokin ominaisvektori.

Nyt siis funktion $h(\boldsymbol{X})$ minimi ehdolla $\boldsymbol{X}'\boldsymbol{X} = \boldsymbol{I}$ saavutetaan, kun

$$\min_{\boldsymbol{X}'\boldsymbol{X}=\boldsymbol{I}_r} h(\boldsymbol{X}) = \prod_{i=1}^{r} \frac{2\sqrt{\lambda_i\lambda_{p-i+1}}}{\lambda_i + \lambda_{p-i+1}} = \prod_{i=1}^{r^*} \frac{2\sqrt{\lambda_i\lambda_{p-i+1}}}{\lambda_i + \lambda_{p-i+1}}, \tag{7.19}$$

missä $r^* \leq r$ tarkoittaa termien lukumäärää joilla yhtälöllä (7.17) on erilliset juuret. Nyt yhtälön (7.8) perusteella yksi ratkaisu antiominaismatriisille $\boldsymbol{X}$ on siis

$$\boldsymbol{X} = (\boldsymbol{x}_1 : \ldots : \boldsymbol{x}_{r^*} : \boldsymbol{q}_{r^*+1} : \ldots : \boldsymbol{q}_r),$$

missä $\boldsymbol{q}_i$ on i:nnettä ominaisarvoa vastaava ominaisvektori. Matriisissa $\boldsymbol{X}$ voidaan $\boldsymbol{q}_i$ valita miten tahansa, kunhan vektoria $\boldsymbol{q}_i$ ei ole käytetty vektoreiden $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{r^*}$ muodostamisessa.

Alussa teimme oletuksen $1 \leq r \leq [p/2]$. Todistus voidaan myös esittää siten, että siihen sisältyy tilanne $[p/2] \leq r \leq p$ (esim. Knott 1975), mutta silloin yhtälö (7.19) saa muodon

$$\min_{\boldsymbol{X}'\boldsymbol{X}=\boldsymbol{I}_r} h(\boldsymbol{X}) = \prod_{i=1}^{\min(r,p-r)} \frac{2\sqrt{\lambda_i\lambda_{p-i+1}}}{\lambda_i + \lambda_{p-i+1}}.$$

Huomautus. On syytä huomata, että yleistetty antiominaisarvo saa yksikertaisen muodon antiominaisarvojen tulona. Tällöin, jos $r = 1$, yleistetty antiominaisarvo on sama kuin pienin antiominaisarvo, eli $\mu_{[1]} = \mu_1$.

Nyt edellä esitetty tulos voidaan esittää yleistettyä antiominaisarvoa koskevana lauseena.

Lause 7.1. *Olkoon matriisi $\boldsymbol{A}_{p\times p}$ reaalinen symmetrinen ja positiivisesti definiitti matriisi, jonka ominaisarvot ovat $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p > 0$ ja niitä vastaavat ominaisvektorit ovat $\boldsymbol{q}_1, \boldsymbol{q}_1, \ldots, \boldsymbol{q}_p$. Tällöin sen r:s yleistetty antiominaisarvo $\mu_{[r]}$ on*

$$\mu_{[r]} = \min_{\boldsymbol{X}} \frac{|\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X}|}{\sqrt{|\boldsymbol{X}'\boldsymbol{A}^2\boldsymbol{X}||\boldsymbol{X}'\boldsymbol{X}|}} = \prod_{i=1}^{\min(r,p-r)} \frac{2\sqrt{\lambda_i\lambda_{p-i+1}}}{\lambda_i + \lambda_{p-i+1}},$$

missä $\boldsymbol{X}_{p\times r}$ on reaalinen matriisi astetta r. Ratkaisun tuottava antiominaismatriisi $\boldsymbol{X}^* = (\boldsymbol{x}_1 : \boldsymbol{x}_2 : \cdots : \boldsymbol{x}_{r^*} : \boldsymbol{p}_{r^*} : \cdots : \boldsymbol{p}_r)$, missä $\boldsymbol{x}_i = \boldsymbol{q}_i \pm \boldsymbol{q}_{p-i+1}$ on matriisin $\boldsymbol{A}$ i:nettä antiominaisarvoa vastaava antiominaisvektori ja $\boldsymbol{p}_i$ voi olla mikä tahansa matriisin $\boldsymbol{A}$ ominaisvektori jota ei ole käytetty antiominaisvektoreiden $\boldsymbol{x}_i$ muodostamisessa.

Esimerkki 7.1. Esimerkit 7.1 ja 7.2 ovat peräisin Khattreen (2002) artikkelista. Olkoon matriisi $\boldsymbol{A} = \mathrm{diag}(1,2,3,4,5)$, ja $r = 3$. Etsitään siis kolmatta yleistettyä antiominaisarvoa ja sitä vastaavaa antiominaismatriisia.

$$\mu_{[3]} = \min_{\boldsymbol{X'X}=\boldsymbol{I_r}} h(\boldsymbol{X}) = \left(\frac{2\sqrt{5}}{6}\right)\left(\frac{2\sqrt{8}}{6}\right)(1) = \frac{\sqrt{40}}{9}$$

ja tätä vastaava antiominaismatriisi $\boldsymbol{X}$ on siis

$$\boldsymbol{X} = \begin{pmatrix} \sqrt{\frac{5}{6}} & 0 & 0 \\ 0 & \sqrt{\frac{2}{3}} & 0 \\ 0 & 0 & \pm 1 \\ 0 & \pm\sqrt{\frac{1}{3}} & 0 \\ \pm\sqrt{\frac{1}{6}} & 0 & 0 \end{pmatrix}.$$

Edeltä nähdään että $\mu_{[2]} = \mu_{[3]}$, sillä matriisilla $\boldsymbol{A}$ on ainoastaan kaksi antiominaisarvoa. Tällöin antiominaismatriisiksi riittää ilmoittaa antiominaismatriisin $\boldsymbol{X}$ kaksi ensimmäistä saraketta.

Esimerkki 7.2. Tarkastellaan toisena esimerkkinä matriisia jolla on useampikertaisia ominaisarvoja. Olkoon matriisi $\boldsymbol{A} = \mathrm{diag}(1,1,2,2,3,3)$, ja $r = 3$. Etsitään siis kolmatta yleistettyä antiominaisarvoa ja sitä vastaavaa antiominaismatriisia. Tällöin

$$\mu_{[3]} = \left(\frac{2\sqrt{3}}{4}\right)\left(\frac{2\sqrt{3}}{4}\right)(1) = \frac{3}{4}$$

ja sitä vastaava antiominaismatriisiksi voidaan valita

$$\boldsymbol{X} = \begin{pmatrix} \frac{\sqrt{3}}{2} & 0 & 0 \\ 0 & \frac{\sqrt{3}}{2} & 0 \\ 0 & 0 & \frac{1}{\sqrt{2}} \\ 0 & 0 & \pm\frac{1}{\sqrt{2}} \\ 0 & \pm\frac{1}{2} & 0 \\ \pm\frac{1}{2} & 0 & 0 \end{pmatrix}$$

tai

$$\boldsymbol{X} = \begin{pmatrix} \frac{\sqrt{3}}{2} & 0 & 0 \\ 0 & \frac{\sqrt{3}}{2} & 0 \\ 0 & 0 & \frac{1}{\sqrt{2}} \\ 0 & 0 & \pm\frac{1}{\sqrt{2}} \\ \pm\frac{1}{2} & 0 & 0 \\ 0 & \pm\frac{1}{2} & 0 \end{pmatrix},$$

sillä matriisilla $\boldsymbol{A}$ on moninkertaisia ominaisarvoja.

7.2 Yleistetty antiominaisarvo ja tehokkuus

Vaikka yleistetyn antiominaisarvon termi on tullut käyttöön vasta Khattreen (2002) artikkelin myötä, niin aikaisemmin tilastotieteessä yleistettyyn antiominaisarvoon ovat törmänneet ainakin Bloomfield ja Watson (1975) ja Knott (1975) tarkastellessaan OLSE:n ja BLUE:n välistä tehokkuutta. Myöhemmin Bartmann ja Bloomfield (1981) päätyivät yleistettyihin antiominaisarvoihin tarkastellessaan korrelaatiota ja tehottomuutta.

Bloomfield, Watson (1975) ja Knott (1975) tarkastelivat tehokkuutta toisistaan riippumatta, ja julkaisivat lähes identtiset todistukset tehokkuuden alarajalle samassa Biometrikan julkaisussa 1975. Kummatkin tarkastelivat mallia $\{\boldsymbol{y},\, \boldsymbol{X\beta},\, \sigma^2\boldsymbol{V}\}$, ja olettivat että mallimatriisi $\boldsymbol{X}$ on täysiasteinen. Tehokkuuden Bloomfield, Watson ja Knott määrittelivät OLSE:n ja BLUE:n yleistettyjen varianssien osamääräksi. Yleistetty varianssi on peräisin Watsonin (1955) artikkelista ja lienee tehokkuustarkasteluissa yleisimmin käytetty varianssin yleistys kovarianssimatriisille. Vaikka yleistetty varianssi on luultavasti yleisin varianssin yleistys kovarianssimatriisille, niin muitakin on olemassa ja niitä voidaan käyttää yhtä hyvin tehokkuustarkastelussa. Esimerkiksi Liu (2000) vertailee useiden eri varianssin yleistysten perusteella laskettuja tehokkuuksia.

Tarkastellaan mallia $\{\boldsymbol{y},\, \boldsymbol{X\beta},\, \sigma^2\boldsymbol{V}\}$, jossa $\boldsymbol{X}_{n\times k}$ täysiasteinen havaintomatriisi. Merkitään pienimmän neliösumman estimaattia $\hat{\boldsymbol{\beta}}$:lla ja parasta lineaarista estimaattia $\tilde{\boldsymbol{\beta}}$:lla. Tällöin

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}, \qquad \tilde{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{y}$$

ja estimaattorien kovarianssimatriisit

$$\operatorname{cov}(\hat{\boldsymbol{\beta}}) = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{V}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}, \qquad \operatorname{cov}(\tilde{\boldsymbol{\beta}}) = (\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}.$$

Määritellään tehokkuus ϕ yleistettyjen varianssien suhteeksi

$$\begin{aligned} \phi = \frac{|\operatorname{var}(\tilde{\boldsymbol{\beta}})|}{|\operatorname{var}(\hat{\boldsymbol{\beta}})|} &= \frac{|(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}|}{|(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{V}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}|} \\ &= \frac{|\boldsymbol{X}'\boldsymbol{X}|^2}{|\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X}||\boldsymbol{X}'\boldsymbol{V}\boldsymbol{X}|}. \end{aligned} \tag{7.20}$$

Olkoon taas $\boldsymbol{V}^{\frac{1}{2}}$ matriisin $\boldsymbol{V}$ neliöjuuri. Tehdään edelliseen yhtälöön (7.20) sijoitus $\boldsymbol{X} = \boldsymbol{V}^{\frac{1}{2}}\boldsymbol{Y}$. Tällöin tehokkuus voidaan kirjoittaa muodossa

$$\phi = \frac{|\boldsymbol{Y}'\boldsymbol{V}^{\frac{1}{2}}\boldsymbol{V}^{\frac{1}{2}}\boldsymbol{Y}|^2}{|\boldsymbol{Y}'\boldsymbol{V}^{\frac{1}{2}}\boldsymbol{Y}^2\boldsymbol{V}^{\frac{1}{2}}\boldsymbol{Y}||\boldsymbol{Y}\boldsymbol{V}^{\frac{1}{2}}\boldsymbol{V}^{-1}\boldsymbol{V}^{\frac{1}{2}}\boldsymbol{Y}|} = \frac{|\boldsymbol{Y}'\boldsymbol{V}\boldsymbol{Y}|^2}{|\boldsymbol{Y}'\boldsymbol{V}^2\boldsymbol{Y}||\boldsymbol{Y}'\boldsymbol{Y}|} = \mu_{[r]}^2,$$

mikä on juuri edellä esitetyn yleistetyn antiominaisarvoyhtälön neliö.

Nyt voidaan siis todeta, että tällä tavalla määritelty OLSE:n tehokkuuden minimi on pienin yleistetyn antiominaisarvon neliö, ja se saavutetaan sitä vastaavalla antiominaismatriisilla.

Eaton (1976) ja Venables (1976) osoittivat että kanonisille korrelaatioille pätee, että $\rho^2 \leq 1-\mu_1^2$. Myöhemmin Bartman sekä Bloomfield (1981) yleistivät, että

$$\prod_{i=1}^{r}(1-\rho_i^2) \geq \mu_{[r]}^2.$$

Kuten Bloomfield, Watson (1975) ja Knott (1975), myös Bartmann ja Bloomfield tarkastelivat mallia $\{\boldsymbol{y},\ \boldsymbol{X\beta},\ \sigma^2\boldsymbol{V}\}$, jossa OLSE:n ja BLUE:n kovarianssimatriisit voidaan, kuten edellä, kirjoittaa muodossa:

$$\operatorname{cov}(\hat{\boldsymbol{\beta}}) = \boldsymbol{X}'\boldsymbol{V}\boldsymbol{X}, \qquad \operatorname{cov}(\tilde{\boldsymbol{\beta}}) = (\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}.$$

Tarkastellaan pienimmän neliösumman sovitteen ja jäännöstermin välistä kanonista korrelaatiota

$$\mathrm{cc}^2(\boldsymbol{Hy}, \boldsymbol{My}).$$

Tällöin

$$(\boldsymbol{H} : \boldsymbol{M})'(\boldsymbol{H} : \boldsymbol{M}) = \boldsymbol{I}_n$$

ja $\boldsymbol{Hy}$:n ja $\boldsymbol{My}$:n yhteisjakauman kovarianssimatriisi on

$$\operatorname{cov}\begin{pmatrix}\boldsymbol{Hy}\\ \boldsymbol{My}\end{pmatrix} = \begin{pmatrix}\boldsymbol{HVH} & \boldsymbol{HVM}\\ \boldsymbol{MVH} & \boldsymbol{MVM}\end{pmatrix}.$$

BLUE:n varianssi voidaan ositettujen matriisien laskusääntöjen perusteella kirjoittaa muodossa

$$\operatorname{cov}(\tilde{\boldsymbol{\beta}}) = (\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1} = \boldsymbol{X}'\boldsymbol{V}\boldsymbol{X} - \boldsymbol{X}'\boldsymbol{V}\boldsymbol{M}(\boldsymbol{M}'\boldsymbol{V}\boldsymbol{M})^{-}\boldsymbol{M}'\boldsymbol{V}\boldsymbol{X}. \tag{7.21}$$

Kerrotaan yhtälö (7.21) vasemmalta ja oikealta matriisilla $(\boldsymbol{X}'\boldsymbol{V}\boldsymbol{X})^{-\frac{1}{2}}$, ja tarkastellaan determinanttia:

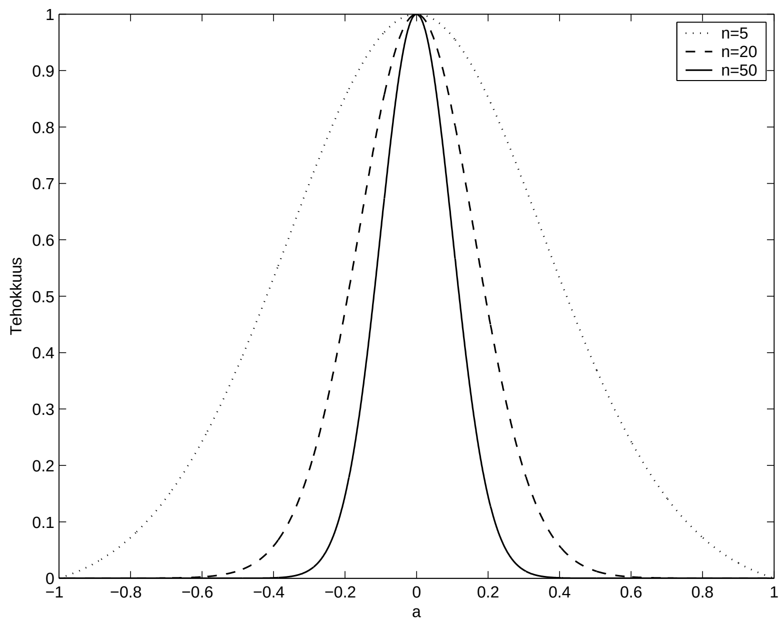
$$\begin{aligned}
&|\boldsymbol{I} - (\boldsymbol{X}'\boldsymbol{V}\boldsymbol{X})^{-\frac{1}{2}}\boldsymbol{X}'\boldsymbol{V}\boldsymbol{M}(\boldsymbol{M}'\boldsymbol{V}\boldsymbol{M})^{-}\boldsymbol{M}'\boldsymbol{V}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{V}\boldsymbol{X})^{-\frac{1}{2}}|\\
&= |(\boldsymbol{X}'\boldsymbol{V}\boldsymbol{X})^{-\frac{1}{2}}(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}(\boldsymbol{X}'\boldsymbol{V}\boldsymbol{X})^{-\frac{1}{2}}|\\
&= \frac{1}{|\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X}||\boldsymbol{X}'\boldsymbol{V}\boldsymbol{X}|}.
\end{aligned} \tag{7.22}$$

Koska oletettiin että $\boldsymbol{X}'\boldsymbol{X} = \boldsymbol{I}$, niin yhtälöstä (7.22) saadaan, että OLSE:n tehokkuuden minimi ϕ_{min} on yleistetyn antiominaisarvon neliö. Tällöin siis tehokkuuden minimin ϕ_{min} voidaan kirjoittaa muodossa:

$$\phi_{min} = |\boldsymbol{I} - (\boldsymbol{X}'\boldsymbol{V}\boldsymbol{X})^{-\frac{1}{2}}\boldsymbol{X}'\boldsymbol{V}\boldsymbol{M}(\boldsymbol{M}'\boldsymbol{V}\boldsymbol{M})^{-}\boldsymbol{M}'\boldsymbol{V}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{V}\boldsymbol{X})^{-\frac{1}{2}}| = \mu_{[r]}^2.$$

Esimerkki 7.3. Tarkastellaan vastaavaa tilannetta kuin esimerkissä 3.1, mutta tällä kertaa tarkastellaan mallia $\{\boldsymbol{y}, \boldsymbol{X\beta}, \boldsymbol{V}\}$, jossa $\boldsymbol{X}$ on $n \times k$-mallimatriisi, ja tehokkuus määritellään Bloomfieldin, Watsonin (1975) ja Knottin (1975) tyyliin kovarianssimatriisien determinanttina.

Kovarianssimatriisi on kuten esimerkissä 3.1. Piirretään minimaalinen tehokkuus $\phi_{min} = \mu_{[p]}^2$ pystyakselille ja vaaka-akselille parametri a, kolmella eri havaintojen määrällä eli n:nnän arvolla.



Kuvio 7.1. Kuvion pystyakselilla on OLSE:n tehokkuuden minimi ϕ_{min} AR(1)-kovarianssirakenteen tapauksessa, kun vaaka-akselilla on autokorrelaatio viiveellä 1.

Kuviosta 7.1 nähdään, että tehokkuuden minimiin vaikuttaa voimakkaasti autokorrelaatio a, sekä havaintojen lukumäärä. Suuremmilla havaintomäärillä autokorrelaation vaikutus tehokkuuden minimiin on voimakkaampaa kuin pienillä havaintomäärillä. Tämä näkyy kuvion voimakkaana huipukkuutena suurilla havaintomäärillä.

Esimerkki 7.4. Edellä esimerkissä 3.3 tarkasteltiin tehokkuutta kun mallimatriisina on vektori $\boldsymbol{x}$. Kuten edellä on jo todettu, tämä vastaa yleistettyä antiominaisarvoa $\mu_{[1]}$. Tasakovarianssimatriisin $\boldsymbol{V}$ ominaisarvot ovat kuitenkin tunnettuja suljetussa muodossa, eli

$$\{\mathrm{ch}(\boldsymbol{V})\} = \begin{cases} 1+(n-1)\rho & \text{kertalukuna 1} \\ 1-\rho & \text{kertalukuna n-1} \end{cases}. \tag{7.23}$$

Tämän perusteella tiedetään, että ensimäinen antiominaisarvo μ_1 on erisuuri kuin yksi ja muut antiominaisarvot ovat ykkösiä. Tällöin myös kaikki yleistetyt antiominaisarvot $\mu_{[i]} = \mu_1$, kun $i = 1, \ldots, r$.

Nyt sijoittamalla tasakorrelaatiomatriisin ominaisarvot (7.23), saadaan Watsonin tehokkuus ϕ muotoon

$$\phi = \frac{4\lambda_1\lambda_n}{(\lambda_1+\lambda_n)^2} = \frac{4(1+(n-2)c-(n-1)c^2)}{(2+nc-2c)^2},$$

missä c on korrelaatio. Kuvio missä tehokkuuden minimi on piirretty korrelaation c funktiona, on siis sama kuin esimerkin 3.3 kuvio 3.4.

Tästä tehokkuuden kannalta heikoimmasta mahdollisesta tapauksesta puuttuu jälleen kerran vakiotermiä vastaava vektori $\mathbf{1} = (1, 1, \ldots, 1)'$ ja tämän vuoksi OLSE ei ole BLUE tasakorrelaatiorakenteen tapauksessa.

Edellä on tarkasteltu Watsonin tehokkuutta lineaarisessa mallissa $\{\boldsymbol{y}, \boldsymbol{X\beta}, \boldsymbol{V}\}$, kun kovarianssimatriisi $\boldsymbol{V}$ on tunnettu. Kuten edellä ollaan todettu, Watsonin tehokkuuden minimi on yleistetyn antiominaisarvon neliö ja se saavutetaan yleistetyllä antiominaismatriisilla. Yleistetty antiominaismatriisi ei kuitenkaan sisällä ykkösvektoria $\mathbf{1} = (1, 1, \ldots, 1)'$, joten tehokkuuden kannalta huonoin havaintomatriisi ei sisällä ollenkaan vakiotermiä ja kulkee siis origon kautta.

Tarkastellaan nyt mallia

$$\mathcal{M}_{12} = \{\boldsymbol{y}, \mathbf{1}\beta_0 + \boldsymbol{X\beta}_1, \boldsymbol{V}\}$$

ja sen "redusoitua" versiota

$$\mathcal{M}_2 = \{\boldsymbol{Cy}, \boldsymbol{CX\beta}_1, \boldsymbol{CVC}\}.$$

missä $\boldsymbol{C} = \boldsymbol{I} - \boldsymbol{J} = \boldsymbol{I} - \frac{1}{n}\mathbf{11}'$ on siis keskistäjämatriisi. Malleissa $\mathcal{M}_{12}$ ja $\mathcal{M}_2$ vektoreiden $\boldsymbol{\beta}_1$ BLUE:t ovat samat, kuten Zyskind ja Martin (1969) ovat osoittaneet. Tällöin myös $\boldsymbol{\beta}_1$ vektoreiden tehokkuudet ovat samat malleissa $\mathcal{M}_{12}$ ja $\mathcal{M}_2$.

Mallista $\mathcal{M}_2$ voidaan helposti laskea $\boldsymbol{\beta}_1$ vektorin tehokkuus tarkastelemalla kovarianssimatriisin $\boldsymbol{CVC}$ yleistettyä antiominaisarvoa. Ongelmaksi kuitenkin osoittautuu, että kovarianssimatriisi $\boldsymbol{CVC}$ on singulaarinen ja tämän vuoksi edellä määriteltyä yleistettyä antiominaisarvoa ei ole olemassa.

Chu, Isotalo, Puntanen ja Styan (2004, 2005) tarkastelevat osavektoreiden tehokkuuksia lineaarisessa mallissa, ja tekevät Watsonin tehokkuuteen seuraavan yleistyksen.

Lause 7.2. *Olkoon matriisi $\boldsymbol{A}_{p\times p}$ reaalinen symmetrinen ja ei-negatiivisesti definiitti, jonka aste on v ja $r+1 \leq v \leq p$. Matriisi $\boldsymbol{A}^+$ on matriisin $\boldsymbol{A}$ Moore-Penrose inverssi ja matriisi $\boldsymbol{X}_{p\times r}$ on astetta r, siten että $1 \leq r \leq p$. Tällöin*

$$\begin{aligned}\min_{\boldsymbol{X}_*} \frac{|\boldsymbol{X}_*'\boldsymbol{A}\boldsymbol{X}_*|}{\sqrt{|\boldsymbol{X}_*'\boldsymbol{A}^2\boldsymbol{X}_*|\cdot|\boldsymbol{X}_*'\boldsymbol{X}_*|}} &= \min_{\boldsymbol{X}} \frac{|\boldsymbol{X}'\boldsymbol{X}|}{\sqrt{|\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X}|\cdot|\boldsymbol{X}'\boldsymbol{A}^+\boldsymbol{X}|}} \\ &= \min_{\boldsymbol{X}'\boldsymbol{X}=\boldsymbol{I}} \frac{1}{\sqrt{|\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X}|\cdot|\boldsymbol{X}'\boldsymbol{A}^+\boldsymbol{X}|}} \\ &= \prod_{i=1}^{\min(r,v-r)} \frac{2\sqrt{\lambda_i\lambda_{v-i+1}}}{\lambda_i+\lambda_{v-i+1}}.\end{aligned}$$

Todistus. Matriiseilla $\boldsymbol{A}$ ja $\boldsymbol{A}^+$ on olemassa ominaisarvohajotelmat

$$\boldsymbol{A} = \boldsymbol{T}_1\boldsymbol{\Lambda}_1\boldsymbol{T}_1' \qquad \boldsymbol{A}^+ = \boldsymbol{T}_1\boldsymbol{\Lambda}_1^{-1}\boldsymbol{T}_1', \tag{7.24}$$

missä $v\times v$ -matriisin $\boldsymbol{\Lambda}_1$ diagonaalilla ovat $\boldsymbol{A}$:n positiiviset ominaisarvot, ja matriisissa $\boldsymbol{T}_1$ niitä vastaavat ominaisvektorit. Kun oletetaan, että $\boldsymbol{X}$:n sarakkeet ovat ortogonaaliset ja että $\mathcal{C}(\boldsymbol{X}) \subset \mathcal{C}(\boldsymbol{A})$, niin $\boldsymbol{X}$ voidaan kirjoittaa muodossa $\boldsymbol{X} = \boldsymbol{T}_1\boldsymbol{P}$, missä $\boldsymbol{P}_{(v\times r)}$ toteuttaa ehdon $\boldsymbol{P}'\boldsymbol{P} = \boldsymbol{I}_r$. Tällöin sijoittamalla yhtälöt (7.24), yleistetty antiominaisarvoyhtälö ei-negatiivisesti definiiteille matriiseille voidaan kirjoittaa muodossa

$$\begin{aligned}\min_{\boldsymbol{X}_*'\boldsymbol{X}_*=I} \frac{|\boldsymbol{X}_*'\boldsymbol{A}\boldsymbol{X}_*|}{\sqrt{|\boldsymbol{X}_*'\boldsymbol{A}^2\boldsymbol{X}_*|\cdot|\boldsymbol{X}_*'\boldsymbol{X}_*|}} &= \min_{\boldsymbol{X}'\boldsymbol{X}=\boldsymbol{I}} \frac{1}{\sqrt{|\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X}|\cdot|\boldsymbol{X}'\boldsymbol{A}^+\boldsymbol{X}|}} \\ &= \min_{\boldsymbol{P}'\boldsymbol{P}=\boldsymbol{I}} \frac{1}{\sqrt{|\boldsymbol{P}'\boldsymbol{\Lambda}_1\boldsymbol{P}|\cdot|\boldsymbol{P}'\boldsymbol{\Lambda}_1^{-1}\boldsymbol{P}|}} \\ &= \prod_{i=1}^{\min(r,v-r)} \frac{2\sqrt{\lambda_i\lambda_{v-i+1}}}{\lambda_i+\lambda_{v-i+1}}.\end{aligned}$$

□

Lauseen 7.2 perusteella yleistetty antiominaisarvo voidaan yleistää koskemaan myön reaalisia symmetrisiä ja ei-negatiivisesti definiittejä matriiseja. Tällöin antiominaisarvot lasketaan kuten positiivisesti definiittien matriisien tapauksessa, mutta ottamalla mukaan ainoastaan nollasta poikkeavat ominaisarvot. Jatkossa myös ei-negatiivisesti definiitin matriisin ei-singulaarisesta osasta laskettuja yleistettyjä antiominaisarvoja kutsutaan yleistetyksi antiominaisarvoksi.

Esimerkki 7.5. Tarkastellaan $\boldsymbol{\beta}_1$ vektorin tehokkuutta mallissa $\mathcal{M}_{12}$, kun kovarianssimatriisilla $\boldsymbol{V}$ on tasakorrelaatiorakenne. Keskistäjämatriisi voidaan kirjoittaa muodossa

$$\boldsymbol{C} = \boldsymbol{I} - \boldsymbol{J} = \boldsymbol{I} - \frac{1}{n}\mathbf{1}\mathbf{1}'$$

ja tasakorrelaatiorakenteen kovarianssimatriisi muodossa

$$\boldsymbol{V} = (1-\rho)\boldsymbol{I} + \rho\mathbf{1}\mathbf{1}' = (1-\rho)\boldsymbol{I} + \rho n\boldsymbol{J},$$

missä ρ on korrelaatiokerroin ja n on havaintojen lukumäärä. Nyt $\boldsymbol{\beta}_1$:n tehokkuuden minimi on matriisin $\boldsymbol{CVC}$ pienimmän yleistetyn antiominaisarvon neliö, kun antiominaisarvot on laskettu matriisin matriisin $\boldsymbol{CVC}$ nollasta poikkeavien ominaisarvojen perusteella.

Suorittamalla kertolasku, saadaan

$$\boldsymbol{CVC} = (1-\rho)\boldsymbol{I} - (1-\rho)\boldsymbol{J} = (1-\rho)\boldsymbol{I} - \frac{1}{n}(1-\rho)\mathbf{1}\mathbf{1}'.$$

Tällöin nähdään, että matriisin $\boldsymbol{CVC}$ alkiot ovat

$$\boldsymbol{CVC} = \begin{cases} (1-\rho)(1-\frac{1}{n}) & \text{diagonaalilla} \\ -\frac{1}{n}(1-\rho) & \text{muualla} \end{cases}.$$

Otetaan yhteiseksi tekijäksi $\frac{1}{(1-\rho)(1-\frac{1}{n})}$, jolloin

$$\boldsymbol{CVC} = \frac{1}{(1-\rho)(1-\frac{1}{n})}\begin{pmatrix} 1 & -\frac{1}{n-1} & \cdots & -\frac{1}{n-1} \\ -\frac{1}{n-1} & 1 & \cdots & \vdots \\ \vdots & \vdots & \ddots & -\frac{1}{n-1} \\ -\frac{1}{n-1} & \cdots & -\frac{1}{n-1} & 1 \end{pmatrix}, \tag{7.25}$$

missä vakiotermi $\frac{1}{(1-\rho)(1-\frac{1}{n})}$ voidaan jättää taas huomiotta. Merkitään matriisilla $\boldsymbol{D}$ yhtälön (7.25) tasakovarianssimatriisia, josta vakiotermi $\frac{1}{(1-\rho)(1-\frac{1}{n})}$ on jätetty huomiotta. Tasakorrelaatiomatriisin ominaisarvot tunnetaan yhtälön (7.23) perusteella, joten

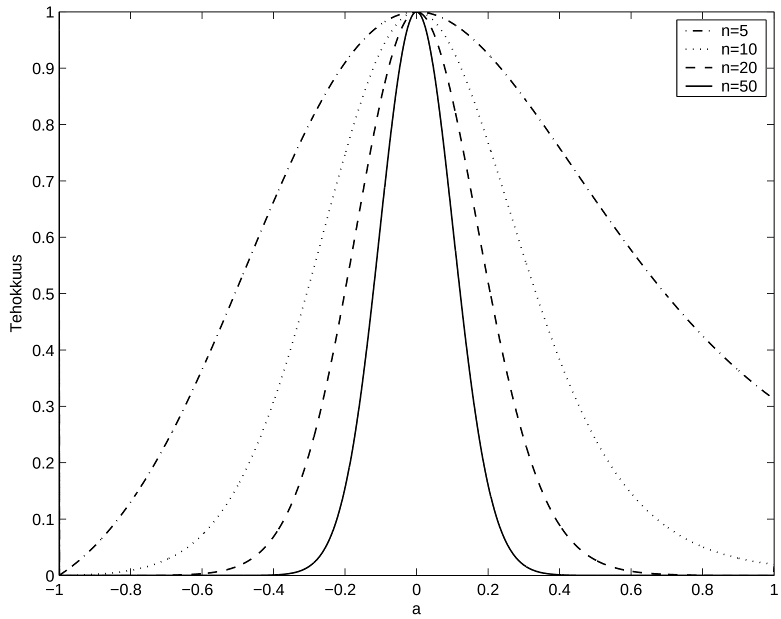
$$\{\mathrm{ch}(\boldsymbol{D})\} = \begin{cases} 0 & \text{kertalukuna 1} \\ \frac{n}{n-1} & \text{kertalukuna n-1} \end{cases}.$$

Koska matriisin $\boldsymbol{D}$ kaikki nollasta poikkeavat ominaisarvot ovat samoja, niin pienin yleistetty antiominaisarvo on yksi.

Nyt on siis osoitettu, että $\boldsymbol{\beta}_1$ vektorin Watsonin tehokkuus mallissa $\mathcal{M}_{12}$ on 1, kun kyseessä on tasakorrelaatiorakenne. Tämä voidaan myös ymmärtää niin, että riippumatta korrelaatiosta tai havaintomatriisista $\boldsymbol{X}$, OLSE:n ja BLUE:n tehokkuudet ovat samat kunhan mallissa on mukana vakiotermi. Vertaamalla tätä kuvion 3.4 tilanteeseen, jossa vakiotermiä ei ole mukana, huomaa kuinka suuri merkitys ykkösvektorilla saattaa olla tehokkuuteen. Yleisemminkin on helppoa todistaa, että OLSE on BLUE tasakorrelaatiorakenteen tapauksessa, kun mallissa on mukana vakiotermi (esim. Puntanen 1999, s. 168).

Esimerkki 7.6. Tarkastellaan $\boldsymbol{\beta}_1$ kerroinvektorin tehokkuutta mallissa $\mathcal{M}_{12}$, kun kovarianssimatriisilla $\boldsymbol{V}$ on AR(1)- ja MA(1)-rakenne. Kuten esimerkissä 7.5, myös nyt $\boldsymbol{\beta}_1$ kerroinvektorin tehokkuus mallissa $\mathcal{M}_{12}$ on sama kuin mallissa $\mathcal{M}_2$ ja kerroinvektorin $\boldsymbol{\beta}_1$ tehokkuus mallissa $\mathcal{M}_2$ voidaan laskea helposti matriisin $\boldsymbol{CVC}$ antiominaisarvojen avulla.

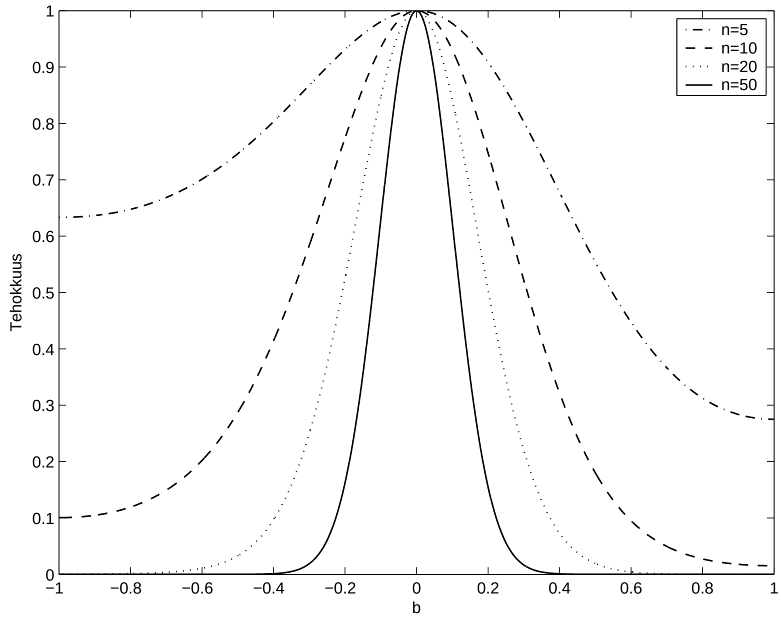
Kuviossa 7.2 pystyakselilla on Watsonin tehokkuuden alaraja, eli pienimmän



Kuvio 7.2. Kuvion pystyakselilla on $\boldsymbol{\beta}_1$ kerroinvektorin Watsonin tehokkuuden alaraja ϕ_{min}, kun kovarianssimatriisilla on AR(1)-rakenne. Vaaka-akselilla on autokorrelaatio a.

yleistetyn antiominaisarvon neliö ja vaaka-akselilla on AR(1)-mallin autokorrelaatio a. Kuvion perusteella negatiivinen autokorrelaatio saattaa vaikuttaa pahimmillaan $\boldsymbol{\beta}_1$ vektorin tehokkuuteen voimakkaammin kuin positiivinen autokorrelaatio. Isoilla havaintomäärillä tämä ero kuitenkin näyttää jäävän merkityksettömäksi.

Kuvio 7.3 vastaa kuviota 7.2 mallille, jossa kovarianssimatriisilla $\boldsymbol{V}$ on MA(1)-rakenne ja parametri b on sama kuin esimerkissä 3.2. Kuvion 7.3 perusteella nähdään, että parametri b vaikuttaa tehokkuuteen epäsymmetrisesti pienillä havaintomäärillä. Positiivisilla b:n arvoilla näyttäisi olevan huonoimmillaan



Kuvio 7.3. Kuvion pystyakselilla on $\boldsymbol{\beta}_1$ kerroinvektorin Watsonin tehokkuuden alaraja ϕ_{min}, kun kovarianssimatriisilla on MA(1)-rakenne. Vaaka-akselilla on parametri b.

enemmän vaikutusta tehokkuuteen kuin negatiivisilla b:n arvoilla. Suurilla havaintomäärillä parametrin b vaikutus on symmetrinen.

7.3 Antiominaisarvot ja matriisin kuntoisuus

Kollineaarisuudella tarkoitetaan että mallimatriisin $\boldsymbol{X}$ havaintovektorit, eli sarakkeet, ovat keskenään lineaarisesti riippuvia. Täydellinen lineaarinen riippuvuus on käytännössä harvinaista, mutta tilanne jossa sarakkeet ovat lähes lineaarisesti riippuvia saattaa olla yhtä haitallinen. Tällöinkin on siis kyse kollineaarisuudesta, vaikkakaan ei täydellisestä. Kollineaarisuus saattaa aiheuttaa ongelmia esimerkiksi regressioanalyysin kanssa. Regressiokertoimien merkit saattavat olla väärät oletettuihin nähden. Selittävien muuttujien t-testisuureen arvot saattavat olla pieniä, ja regressiokertoimet saattavat olla herkkiä havaintojen tai selittävien muuttujien poistamiselle. Aina kollineaarisuus ei kuitenkaan aiheuta mitään näistä ongelmista, mutta on tärkeää tunnistaa kollineaarisuus jos sitä esiintyy aineistossa. (Belsley 1991)

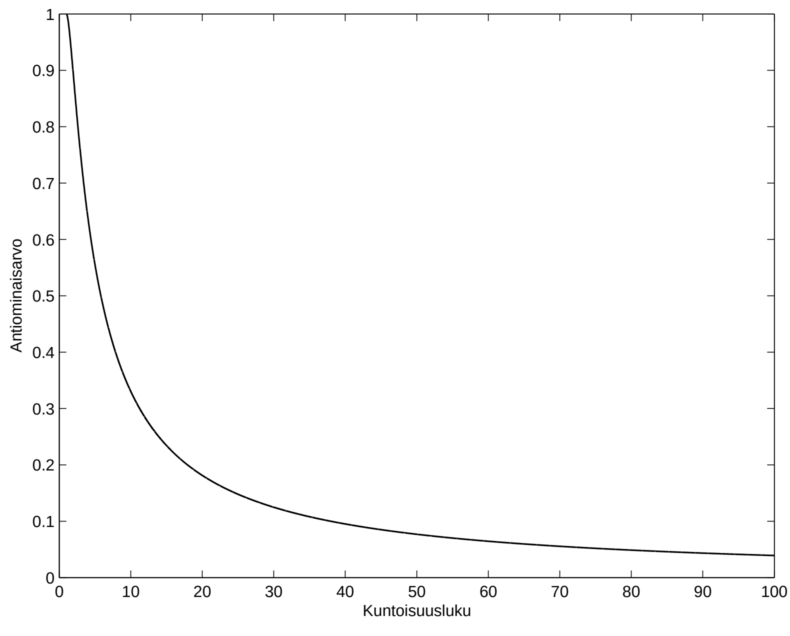
Kollineaarisuus voidaan havaita esimerkiksi tarkastelemalla VIF (*variance inflation factor*)-kertoimia, $\boldsymbol{X}$:n singulaariarvoja, matriisin $\boldsymbol{X}'\boldsymbol{X}$ determinanttia, kuntoisuuslukua '*condition number*' tai matriisin $\boldsymbol{X}$ kuntoisuus indeksejä '*condition indexes*'. Enemmän tietoa näiden käytöstä löytää esimerkiksi Belsleyn (1991) kirjasta.

Kuntoisuusluku κ on matriisin $\boldsymbol{X}$ suurimman ja pienimmän singulaariarvon suhde. Belsleyn (1991, s. 50) kirjassa perustellaan täsmällisemmin miksi kuntoisuusluku on järkevä mitta aineiston kuntoisuudelle. Wangin ja Ipn (1999) artikkelissa pienin antiominaisarvo esiintyy myös muodossa

$$\mu_1 = \frac{2\sqrt{\lambda_1\lambda_p}}{\lambda_1 + \lambda_p} = \sqrt{\lambda_p}^{-2}\frac{2\sqrt{\frac{\lambda_1\lambda_p}{\lambda_p^2}}}{\lambda_1\lambda_p} = \frac{2\sqrt{\frac{\lambda_1}{\lambda_p}}}{\frac{\lambda_1+\lambda_p}{\lambda_p}} = \frac{2\sqrt{\kappa}}{\kappa+1},$$

missä κ on siis kuntoisuusluku. Tämän vuoksi on siis ilmeistä että matriisin $\boldsymbol{X}'\boldsymbol{X}$ pienintä antiominaisarvoa voidaan käyttää kuten kuntoisuuslukua kollineaarisuuden havainnointiin.

Kuntoisuusluvulle tai -indeksille ei ole olemassa yksiselitteistä raja-arvoa, jonka jälkeen kyse olisi kollineaarisuudesta. Belsley (1991, s. 56) ja Khuri (2003, s. 47) kuitenkin esittävät, että kuntoisuusluvun ollessa alle 10 ongelmat ovat vähäisiä, ja kuntoisuusluvut 10-30 viittaavat jo vahvan kollineaarisuuden mahdollisuuteen. Yli 30 olevilla kuntoisuusluvuilla voidaan jo päätellä että aineistossa on jo vahvoja riippuvuuksia. Kuntoisuusluvun ollessa suurempi kuin kolmekummentä, aineistoa tulisi siltä osin tarkastella jo uusiksi. Kuviosta 7.4 nähdään, että kuntoisuusluvun kasvaessa antiominaisarvo pienenee. Vastaavat raja-arvot pienimmälle antiominaisarvolle ovat taulukossa 7.1.



Kuvio 7.4. Pystyakselilla on pienin antiominaisarvo μ_1, ja vaaka-akselilla kuntoisuusluku κ.

Antiominaisarvoa voidaan siis käyttää matriisin kuntoisuuden arviointiin kuten

Taulukko 7.1. Antiominaisarvon ja kuntoisuusluvun vastaavuus.

Kollineaarisuuden riski	Kuntoisuusluku κ	Antiominaisarvo μ_1
Vähäinen	0-10	$1 - 0.75$
Kohtalainen	10-30	$0.58 - 0.35$
Suuri	30-	$0.35-$

perinteistä kuntoisuuslukua. Mitään tulkinnallista perustetta antiominaisarvon käytölle ei vielä tiettävästi ole johdettu, mutta Khattree (2003) kertoo tekevänsä tutkimuksia asiasta.

Khattree (2003) arvelee myös, että yleistettyä antiominaisarvoa voitaisiin käyttää samassa tarkoituksessa. Tässä käytössä yleistetty antiominaisarvo tiivistäisi sopivasti muidenkin kuin pienimmän antiominaisarvon sisältämän informaation yhteen lukuun. Perinteinen kuntoisuusluku on ainoastaan pienimmän ja suurimman ominaisarvon suhde, ja ei tällöin sisällä mitään informaatiota muista ominaisarvoista kuten seuraavasta esimerkistä 7.7 voidaan huomata.

Esimerkki 7.7. Tarkastellaan kahta reaalista symmetristä ja positiivisesti definiittiä matriisia $\boldsymbol{A}_1$ ja $\boldsymbol{A}_2$.

$$\boldsymbol{A}_1 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 5 \end{pmatrix} \qquad \boldsymbol{A}_2 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 5 \end{pmatrix}.$$

Nyt siis kuntoisuusluvun mielessä edellä olevat matriisit $\boldsymbol{A}_1$ ja $\boldsymbol{A}_2$ ovat samanlaisia, sillä molemmilla suurimman ja pienimmän ominaisarvon suhde on sama

$$\kappa(\boldsymbol{A}_1) = \frac{\lambda_1}{\lambda_4} = \frac{5}{1} = 5 = \kappa(\boldsymbol{A}_2).$$

Yleistetty antiominaisarvo erottelee matriisit $\boldsymbol{A}_1$ ja $\boldsymbol{A}_2$ paremmin kuntoisuuden mukaan.

$$\begin{aligned} \mu_{[2]}(\boldsymbol{A}_1) &= \frac{2\sqrt{5 \cdot 1}}{5+1} \cdot \frac{2\sqrt{3 \cdot 2}}{3+2} = \frac{2\sqrt{5}}{6} \cdot \frac{2\sqrt{6}}{5} \approx 0.7303, \\ \mu_{[2]}(\boldsymbol{A}_2) &= \frac{2\sqrt{5 \cdot 1}}{5+1} \cdot \frac{2\sqrt{4 \cdot 2}}{4+2} = \frac{\sqrt{5}}{3} \cdot \frac{2\sqrt{2}}{3} \approx 0.7027. \end{aligned}$$

Yleistetty antiominaisarvo on antiominaisarvojen tulo, ja tällöin mitä pienempi yleistetty antiominaisarvo on, niin sitä enemmän on viitteitä kuntoisuusongelmasta.

Edellä jo todettiin, että yleistetyn antiominaisarvon käyttöön kuntoisuuden arvioinnissa ei ainakaan vielä ole löytynyt teoreettisia perusteita. Jos teoreettinen perusta tälle löytyy, niin yleistettyä antiominaisarvoa voidaan tilastotieteessä kuntoisuuden mittamisen lisäksi käyttää, kollineaarisuuden identifiointiin, vaikuttavien havaintojen etsimiseen ja erilasten tilastollisten mallien herkkyysanalyysiin. (Khattree 2003)

7.4 Yleistetty antiominaisarvo ja kovarianssimatriisin testaaminen

Olkoon $\boldsymbol{S}_{p\times p}$ otoskovarianssimatriisi, joka noudattaa Wishartin jakaumaa vapausartein f ja odotusarvolla $E(\boldsymbol{S}) = \boldsymbol{\Sigma}$:

$$f\boldsymbol{S} \sim W_p(f, \boldsymbol{\Sigma}).$$

Halutaan testataan hypoteesia

$$\mathrm{H}_0 : \boldsymbol{\Sigma} = \sigma^2 \boldsymbol{I}_p, \qquad \text{missä } \sigma^2 \text{ on tuntematon,}$$

ja vastahypoteesi on

$$\mathrm{H}_1 : \boldsymbol{\Sigma} \text{ tuntematon.}$$

Venables (1976) esitti, että H_0 on tosi, jos vektori $\boldsymbol{h}_{p\times 1}$ on jokin matriisin $\boldsymbol{\Sigma}$ ominaisvektori. Yleisemmin voidaan osoittaa, että H_0 on tosi jos $\boldsymbol{\Sigma}$:n aliavaruus $\mathcal{A}$ on suljettu $\boldsymbol{\Sigma}$:n suhteen, eli jokaisella vektorilla $\boldsymbol{h} \in \mathcal{A}$, myös $\boldsymbol{\Sigma h} \in \mathcal{A}$. Tämä on siis välttämätön ja riittävä ehto, että nollahypoteesit H_0 ja H_1 voidaan kirjoittaa muodossa:

$$\mathrm{H}_0 : \bigcap_{\mathcal{A}} \mathrm{H}_0(\mathcal{A}) \quad \text{ja} \quad \mathrm{H}_1 : \bigcup_{\mathcal{A}} \mathrm{H}_0(\mathcal{A}),$$

missä sekä leikkausjoukko että unioni ovat r –ulotteisia $\boldsymbol{\Sigma}$:n aliavaruuksia. Koska molemmat $\mathcal{A}$ sekä $\mathcal{A}^{\perp}$ ovat joko suljettuja, tai eivät ole suljettuja $\boldsymbol{\Sigma}$:n suhteen, niin voidaan yleisyyttä menettämättä olettaa, että $1 \leq r \leq [p/2]$.

Olkoon matriisit $\boldsymbol{X}_{(p\times r)}$ ja $\boldsymbol{Y}_{(p\times(p-r))}$ ortonormaaleja kantoja joukoille $\mathcal{A}$ sekä $\mathcal{A}^{\perp}$, ja

$$\boldsymbol{Z}_{(p\times p)} = (\boldsymbol{X} : \boldsymbol{Y}).$$

Nyt uskottavuussuhdetestin testisuure us voidaan kirjoittaa muodossa:

$$US(\mathcal{A}) = \frac{|\boldsymbol{Z}'\boldsymbol{S}\boldsymbol{Z}|}{|\boldsymbol{X}'\boldsymbol{S}\boldsymbol{X}||\boldsymbol{Y}'\boldsymbol{S}\boldsymbol{Y}|}. \tag{7.26}$$

Toisaalta tiedetään, että

$$|\boldsymbol{Z}'\boldsymbol{S}\boldsymbol{Z}| = |\boldsymbol{Y}'\boldsymbol{S}\boldsymbol{Y}||\boldsymbol{X}'\boldsymbol{S}\boldsymbol{X} - \boldsymbol{X}'\boldsymbol{S}\boldsymbol{Y}(\boldsymbol{X}'\boldsymbol{S}\boldsymbol{X})^{-1}\boldsymbol{Y}'\boldsymbol{S}\boldsymbol{X}| \tag{7.27}$$

$$= \frac{|\boldsymbol{Y}'\boldsymbol{S}\boldsymbol{Y}|}{|\boldsymbol{X}'\boldsymbol{S}^{-1}\boldsymbol{X}|}. \tag{7.28}$$

Sijoittamalla yhtälö (7.28) yhtälöön (7.26), saadaan

$$US(\mathcal{A}) = (|\boldsymbol{X}'\boldsymbol{S}\boldsymbol{X}||\boldsymbol{X}'\boldsymbol{S}^{-1}\boldsymbol{X}|)^{-1}$$

Valitaan että

$$us(\mathcal{A}) = (US(\mathcal{A}))^{-1} = |\boldsymbol{X}'\boldsymbol{S}\boldsymbol{X}||\boldsymbol{X}'\boldsymbol{S}^{-1}\boldsymbol{X}|,$$

jolloin uskottavuussuhdetestin testisuureen arvo saadaan maksimointitehtävä-nä

$$us^*(\mathcal{A}) = \max_{\mathcal{A}} us(\mathcal{A}). \tag{7.29}$$

Maksimointitehtävä (7.29) voidaan kuitenkin esittää minimointitehtävänä

$$us^*(\mathcal{A}) = \min_{\mathcal{A}}(us(\mathcal{A}))^{-1} = \min_{\boldsymbol{X}} \frac{1}{|\boldsymbol{X}'\boldsymbol{S}\boldsymbol{X}||\boldsymbol{X}'\boldsymbol{S}^{-1}\boldsymbol{X}|} = \mu_{[r]}^2,$$

mikä yhtyy matriisin $\boldsymbol{\Sigma}$ pienimmän yleistetyn antiominaisarvon etsimiseen. Tällöin nollahypoteesiä H_0 vastaava testisuure on matriisin $\boldsymbol{\Sigma}$ pienimmän yleistetyn antiominaisarvon $\mu_{[r]}$ neliö, ja nollahypoteesi hylätään liian pienillä $\mu_{[r]}^2$:n arvoilla.

8 Yhteenveto

Tässä tutkielmassa ollaan rajoituttu tarkastelemaan reaalisen, symmetrisen ja ei-negatiivisesti definiittien matriisien antiominaisarvoja ja -vektoreita sekä niiden sovellutuksia tilastotieteessä. Antiominaisarvo on terminä uusi, mutta tilastotieteen sovellutuksia sille on jo Fruchtin (1943) ja Kantorovichin (1948) julkaisuista lähtien. Tässä tutkielmassa siis otetaan käyttöön uudet termit antiominaisarvo ja antiominaisvektori, ja kerätään jo pitkään tilastotieteen käytössä olleita tuloksia tästä lähtökohdasta.

Tässä tutkimuksessa mainitaan vain hyvin lyhyesti mistä antiominaisarvo on saanut nimensä, ja termi antiominaisarvo otetaan käyttöön määritelmänä. Antiominaisarvot ja -vektorit voidaan esittää ominaisarvojen ja ominaisvektoreiden funktiona, mikä tutkielmassa johdetaan täsmällisti. Tutkielmassa johdetaan OLSE:n ja BLUE:n välisen tehokkuuden alaraja, joka on pienimmän antiominaisarvon neliö lineaarisessa mallissa kun mallimatriisina on vektori. Tehokkuutta tarkastellaan esimerkkien avulla erikoistapauksissa, joissa kovarianssimatriisilla on AR(1), MA(1) ja tasakorrelaatiorakenne.

Koska antiominaisarvo voidaan tulkita vektoreiden välisen kulman kosiniksi, niin antiominaisarvolle löytyy myös tulkinta korrelaatiokertoimen teoreettisena alarajana. Korrelaatiokertoimen ja antiominaisarvon yhteys käsitellään tässä tutkimuksessa kohtuullisen tarkasti, mutta enemmän painoa tarkastelussa annetaan kanonisille korrelaatioille. Kanonisten korrelaatioiden ja antiominaisnarvojen yhteys johdetaan täsmällisesti ja esitetään joitakin esimerkkejä.

Tutkielmassa otetaan käyttöön vakiintumattomana terminä yleistetty antiominaisarvo, ja johdetaan sille samat tulokset kuin antiominaisarvolle. Yleistetyn antiominaisarvon käsitteen avulla päästään suoraan tarkastelemaan Watsonin (1955) tehokkuuden minimiä lineaarisessa mallissa Bloomfieldin ja Watsonin (1975) sekä Knottin (1975) tapaan. Lisäksi tutkielmassa esitellään Watsonin tehokkuuden minimi Bartmanin ja Bloomfieldin (1981) tapaan. Watsonin tehokkuus yleistetään ei-negatiivisesti definiiteille kovarianssimatriiseille, jolloin pystytään tarkastelemaan osavektoreiden tehokkuuksia. Tutkielmassa käydään myös lyhyt keskustelu antiominaisarvon ja yleistetyn antiominaisarvon käyttämisestä matriisin kuntoisuuden arvioinnissa. Lopuksi vielä johdetaan uskottavuussuhdetesti kovarianssimatriisin identtisyyden testaamiseen, jolloin testisuureeksi saadaan yleistetyn antiominaisarvon neliö.

Tällä hetkellä tunnetuista antiominaisarvon ja yleistetyn antiominaisarvon sovellutuksista ehkä merkittävimmät liityvät OLSE:n ja BLUE:n suhteelliseen tehokkuutteen. Antiominaisarvo on kuitenkin terminä suhteellisen nuori, ja muita merkittäviä tilastotieteen sovellutuksia varmaankin vielä löytyy. Tässä tutkimuksessa ei tarkasteltu muita kuin reaalisia, symmetrisiä ja ei-negatiivisesti definiittejä matriiseja, joten iso osa mahdollisista sovellutuksista jää jo tämän rajauksen vuoksi tarkastelun ulkopuolelle. Tilastotieteen sovellutuksia muille kuin reaalisille, symmetrisille ja positiivisesti definiiteille ei ole tutkittu kovinkaan paljon, joten siinä olisi yksi mielenkiintoinen aihepiiri jatkotutkimukselle. Myöskin antiominaisarvon ja yleistetyn antiominaisarvon käyttäminen matriisin kuntoisuuden mittaamisessa jää tutkielmassa lyhyen esittelyn varaan ja antaa myös aihetta jatkotutkimukselle.

Lähdeluettelo

Anderson, T. W. (2003), *An Introduction to Multivariate Statistical Analysis*, Third Edition, Wiley.

Bartmann, F. C. & Bloomfield, P. (1981), "Ineffiency and correlation", *Biometrika*, 68, 1. 67–71.

Belsley, D. A. (1991), *Conditioning Diagnostics, Collinearity and Weak Data in Regression*, Wiley.

Bhatia, R. (1999), "The matrix version of inequalities of Kantorovich and Wielandt", unpublished manuscript. *Indian Statistical Institute–Delhi Centre*, 2pp. October 1999.

Bhatia, R. & Davis C. (2000), "Matrix operator versions of the Schwarz inequality", preprint, Indian Statistical Institute–Delhi Centre, 10pp. To be published in *Communications in Mathematical Physics*.

Bloomfield, P. & Watson G. (1975), "The inefficiency of least squares", *Biometrika*, 62, 121–128.

Chu, K. L., Isotalo J., Puntanen S., Styan G. P. H. (2004), "On decomposing the Watson efficiency of ordinary least squares in a partitioned weakly singular linear model", *Sankhyā*, 66, 634–651.

Chu, K. L., Isotalo J., Puntanen S., Styan G. P. H. (2005), "Some further results concerning the decomposition of the Watson efficiency in partitioned linear models", *Sankhyā*, 67, 74–89.

Davis, C. (1980), "Extending the Kantorovich inequalities to normal matrices", *Linear Algebra and its Applications*, 31, 173–177.

Drury, S. W., Liu, S., Lu, C., Puntanen S. & Styan G. P. H. (2000), "Some Comments on Several Matrix Inequalities with Applications to Canonical Correlations: Historical Backround and Recent Developments", *Dept. of Mathematics, Statistics & Philosophy, University of Tampere, Finland*, REPORT A332.

Eaton, M. L. (1976), "A Maximization problem and its application to canonical correlation", *Journal of Multivariate Analysis*, 6, 422–425.

Frucht, R. W. (1943), "Sobre algunas desigualdades: Observacación relativa a la solución del Problema N° 21, indicada por el Ing. Ernesto M. Saleme", *MathematicœNotœ–Boletin del Instituto de Matemática "Beppo Levi" (Rosario)*, 3, 41–46.

Gustafson, K. E. (1972), "Antieigenvalue inequalities in operator theory", in Inequalities III, Los Angeles, 1969, (O. Shisha, ed.), Academic Press, pp. 115–119.

Gustafson, K. E. (2002), "Operator trigonometry of statistics and econometrics", *Linear Algebra and its Applications*, 354, 141–158.

Gustafson, K. E. (2005), "The geometry of statistical efficiency", Research Letters in the Information and Mathematical Sciences, 2005, Vol. 8, pp. 105–121.

Gustafson, K. E. & Rao, D. K. M. (1997), *Numerical Range: The Field of Values of Linear Operators and Matrices*, Springer-Verlag.

Householder, A. S. (1974), *The Theory of Matrices in Numerical Analysis*, reprint ed. Dover, New York, Section 3.4, 81–85.

Kantorovich, L. V. (1948), "Funkcional'nyi analiz i prikladnaya matematika", *Uspekhi Matematičeskĭĭ Nauk, Novaya Seriya*, 3(6/28), 89–185.

Khattree, R. (2001), "On calculation of antieigenvalues and antieigenvectors", *Journal of Interdisciplinary Mathematics*, 4, 195–199.

——— (2002), "Generalized antieigenvalues of order r", *American Journal of Mathematical Management Science*, 22, 89–98.

——— (2003), "Antieigenvalues and Antieigenvectors in statistics", *Journal of Statistical Planning and Inference*, 104, 131–144.

Khuri, A. I. (2003), *dvanced Calculus with Applications in Statistics*, Second Edition, Wiley.

Knott, M. (1975), "On the minimum efficiency of least squares", *Biometrika*, 62, 1, 129–132.

Liu, S. (2000), "Efficiency comparisons between two estimators based on matrix determinant Kantorovich-type inequalities", *Metrika*, 51, 145–155.

Pták, V. (1995), "The Kantorovich Inequality", *The American Mathematical Monthly*, Vol. 102, 820–821.

Puntanen, S. (1999), "Regressioanalyysi I–II", *Matematiikan, tilastotieteen ja filosofian laitos, Tampereen yliopisto*, B48–B49.

Puntanen, S. (2001), "Yleistetyistä käänteismatriiseista ja niiden sovellutuksista lineaarisissa tilastollisissa malleissa", *Matematiikan, tilastotieteen ja filosofian laitos, Tampereen yliopisto*, B55.

Puntanen, S. & Styan, G. P. H. (2003), "Matrix Tricks for Linear Statistical Models: Our Personal Top Thirteen", *Dept. of Mathematics, Statistics & Philosophy, University of Tampere, Finland*, REPORT A345.

Rao, C. R (1973), "Linear Statistical Inference and its Applications", 2nd edition, Wiley, New Yourk.

Rao, C. R. (2005), "Antieigenvalues and antisingularvalues of a matrix and applications to problems in statistics ", Research Letters in the Information and Mathematical Sciences, 2005, Vol. 8, pp. 53–76.

Rao, C. R & Rao, C. V. (1987), "Stationary values of the product of two Raleigh quotents: homologous canonical correlations", *Sankhyā*, Series B 37, 135–138.

Seber, G. A. F. (1984), "Multivariate Observations", Wiley

Wang, S. & Ip, W. (1999), "A matrix version of the Wiedlandt inequality and its applications to statistics", *Linear Algebra and its Applications*, 296, 171–181.

Watson, G. S. (1955), "Serial correlation in regression analysis", *Biometrika*, 42, 327–341.

Watson, G. S., Alpargu, G. & Styan, G. P. H. (1997), "Some comments on six inequalities associated with the inefficiency of ordinary least squares with one regressor", *Linear Algebra and its Applications*, 264, 13–53.

Venables, W. (1976), "Some implications of the union-intersection principle for test of sphericity", *Journal of Multivariate Analysis*, 6, 185–190.

Zhang, F. (1999), "Equivalence of the vector type Wielandt inequality and the Kantorovich inequality", Preprint. *Dept. of Math, Science and Technology, Nova Southeastern University, Fort Lauderdale, Florida*, 4pp. Submitted for publication.

Zhang, F. (2001), "Equivalence of the Wielandt inequality and the Kantorovich inequality", *Linear and Multilinear Algebra*, 3. 275–279.

Zyskind, G. & Martin, F. B. (1969), "On best linear estimation and general Gauss-Markov theorem in linear models with arbitrary nonnegative covariance structure", *SIAM Journal on Applied Mathematics*, 17. 1190–1202.

Liite:

Lause L.1 (Cauchyn–Schwarzin epäyhtälö)**.** *Jos* $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$*, niin*

$$(\boldsymbol{x}'\boldsymbol{y})^2 \leq \boldsymbol{x}'\boldsymbol{x} \cdot \boldsymbol{y}'\boldsymbol{y} \qquad \forall \boldsymbol{x}, \boldsymbol{y}. \tag{K.1}$$

Yhtäsuuruus pitää paikkansa kun $\boldsymbol{x}$ *ja* $\boldsymbol{y}$ *ovat lineaarisesti riippuvia. Olkoon* $\boldsymbol{A}_{n\times n}$ *symmetrinen ja positiivisesti definiitti matriisi, ja* $\boldsymbol{A}^-$ *jokin mielivaltainen* $\boldsymbol{A}$*:n g-inverssi. Tällöin*

$$(\boldsymbol{x}'\boldsymbol{y})^2 \leq \boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} \cdot \boldsymbol{y}'\boldsymbol{A}^-\boldsymbol{y}, \qquad \boldsymbol{y} \in \mathcal{C}(\boldsymbol{A}). \tag{K.2}$$

Yhtäsuuruus pitään pitää paikkansa kun $\boldsymbol{A}\boldsymbol{x} = \lambda\boldsymbol{y}, \quad \lambda \in \mathbb{R}.$

Todistus löytyy monista lähteistä, mutta seurataan Puntasen ja Styanin (2003) todistusta.

Todistus. Todistetaan epäyhtälö ortogonaaliprojeketorien ominaisuuksien perusteella. Merkitään $\boldsymbol{P}_{\boldsymbol{x}^\perp}$:llä ortogonaaliprojektoria $\boldsymbol{x}$:n sarakeavaruuden $\mathcal{C}(\boldsymbol{x})$ ortogonaalikomplementtiin $\mathcal{C}(\boldsymbol{x})^\perp$. Havaitsemme että jos $\boldsymbol{x} = \boldsymbol{0}$ tai $\boldsymbol{y} = \boldsymbol{0}$, niin yhtälössä (K.1) vallitsee yhtäsuuruus. Oletetaan nyt että $\boldsymbol{x} \neq \boldsymbol{0}$.

Tiedetään että vektorin $\boldsymbol{P}_{\boldsymbol{x}^\perp}\boldsymbol{y}$ tavanomainen normi voidaan kirjoittaa muodossa

$$||\boldsymbol{P}_{\boldsymbol{x}^\perp}\boldsymbol{y}||^2 = \boldsymbol{y}'\boldsymbol{P}'_{\boldsymbol{x}^\perp}\boldsymbol{P}_{\boldsymbol{x}^\perp}\boldsymbol{y} = \boldsymbol{y}'\boldsymbol{P}_{\boldsymbol{x}^\perp}\boldsymbol{y} = \boldsymbol{y}'(\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{x}})\boldsymbol{y} \geq 0. \tag{K.3}$$

Suorittamalla matriisikertolasku saadaan

$$\begin{aligned} \boldsymbol{y}'\boldsymbol{y} - \boldsymbol{y}'\boldsymbol{x}(\boldsymbol{x}'\boldsymbol{x})^{-1}\boldsymbol{x}'\boldsymbol{y} &\geq 0 \\ \frac{(\boldsymbol{x}'\boldsymbol{y})^2}{\boldsymbol{x}'\boldsymbol{x}} &\leq \boldsymbol{y}'\boldsymbol{y}, \end{aligned}$$

josta väite seuraa. Yhtäsuuruuden ollessa voivassa yhtälö (K.3) voidaan kirjoittaa muodossa

$$\boldsymbol{y}'(\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{x}})\boldsymbol{y} = 0.$$

Tämä pätee ainoastaan silloin kuin

$$\boldsymbol{y} - \boldsymbol{P}_{\boldsymbol{x}}\boldsymbol{y} = \boldsymbol{0},$$

eli vektori $\boldsymbol{y}$ kuuluu $\boldsymbol{x}$:n sarakeavaruuteen $\mathcal{C}(\boldsymbol{x})$. Tällöin

$$\exists \lambda \in \mathbb{R} : \boldsymbol{y} = \lambda \boldsymbol{x},$$

eli $\boldsymbol{x}$ ja $\boldsymbol{y}$ ovat lineaarisesti riippuvia.
Epäyhtälön toinen muoto (K.2) voidaan todistaa sijoittamalla epäyhtälön ensimmäiseen (K.1) muotoon, että

$$\boldsymbol{x} = \boldsymbol{A}^{\frac{1}{2}}\boldsymbol{u}, \quad \boldsymbol{y} = (\boldsymbol{A}^{\frac{1}{2}})^{+}\boldsymbol{v},$$

missä $\boldsymbol{u}$ ja $\boldsymbol{v}$ ovat n alkioisia reaalisia vektoreita, $\boldsymbol{A}^{\frac{1}{2}}$ on symmetrinen matriisi joka toteuttaa yhtälön $\boldsymbol{A}^{\frac{1}{2}}\boldsymbol{A}^{\frac{1}{2}} = \boldsymbol{A}$, sekä $(\boldsymbol{A}^{\frac{1}{2}})^{+}$:lla tarkoitetaan matriisin $\boldsymbol{A}^{\frac{1}{2}}$ Moore–Penrose -inverssiä. Lisää Moore–Penrose -inverssistä voi lukea esimerkiksi Puntasen (2001) artikkelista.

Tällöin yhtälö (K.1) voidaan esittää muodossa

$$\begin{aligned} (\boldsymbol{u}'\boldsymbol{v})^2 &\leq \boldsymbol{u}'\boldsymbol{u} \cdot \boldsymbol{v}'\boldsymbol{v}, \\ (\boldsymbol{x}'\boldsymbol{A}^{\frac{1}{2}}(\boldsymbol{A}^{\frac{1}{2}})^{+}\boldsymbol{y})^2 &\leq \boldsymbol{x}'\boldsymbol{A}^{\frac{1}{2}}\boldsymbol{A}^{\frac{1}{2}}\boldsymbol{x} \cdot \boldsymbol{y}'(\boldsymbol{A}^{\frac{1}{2}})^{+}(\boldsymbol{A}^{\frac{1}{2}})^{+}\boldsymbol{y}, \\ (\boldsymbol{x}'\boldsymbol{P}_{\boldsymbol{A}^{\frac{1}{2}}}\boldsymbol{y})^2 &\leq \boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} \cdot \boldsymbol{y}'\boldsymbol{A}^{+}\boldsymbol{y}. \end{aligned} \tag{K.4}$$

Nyt ominaisarvohajotelman perusteella

$$(\boldsymbol{A}^{\frac{1}{2}})^{+} = (\boldsymbol{T}_1 : \boldsymbol{T}_0)\begin{pmatrix} \boldsymbol{\Lambda}_1^{-\frac{1}{2}} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}\begin{pmatrix} \boldsymbol{T}_1' \\ \boldsymbol{T}_0' \end{pmatrix} = \boldsymbol{T}_1\boldsymbol{\Lambda}_1^{-\frac{1}{2}}\boldsymbol{T}_1' \tag{K.5}$$

ja

$$\boldsymbol{A}^{\frac{1}{2}} = \boldsymbol{T}_1\boldsymbol{\Lambda}_1^{\frac{1}{2}}\boldsymbol{T}_1'. \tag{K.6}$$

Tällöin yhtälöiden (K.5) ja (K.6) perusteella

$$\boldsymbol{P}_{\boldsymbol{A}^{\frac{1}{2}}} = \boldsymbol{A}^{\frac{1}{2}}(\boldsymbol{A}^{\frac{1}{2}})^{+} = \boldsymbol{T}_1\boldsymbol{\Lambda}_1^{\frac{1}{2}}\boldsymbol{T}_1'\boldsymbol{T}_1\boldsymbol{\Lambda}_1^{-\frac{1}{2}}\boldsymbol{T}_1' = \boldsymbol{T}_1\boldsymbol{T}_1' = \boldsymbol{P}_{\boldsymbol{T}_1} = \boldsymbol{P}_{\boldsymbol{A}},$$

sillä

$$\boldsymbol{P}_{\boldsymbol{A}} = \boldsymbol{A}\boldsymbol{A}^{+} = \boldsymbol{T}_1\boldsymbol{\Lambda}_1\boldsymbol{T}_1'\boldsymbol{T}_1\boldsymbol{\Lambda}_1^{-1}\boldsymbol{T}_1' = \boldsymbol{T}_1\boldsymbol{T}_1'.$$

Tällöin yhtälö (K.4) saadaan muotoon

$$(\boldsymbol{x}'\boldsymbol{P}_{\boldsymbol{A}}\boldsymbol{y})^2 \leq \boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} \cdot \boldsymbol{y}'\boldsymbol{A}^{+}\boldsymbol{y}. \tag{K.7}$$

Koska $\boldsymbol{y} \in \mathcal{C}(\boldsymbol{A})$, niin $\boldsymbol{P}_{\boldsymbol{A}}\boldsymbol{y} = \boldsymbol{y}$ ja $\boldsymbol{y}'\boldsymbol{A}^{-}\boldsymbol{y}$ on riippumaton $\boldsymbol{A}^{-}$:n valinnasta. Tällöin epäyhtälöstä (K.7) saadaan väite

$$(\boldsymbol{x}'\boldsymbol{y})^2 \leq \boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} \cdot \boldsymbol{y}'\boldsymbol{A}^{-}\boldsymbol{y}.$$

Yhtälö (K.7) voidaan kirjoittaa myös muodossa

$$(\boldsymbol{x}'\boldsymbol{A}^{\frac{1}{2}}(\boldsymbol{A}^{\frac{1}{2}})^+\boldsymbol{y})^2 \leq \boldsymbol{x}'\boldsymbol{A}^{\frac{1}{2}}\boldsymbol{A}^{\frac{1}{2}}\boldsymbol{x} \cdot \boldsymbol{y}'(\boldsymbol{A}^{\frac{1}{2}})^+(\boldsymbol{A}^{\frac{1}{2}})^+\boldsymbol{y},$$

jolloin yhtäsuuruuden nähdään olevan voimassa kun

$$\boldsymbol{A}^{\frac{1}{2}}\boldsymbol{x} = \lambda(\boldsymbol{A}^+)^{\frac{1}{2}}\boldsymbol{y}, \quad \lambda \in \mathbb{R}. \tag{K.8}$$

Kerrotaan yhtälö (K.8) vasemmalta $\boldsymbol{A}^{\frac{1}{2}}$:lla niin

$$\boldsymbol{A}\boldsymbol{x} = \lambda \boldsymbol{P}_{\boldsymbol{A}}\boldsymbol{y}.$$

Koska $\boldsymbol{y} \in \mathcal{C}(\boldsymbol{A})$, niin yhtäsuuruudelle saadaan ehto

$$\boldsymbol{A}\boldsymbol{x} = \lambda\boldsymbol{y}.$$

□

Apulause L.2. *Olkoon $0 < t_1 \leq t_2 \leq \ldots \leq t_n$ joukko reaalilukuja, sekä $a_{i,j}$ ja $g_{i,j}$ lukujen t_i ja t_j välinen aritmeettinen ja geometrinen keskiarvo. Tällöin*

$$\frac{a_{1,4}}{g_{1,4}} \geq \frac{a_{2,3}}{g_{2,3}}.$$

Todistus. Kahden luvun aritmeettisen ja geometrisen keskiarvon suhde voidaan kijoittaa muodossa

$$\frac{a_{1,4}}{g_{1,4}} = \frac{\frac{1}{2}(t_1 + t_n)}{\sqrt{t_1 t_n}} = \frac{1}{2}\left(\frac{t_1}{\sqrt{t_1 t_n}} + \frac{t_n}{\sqrt{t_1 t_n}}\right) = \frac{1}{2}\left(\frac{\sqrt{t_1}}{\sqrt{t_n}} + \frac{\sqrt{t_n}}{\sqrt{t_1}}\right).$$

Merkitään, että

$$\frac{\sqrt{t_1}}{\sqrt{t_n}} = f,$$

Tarkastellaan epäyhtälöä

$$d + \frac{1}{d} \leq f + \frac{1}{f} \tag{K.9}$$

$$d^2 + 1 \leq \left(f + \frac{1}{f}\right)d$$

$$d^2 - \left(f + \frac{1}{f}\right)d + 1 \leq 0, \tag{K.10}$$

jossa d on muuttuja ja $f = \frac{\sqrt{t_1}}{\sqrt{t_n}}$.

Koska yhtälö (K.10) on toisen asteen yhtälö, niin tiedetään että sen juurien summa on $d_1 + d_2 = f + \frac{1}{f}$ ja tulo $d_1 d_2 = 1$. Näistä yhtälöistä voidaan ratkaista epäyhtälöä (K.10) vastaavan yhtälön juuret jotka ovat $d_1 = f$ ja $d_2 = \frac{1}{f}$. Koska yhtälö (K.10) on muodoltaan ylöspäin aukeava paraabeli, niin epäyhtälön (K.9) voidaan todeta pitävän paikkansa aina kun $d \in (f, \frac{1}{f})$. □

Lause L.3 (Fruchtin–Kantorovichin epäyhtälö). *Olkoon $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ on symmetrinen ja positiivisesti definiitti matriisi, jonka suurimmalle ja pienimmälle ominaisarvolle pätee yhtälö $\lambda_1 \geq \cdots \geq \lambda_n \geq 0$. Tällöin*

$$\frac{(\boldsymbol{y}'\boldsymbol{y})^2}{\boldsymbol{y}'\boldsymbol{A}\boldsymbol{y}\cdot\boldsymbol{y}'\boldsymbol{A}^{-1}\boldsymbol{y}} \geq \frac{4\lambda_1\lambda_n}{(\lambda_1+\lambda_n)^2}. \tag{K.11}$$

Fruchtin–Kantorovichin epäyhtälön alkuperäisiä todistuksia ovat esittäneet Frucht (1943) ja Kantorovich (1948), mutta seurataan jatkossa Vlastimil Ptákin (1995) todistusta.

Todistus. Ominaisarvohajotelman avulla yhtälö (K.11) voidaan esittää muodossa

$$\frac{(\boldsymbol{y}'\boldsymbol{y})^2}{\boldsymbol{y}'\boldsymbol{A}\boldsymbol{y}\cdot\boldsymbol{y}'\boldsymbol{A}^{-1}\boldsymbol{y}} = \frac{(\boldsymbol{y}'\boldsymbol{D}\boldsymbol{D}'\boldsymbol{y})^2}{\boldsymbol{y}'\boldsymbol{D}\boldsymbol{\Lambda}\boldsymbol{D}'\boldsymbol{y}\cdot\boldsymbol{y}'\boldsymbol{D}\boldsymbol{\Lambda}^{-1}\boldsymbol{D}'\boldsymbol{y}} = \frac{(\boldsymbol{z}'\boldsymbol{z})^2}{\boldsymbol{z}'\boldsymbol{\Lambda}\boldsymbol{z}\cdot\boldsymbol{z}'\boldsymbol{\Lambda}^{-1}\boldsymbol{z}}, \tag{K.12}$$

kun viimeisessä yhtäsuuruudessa ollaan tehty sijoitus $\boldsymbol{z} = \boldsymbol{D}'\boldsymbol{y}$. Yhtälössä (K.12) vektori $\boldsymbol{z}$ voidaan kertoa jollakin vakiolla $\alpha \in \mathbb{R}$, koska

$$\begin{aligned}
\frac{(\alpha\boldsymbol{z}'\alpha\boldsymbol{z})^2}{\alpha\boldsymbol{z}'\boldsymbol{\Lambda}\alpha\boldsymbol{z}\cdot\alpha\boldsymbol{z}'\boldsymbol{\Lambda}^{-1}\alpha\boldsymbol{z}} &= \frac{(\alpha^2\boldsymbol{z}'\boldsymbol{z})^2}{\alpha^4\boldsymbol{z}'\boldsymbol{\Lambda}\boldsymbol{z}\cdot\boldsymbol{z}'\boldsymbol{\Lambda}^{-1}\boldsymbol{z}} = \frac{(\boldsymbol{z}'\boldsymbol{z})^2}{\boldsymbol{z}'\boldsymbol{\Lambda}\boldsymbol{z}\cdot\boldsymbol{z}'\boldsymbol{\Lambda}^{-1}\boldsymbol{z}} \\
&= \frac{(\sum_{1=1}^p z_i^2)^2}{(\sum_{1=1}^p z_i^2\lambda_i)\left(\sum_{1=1}^p z_i^2\lambda_i^{-1}\right)} \\
&= \frac{(\sum_{1=1}^p x_i)^2}{(\sum_{1=1}^p x_i\lambda_i)\left(\sum_{1=1}^p x_i\lambda_i^{-1}\right)},
\end{aligned} \tag{K.13}$$

missä $x_i = z_i^2$.

Valitaan α siten, että $\sum_{1=1}^p x_i = 1$, jolloin edellisestä yhtälöstä (K.13) alkuperäinen väite saadaan muotoon

$$\left(\sum_{1=1}^p x_i\lambda_i\right)\left(\sum_{1=1}^p x_i\lambda_i^{-1}\right) \leq \mu_p^2 = \frac{a^2}{g^2}, \tag{K.14}$$

missä a on suurimman ja pienimmän ominaisarvon aritmeettinen keskiarvo, ja g on vastaava geometrinen keskiarvo. Eli

$$\begin{aligned}
a &= \frac{1}{2}(\lambda_1 - \lambda_n) \\
g &= (\lambda_1\lambda_n)^{\frac{1}{2}}.
\end{aligned}$$

Koska yhtälössä (K.14) voidaan ominaisarvot λ_i kertoa sopivalla reaaliluvulla α siten, että $\lambda_1\lambda_n = 1$, niin voidaan olettaa että $g = 1$. Tällöin voidaan siis olettaa, että $\lambda_n = \frac{1}{\lambda_1}$.

Tarkastellaan epäyhtälöä

$$d + \frac{1}{d} \leq f + \frac{1}{f} \tag{K.15}$$

$$d^2 + 1 \leq \left(f + \frac{1}{f}\right) d$$

$$d^2 - \left(f + \frac{1}{f}\right) d + 1 \leq 0, \tag{K.16}$$

jossa d on muuttuja ja f on reaalinen vakio.

Koska yhtälö (K.16) on toisen asteen yhtälö, niin tiedetään että sen juurien summa on $d_1 + d_2 = f + \frac{1}{f}$ ja tulo $d_1 d_2 = 1$. Näistä yhtälöistä voidaan ratkaista epäyhtälöä (K.16) vastaavan yhtälön juuret jotka ovat $d_1 = f$ ja $d_2 = \frac{1}{f}$. Koska yhtälö (K.16) on muodoltaan ylöspäin aukeava paraabeli, niin epäyhtälön (K.15) voidaan todeta pitävän paikkansa aina kun $d \in (f, \frac{1}{f})$.

Tämän perusteella tiedetään että

$$\sum_{i=1}^{n} x_i \lambda_i + \sum_{1=1}^{n} x_i \lambda_i^{-1} \leq \lambda_1 + \frac{1}{\lambda_1} = 2a. \tag{K.17}$$

Sijoittamalla yhtälö (K.17) yhtälöön (K.14) saadaan

$$\left[\left(\sum_{1=1}^{p} x_i \lambda_i\right)\left(\sum_{1=1}^{p} x_i \lambda_i^{-1}\right)\right]^{\frac{1}{2}} \leq a$$

$$\left[\left(\sum_{1=1}^{p} x_i \lambda_i\right)\left(\sum_{1=1}^{p} x_i \lambda_i^{-1}\right)\right]^{\frac{1}{2}} \leq \frac{1}{2}\left(\sum_{i=1}^{n} x_i \lambda_i + \sum_{1=1}^{n} x_i \lambda_i^{-1}\right),$$

mikä pitää paikkansa, sillä geometrinen keskiarvo on aina pienempi tai yhtäsuuri kuin aritmeettinen keskiarvo.

□

Lause L.4 (Tulon $\boldsymbol{AB}$ nollasta poikkeavat ominaisarvot)**.** *Olkoon* $\boldsymbol{A} \in \mathbb{R}^{n \times m}$ *ja* $\boldsymbol{B} \in \mathbb{R}^{m \times n}$*. Tällöin matriiseilla* $\boldsymbol{AB} \in \mathbb{R}^{n \times n}$ *ja* $\boldsymbol{BA} \in \mathbb{R}^{m \times m}$ *on samat nollastapoikkeavat ominaisarvot, eli*

$$nzch(\boldsymbol{AB}) = nzch(\boldsymbol{BA}). \tag{K.18}$$

Jos $\boldsymbol{AB}$ *ja* $\boldsymbol{BA}$ *ovat neliömatriiseja, niin ominaisarvot ovat identtiset.*

Todistus. Todistuksen on osoittanut esimerkiksi Seber (1984). Olkoon λ matriisin $\boldsymbol{AB}$ nollasta poikkeava ominaisarvo. Tällöin on olemassa sellainen vektori

$\boldsymbol{u} \neq \mathbf{0}$ siten että $\boldsymbol{ABu} = \lambda \boldsymbol{u}$. Jos yhtälö kerrotaan vasemmalta $\boldsymbol{B}$:llä, niin saadaan että $\boldsymbol{BABu} = \lambda \boldsymbol{Bu}$. Tällöin voidaan merkitä, että $\boldsymbol{v} = \boldsymbol{Bu} \neq \mathbf{0}$, kunhan $\boldsymbol{ABu} \neq \mathbf{0}$. Tällöin λ on $\boldsymbol{BA}$:n ominaisarvo. Matriiseiden $\boldsymbol{A}$ ja $\boldsymbol{B}$ roolit todistuksessa voidaan vaihtaa.

□

Lause L.5. *(simultaanisesti diagonalisoituvuus) Olkoon matriisit $\boldsymbol{A}$ ja $\boldsymbol{B}$ symmetrisiä $n \times n$ matriiseja. Tällöin*

$$\exists \boldsymbol{P} : \quad \boldsymbol{P}'\boldsymbol{AP} = \boldsymbol{D}, \quad \boldsymbol{P}'\boldsymbol{BP} = \boldsymbol{\Lambda} \quad \Leftrightarrow \quad \boldsymbol{AB} = \boldsymbol{BA}, \tag{K.19}$$

missä $\boldsymbol{D}$ ja $\boldsymbol{\Lambda}$ ovat diagonaalimatriiseja.

Todistus. Todistus sivuutetaan, mutta se läydettävissä esimerkiksi Raon (1973) kirjasta. □