Matriisilaskentaa tilastotieteilijöille

Ilkka Mellin

3. Edistyneempää matriisilaskentaa

Matriisilaskentaa tilastotieteilijöille

3. Edistyneempää matriisilaskentaa

3. Edistyneempää matriisilaskentaa

3.1. Cholesky-hajotelma ja QR-hajotelma

Lineaaristen mallien ja *monimuuttujamenetelmien* teoriassa on usein hyötyä seuraavassa esitettävistä kahdesta *matriisihajotelmasta*.

Cholesky–hajotelma

Lause 3.1.1.

Olkoon $\mathbf{A}$ *positiivisesti definiitti* $m \times m$-matriisi. Tällöin on olemassa *epäsingulaarinen alakolmiomatriisi* $\mathbf{L}$ ja *epäsingulaarinen yläkolmiomatriisi* $\mathbf{U}$ siten, että

$$\mathbf{A} = \mathbf{LL}' = \mathbf{UU}'$$

Todistus :

Koska matriisi $\mathbf{A}$ on oletettu positiivisesti definiitiksi, se on *symmetrinen* ja *epäsingulaarinen*. Lisäksi lauseen 2.7.1. mukaan kaikki matriisin $\mathbf{A}$ alimatriisit, jotka saadaan poistamalla matriisista $\mathbf{A}$ r kappaletta toisiaan vastaavaa riviä ja saraketta ovat positiivisesti definiittejä.

Tarkastelemme alla matriisin $\mathbf{A}$ esittämistä alakolmiomatriisin $\mathbf{L}$ ja sen transpoosin $\mathbf{L}'$ tulona. Matriisin $\mathbf{A}$ esittäminen yläkolmiomatriisin $\mathbf{U}$ ja sen transpoosin $\mathbf{U}'$ tulona tapahtuu vastaavalla tavalla.

Esitämme pedagogisista syistä alla kaksi erilaista todistusta lauseelle.

Todistus 1: Matriisin A ja tulon LL′ alkioiden suora vertailu.

Olkoon $\mathbf{L}$ alakolmiomatriisi:

$$\mathbf{L} = \begin{bmatrix} l_{11} & 0 & 0 & \cdots & 0 \\ l_{21} & l_{22} & 0 & \cdots & 0 \\ l_{31} & l_{32} & l_{33} & \vdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ l_{m1} & l_{m2} & l_{m3} & \cdots & l_{mm} \end{bmatrix}$$

Asetetaan matriisiyhtälö

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1m} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2m} \\ a_{31} & a_{32} & a_{33} & \vdots & a_{2m} \\ \vdots & \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \cdots & a_{mm} \end{bmatrix} = \begin{bmatrix} l_{11} & 0 & 0 & \cdots & 0 \\ l_{21} & l_{22} & 0 & \cdots & 0 \\ l_{31} & l_{32} & l_{33} & \vdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ l_{m1} & l_{m2} & l_{m3} & \cdots & l_{mm} \end{bmatrix} \begin{bmatrix} l_{11} & l_{21} & l_{31} & \cdots & l_{m1} \\ 0 & l_{22} & l_{32} & \cdots & l_{m2} \\ 0 & 0 & l_{33} & \vdots & l_{m3} \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & l_{mm} \end{bmatrix} = \mathbf{LL}'$$

Siten matriisin $\mathbf{L}$ alkioiden l_{ij} on toteutettava seuraavat yhtälöt:

$$a_{11} = l_{11}^2,\ a_{12} = l_{11}l_{21},\ a_{13} = l_{11}l_{31}, \cdots,\ a_{1m} = l_{11}l_{m1}$$

$$a_{21} = l_{21}l_{11},\ a_{22} = l_{21}^2 + l_{22}^2,\ a_{23} = l_{21}l_{31} + l_{22}l_{32}, \cdots,\ a_{2m} = l_{21}l_{m1} + l_{22}l_{m2}$$

$$a_{31} = l_{31}l_{11},\ a_{32} = l_{31}l_{21} + l_{32}l_{22},\ a_{33} = l_{31}^2 + l_{32}^2 + l_{33}^2, \cdots, a_{3m} = l_{31}l_{m1} + l_{32}l_{m2} + l_{33}l_{m3}$$

$$\cdots$$

$$a_{m1} = l_{m1}l_{11}, a_{m2} = l_{m1}l_{21} + l_{m2}l_{22}, a_{m3} = l_{m1}l_{31} + l_{m2}l_{32} + l_{m3}l_{33}, \cdots,\ a_{mm} = \sum\nolimits_{j=1}^{m} l_{mj}^2$$

Tämä yhtälöryhmä on matriisin **L** tuntemattomien alkioiden l_{ij} suhteen *rekursiivinen* ja siten alkiot l_{ij} saadaan ratkaistuksi yhtälöryhmästä rivi riviltä käyttämällä hyväksi aikaisemmin saatuja ratkaisuja.

Yhtälöryhmän ensimmäiseltä riviltä saadaan ratkaistuksi matriisin **L** 1. sarakkeen alkiot $l_{11}, l_{21}, \ldots , l_{m1}$:

$$l_{11} = \pm\sqrt{a_{11}}, \; l_{21} = \frac{a_{12}}{l_{11}}, \; l_{31} = \frac{a_{13}}{l_{11}}, \ldots, \; l_{m1} = \frac{a_{1m}}{l_{11}}$$

Yhtälöryhmän toiselta riviltä saadaan ratkaistuksi matriisin **L** alkio l_{21} ja matriisin **L** 2. sarakkeen alkiot $l_{22}, l_{32}, \ldots , l_{m2}$:

$$l_{21} = \frac{a_{21}}{l_{11}},$$

$$l_{22} = \pm\left(\frac{a_{22}a_{11} - a_{21}^2}{a_{11}}\right)^{1/2}, \; l_{i2} = \frac{a_{2i} - l_{21}l_{i1}}{l_{22}}, \; i = 3, 4, \ldots, m$$

Yhtälöryhmän kolmannelta riviltä saadaan ratkaistuksi matriisin **L** alkiot l_{31} ja l_{32} matriisin **L** 3. sarakkeen alkiot $l_{33}, l_{43}, \ldots , l_{m3}$:

$$l_{31} = \frac{a_{31}}{l_{11}}, \; l_{32} = \frac{a_{32} - l_{21}l_{31}}{l_{22}}$$

$$l_{33} = \pm\left(a_{33} - \frac{a_{31}^2}{l_{31}^2} - \frac{(a_{32} - l_{21}l_{31})^2}{l_{22}^2}\right)^{1/2}, \; l_{i3} = \frac{a_{3i} - l_{31}l_{i1} - l_{32}l_{i2}}{l_{33}}, \; i = 4, 5, \ldots, m$$

Jatkamalla kuvattua menettelyä, saadaan loputkin matriisin **L** alkiot ratkaistuksi.

Huomautus 1:

Olkoon

$$\mathbf{B}_k, \; k = 1, 2, \ldots, m$$

matriisin **A** $k\times k$-alimatriisi, joka saadaan poistamalla matriisista **A** sen $(m-k)$ viimeistä riviä ja saraketta (huomaa, että $\mathbf{B}_m = \mathbf{A}$).

Koska matriisi **A** on positiivisesti definiitti, niin myös sen alimatriisit $\mathbf{B}_k$ ovat positiivisesti definiittejä, jolloin niiden determinantit ovat *positiivisia*:

$$\det(\mathbf{B}_k) > 0$$

Kun $k = 1$, niin

$$\mathbf{B}_1 = [a_{11}]$$

ja

$$\det(\mathbf{B}_1) = a_{11} > 0$$

Siten alkio l_{11} ja edelleen alkiot $l_{21}, \ldots , l_{m1}$ saadaan ratkaistuksi yhtälöryhmän ensimmäiseltä riviltä.

Kun $k = 2$, niin

$$\mathbf{B}_2 = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

ja

$$\det(\mathbf{B}_2) = a_{11}a_{22} - a_{12}^2 > 0$$

Koska lisäksi edellä todettiin, että $a_{11} > 0$, niin

$$\frac{a_{22}a_{11} - a_{21}^2}{a_{11}} > 0$$

ja siten alkio l_{22} ja edelleen alkiot $l_{32}, \ldots , l_{m2}$ saadaan ratkaistuksi yhtälöryhmän toiselta riviltä.

Tapaukset $k = 3, 4, \ldots , m$ voidaan käsitellä vastaavalla tavalla.

Huomautus 2:

Koska matriisi **A** on *symmetrinen*, toiselta riviltä saatava ratkaisu alkiolle l_{21} *ei ole ristiriidassa* ensimmäiseltä riviltä saadun ratkaisun kanssa:

$$l_{21} = \frac{a_{21}}{l_{11}} = \frac{a_{12}}{l_{11}}$$

Koska matriisi **A** on *symmetrinen*, kolmannelta riviltä saatavat ratkaisut alkioille l_{21} ja l_{32} *eivät ole ristiriidassa* ensimmäiseltä ja toiselta riviltä saatujen ratkaisujen kanssa:

$$l_{31} = \frac{a_{31}}{l_{11}} = \frac{a_{13}}{l_{11}}$$

$$l_{32} = \frac{a_{23} - l_{21}l_{31}}{l_{22}} = \frac{a_{32} - l_{21}l_{31}}{l_{22}}$$

Vastaava ilmiö tulee esiin myös muilla riveillä.

Todetaan lopuksi, että lauseen 2.5.2. mukaan matriisi **L** on oltava epäsingulaarinen, koska matriisi **A** on positiivisesti definiittinä matriisina epäsingulaarinen.

Todistus 2: Induktio-todistus.

Koska **A** on symmetrinen $m\times m$-matriisi, se voidaan esittää muodossa

$$\mathbf{A}_m = \begin{bmatrix} \mathbf{A}_{m-1} & \mathbf{c} \\ \mathbf{c}' & a_{mm} \end{bmatrix}$$

jossa **c** on $(m–1)$-vektori.

Asetetaan induktio-oletus:

$$\mathbf{A}_{m-1} = \mathbf{L}_{m-1}\mathbf{L}'_{m-1}$$

jossa $\mathbf{L}_{m-1}$ on epäsingulaarinen alakolmiomatriisi.

Olkoon

$$\mathbf{L}_m = \begin{bmatrix} \mathbf{L}_{m-1} & \mathbf{0} \\ \mathbf{l}' & \lambda \end{bmatrix}$$

jossa **l** on toistaiseksi tuntematon (m–1)-vektori ja λ on toistaiseksi tuntematon reaaliluku.

Muodostetaan matriisiyhtälö

$$\mathbf{A}_m = \mathbf{L}_m \mathbf{L}'_m$$

eli

$$\begin{bmatrix} \mathbf{L}_{m-1}\mathbf{L}'_{m-1} & \mathbf{L}_{m-1}\mathbf{l} \\ \mathbf{l}'\mathbf{L}'_{m-1} & \mathbf{l}'\mathbf{l} + \lambda^2 \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{m-1} & \mathbf{c} \\ \mathbf{c}' & a_{mm} \end{bmatrix}$$

Siten saamme (m–1)-vektorin **l** ja reaaliluvun λ ratkaisemiseksi yhtälöt

$$\mathbf{L}_{m-1}\mathbf{l} = \mathbf{c}$$

$$\mathbf{l}'\mathbf{l} + \lambda^2 = a_{mm}$$

Koska $\mathbf{L}_{m-1}$ on induktio-oletuksen mukaan epäsingulaarinen, niin vektorin **l** ratkaisuksi saadaan

$$\mathbf{l} = \mathbf{L}_{m-1}^{-1}\mathbf{c}$$

Edelleen

$$\begin{aligned} \lambda^2 &= a_{mm} - \mathbf{l}'\mathbf{l} \\ &= a_{mm} - \mathbf{c}'(\mathbf{L}_{m-1}^{-1})'\mathbf{L}_{m-1}^{-1}\mathbf{c} = a_{mm} - \mathbf{c}'(\mathbf{L}'_{m-1})^{-1}\mathbf{L}_{m-1}^{-1}\mathbf{c} = a_{mm} - \mathbf{c}'(\mathbf{L}_{m-1}\mathbf{L}'_{m-1})^{-1}\mathbf{c} \\ &= a_{mm} - \mathbf{c}'\mathbf{A}_{m-1}^{-1}\mathbf{c} > 0 \end{aligned}$$

mikä seuraa siitä, että

$$|\mathbf{A}_m| = |a_{mm} - \mathbf{c}'\mathbf{A}_{m-1}^{-1}\mathbf{c}||\mathbf{A}_{m-1}| > 0$$

ja

$$|\mathbf{A}_{m-1}| > 0$$

koska matriisi $\mathbf{A}_m$ on positiivisesti definiitti.

Siten saamme myös reaaliluvun λ ratkaistuksi:

$$\lambda = \pm\sqrt{a_{mm} - \mathbf{c}'\mathbf{A}_{m-1}^{-1}\mathbf{c}}$$

■

Huomautus 1:

Positiivisesti definiitin matriisin **A** esitystä epäsingulaarisen alakolmiomatriisin **L** ja sen transpoosin **L** tai epäsingulaaristen yläkolmiomatriisi **U** ja sen transpoosin tulona kutsutaan matriisin **A Cholesky-hajotelmaksi**.

Huomautus 2:

Cholesky-hajotelma *ei ole yksikäsitteinen*, mutta jos esitämme lisävaatimuksen, että matriisin **L** (tai matriisin **U**) kaikkien diagonaalialkioiden on oltava *positiivisia*, Cholesky-hajotelma *on yksikäsitteinen*.

QR-hajotelma

Lause 3.1.2.

Olkoon **A** $m\times n$-matriisi, jossa $m \geq n$ ja olkoon matriisin **A** *aste* $r(\mathbf{A}) = n$. Tällöin on olemassa $m\times n$-matriisi **Q**, jonka sarakkeet ovat *ortonormaalisia* eli $\mathbf{Q}'\mathbf{Q} = \mathbf{I}$, ja *epäsingulaarinen* $n\times n$*-yläkolmiomatriisi* **R** siten, että

$$\mathbf{A} = \mathbf{QR}$$

Todistus:

Matriisin **A** QR-hajotelma saadaan ortogonalisoimalla matriisin **A** sarakevektorit *Gramin ja Schmidtin ortogonalisointi-menetelmällä*; ks. lausetta 1.6.4.

Asetetaan ortogonalisoinnin tuloksena saatavat ortonormaalit vektorit matriisin **Q** sarakkeiksi. Matriisi **Q** saadaan kertomalla matriisi **A** epäsingulaarisella yläkolmiomatriisilla **T** oikealta:

$$\mathbf{AT} = \mathbf{Q}$$

$n\times n$-matriisi **T** on epäsingulaarinen, koska sen aste on n, mikä seuraa siitä, että sekä matriisin **A** että matriisin **Q** aste on n.

Koska matriisi **T** on yläkolmiomatriisi, niin myös sen käänteismatriisi

$$\mathbf{T}^{-1} = \mathbf{R}$$

on yläkolmiomatriisi. Siten

$$\mathbf{A} = \mathbf{QT}^{-1} = \mathbf{QR}$$

■

Huomautus 1:

Matriisin **A** esitystä sarakkeiltaan ortonormaalisen matriisin **Q** ja epäsingulaarisen yläkolmiomatriisin **R** tulona kutsutaan matriisin **A** **QR-hajotelmaksi**.

Huomautus 2:

Matriisin **A** sarakevektoreiden on oltava *lineaarisesti riippumattomia*, jotta matriisilla **A** olisi QR-hajotelma.

3.2. Yleistetyt käänteismatriisit

Singulaarisella matriisilla ei ole käänteismatriisia. Tarkastelemme tässä luvussa käänteismatriisin käsitteen yleistämistä ns. **yleistetyksi käänteismatriisiksi**, joka voidaan määrätä kaikille matriiseille. Matriisin **A** yleistettyä käänteismatriisia ei voida selvästikään määritellä epäsingulaarisen matriisin **C** käänteismatriisin **B** määrittelevällä yhtälöllä

$$\mathbf{CB} = \mathbf{BC} = \mathbf{I}$$

mutta sillä voi silti olla (järkevästi määriteltynä) riittävä määrä epäsingulaarisen matriisin käänteismatriisin muista ominaisuuksista ollakseen hyödyllinen käsite.

Käänteismatriisin käsitteen yleistäminen singulaarisille matriiseilla voi tapahtua usealla eri tavalla riippuen siitä, mitä ehtoja tarkasteltavan singulaarisen matriisin oletetaan toteuttavan ja mitä ehtoja yleistetyn käänteismatriisin halutaan toteuttavan.

Rajoitumme tässä esityksessä tarkastelemaan yleistettyä käänteismatriisia, joka on olemassa jokaiselle matriisille ja on lisäksi yksikäsitteinen. Tätä yleistettyä käänteismatriisia kutsutaan **Moore-Penrose-inverssiksi**.

Moore-Penrose-inverssi

Lause 3.2.1.

Jokaiselle $m\times n$-matriisille $\mathbf{A}$ on olemassa *yksikäsitteinen* $n\times m$-matriisi $\mathbf{A}^+$, joka toteuttaa seuraavat ehdot:

(i) $\mathbf{AA}^+\mathbf{A} = \mathbf{A}$

(ii) $\mathbf{A}^+\mathbf{AA}^+ = \mathbf{A}^+$

(iii) $(\mathbf{A}^+\mathbf{A})' = \mathbf{A}^+\mathbf{A}$

(iv) $(\mathbf{AA}^+)' = \mathbf{AA}^+$

Todistus:

Jos $\mathbf{A} = \mathbf{0}$, niin $\mathbf{A}^+ = \mathbf{0}'$. Oletetaan siis, että

$$r(\mathbf{A}) = r > 0$$

Tällöin on olemassa $m\times r$-matriisi $\mathbf{C}_1$, jonka aste on $r(\mathbf{C}_1) = r$ ja $r\times n$-matriisi $\mathbf{C}_2$, jonka aste $r(\mathbf{C}_2) = r$ siten, että

$$\mathbf{A} = \mathbf{C}_1\mathbf{C}_2$$

Sivuutamme tämän todistamisen.

Määritellään $n\times m$-matriisi

(1) $$\mathbf{A}^+ = \mathbf{C}_2'(\mathbf{C}_2\mathbf{C}_2')^{-1}(\mathbf{C}_1'\mathbf{C}_1)^{-1}\mathbf{C}_1'$$

Todetaan ensin, että

(2) $$\mathbf{A}^+\mathbf{A} = \mathbf{C}_2'(\mathbf{C}_2\mathbf{C}_2')^{-1}(\mathbf{C}_1'\mathbf{C}_1)^{-1}\mathbf{C}_1'\mathbf{C}_1\mathbf{C}_2 = \mathbf{C}_2'(\mathbf{C}_2\mathbf{C}_2')^{-1}\mathbf{C}_2$$

ja

(3) $$\mathbf{AA}^+ = \mathbf{C}_1\mathbf{C}_2\mathbf{C}_2'(\mathbf{C}_2\mathbf{C}_2')^{-1}(\mathbf{C}_1'\mathbf{C}_1)^{-1}\mathbf{C}_1' = \mathbf{C}_1(\mathbf{C}_1'\mathbf{C}_1)^{-1}\mathbf{C}_1'$$

(i) Soveltamalla kaavaa (2) nähdään, että

$$\mathbf{AA}^+\mathbf{A} = \mathbf{C}_1\mathbf{C}_2\mathbf{C}_2'(\mathbf{C}_2\mathbf{C}_2')^{-1}\mathbf{C}_2 = \mathbf{C}_1\mathbf{C}_2 = \mathbf{A}$$

(ii) Soveltamalla kaavoja (1) ja (3) nähdään, että

$$\begin{aligned}\mathbf{A}^+\mathbf{AA}^+ &= \mathbf{C}_2'(\mathbf{C}_2\mathbf{C}_2')^{-1}(\mathbf{C}_1'\mathbf{C}_1)^{-1}\mathbf{C}_1'\mathbf{C}_1(\mathbf{C}_1'\mathbf{C}_1)^{-1}\mathbf{C}_1' \\ &= \mathbf{C}_2'(\mathbf{C}_2\mathbf{C}_2')^{-1}(\mathbf{C}_1'\mathbf{C}_1)^{-1}\mathbf{C}_1' = \mathbf{A}^+\end{aligned}$$

(iii) Matriisi $\mathbf{A}^+\mathbf{A}$ on symmetrinen, koska transpoosin ominaisuuksista seuraa, että

$$(\mathbf{A}^+\mathbf{A})' = \mathbf{C}_2'(\mathbf{C}_2\mathbf{C}_2')^{-1}\mathbf{C}_2 = \mathbf{A}^+\mathbf{A}$$

(iv) Matriisi $\mathbf{AA}^+$ on symmetrinen, koska transpoosin ominaisuuksista seuraa, että

$$(\mathbf{AA}^+)' = \mathbf{C}_1(\mathbf{C}_1'\mathbf{C}_1)^{-1}\mathbf{C}_1' = \mathbf{AA}^+$$

Siten jokaiselle $m\times n$-matriisille **A** *on olemassa* matriisi $\mathbf{A}^+$, joka toteuttaa ehdot (i)-(iv).

Todistetaan vielä, että matriisi $\mathbf{A}^+$ on *yksikäsitteinen*. Oletetaan siksi, että sekä matriisi $\mathbf{A}^+$ että matriisi $\mathbf{B}^+$ toteuttavat ehdot (i)-(iv).

Siten erityisesti

$$\mathbf{AA}^+\mathbf{A} = \mathbf{A}$$

Kertomalla tämä yhtälö oikealta matriisilla $\mathbf{B}^+$ saadaan yhtälö

$$\mathbf{AA}^+\mathbf{AB}^+ = \mathbf{AB}^+$$

Vastaavasti kertomalla yhtälö

$$\mathbf{AB}^+\mathbf{A} = \mathbf{A}$$

vasemmalta matriisilla $\mathbf{A}^+$ saadaan yhtälö

$$\mathbf{AB}^+\mathbf{AA}^+ = \mathbf{AA}^+$$

Koska matriisit $\mathbf{AB}^+$ ja $\mathbf{AA}^+$ ovat ehdon (iv) mukaan symmetrisiä, niin

$$\mathbf{AB}^+ = \mathbf{AA}^+\mathbf{AB}^+ = ((\mathbf{AA}^+)(\mathbf{AB}^+))' = (\mathbf{AB}^+)'(\mathbf{AA}^+)' = \mathbf{AB}^+\mathbf{AA}^+ = \mathbf{AA}^+$$

Samaan tapaan voidaan osoittaa, että

$$\mathbf{B}^+\mathbf{A} = \mathbf{A}^+\mathbf{A}$$

Kertomalla yhtälö $\mathbf{AB}^+ = \mathbf{AA}^+$ vasemmalta matriisilla $\mathbf{B}^+$ saadaan yhtälöketju

$$\mathbf{B}^+ = \mathbf{B}^+\mathbf{AB}^+ = \mathbf{B}^+\mathbf{AA}^+ = \mathbf{A}^+\mathbf{AA}^+ = \mathbf{A}$$

joten matriisi $\mathbf{A}^+$ on yksikäsitteinen.

■

Kutsumme lauseessa 3.2.1 määriteltyä matriisia $\mathbf{A}^+$ matriisin **A Moore-Penrose-inverssiksi**. Moore-Penrose-inverssi voidaan määritellä myös matriisin *pääakselihajotelman* (ks. lausetta 3.2.4.) tai matriisin *singulaariarvohajotelman* avulla (ks. lausetta 3.3.4).

Lause 3.2.2.

Olkoon matriisin **A** yleistetty käänteismatriisi $\mathbf{A}^+$. Tällöin matriisin **A** transpoosin $\mathbf{A}'$ yleistetty käänteismatriisi on $(\mathbf{A}^+)'$.

Todistus:

Lause seuraa transponoimalla matriisin **A** yleistetyn käänteismatriisin määrittelevät ehdot (i)-(iv) lauseessa 3.2.1.

■

Moore-Penrose-inverssin ominaisuuksia

Lause 3.2.3.

Jos matriisi **A** on *epäsingulaarinen*, niin

$$\mathbf{A}^+ = \mathbf{A}^{-1}$$

Todistus:

Jos matriisi **A** on epäsingulaarinen, sillä on *yksikäsitteinen* käänteismatriisi $\mathbf{A}^{-1}$, joka toteuttaa ehdon

$$\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$$

Suoraan laskemalla nähdään, että matriisi $\mathbf{A}^{-1}$ toteuttaa myös lauseen 3.2.1. ehdot (i)-(iv), joten $\mathbf{A}^{-1}$ on epäsingulaarisen matriisin **A** yleistetty käänteismatriisi:

$$\mathbf{A}^{+} = \mathbf{A}^{-1}$$

■

Lause 3.2.4.

Olkoon **A** $m\times n$-matriisi ja $r(\mathbf{A}) = r$. Olkoon

$$\mathbf{A}'\mathbf{A} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}'$$

matriisin $\mathbf{A}'\mathbf{A}$ *pääakselihajotelma*, jossa

$$\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_r, 0, 0, \ldots, 0)$$

on matriisin $\mathbf{A}'\mathbf{A}$ *ominaisarvojen* muodostama *diagonaalimatriisi* ja **Q** on vastaavien *ominaisvektoreiden* muodostama *ortogonaalinen* matriisi, jossa ominaisvektorit ovat sarakkeina. Tällöin matriisin **A** *yleistetty käänteismatriisi* on muotoa

$$\mathbf{A}^{+} = (\mathbf{A}'\mathbf{A})^{+}\mathbf{A}'$$

missä

$$(\mathbf{A}'\mathbf{A})^{+} = \mathbf{Q}\mathbf{\Lambda}^{+}\mathbf{Q}'$$

ja

$$\mathbf{\Lambda}^{+} = \mathrm{diag}(\lambda_1^{-1}, \lambda_2^{-1}, \ldots, \lambda_r^{-1}, 0, 0, \ldots, 0)$$

Todistus:

Koska olemme olettaneet, että $r(\mathbf{A}) = r$, niin lauseen 2.5.2. mukaan myös $r(\mathbf{A}'\mathbf{A}) = r$. Siten matriisilla $\mathbf{A}'\mathbf{A}$ on pääakselihajotelma

$$\mathbf{A}'\mathbf{A} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}'$$

jossa

$$\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_r, 0, 0, \ldots, 0)$$

on matriisin $\mathbf{A}'\mathbf{A}$ ominaisarvojen muodostama diagonaalimatriisi ja **Q** on vastaavien ominaisvektoreiden muodostama ortogonaalinen matriisi, jossa ominaisvektorit ovat sarakkeina.

Määritellään matriisi

$$\mathbf{B} = (\mathbf{A}'\mathbf{A})^{+}\mathbf{A}'$$

jossa

$$(\mathbf{A}'\mathbf{A})^{+} = \mathbf{Q}\mathbf{\Lambda}^{+}\mathbf{Q}'$$

ja

$$\mathbf{\Lambda}^{+} = \mathrm{diag}(\lambda_1^{-1}, \lambda_2^{-1}, \ldots, \lambda_r^{-1}, 0, 0, \ldots, 0)$$

Suoraan laskemalla nähdään, että matriisi **B** toteuttaa lauseen 3.2.1. ehdot (i)-(iv), joten **B** on matriisin **A** yleistetty käänteismatriisi:

$$\mathbf{B} = (\mathbf{A}'\mathbf{A})^{+}\mathbf{A}' = \mathbf{A}^{+}$$

■

Lause 3.2.5.

Olkoon **A** $m\times n$-matriisi ja $r(\mathbf{A}) = n$. Tällöin

$$\mathbf{A}^{+} = (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'$$

Todistus:

Olkoon **A** $m\times n$-matriisi. Jos $r(\mathbf{A}) = n$, niin lauseen 2.5.2. mukaan myös $r(\mathbf{A}'\mathbf{A}) = r$. Siten matriisi $\mathbf{A}'\mathbf{A}$ on epäsingulaarinen ja lauseen 3.2.3. mukaan

$$(\mathbf{A}'\mathbf{A})^{+} = (\mathbf{A}'\mathbf{A})^{-1}$$

ja lauseen 3.2.4. mukaan

$$\mathbf{A}^{+} = (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'$$

■

Lause 3.2.6.

Olkoon $\mathbf{A}^{+}$ matriisin **A** yleistetty käänteismatriisi. Tällöin matriisit $\mathbf{A}^{+}\mathbf{A}$ ja $\mathbf{A}\mathbf{A}^{+}$ ovat *symmetrisiä* ja *idempotentteja* eli *projektioita*.

Todistus:

Olkoon $\mathbf{A}^{+}$ matriisin **A** yleistetty käänteismatriisi. Lauseen 3.2.1. kohdissa (iii) ja (iv) on todettu, että matriisit $\mathbf{A}^{+}\mathbf{A}$ ja $\mathbf{A}\mathbf{A}^{+}$ ovat symmetrisiä.

Suoraan laskemalla nähdään, että matriisit $\mathbf{A}^{+}\mathbf{A}$ ja $\mathbf{A}\mathbf{A}^{+}$ ovat idempotentteja:

Lauseen 3.2.1. kohdasta (ii) seuraa, että

$$(\mathbf{A}^{+}\mathbf{A})^2 = \mathbf{A}^{+}\mathbf{A}\mathbf{A}^{+}\mathbf{A} = \mathbf{A}^{+}\mathbf{A}$$

Lauseen 3.2.1. kohdasta (i) seuraa, että

$$(\mathbf{A}\mathbf{A}^{+})^2 = \mathbf{A}\mathbf{A}^{+}\mathbf{A}\mathbf{A}^{+} = \mathbf{A}\mathbf{A}^{+} \;\; (\mathbf{A}\mathbf{A}^{+})^2 = \mathbf{A}\mathbf{A}^{+}\mathbf{A}\mathbf{A}^{+} = \mathbf{A}\mathbf{A}^{+}$$

Siten matriisit $\mathbf{A}^{+}\mathbf{A}$ ja $\mathbf{A}\mathbf{A}^{+}$ ovat projektioita.

■

Lause 3.2.7.

Olkoon matriisisi **P** *symmetrinen* ja *idempotentti* eli *projektio*. Tällöin matriisin **P** *yleistetty käänteismatriisi* on matriisi **P** itse:

$$\mathbf{P}^{+} = \mathbf{P}$$

Todistus:

Olkoon matriisi **P** symmetrinen ja idempotentti eli projektio.

Tällöin

$$\mathbf{P}' = \mathbf{P}$$

ja

$$\mathbf{P}^2 = \mathbf{PP} = \mathbf{P}$$

Suoraan laskemalla nähdään, että matriisi **P** toteuttaa lauseen 3.2.1. ehdot (i)-(iv), joten **P** on matriisin **P** yleistetty käänteismatriisi:

$$\mathbf{P}^+ = \mathbf{P}$$

■

Sovellus: Lineaarinen regressiomalli

Olkoon

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

tavanomainen **lineaarinen regressiomalli**, jossa

$\mathbf{y}$ = **selitettävän muuttujan** *y havaittujen* ja *satunnaisten* arvojen y_i, $i = 1, 2, \dots, n$ muodostama n-vektori,

$\mathbf{X}$ = **selittävien muuttujien** x_j, $j = 1, 2, \dots, p$ *havaittujen* ja *kiinteiden* (*ei-satunnaisten*) arvojen x_{ij}, $i = 1, 2, \dots, n$, $j = 1, 2, \dots, p$ muodostama $n\times p$-matriisi, $n \ge p$, $r(\mathbf{X}) \le p$,

$\boldsymbol{\beta}$ = *tuntemattomien* ja *kiinteiden* **regressiokertoimien** β_j, $j = 1, 2, \dots, p$ muodostama p-vektori,

$\boldsymbol{\varepsilon}$ = **jäännös-** eli **virhetermin** ε *ei-havaittujen* ja *satunnaisten* arvojen ε_i, $i = 1, 2, \dots, n$ muodostama n-vektori.

Regressiokertoimien vektori $\boldsymbol{\beta}$ estimoidaan tavallisesti **pienimmän neliösumman menetelmällä**, jossa $\boldsymbol{\beta}$ pyritään valitsemaan siten, että neliösumma

$$(1) \qquad \sum_{i=1}^{n} \varepsilon_i^2 = \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = \mathbf{y}'\mathbf{y} - 2\boldsymbol{\beta}'\mathbf{X}'\mathbf{y} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta}$$

minimoituu.

Etsitään neliösumman (1) minimi *derivoimalla* neliösumma (1) vektorin $\boldsymbol{\beta}$ suhteen ja merkitsemällä derivaatta nollaksi. Soveltamalla matriisien derivointisääntöjä (ks. kappaletta 3.5.) saadaan *normaaliyhtälö*

$$(2) \qquad \frac{\partial}{\partial \boldsymbol{\beta}} \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} = -2\mathbf{X}'\mathbf{y} + 2\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = \mathbf{0}$$

vektorin $\boldsymbol{\beta}$ ratkaisemiseksi.

Jos matriisin **X** sarakkeet ovat *lineaarisesti riippumattomia* eli $r(\mathbf{X}) = p$, niin matriisi $\mathbf{X}'\mathbf{X}$ on *epäsingulaarinen* ja *normaaliyhtälöllä* (2) *on yksikäsitteinen ratkaisu*. Ratkaisuna on tavanomainen **pienimmän neliösumman estimaattori**

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

Normaaliyhtälön (2) ratkaisu **b** tuottaa todellakin neliösumman (1) *minimin*, koska

$$\frac{\partial^2}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'}\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} = 2\mathbf{X}'\mathbf{X}$$

ja matriisi $\mathbf{X}'\mathbf{X}$ on *positiivisesti definiitti*, jos matriisin **X** sarakkeet ovat *lineaarisesti riippumattomia* eli $r(\mathbf{X}) = p$.

Oletetaan nyt, että $r(\mathbf{X}) = r < p$. Tällöin matriisi $\mathbf{X}'\mathbf{X}$ on *singulaarinen* ja regressiokertoimien vektorin β *tavanomainen pienimmän neliösumman estimaattori ei ole olemassa*.

Olkoon

$$\mathbf{X}'\mathbf{X} = \mathbf{Q}\boldsymbol{\Lambda}\mathbf{Q}'$$

matriisin $\mathbf{X}'\mathbf{X}$ *pääakselihajotelma*, jossa

$$\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots , \lambda_r, 0, 0, \dots, 0)$$

on matriisin $\mathbf{X}'\mathbf{X}$ *ominaisarvojen* muodostama diagonaalimatriisi ja **Q** on vastaavien *ominaisvektoreiden* muodostama ortogonaalinen matriisi, jossa ominaisvektorit ovat sarakkeina.

Valitaan regressiokertoimien vektorin β *estimaattoriksi p*-vektori

$$\mathbf{b}^+ = \mathbf{X}^+\mathbf{y} = (\mathbf{X}'\mathbf{X})^+\mathbf{X}'\mathbf{y}$$

jossa

$$(\mathbf{X}'\mathbf{X})^+ = \mathbf{Q}\boldsymbol{\Lambda}^+\mathbf{Q}'$$

ja

$$\boldsymbol{\Lambda}^+ = \text{diag}(\lambda_1^{-1}, \lambda_2^{-1}, \dots, \lambda_r^{-1}, 0, 0, \dots, 0)$$

Estimaattori $\mathbf{b}^+$ on vektorin β **pääkomponenttiestimaattori**; ks. monisteen **Monimuuttujamenetelmät** lukua **Pääkomponenttianalyysi**.

Jos matriisin **X** sarakkeet ovat *lineaarisesti riippumattomia* eli $r(\mathbf{X}) = r = p$, niin

$$\mathbf{b}^+ = \mathbf{b}$$

3.3. Singulaariarvohajotelma

Olkoon **A** mielivaltainen $m\times n$-matriisi. Symmetrisen neliömatriisin *pääakselihajotelmaa vastaa* yleisessä tapauksessa ns. *singulaariarvohajotelma*.

Singulaariarvohajotelma

Lause 3.3.1.

Olkoon **A** $m\times n$-matriisi, $m \geq n$, $r(\mathbf{A}) = r$. Tällöin on olemassa $m\times n$-matriisi **P**, $n\times n$-*diagonaalimatriisi* **L** ja $n\times n$-matriisi **Q** siten, että

$$\mathbf{A} = \mathbf{PLQ}'$$

Lisäksi matriisin **P** sarakkeet ovat *ortonormaalisia* eli

$$\mathbf{P}'\mathbf{P} = \mathbf{I}$$

ja matriisi **Q** on *ortogonaalinen* eli

$$\mathbf{Q}'\mathbf{Q} = \mathbf{Q}\mathbf{Q}' = \mathbf{I}$$

Todistus:

Olkoon $\mathbf{A}$ $m\times n$-matriisi, $m \geq n$, $r(\mathbf{A}) = r$, jolloin $r(\mathbf{A}'\mathbf{A}) = r$. Tällöin matriisin $\mathbf{A}'\mathbf{A}$ pääakselihajotelma on muotoa

$$\mathbf{A}'\mathbf{A} = \mathbf{Q}\mathbf{L}^2\mathbf{Q}'$$

jossa

$$\mathbf{L}^2 = \begin{bmatrix} \mathbf{L}_1^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$$

ja

$$\mathbf{L}_1^2 = \mathrm{diag}(l_1^2, l_2^2, \ldots, l_r^2)$$

on matriisin $\mathbf{A}'\mathbf{A}$ positiivisten ominaisarvojen $l_1^2, l_2^2, \ldots, l_r^2$ muodostama diagonaalimatriisi ja $\mathbf{Q}$ on matriisin $\mathbf{A}'\mathbf{A}$ ominaisvektoreiden muodostama ortogonaalinen $n\times n$-matriisi:

$$\mathbf{Q}'\mathbf{Q} = \mathbf{Q}\mathbf{Q}' = \mathbf{I}$$

Ositetaan matriisi $\mathbf{Q}$ sarakkeittensa suhteen seuraavalla tavalla:

$$\mathbf{Q} = [\mathbf{Q}_1 \ \mathbf{Q}_2]$$

jossa $n\times r$-matriisin $\mathbf{Q}_1$ sarakkeina on matriisin $\mathbf{A}'\mathbf{A}$ positiivisia ominaisarvoja $l_1^2, l_2^2, \ldots, l_r^2$ vastaavat ominaisvektorit ja $n\times(n-r)$-matriisin $\mathbf{Q}_2$ sarakkeina on matriisin nollaominaisarvoja vastaavat ominaisvektorit.

Siten matriisin $\mathbf{A}'\mathbf{A}$ pääakselihajotelma voidaan kirjoittaa muotoon

$$\mathbf{A}'\mathbf{A} = \mathbf{Q}_1\mathbf{L}_1^2\mathbf{Q}_1'$$

josta seuraa yhtälö

$$\mathbf{Q}_1'\mathbf{A}'\mathbf{A}\mathbf{Q}_1 = \mathbf{L}_1^2$$

Määritellään $m\times r$-matriisi $\mathbf{P}_1$ seuraavalla kaavalla:

$$\mathbf{P}_1 = \mathbf{A}\mathbf{Q}_1\mathbf{L}_1^{-1}$$

jossa siis

$$\mathbf{L}_1 = \mathrm{diag}(l_1, l_2, \ldots, l_r)$$

Matriisin $\mathbf{P}_1$ sarakkeet ovat ortonormaalisia, koska

$$\mathbf{P}_1'\mathbf{P}_1 = \mathbf{L}_1^{-1}\mathbf{Q}_1\mathbf{A}'\mathbf{A}\mathbf{Q}_1\mathbf{L}_1^{-1} = \mathbf{L}_1^{-1}\mathbf{L}_1^2\mathbf{L}_1^{-1} = \mathbf{I}_r$$

Koska olemme olettaneet, että $m \geq n$, matriisi $\mathbf{P}_1$ voidaan täydentää Gramin ja Schmidtin ortogonalisointimenetelmällä ortogonaaliseksi $m\times m$-matriisiksi

$$\mathbf{P}^* = [\mathbf{P}_1 \ \mathbf{P}_2]$$

Kertomalla yhtälö $\mathbf{P}_1 = \mathbf{A}\mathbf{Q}_1\mathbf{L}_1^{-1}$ oikealta matriisilla $\mathbf{L}_1\mathbf{Q}_1'$ saadaan matriisille $\mathbf{A}$ esitysmuoto

$$\begin{aligned}\mathbf{A} &= \mathbf{P}_1\mathbf{L}_1\mathbf{Q}_1' \\ &= \begin{bmatrix}\mathbf{P}_1 & \mathbf{P}_2\end{bmatrix}\begin{bmatrix}\mathbf{L}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0}\end{bmatrix}\begin{bmatrix}\mathbf{Q}_1' \\ \mathbf{Q}_2'\end{bmatrix} \\ &= \mathbf{P}^*\mathbf{L}^*\mathbf{Q}'\end{aligned}$$

jossa $n\times n$-matriisi **Q** on määritelty edellä ja $\mathbf{L}^*$ on $m\times n$-matriisi, jonka vasemman yläkulman lohkon muodostaa matriisi $\mathbf{L}_1$. Koska matriisin $\mathbf{L}^*$ $(m-n)$ viimeistä riviä ovat nollia, matriisi **A** voidaan esittää myös muodossa

$$\mathbf{A} = \mathbf{PLQ}'$$

jossa $m\times n$-matriisi **P** saadaan ottamalla matriisista $\mathbf{P}^*$ n ensimmäistä saraketta, $n\times n$-diagonaalimatriisi **L** ja $n\times n$-matriisi **Q** on määritelty edellä. Lisäksi matriisin **P** sarakkeet ovat ortonormaalisia ja matriisi **Q** on ortogonaalinen.

■

Kutsumme lauseessa 3.3.1 määriteltyä matriisihajotelmaa $\mathbf{A} = \mathbf{PLQ}'$ matriisin **A** **singulaariarvohajotelmaksi**.

Lauseessa 3.3.1. on oletettu, että **A** on $m\times n$-matriisi, jossa $m \geq n$. Jos $m \leq n$, niin matriisin **A** singulaariarvohajotelma saadaan transponoimalla matriisin **A´** singulaariarvohajotelma

$$\mathbf{A}' = \mathbf{PLQ}'$$

eli tällöin matriisin **A** singulaariarvohajotelma on muotoa

$$\mathbf{A} = \mathbf{QLP}'$$

Singulaariarvot

Matriisin **A** singulaariarvohajotelman $\mathbf{A} = \mathbf{PLQ}'$ *diagonaalimatriisin*

$$\mathbf{L} = \mathrm{diag}(l_1 , l_2 , \ldots , l_r)$$

diagonaalialkioita kutsutaan matriisin **A** **singulaariarvoiksi**.

Singulaariarvohajotelman tulkinta

Jokainen lineaarikuvaus voidaan esittää yhdistämällä *kierto*, *venytys* ja *kierto* (kierrot yhdistettyinä mahdollisiin *peilauksiin*), koska kuvausta vastaava matriisi **A** voidaan esittää kiertoa vastaavan ortogonaalisen matriisin **Q´**, venytystä vastaavan diagonaalimatriisin **L** ja toista kiertoa vastaavan ortogonaalisen matriisin **P** tulona.

Singulaariarvohajotelman ominaisuuksia

Olkoon **A** $m\times n$-matriisi, $m \geq n$, $r(\mathbf{A}) = r$ ja olkoon

$$\mathbf{A} = \mathbf{PLQ}'$$

matriisin **A** *singulaariarvohajotelma*, missä

P = *sarakeiltaan ortonormaalinen* $m\times n$-matriisi: $\mathbf{P}'\mathbf{P} = \mathbf{I}$

Q = *ortogonaalinen* $n\times n$-matriisi: $\mathbf{Q}'\mathbf{Q} = \mathbf{QQ}' = \mathbf{I}$

L = $\mathrm{diag}(l_1, l_2, \ldots , l_r, 0, \ldots , 0)$ on $n\times n$-diagonaalimatriisi

Olkoot

$$\mathbf{P} = [\mathbf{p}_1\ \mathbf{p}_2\ \ldots\ \mathbf{p}_n]$$

$$\mathbf{Q} = [\mathbf{q}_1\ \mathbf{q}_2\ \ldots\ \mathbf{q}_n]$$

missä

$\mathbf{p}_i$ = matriisin $\mathbf{P}$ *i. sarakevektori*, $i = 1, 2, \ldots, n$

$\mathbf{q}_i$ = matriisin $\mathbf{Q}$ *i. sarakevektori*, $i = 1, 2, \ldots, n$

Lause 3.3.2.

Olkoon $\mathbf{A}$ $m\times n$-matriisi, $m \geq n$, $r(\mathbf{A}) = r$. Matriisi $\mathbf{A}$ voidaan aina esittää positiivisten singulaariarvojensa l_i , $i = 1, 2, \ldots, r$ ja niitä vastaavien vektoreiden $\mathbf{p}_i$ ja $\mathbf{q}_i$, $i = 1, 2, \ldots, r$ avulla muodossa

$$\mathbf{A} = \sum_{i=1}^{r} l_i \mathbf{p}_i \mathbf{q}_i'$$

Lause 3.3.3.

Olkoon $\mathbf{A}$ $m\times n$-matriisi, $r(\mathbf{A}) = r$ ja olkoon

$$\mathbf{A} = \mathbf{PLQ}'$$

matriisin $\mathbf{A}$ *singulaariarvohajotelma*. Tällöin pätee:

(i) $\mathbf{P}'\mathbf{AQ} = \mathbf{L}$

(ii) $\mathbf{A}'\mathbf{A} = \mathbf{QL}^2\mathbf{Q}'$

(iii) $\mathbf{AA}' = \mathbf{PL}^2\mathbf{P}'$

Todistus:

(i) Kertomalla matriisin $\mathbf{A}$ singulaariarvohajotelma $\mathbf{A} = \mathbf{PLQ}'$ vasemmalta matriisilla $\mathbf{P}'$ ja oikealta matriisilla $\mathbf{Q}$ saadaan yhtälö

$$\mathbf{P}'\mathbf{AQ} = \mathbf{P}'\mathbf{PLQ}'\mathbf{Q} = \mathbf{L}$$

(ii) $\mathbf{AA}' = \mathbf{PLQ}'\mathbf{QLP}' = \mathbf{PL}^2\mathbf{P}'$

(iii) $\mathbf{A}'\mathbf{A} = \mathbf{QLP}'\mathbf{PLQ}' = \mathbf{QL}^2\mathbf{Q}'$

■

Huomautus 1:

Matriiseilla $\mathbf{A}'\mathbf{A}$ ja $\mathbf{AA}'$ on Lauseen 3.3.3. kohtien (ii) ja (iii) mukaan *samat ominaisarvot*, mutta *eri ominaisvektorit*.

Huomautus 2:

Matriisien $\mathbf{A}'\mathbf{A}$ ja $\mathbf{AA}'$ ominaisarvoina ovat Lauseen 3.3.3. kohtien (ii) ja (iii) mukaan matriisin $\mathbf{A}$ *singulaariarvojen neliöt*.

Lause 3.3.4.

Olkoon $\mathbf{A}$ $m\times n$-matriisi, $m \geq n$, $r(\mathbf{A}) = r$ ja olkoon

$$\mathbf{A} = \mathbf{PLQ}'$$

matriisin $\mathbf{A}$ *singulaariarvohajotelma*. Tällöin matriisin $\mathbf{A}$ *Moore-Penrose-inverssi* $\mathbf{A}^+$ voidaan esittää muodossa

$$\mathbf{A}^{+} = \mathbf{Q}\mathbf{L}^{+}\mathbf{P}'$$

jossa

$$\mathbf{L}^{+} = \mathrm{diag}(l_1^{-1}, l_2^{-1}, \ldots, l_r^{-1}, 0, 0, \ldots, 0)$$

Todistus:

Olkoon **A** $m\times n$-matriisi, $m \geq n$, $r(\mathbf{A}) = r$ ja olkoon

$$\mathbf{A} = \mathbf{P}\mathbf{L}\mathbf{Q}'$$

matriisin **A** *singulaariarvohajotelma*.

Suoraan laskemalla nähdään, että matriisi $\mathbf{Q}\mathbf{L}^{+}\mathbf{P}'$ toteuttaa lauseen 3.2.1. ehdot, joten matriisi $\mathbf{Q}\mathbf{L}^{+}\mathbf{P}'$ on matriisin **A** Moore-Penrose-inverssi:

$$\mathbf{A}^{+} = \mathbf{Q}\mathbf{L}^{+}\mathbf{P}'$$

■

Sovellus: Lineaarinen regressiomalli

Olkoon

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

kappaleessa 3.2. määritelty tavanomainen **lineaarinen regressiomalli**, jossa **y** on *selitettävän muuttujan y* havaittujen ja satunnaisten arvojen muodostama n-vektori, **X** on *selittävien muuttujien* x_j , $j = 1, 2, \ldots , p$ havaittujen ja ei-satunnaisten arvojen muodostama $n\times p$-matriisi, $n \geq p$, $r(\mathbf{X}) \leq p$, $\boldsymbol{\beta}$ on *tuntemattomien regressiokertoimien* muodostama p-vektori ja $\boldsymbol{\varepsilon}$ on *jäännös*- eli *virhetermin* ε ei-havaittujen ja satunnaisten arvojen muodostama n-vektori.

Jos matriisi **X** on *täysiasteinen* eli $r(\mathbf{X}) = p$, niin vektorin $\boldsymbol{\beta}$ tavanomaiseksi **pienimmän neliösumman estimaattoriksi** saadaan (ks. kappaletta 3.2.)

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

Jos matriisi **X** *ei ole täysiasteinen* eli $r(\mathbf{X}) = r < p$, niin *tavanomainen pienimmän neliösumman estimaattori ei ole olemassa*. Tällöin regressiokertoimien vektorin $\boldsymbol{\beta}$ *estimaattoriksi* voidaan valita vektorin $\boldsymbol{\beta}$ **pääkomponenttiestimaattori**

$$\mathbf{b}^{+} = \mathbf{X}^{+}\mathbf{y}$$

jossa $\mathbf{X}^{+}$ on matriisin **X** *Moore-Penrose-inverssi*.

Lauseen 3.3.4. mukaan vektorin $\boldsymbol{\beta}$ pääkomponenttiestimaattori $\mathbf{b}^{+}$ voidaan esittää matriisin **X** *singulaariarvohajotelman*

$$\mathbf{X} = \mathbf{P}\mathbf{L}\mathbf{Q}'$$

avulla muodossa

$$\mathbf{b}^{+} = \mathbf{X}^{+}\mathbf{y} = \mathbf{Q}\mathbf{L}^{+}\mathbf{P}'\mathbf{y}$$

Jos matriisin **X** sarakkeet ovat *lineaarisesti riippumattomia* eli $r(\mathbf{X}) = r = p$, niin

$$\mathbf{b}^{+} = \mathbf{b}$$

3.4. Matriisien kääntäminen

Lineaaristen mallien ja *monimuuttujamenetelmien* teoriassa on usein hyötyä seuraavasta *käänteismatriisin laskukaavasta* ja sen erikoistapauksista.

Käänteismatriisin laskukaavat

Lause 3.4.1.

Olkoon $\mathbf{A}$ *epäsingulaarinen* $m\times m$-matriisi, $\mathbf{x}$ ja $\mathbf{y}$ kaksi m-vektoria sekä $k \in \mathbb{R}$. Tällöin

$$(\mathbf{A}+k\mathbf{x}\mathbf{y}')^{-1} = \mathbf{A}^{-1} - \frac{k\mathbf{A}^{-1}\mathbf{x}\mathbf{y}'\mathbf{A}^{-1}}{1+k\mathbf{y}'\mathbf{A}^{-1}\mathbf{x}}$$

Todistus:

Merkitään

$$c = \frac{k}{1+k\mathbf{y}'\mathbf{A}^{-1}\mathbf{x}}$$

jolloin

$$\mathbf{A}^{-1} - \frac{k\mathbf{A}^{-1}\mathbf{x}\mathbf{y}'\mathbf{A}^{-1}}{1+k\mathbf{y}'\mathbf{A}^{-1}\mathbf{x}} = \mathbf{A}^{-1} - c\mathbf{A}^{-1}\mathbf{x}\mathbf{y}'\mathbf{A}^{-1}$$

Suoraan laskemalla saadaan

$$\begin{aligned}
\left[\mathbf{A}+k\mathbf{x}\mathbf{y}'\right]&\left[\mathbf{A}^{-1} - c\mathbf{A}^{-1}\mathbf{x}\mathbf{y}'\mathbf{A}^{-1}\right] \\
&= \mathbf{I} - c\mathbf{x}\mathbf{y}'\mathbf{A}^{-1} + k\mathbf{x}\mathbf{y}'\mathbf{A}^{-1} - ck\mathbf{x}\mathbf{y}'\mathbf{A}^{-1}\mathbf{x}\mathbf{y}'\mathbf{A}^{-1} \\
&= \mathbf{I} + \mathbf{x}\left[k - c(1+k\mathbf{y}'\mathbf{A}^{-1}\mathbf{x})\right]\mathbf{y}'\mathbf{A}^{-1} \\
&= \mathbf{I} + \left[k - c(1+k\mathbf{y}'\mathbf{A}^{-1}\mathbf{x})\right]\mathbf{x}\mathbf{y}'\mathbf{A}^{-1} \\
&= \mathbf{I}
\end{aligned}$$

koska

$$c(1+k\mathbf{y}'\mathbf{A}^{-1}\mathbf{x}) = k$$

Vastaavalla tavalla saadaan

$$\left[\mathbf{A}^{-1} - c\mathbf{A}^{-1}\mathbf{x}\mathbf{y}'\mathbf{A}^{-1}\right]\left[\mathbf{A}+k\mathbf{x}\mathbf{y}'\right] = \mathbf{I}$$

Yhdistämällä saadut tulokset nähdään, että

$$(\mathbf{A}+k\mathbf{x}\mathbf{y}')^{-1} = \mathbf{A}^{-1} - c\mathbf{A}^{-1}\mathbf{x}\mathbf{y}'\mathbf{A}^{-1}$$

jossa

$$c = k/(1+k\mathbf{y}'\mathbf{A}^{-1}\mathbf{x})$$

■

Lause 3.4.2.

Olkoon **A** *epäsingulaarinen* $m\times m$-matriisi, **C** *epäsingulaarinen* $n\times n$-matriisi ja **B** $m\times n$-matriisi. Tällöin

$$
\begin{aligned}
(\mathbf{A}+\mathbf{BCB}')^{-1} &= \mathbf{A}^{-1}-\mathbf{A}^{-1}\mathbf{B}(\mathbf{B}'\mathbf{A}^{-1}\mathbf{B}+\mathbf{C}^{-1})^{-1}\mathbf{B}'\mathbf{A}^{-1} \\
&= \mathbf{A}^{-1}-\mathbf{A}^{-1}\mathbf{B}(\mathbf{B}'\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{B}'\mathbf{A}^{-1} \\
&\quad +\mathbf{A}^{-1}\mathbf{B}(\mathbf{B}'\mathbf{A}^{-1}\mathbf{B})^{-1}\left[(\mathbf{B}'\mathbf{A}^{-1}\mathbf{B})^{-1}+\mathbf{C}\right]^{-1}(\mathbf{B}'\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{B}'\mathbf{A}^{-1}
\end{aligned}
$$

Todistus:

Suoraan laskemalla saadaan

$$
\begin{aligned}
&\left[\mathbf{A}+\mathbf{BCB}'\right]\left[\mathbf{A}^{-1}-\mathbf{A}^{-1}\mathbf{B}(\mathbf{B}'\mathbf{A}^{-1}\mathbf{B}+\mathbf{C}^{-1})^{-1}\mathbf{B}'\mathbf{A}^{-1}\right] \\
&= \mathbf{I}-\mathbf{B}(\mathbf{B}'\mathbf{A}^{-1}\mathbf{B}+\mathbf{C}^{-1})^{-1}\mathbf{B}'\mathbf{A}^{-1} \\
&\quad +\mathbf{BCB}'\mathbf{A}^{-1}-\mathbf{BCB}'\mathbf{A}^{-1}\mathbf{B}(\mathbf{B}'\mathbf{A}^{-1}\mathbf{B}+\mathbf{C}^{-1})^{-1}\mathbf{B}'\mathbf{A}^{-1} \\
&= \mathbf{I}-\left[\mathbf{B}-\mathbf{BC}(\mathbf{B}'\mathbf{A}^{-1}\mathbf{B}+\mathbf{C}^{-1})+\mathbf{BCB}'\mathbf{A}^{-1}\mathbf{B}\right](\mathbf{B}'\mathbf{A}^{-1}\mathbf{B}+\mathbf{C}^{-1})^{-1} \\
&= \mathbf{I}-\left[\mathbf{B}-\mathbf{BCB}'\mathbf{A}^{-1}\mathbf{B}-\mathbf{BCC}^{-1}+\mathbf{BCB}'\mathbf{A}^{-1}\mathbf{B}\right](\mathbf{B}'\mathbf{A}^{-1}\mathbf{B}+\mathbf{C}^{-1})^{-1} \\
&= \mathbf{I}-\left[\mathbf{B}-\mathbf{BCB}'\mathbf{A}^{-1}\mathbf{B}-\mathbf{B}+\mathbf{BCB}'\mathbf{A}^{-1}\mathbf{B}\right](\mathbf{B}'\mathbf{A}^{-1}\mathbf{B}+\mathbf{C}^{-1})^{-1}=\mathbf{I}
\end{aligned}
$$

Vastaavalla tavalla saadaan

$$
\left[\mathbf{A}^{-1}-\mathbf{A}^{-1}\mathbf{B}(\mathbf{B}'\mathbf{A}^{-1}\mathbf{B}+\mathbf{C}^{-1})^{-1}\mathbf{B}'\mathbf{A}^{-1}\right]\left[\mathbf{A}+\mathbf{BCB}'\right]=\mathbf{I}
$$

Yhdistämällä saadut tulokset nähdään, että

$$
(\mathbf{A}+\mathbf{BCB}')^{-1}=\mathbf{A}^{-1}-\mathbf{A}^{-1}\mathbf{B}(\mathbf{B}'\mathbf{A}^{-1}\mathbf{B}+\mathbf{C}^{-1})^{-1}\mathbf{B}'\mathbf{A}^{-1}
$$

Kaavan toinen muoto saadaan todistetuksi juuri johdettua kaavaa lausekkeeseen

$$
(\mathbf{B}'\mathbf{A}^{-1}\mathbf{B}+\mathbf{C}^{-1})^{-1}
$$

■

Lauseen 3.4.2. erikoistapauksena saadaan seuraava lause:

Lause 3.4.3.

(i) Olkoon **A** *epäsingulaarinen* $m\times m$-matriisi ja **B** $m\times n$-matriisi. Tällöin

$$
\begin{aligned}
(\mathbf{A}+\mathbf{BB}')^{-1} &= \mathbf{A}^{-1}-\mathbf{A}^{-1}\mathbf{B}(\mathbf{B}'\mathbf{A}^{-1}\mathbf{B}+\mathbf{I})^{-1}\mathbf{B}'\mathbf{A}^{-1} \\
&= \mathbf{A}^{-1}-\mathbf{A}^{-1}\mathbf{B}(\mathbf{B}'\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{B}'\mathbf{A}^{-1} \\
&\quad +\mathbf{A}^{-1}\mathbf{B}(\mathbf{B}'\mathbf{A}^{-1}\mathbf{B})^{-1}\left[(\mathbf{B}'\mathbf{A}^{-1}\mathbf{B})^{-1}+\mathbf{I}\right]^{-1}(\mathbf{B}'\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{B}'\mathbf{A}^{-1}
\end{aligned}
$$

(ii) Olkoon **A** *epäsingulaarinen* $m\times m$-matriisi ja **x** m-vektori. Tällöin

$$
\begin{aligned}
(\mathbf{A}+\mathbf{xx}')^{-1} &= \mathbf{A}^{-1}-\mathbf{A}^{-1}\mathbf{x}(\mathbf{x}'\mathbf{A}^{-1}\mathbf{x}+1)^{-1}\mathbf{x}'\mathbf{A}^{-1} \\
&= \mathbf{A}^{-1}-\mathbf{A}^{-1}\mathbf{x}(\mathbf{x}'\mathbf{A}^{-1}\mathbf{x})^{-1}\mathbf{x}'\mathbf{A}^{-1} \\
&\quad +\mathbf{A}^{-1}\mathbf{x}(\mathbf{x}'\mathbf{A}^{-1}\mathbf{x})^{-1}\left[(\mathbf{x}'\mathbf{A}^{-1}\mathbf{x})^{-1}+1\right]^{-1}(\mathbf{x}'\mathbf{A}^{-1}\mathbf{x})^{-1}\mathbf{x}'\mathbf{A}^{-1}
\end{aligned}
$$

3.5. Matriisien derivointi

1. kertaluvun osittaisderivaattaoperaattoreiden vektori

Olkoon $\mathbf{x} = (x_1, x_2, \dots, x_p)$ p-vektori. Tällöin **1. kertaluvun osittaisderivaatta-operaattoreiden vektori** on p-vektori

$$\mathbf{D}_\mathbf{x} = \frac{\partial}{\partial \mathbf{x}} = \left(\frac{\partial}{\partial x_1}, \frac{\partial}{\partial x_2}, \cdots, \frac{\partial}{\partial x_p} \right)$$

missä

$$\frac{\partial}{\partial x_i} = 1.\ \textit{kertaluvun osittaisderivaattaoperaattori muuttujan } x_i \textit{ suhteen},$$

$$i = 1, 2, \dots, p$$

Vektorin $\mathbf{D}_\mathbf{x}$ *transpoosi* on $1 \times p$-matriisi

$$\mathbf{D}'_\mathbf{x} = \frac{\partial}{\partial \mathbf{x}'} = \begin{bmatrix} \dfrac{\partial}{\partial x_1} & \dfrac{\partial}{\partial x_2} & \cdots & \dfrac{\partial}{\partial x_p} \end{bmatrix}$$

Funktion gradientti

Olkoon

$$y = f(\mathbf{x}) = f(x_1, x_2, \dots, x_p)$$

avaruuden $\mathbb{R}^p$ reaaliarvoinen funktio:

$$f : \mathbb{R}^p \to \mathbb{R}$$

Tällöin p-vektori

$$\mathbf{D}_\mathbf{x} f(\mathbf{x}) = \frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} = \left(\frac{\partial f(\mathbf{x})}{\partial x_1}, \frac{\partial f(\mathbf{x})}{\partial x_2}, \cdots, \frac{\partial f(\mathbf{x})}{\partial x_p} \right)$$

on funktion f **gradientti** pisteessä $\mathbf{x}$. Gradienttivektori osoittaa pisteestä $\mathbf{x}$ siihen suuntaan, jossa funktio f kasvaa nopeimmin.

2. kertaluvun osittaisderivaattaoperaattoreiden matriisi

Olkoon $\mathbf{x} = (x_1, x_2, \dots, x_p)$ p-vektori. Tällöin **2. kertaluvun osittaisderivaatta-operaattoreiden matriisi** on $p \times p$-matriisi

$$\mathbf{D}^2_\mathbf{x} = \mathbf{D}_\mathbf{x}\mathbf{D}'_\mathbf{x} = \frac{\partial}{\partial \mathbf{x}} \frac{\partial}{\partial \mathbf{x}'} = \frac{\partial^2}{\partial \mathbf{x} \partial \mathbf{x}'} = \begin{bmatrix} \dfrac{\partial^2}{\partial x_1^2} & \dfrac{\partial^2}{\partial x_1 \partial x_2} & \cdots & \dfrac{\partial^2}{\partial x_1 \partial x_p} \\ \dfrac{\partial^2}{\partial x_2 \partial x_1} & \dfrac{\partial^2}{\partial x_2^2} & \cdots & \dfrac{\partial^2}{\partial x_2 \partial x_p} \\ \vdots & \vdots & & \vdots \\ \dfrac{\partial^2}{\partial x_p \partial x_1} & \dfrac{\partial^2}{\partial x_p \partial x_2} & \cdots & \dfrac{\partial^2}{\partial x_p^2} \end{bmatrix}$$

missä

$$\frac{\partial^2}{\partial x_i \partial x_j} = 2.\ \textit{kertaluvun osittaisderivaattaoperaattori muuttujien } x_i \text{ ja } x_j$$

suhteen, $i, j = 1, 2, \dots, p$

1. kertaluvun osittaisderivaattaoperaattoreiden matriisi

Olkoon $\mathbf{X} = [x_{ij}]$ $n \times p$-matriisi. Tällöin **1. kertaluvun osittaisderivaattaoperaattoreiden matriisi on** $n \times p$-matriisi

$$\frac{\partial}{\partial \mathbf{X}} = \begin{bmatrix} \frac{\partial}{\partial x_{11}} & \frac{\partial}{\partial x_{12}} & \cdots & \frac{\partial}{\partial x_{1p}} \\ \frac{\partial}{\partial x_{21}} & \frac{\partial}{\partial x_{22}} & \cdots & \frac{\partial}{\partial x_{2p}} \\ \vdots & \vdots & & \vdots \\ \frac{\partial}{\partial x_{n1}} & \frac{\partial}{\partial x_{n2}} & \cdots & \frac{\partial}{\partial x_{np}} \end{bmatrix}$$

missä

$$\frac{\partial}{\partial x_{ij}} = 1.\ \textit{kertaluvun osittaisderivaattaoperaattori muuttujan } x_{ij} \textit{ suhteen},$$

$i = 1, 2, \dots, n, j = 1, 2, \dots, p$

Seuraavassa esitetään joukko hyödyllisiä matriisilausekkeiden derivointikaavoja.

Vektoreiden lineaarikombinaatioiden derivointi

Lause 3.5.1.

Olkoot **a** ja **x** p-vektoreita. Tällöin

(i) $$\frac{\partial}{\partial \mathbf{x}}(\mathbf{a}'\mathbf{x}) = \mathbf{a}$$

(ii) $$\frac{\partial}{\partial \mathbf{x}'}(\mathbf{a}'\mathbf{x}) = \mathbf{a}'$$

Todistus:

Olkoot $\mathbf{a} = (a_1, a_2, \dots, a_p)$ ja $\mathbf{x} = (x_1, x_2, \dots, x_p)$ p-vektoreita.

(i) Koska

$$\mathbf{a}'\mathbf{x} = \sum_{i=1}^{p} a_i x_i$$

niin

$$\frac{\partial}{\partial x_i}(\mathbf{a}'\mathbf{x}) = a_i\ ,\ i = 1, 2, \dots, p$$

ja siten

$$\frac{\partial}{\partial \mathbf{x}}(\mathbf{a}'\mathbf{x}) = (a_1, a_2, \ldots, a_p) = \mathbf{a}$$

(ii) Kohta (ii) seuraa kohdasta (i) transponoimalla.

■

Vektorin normin neliön derivointi

Lause 3.5.2.

Olkoon $\mathbf{x}$ p-vektori. Tällöin

$$\frac{\partial}{\partial \mathbf{x}}(\mathbf{x}'\mathbf{x}) = 2\mathbf{x}$$

Todistus:

Olkoon $\mathbf{x} = (x_1, x_2, \ldots , x_p)$ p-vektori. Koska

$$\mathbf{x}'\mathbf{x} = \sum_{i=1}^{p} x_i^2$$

niin

$$\frac{\partial}{\partial x_i}(\mathbf{x}'\mathbf{x}) = 2x_i \text{ , } i = 1, 2, \ldots, p$$

ja siten

$$\frac{\partial}{\partial \mathbf{x}}(\mathbf{x}'\mathbf{x}) = (2x_1, 2x_2, \ldots, 2x_p) = 2\mathbf{x}$$

■

Huomaa, että

$$\|\mathbf{x}\| = \sqrt{\mathbf{x}'\mathbf{x}}$$

on vektorin $\mathbf{x}$ *normi* eli *pituus*.

Bilineaarimuotojen ja neliömuotojen derivointi

Lause 3.5.3.

Olkoot $\mathbf{A}$ $n\times p$-matriisi ja $\mathbf{y}$ p-vektori. Tällöin

$$\frac{\partial}{\partial \mathbf{y}}(\mathbf{A}\mathbf{y}) = \mathbf{A}'$$

Todistus:

Olkoot $\mathbf{A} = [a_{ij}]$ $n\times p$-matriisi ja $\mathbf{y} = (y_1, y_2, \ldots , y_p)$ p-vektori. Tällöin $\mathbf{Ay}$ on n-vektori, jonka i. alkio on

$$[\mathbf{Ay}]_i = \sum_{j=1}^{p} a_{ij} y_j \text{ , } i = 1, 2, \ldots, n$$

Koska

$$\frac{\partial}{\partial y_k}[\mathbf{Ay}]_i = a_{ik} \ , k = 1,2,\ldots,p \, , i = 1,2,\ldots,n$$

niin

$$\frac{\partial}{\partial \mathbf{y}}[\mathbf{Ay}]_i = (a_{i1}, a_{i2}, \ldots, a_{ip}) \ , i = 1,2,\ldots,n$$

ja siten

$$\frac{\partial}{\partial \mathbf{y}}(\mathbf{Ay}) = \mathbf{A}'$$

■

Lause 3.5.4.

Olkoot **A** $n \times p$-matriisi, **x** n-vektori ja **y** p-vektori. Tällöin

(i) $$\frac{\partial}{\partial \mathbf{x}}(\mathbf{x}'\mathbf{Ay}) = \mathbf{Ay}$$

(ii) $$\frac{\partial}{\partial \mathbf{y}}(\mathbf{x}'\mathbf{Ay}) = \mathbf{A}'\mathbf{x}$$

Todistus:

Olkoon **A** $n \times p$-matriisi, **x** n-vektori ja **y** p-vektori.

(i) Olkoon

$$\mathbf{a} = \mathbf{Ay}$$

Tällöin

$$\mathbf{x}'\mathbf{Ay} = \mathbf{x}'\mathbf{Ay} = \mathbf{x}'\mathbf{a} = \mathbf{a}'\mathbf{x}$$

Lauseen 3.5.1. kohdasta (i) seuraa, että

$$\frac{\partial}{\partial \mathbf{x}}(\mathbf{x}'\mathbf{Ay}) = \frac{\partial}{\partial \mathbf{x}}(\mathbf{a}'\mathbf{x}) = \mathbf{a} = \mathbf{Ay}$$

(ii) Kohta (ii) seuraa kohdasta (i), koska

$$\mathbf{x}'\mathbf{Ay} = (\mathbf{x}'\mathbf{Ay})' = \mathbf{y}'\mathbf{A}'\mathbf{x}$$

■

Lause 3.5.5.

Olkoon **A** $p \times p$-matriisi ja **x** p-vektori. Tällöin

(i) $$\frac{\partial}{\partial \mathbf{x}}(\mathbf{x}'\mathbf{Ax}) = (\mathbf{A} + \mathbf{A}')\mathbf{x}$$

(ii) $$\frac{\partial^2}{\partial \mathbf{x} \partial \mathbf{x}'}(\mathbf{x}'\mathbf{Ax}) = \mathbf{A} + \mathbf{A}'$$

Todistus:

Olkoon $\mathbf{A} = [a_{ij}]$ $p{\times}p$-matriisi ja $\mathbf{x} = (x_1, x_2, \ldots , x_p)$ p-vektori. Tällöin

$$\mathbf{x}'\mathbf{A}\mathbf{x} = \sum_{i=1}^{p}\sum_{j=1}^{p} a_{ij}x_i x_j$$

(i) Suoraan laskemalla saadaan

$$\frac{\partial}{\partial x_k}(\mathbf{x}'\mathbf{A}\mathbf{x}) = \sum_{j=1}^{p} a_{kj}x_j + \sum_{i=1}^{p} a_{ik}x_i \ , k = 1, 2, \ldots, p$$

ja siten

$$\frac{\partial}{\partial \mathbf{x}}(\mathbf{x}'\mathbf{A}\mathbf{x}) = \mathbf{A}\mathbf{x} + \mathbf{A}'\mathbf{x} = (\mathbf{A} + \mathbf{A}')\mathbf{x}$$

(ii) Tulos seuraa kohdan (i) tuloksesta ja lauseesta 3.5.3.

■

Lause 3.5.6.

Olkoon $\mathbf{A}$ *symmetrinen* $p{\times}p$-matriisi ja $\mathbf{x}$ p-vektori. Tällöin

(i) $$\frac{\partial}{\partial \mathbf{x}}(\mathbf{x}'\mathbf{A}\mathbf{x}) = 2\mathbf{A}\mathbf{x}$$

(ii) $$\frac{\partial^2}{\partial \mathbf{x}\partial \mathbf{x}'}(\mathbf{x}'\mathbf{A}\mathbf{x}) = 2\mathbf{A}$$

Todistus:

Olkoon $\mathbf{A} = [a_{ij}]$ symmetrinen $p{\times}p$-matriisi ja $\mathbf{x} = (x_1, x_2, \ldots , x_p)$ p-vektori.

(i) Koska matriisi $\mathbf{A}$ on oletettu symmetriseksi, niin

$$\mathbf{A}' = \mathbf{A}$$

Siten tulos seuraa suoraan lauseesta 3.5.5.

(ii) Tulos seuraa kohdan (i) tuloksesta ja lauseesta 3.5.3.

■

Lause 3.5.7.

Olkoon $\mathbf{X}$ $n{\times}p$-matriisi, $\mathbf{a}$ n-vektori ja $\mathbf{b}$ p-vektori. Tällöin

$$\frac{\partial}{\partial \mathbf{X}}(\mathbf{a}'\mathbf{X}\mathbf{b}) = \mathbf{a}\mathbf{b}'$$

Todistus:

Olkoon $\mathbf{X} = [x_{ij}]$ $n{\times}p$-matriisi, $\mathbf{a} = (a_1, a_2, \ldots , a_n)$ n-vektori ja $\mathbf{b} = (b_1, b_2, \ldots , b_p)$ p-vektori. Tällöin

$$\mathbf{a}'\mathbf{X}\mathbf{b} = \sum_{i=1}^{n}\sum_{j=1}^{p} x_{ij}a_i b_j$$

Koska

$$\frac{\partial}{\partial x_{ij}}(\mathbf{a}'\mathbf{X}\mathbf{b}) = a_i b_j \text{ , } i, j = 1, 2, \ldots, p$$

niin

$$\frac{\partial}{\partial \mathbf{X}}(\mathbf{a}'\mathbf{X}\mathbf{b}) = \mathbf{a}\mathbf{b}'$$

■

Determinantin derivointi

Lause 3.5.8.

Olkoon **X** $p\times p$-matriisi. Tällöin

$$\frac{\partial|\mathbf{X}|}{\partial \mathbf{X}} = [\mathrm{adj}(\mathbf{X}')]$$

Todistus:

Olkoon $\mathbf{X} = [x_{ij}]$ $p\times p$-matriisi.

Palautetaan mieleen, että matriisin **X** *adjungoitu matriisi*

$$\mathrm{adj}(\mathbf{X}) = \begin{bmatrix} X_{11} & X_{21} & \cdots & X_{p1} \\ X_{12} & X_{22} & \cdots & X_{p2} \\ \vdots & \vdots & & \vdots \\ X_{1p} & X_{2p} & \cdots & X_{pp} \end{bmatrix}$$

on $p\times p$-matriisi, joka saadaan korvaamalla matriisin **X** i. rivin ja j. sarakkeen alkio x_{ij} alkion x_{ji} *kofaktorilla*

$$X_{ji} = (-1)^{j+i}|\mathbf{X}(j,i)| \text{ , } j, i = 1, 2, \ldots, p$$

jossa $\mathbf{X}(j, i)$ on matriisin **X** $(p-1)\times(p-1)$-alimatriisi, joka saadaan poistamalla matriisista **X** j. rivi ja i. sarake.

Kehittämällä matriisin **X** determinantti $|\mathbf{X}|$ sen i. rivin suhteen (ks. lause 2.3.2.) saadaan lauseke

$$|\mathbf{X}| = \sum_{j=1}^{m} x_{ij} X_{ij}$$

Siten

$$\frac{\partial}{\partial x_{ij}}|\mathbf{X}| = X_{ij}$$

ja edelleen

$$\frac{\partial|\mathbf{X}|}{\partial \mathbf{X}} = [\mathrm{adj}(\mathbf{X})]' = [\mathrm{adj}(\mathbf{X}')]$$

■

Lause 3.5.9.

Olkoon **X** $p\times p$-matriisi. Tällöin

$$\frac{\partial \log|\mathbf{X}|}{\partial \mathbf{X}} = (\mathbf{X}')^{-1}$$

Todistus:

Lauseista 3.5.8., 2.3.4. ja 2.4.3. seuraa, että

$$\frac{\partial \log|\mathbf{X}|}{\partial \mathbf{X}} = \frac{1}{|\mathbf{X}|}\cdot\frac{\partial|\mathbf{X}|}{\partial \mathbf{X}} = \frac{1}{|\mathbf{X}|}\mathrm{adj}(\mathbf{X}') = \frac{1}{|\mathbf{X}'|}\mathrm{adj}(\mathbf{X}') = (\mathbf{X}')^{-1}$$

■

Matriisin käänteismatriisin derivointi

Lause 3.5.10.

Olkoon $\mathbf{X} = [x_{ij}]$ *epäsingulaarinen* $p\times p$-matriisi. Tällöin

$$\frac{\partial \mathbf{X}^{-1}}{\partial x_{ij}} = -\mathbf{X}^{-1}\mathbf{E}_{ij}\mathbf{X}^{-1}$$

jossa $\mathbf{E}_{ij}$ on $p\times p$-matriisi, jonka ainoa nollasta poikkeava alkio on ykkönen paikassa (i, j).

Todistus:

Olkoon $\mathbf{X} = [x_{ij}]$ epäsingulaarinen $p\times p$-matriisi. Tällöin

$$\mathbf{X}\mathbf{X}^{-1} = \mathbf{I}$$

Siten

$$\frac{\partial \mathbf{X}\mathbf{X}^{-1}}{\partial x_{ij}} = \frac{\partial \mathbf{X}}{\partial x_{ij}}\mathbf{X}^{-1} + \mathbf{X}\frac{\partial \mathbf{X}^{-1}}{\partial x_{ij}} = \mathbf{0}\ ,\ i, j = 1, 2, \ldots, p$$

joten

$$\mathbf{X}\frac{\partial \mathbf{X}^{-1}}{\partial x_{ij}} = -\frac{\partial \mathbf{X}}{\partial x_{ij}}\mathbf{X}^{-1}\ ,\ i, j = 1, 2, \ldots, p$$

ja edelleen

$$\frac{\partial \mathbf{X}^{-1}}{\partial x_{ij}} = -\mathbf{X}^{-1}\frac{\partial \mathbf{X}}{\partial x_{ij}}\mathbf{X}^{-1}\ ,\ i, j = 1, 2, \ldots, p$$

Koska

$$\left[\frac{\partial \mathbf{X}}{\partial x_{ij}}\right]_{rs} = \begin{cases} 1\ , \text{kun } r = i, s = j \\ 0\ , \text{muulloin} \end{cases}$$

niin

$$\frac{\partial \mathbf{X}}{\partial x_{ij}} = \mathbf{E}_{ij}$$

Siten

$$\frac{\partial \mathbf{X}^{-1}}{\partial x_{ij}} = -\mathbf{X}^{-1}\mathbf{E}_{ij}\mathbf{X}^{-1}$$

■

Matriisin jäljen derivointi

Lause 3.5.11.

Olkoon $\mathbf{X}$ $p\times p$-matriisi. Tällöin

$$\frac{\partial}{\partial \mathbf{X}} \mathrm{tr}(\mathbf{X}) = \mathbf{I}$$

Todistus:

Olkoon $\mathbf{X} = [x_{ij}]$ $p\times p$-matriisi. Tällöin

$$\mathrm{tr}(\mathbf{X}) = \sum_{i=1}^{p} x_{ii}$$

Koska

$$\frac{\partial\, \mathrm{tr}(\mathbf{X})}{\partial x_{ij}} = \begin{cases} 1\,, & i = j \\ 0\,, & i \neq j \end{cases}$$

niin

$$\frac{\partial}{\partial \mathbf{X}} \mathrm{tr}(\mathbf{X}) = \mathbf{I}$$

■

Lause 3.5.12.

Olkoon $\mathbf{A}$ $n\times p$-matriisi ja $\mathbf{X}$ $p\times n$-matriisi. Tällöin

$$\frac{\partial}{\partial \mathbf{X}} \mathrm{tr}(\mathbf{AX}) = \mathbf{A}'$$

Todistus:

Olkoon $\mathbf{A} = [a_{ij}]$ $n\times p$-matriisi ja $\mathbf{X} = [x_{ij}]$ $p\times n$-matriisi. Tällöin

$$[\mathbf{AX}]_{ij} = \sum_{k=1}^{p} a_{ik}x_{kj}\ ,\ i, j = 1, 2, \ldots, n$$

ja

$$\mathrm{tr}(\mathbf{AX}) = \sum_{i=1}^{n}\sum_{k=1}^{p} a_{ik}x_{ki}$$

Koska

$$\frac{\partial\,\mathrm{tr}(\mathbf{AX})}{\partial x_{ij}} = a_{ji}\ ,\ i = 1,2,\ldots,n,\ j = 1,2,\ldots,p$$

niin

$$\frac{\partial}{\partial \mathbf{X}}\mathrm{tr}(\mathbf{AX}) = \mathbf{A}'$$

■

Lause 3.5.13.

Olkoon **A** $n\times p$-matriisi, **X** $p\times p$-matriisi ja **B** $p\times n$-matriisi. Tällöin

$$\frac{\partial}{\partial \mathbf{X}}\mathrm{tr}(\mathbf{AXB}) = \mathbf{A}'\mathbf{B}'$$

Todistus:

Todetaan ensin, että transpoosin ominaisuuksien perusteella

$$\mathrm{tr}(\mathbf{AXB}) = \mathrm{tr}(\mathbf{BAX})$$

Merkitään

$$\mathbf{BA} = \mathbf{C}$$

Soveltamalla lausetta 3.5.11. nähdään, että

$$\frac{\partial}{\partial \mathbf{X}}\mathrm{tr}(\mathbf{CX}) = \mathbf{C}' = (\mathbf{AB})' = \mathbf{B}'\mathbf{A}'$$

■

Lause 3.5.14.

Olkoon **A** $m\times m$-matriisi ja **X** $m\times p$-matriisi. Tällöin

$$\frac{\partial}{\partial \mathbf{X}}\mathrm{tr}(\mathbf{X}'\mathbf{AX}) = \mathbf{A}'\mathbf{X} + \mathbf{AX}$$

Todistus:

Olkoon $\mathbf{A} = [a_{ij}]$ $m\times m$-matriisi ja $\mathbf{X} = [x_{ij}]$ $m\times p$-matriisi. Tällöin

$$[\mathbf{AX}]_{ls} = \sum_{k=1}^{m} a_{lk}x_{ks}\ ,\ l = 1,2,\ldots,m\ ,\ s = 1,2,\ldots,p$$

ja

$$[\mathbf{X}'\mathbf{AX}]_{rs} = \sum_{l=1}^{m} x_{lr}[\mathbf{AX}]_{ls} = \sum_{l=1}^{m} x_{lr}\sum_{k=1}^{m} a_{lk}x_{ks} = \sum_{l=1}^{m}\sum_{k=1}^{m} a_{lk}x_{lr}x_{ks}\ ,\ r,s = 1,2,\ldots,p$$

jolloin

$$\mathrm{tr}(\mathbf{X}'\mathbf{AX}) = \sum_{r=1}^{p}\sum_{l=1}^{m}\sum_{k=1}^{m} a_{lk}x_{lr}x_{kr}$$

Koska

$$\frac{\partial}{\partial x_{ij}}\operatorname{tr}(\mathbf{X}'\mathbf{A}\mathbf{X}) = \frac{\partial}{\partial x_{ij}}\sum_{r=1}^{p}\sum_{l=1}^{m}\sum_{k=1}^{m} a_{lk}x_{lr}x_{kr}$$

$$= \sum_{l=1}^{m} a_{li}x_{lj} + \sum_{k=1}^{m} a_{ik}x_{kj} \text{ , } i = 1, 2, \ldots, m \text{ , } j = 1, 2, \ldots, p$$

niin

$$\frac{\partial}{\partial \mathbf{X}}\operatorname{tr}(\mathbf{X}'\mathbf{A}\mathbf{X}) = \mathbf{A}'\mathbf{X} + \mathbf{A}\mathbf{X}$$

■

Yleistyksenä lauseelle 3.5.14. voidaan todistaa:

Lause 3.5.15.

Olkoon **A** $m\times m$-matriisi, **B** $p\times p$-matriisi ja **X** $m\times p$-matriisi. Tällöin

$$\frac{\partial}{\partial \mathbf{X}}\operatorname{tr}(\mathbf{B}\mathbf{X}'\mathbf{A}\mathbf{X}) = \mathbf{A}'\mathbf{X}\mathbf{B}' + \mathbf{A}\mathbf{X}\mathbf{B}$$

Todistus:

Olkoon $\mathbf{A} = [a_{ij}]$ $m\times m$-matriisi, $\mathbf{B} = [b_{ij}]$ $p\times p$-matriisi ja $\mathbf{X} = [x_{ij}]$ $m\times p$-matriisi. Tällöin

$$\operatorname{tr}(\mathbf{B}\mathbf{X}'\mathbf{A}\mathbf{X}) = \sum_{s=1}^{p}\sum_{r=1}^{p}\sum_{l=1}^{m}\sum_{k=1}^{m} a_{lk}b_{sr}x_{lr}x_{ks}$$

Koska

$$\frac{\partial}{\partial x_{ij}}\operatorname{tr}(\mathbf{B}\mathbf{X}'\mathbf{A}\mathbf{X}) = \frac{\partial}{\partial x_{ij}}\sum_{s=1}^{p}\sum_{r=1}^{p}\sum_{l=1}^{m}\sum_{k=1}^{m} a_{lk}b_{sr}x_{lr}x_{ks}$$

$$= \sum_{r=1}^{m}\sum_{l=1}^{m} a_{li}b_{jr}x_{lr} + \sum_{s=1}^{m}\sum_{k=1}^{m} a_{ik}b_{sj}x_{ks} \text{ , } i = 1, 2, \ldots m \text{ , } j = 1, 2, \ldots, p$$

niin

$$\frac{\partial}{\partial \mathbf{X}}\operatorname{tr}(\mathbf{B}\mathbf{X}'\mathbf{A}\mathbf{X}) = \mathbf{A}'\mathbf{X}\mathbf{B}' + \mathbf{A}\mathbf{X}\mathbf{B}$$

■

3.6. Bilineaari- ja neliömuotojen maksimointi

Bilineaarimuodot

Olkoon $\mathbf{A} = [a_{ij}]$ $n\times p$-matriisi, $r(\mathbf{A}) = r$, $\mathbf{x} = (x_1, x_2, \ldots , x_n)$ n-vektori ja $\mathbf{y} = (y_1, y_2, \ldots , y_p)$ p-vektori. Kutsumme reaalilukua

$$\mathbf{x}'\mathbf{A}\mathbf{y} = \sum_{i=1}^{n}\sum_{j=1}^{p} a_{ij}x_iy_j$$

matriisin **A** **bilineaarimuodoksi**.

Bilineaarimuotojen maksimointi

Maksimoidaan matriisin **A** *bilineaarimuoto* $\mathbf{x}'\mathbf{A}\mathbf{y}$ ehdoilla

$$\|\mathbf{x}\| = \|\mathbf{y}\| = 1$$

Olkoon

$$\mathbf{A} = \mathbf{PLQ}'$$

matriisin **A** *singulaariarvohajotelma*. Tällöin

$\mathbf{P}$ = sarakkeidensa suhteen *ortonormaalinen* $n\times p$-matriisi: $\mathbf{P}'\mathbf{P} = \mathbf{I}$

$\mathbf{Q}$ = *ortogonaalinen* $p\times p$-matriisi: $\mathbf{Q}'\mathbf{Q} = \mathbf{I}$

$\mathbf{L}$ = $\mathrm{diag}(l_1, l_2, \ldots, l_r, 0, \ldots, 0)$ on $p\times p$-diagonaalimatriisi

Olkoot

$$\mathbf{P} = [\mathbf{p}_1\ \mathbf{p}_2\ \ldots\ \mathbf{p}_p]$$

$$\mathbf{Q} = [\mathbf{q}_1\ \mathbf{q}_2\ \ldots\ \mathbf{q}_p]$$

missä

$\mathbf{p}_i$ = matriisin **P** *i. sarakevektori*, $i = 1, 2, \ldots, p$

$\mathbf{q}_i$ = matriisin **Q** *i. sarakevektori*, $i = 1, 2, \ldots, p$

Matriisien **P**, **Q** ja **L** sarakkeet voidaan aina järjestää niin, että

$$l_1 \geq l_2 \geq \ldots \geq l_r > 0$$

Lause 3.6.1.

Matriisin **A** *bilineaarimuodon*

$$\mathbf{x}'\mathbf{A}\mathbf{y}$$

maksimi muuttujien **x** ja **y** suhteen ehtojen

$$\|\mathbf{x}\| = \|\mathbf{y}\| = 1$$

pätiessä on matriisin **A** *suurin singulaariarvo* l_1 ja *maksimi saavutetaan*, kun

$$\mathbf{x} = \mathbf{p}_1$$

$$\mathbf{y} = \mathbf{q}_1$$

missä $\mathbf{p}_1$ on matriisin **A** *suurinta singulaariarvoa* l_1 vastaava matriisin **P** sarake ja $\mathbf{q}_1$ on vastaava matriisin **Q** sarake.

Todistus:

Koska haluamme maksimoida bilineaarimuodon

$$\mathbf{x}'\mathbf{A}\mathbf{y}$$

ehtojen $\|\mathbf{x}\| = \|\mathbf{y}\| = 1$ pätiessä, sovellamme *Lagrangen menetelmää* sidottujen ääri-arvojen määräämisessä. Maksimoitava funktio on siten muotoa

$$u = \mathbf{x}'\mathbf{A}\mathbf{y} - \tfrac{1}{2}k(\mathbf{x}'\mathbf{x} - 1) - \tfrac{1}{2}l(\mathbf{y}'\mathbf{y} - 1)$$

jossa k ja l ovat Lagrangen kertoimet.

Derivoidaan u vuorollaan muuttujien $\mathbf{x}$, $\mathbf{y}$, k ja l suhteen ja merkitään derivaatat nolliksi. Saamme lauseiden 3.5.4. ja 3.5.2. nojalla muuttujien $\mathbf{x}$, $\mathbf{y}$, k ja l ratkaisemiseksi normaaliyhtälöt

(1) $$\frac{\partial u}{\partial \mathbf{x}} = \mathbf{Ay} - k\mathbf{x} = 0$$

(2) $$\frac{\partial u}{\partial \mathbf{y}} = \mathbf{A}'\mathbf{x} - l\mathbf{y} = 0$$

(3) $$\frac{\partial u}{\partial k} = \tfrac{1}{2}(\mathbf{x}'\mathbf{x} - 1) = 0$$

(4) $$\frac{\partial u}{\partial l} = \tfrac{1}{2}(\mathbf{y}'\mathbf{y} - 1) = 0$$

Kertomalla yhtälö (1) vasemmalta vektorin $\mathbf{x}$ transpoosilla saadaan yhtälön (3) nojalla yhtälö

(5) $$\mathbf{x}'\mathbf{Ay} = k$$

Transponoimalla yhtälö (2) saadaan yhtälö

(6) $$\mathbf{x}'\mathbf{A} - l\mathbf{y}' = 0$$

Kertomalla yhtälö (6) oikealta vektorilla $\mathbf{y}$ saadaan yhtälön (4) nojalla yhtälö

(7) $$\mathbf{x}'\mathbf{Ay} = l$$

Yhtälöistä (5) ja (7) nähdään, että

$$k = l = \mathbf{x}'\mathbf{Ay}$$

Yhtälöistä (5) ja (7) nähdään myös, että bilineaarimuoto $\mathbf{x}'\mathbf{Ay}$ maksimoituu valinnoilla

$$\mathbf{x} = \mathbf{p}_1$$

$$\mathbf{y} = \mathbf{q}_1$$

jossa vektorit $\mathbf{p}_1$ ja $\mathbf{q}_1$ vastaavat matriisin $\mathbf{A}$ suurinta singulaariarvoa l_1 ja maksimin arvo on

$$\mathbf{x}'\mathbf{Ay} = l_1$$

■

Käyttämällä *Lagrangen menetelmää* hyvin samaan tapaan kuin lauseen 3.6.1. todistuksessa voidaan todistaa seuraava lause:

Lause 3.6.2.

Matriisin $\mathbf{A}$ *bilineaarimuodon* $\mathbf{x}'\mathbf{Ay}$ **maksimi** muuttujien $\mathbf{x}$ ja $\mathbf{y}$ suhteen ehtojen

$$||\mathbf{x}|| = ||\mathbf{y}|| = 1$$

$$\mathbf{x}'\mathbf{p}_1 = \cdots = \mathbf{x}'\mathbf{p}_i = 0$$

$$\mathbf{y}'\mathbf{q}_1 = \cdots = \mathbf{y}'\mathbf{q}_i = 0$$

$$i = 1, 2, \ldots , r$$

pätiessä on matriisin $\mathbf{A}$ (i+1). *singulaariarvo* l_{i+1} ja *maksimi saavutetaan*, kun

$\mathbf{x} = \mathbf{p}_{i+1}$

$\mathbf{y} = \mathbf{q}_{i+1}$

missä $\mathbf{p}_{i+1}$ on matriisin **A** (i+1). *singulaariarvoa* l_{i+1} vastaava matriisin **P** sarake ja $\mathbf{q}_{i+1}$ on vastaava matriisin **Q** sarake.

Neliömuodot

Olkoon $\mathbf{A} = [a_{ij}]$ *symmetrinen* $p\times p$-matriisi ja $\mathbf{x} = (x_1, x_2, \ldots , x_p)$ p-vektori. Kutsumme reaalilukua

$$\mathbf{x}'\mathbf{A}\mathbf{x} = \sum_{i=1}^{p}\sum_{j=1}^{p} a_{ij}x_i x_j$$

matriisin **A** **neliömuodoksi**.

Neliömuotojen maksimointi

Maksimoidaan matriisin **A** neliömuoto $\mathbf{x}'\mathbf{A}\mathbf{x}$ ehdolla

$$\|\mathbf{x}\| = 1$$

Olkoon

$$\mathbf{A} = \mathbf{Q}\boldsymbol{\Lambda}\mathbf{Q}'$$

matriisin **A** *pääakselihajotelma*. Tällöin **Q** on *ortogonaalinen* $p\times p$-matriisi:

$$\mathbf{Q}'\mathbf{Q} = \mathbf{Q}\mathbf{Q}' = \mathbf{I}$$

ja

$$\Lambda = \mathrm{diag}(\lambda_1, \lambda_2, \ldots , \lambda_p)$$

on $p\times p$-diagonaalimatriisi.

Olkoon

$$\mathbf{Q} = [\mathbf{q}_1\ \mathbf{q}_2\ \ldots\ \mathbf{q}_p]$$

missä

$\mathbf{q}_i$ = matriisin **Q** *i*. *sarakevektori*, $i = 1, 2, \ldots , p$

Matriisien **Q** ja Λ sarakkeet voidaan aina järjestää niin, että

$$\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p$$

Lause 3.6.3.

Matriisin **A** *neliömuodon*

$$\mathbf{x}'\mathbf{A}\mathbf{x}$$

maksimi muuttujan **x** suhteen ehdon

$$\|\mathbf{x}\| = 1$$

pätiessä on matriisin **A** *suurin ominaisarvo* λ_1 ja maksimi saavutetaan, kun

$$\mathbf{x} = \mathbf{q}_1$$

missä $\mathbf{q}_1$ on matriisin **A** *suurinta ominaisarvoa* λ_1 vastaava *ominaisvektori*.

Todistus:

Koska haluamme maksimoida neliömuodon

$$\mathbf{x}'\mathbf{A}\mathbf{x}$$

ehdon $\|\mathbf{x}\| = 1$ pätiessä sovellamme *Lagrangen menetelmää* sidottujen ääriarvojen määräämisessä. Maksimoitava funktio on siten muotoa

$$v = \mathbf{x}'\mathbf{A}\mathbf{x} - \lambda(\mathbf{x}'\mathbf{x} - 1)$$

jossa λ on Lagrangen kerroin.

Derivoidaan v vuorollaan muuttujien $\mathbf{x}$ ja suhteen ja merkitään derivaatat nolliksi. Saamme lauseiden 3.5.6. ja 3.5.2. nojalla muuttujien $\mathbf{x}$ ja λ ratkaisemiseksi normaaliyhtälöt

(1) $$\frac{\partial u}{\partial \mathbf{x}} = 2\mathbf{A}\mathbf{x} - 2\lambda\mathbf{x} = 0$$

(2) $$\frac{\partial u}{\partial \lambda} = \mathbf{x}'\mathbf{x} - 1 = 0$$

Kertomalla yhtälö (1) vasemmalta vektorin $\mathbf{x}$ transpoosilla saadaan yhtälön (2) nojalla yhtälö

(3) $$\mathbf{x}'\mathbf{A}\mathbf{x} = \lambda$$

Yhtälöstä (3) nähdään välittömästi, että neliömuoto $\mathbf{x}'\mathbf{A}\mathbf{x}$ maksimoituu valinnalla

$$\mathbf{x} = \mathbf{q}_1$$

jossa vektori $\mathbf{q}_1$ vastaa matriisin $\mathbf{A}$ suurinta ominaisarvoa λ_1 ja maksimin arvo on

$$\mathbf{x}'\mathbf{A}\mathbf{x} = \lambda_1$$

■

Käyttämällä *Lagrangen menetelmää* hyvin samaan tapaan kuin lauseen 3.6.3. todistuksessa voidaan todistaa seuraava lause:

Lause 3.6.4.

Matriisin $\mathbf{A}$ *neliömuodon*

$$\mathbf{x}'\mathbf{A}\mathbf{x}$$

maksimi muuttujan $\mathbf{x}$ suhteen ehtojen

$$\|\mathbf{x}\| = 1$$

$$\mathbf{x}'\mathbf{q}_1 = \cdots = \mathbf{x}'\mathbf{q}_i = 0$$

$$i = 1, 2, \ldots, p$$

pätiessä on matriisin $\mathbf{A}$ (i+1). *ominaisarvo* λ_{i+1} ja *maksimi saavutetaan*, kun

$$\mathbf{x} = \mathbf{q}_{i+1}$$

missä $\mathbf{q}_{i+1}$ on matriisin $\mathbf{A}$ (i+1). *ominaisarvoa* λ_{i+1} vastaava *ominaisvektori*.

3.7. Kroneckerin tulo

Kroneckerin tulo

Olkoon $\mathbf{A} = [a_{ij}]$ $m\times n$-matriisi ja $\mathbf{B}$ $p\times q$-matriisi. Matriisien $\mathbf{A}$ ja $\mathbf{B}$ **Kroneckerin tulo** eli **tensoritulo** on $mp\times nq$-matriisi

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

Laskusääntöjä Kroneckerin tulolle

Lause 3.7.1.

Oletetaan, että so. matriisien dimensiot ovat sellaisia, että seuraavien kaavojen laskutoimitukset ovat luvallisia.

(i) $c(\mathbf{A}\otimes\mathbf{B}) = (c\mathbf{A})\otimes\mathbf{B} = \mathbf{A}\otimes(c\mathbf{B})$

(ii) $\mathbf{A}\otimes(\mathbf{B}+\mathbf{C}) = \mathbf{A}\otimes\mathbf{B}+\mathbf{A}\otimes\mathbf{C}$

(iii) $(\mathbf{A}+\mathbf{B})\otimes\mathbf{C} = \mathbf{A}\otimes\mathbf{C}+\mathbf{B}\otimes\mathbf{C}$

(iv) $\mathbf{A}\otimes(\mathbf{B}\otimes\mathbf{C}) = (\mathbf{A}\otimes\mathbf{B})\otimes\mathbf{C}$

(v) $(\mathbf{A}\otimes\mathbf{B})(\mathbf{C}\otimes\mathbf{D}) = (\mathbf{AC})\otimes(\mathbf{BD})$

(vi) $(\mathbf{A}\otimes\mathbf{B})' = \mathbf{A}'\otimes\mathbf{B}'$

(vii) $\mathrm{tr}(\mathbf{A}\otimes\mathbf{B}) = \mathrm{tr}(\mathbf{A})\,\mathrm{tr}(\mathbf{B})$

(viii) Olkoot $\mathbf{A}$ ja $\mathbf{B}$ *epäsingulaarisia* matriiseja. Tällöin niiden Kroneckerin tulo $\mathbf{A}\otimes\mathbf{B}$ on epäsingulaarinen ja

$$(\mathbf{A}\otimes\mathbf{B})^{-1} = \mathbf{A}^{-1}\otimes\mathbf{B}^{-1}$$

(ix) Olkoon $\mathbf{A}$ *epäsingulaarinen* $m\times m$-matriisi ja $\mathbf{B}$ *epäsingulaarinen* $p\times p$-matriisi. Tällöin

$$|\mathbf{A}\otimes\mathbf{B}| = |\mathbf{A}|^p\,|\mathbf{B}|^m$$

Todistus:

(i) Oletetaan, että $\mathbf{A} = [a_{ij}]$ on $m\times n$-matriisi ja $\mathbf{B} = [b_{kl}]$ on $p\times q$-matriisi. Matriisin

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

yleinen alkio on muotoa $a_{ij}b_{kl}$. Reaalilukujen kertolaskun liitäntälain mukaan

$$c(a_{ij}b_{kl}) = (ca_{ij})b_{kl} = a_{ij}(cb_{kl})$$

Koska ca_{ij} on matriisin $c\mathbf{A}$ yleinen alkio ja cb_{kl} on matriisin $c\mathbf{B}$ yleinen alkio, niin

$$\begin{aligned}
c(\mathbf{A}\otimes\mathbf{B}) &= c\begin{bmatrix} a_{11}\mathbf{B} & \cdots & a_{1n}\mathbf{B} \\ \vdots & & \vdots \\ a_{m1}\mathbf{B} & \cdots & a_{mn}\mathbf{B} \end{bmatrix} \\
&= \begin{bmatrix} (ca_{11})\mathbf{B} & \cdots & (ca_{1n})\mathbf{B} \\ \vdots & & \vdots \\ (ca_{m1})\mathbf{B} & \cdots & (ca_{mn})\mathbf{B} \end{bmatrix} \\
&= (c\mathbf{A})\otimes\mathbf{B}
\end{aligned}$$

ja

$$\begin{aligned}
c(\mathbf{A}\otimes\mathbf{B}) &= c\begin{bmatrix} a_{11}\mathbf{B} & \cdots & a_{1n}\mathbf{B} \\ \vdots & & \vdots \\ a_{m1}\mathbf{B} & \cdots & a_{mn}\mathbf{B} \end{bmatrix} \\
&= \begin{bmatrix} a_{11}(c\mathbf{B}) & \cdots & a_{1n}(c\mathbf{B}) \\ \vdots & & \vdots \\ a_{m1}(c\mathbf{B}) & \cdots & a_{mn}(c\mathbf{B}) \end{bmatrix} \\
&= \mathbf{A}\otimes(c\mathbf{B})
\end{aligned}$$

Siten

$$c(\mathbf{A}\otimes\mathbf{B}) = (c\mathbf{A})\otimes\mathbf{B} = \mathbf{A}\otimes(c\mathbf{B})$$

(ii) Olkoon $\mathbf{A} = [a_{ij}]$ $m\times n$-matriisi ja $\mathbf{B}$ ja $\mathbf{C}$ $p\times q$-matriiseja. Ositettujen matriisien laskusäännöistä seuraa, että

$$\begin{aligned}
\mathbf{A}\otimes(\mathbf{B}+\mathbf{C}) &= \begin{bmatrix} a_{11}(\mathbf{B}+\mathbf{C}) & \cdots & a_{1n}(\mathbf{B}+\mathbf{C}) \\ \vdots & & \vdots \\ a_{m1}(\mathbf{B}+\mathbf{C}) & \cdots & a_{mn}(\mathbf{B}+\mathbf{C}) \end{bmatrix} \\
&= \begin{bmatrix} a_{11}\mathbf{B}+a_{11}\mathbf{C} & \cdots & a_{1n}\mathbf{B}+a_{1n}\mathbf{C} \\ \vdots & & \vdots \\ a_{m1}\mathbf{B}+a_{m1}\mathbf{C} & \cdots & a_{mn}\mathbf{B}+a_{mn}\mathbf{C} \end{bmatrix} \\
&= \begin{bmatrix} a_{11}\mathbf{B} & \cdots & a_{1n}\mathbf{B} \\ \vdots & & \vdots \\ a_{m1}\mathbf{B} & \cdots & a_{mn}\mathbf{B} \end{bmatrix} + \begin{bmatrix} a_{11}\mathbf{C} & \cdots & a_{1n}\mathbf{C} \\ \vdots & & \vdots \\ a_{m1}\mathbf{C} & \cdots & a_{mn}\mathbf{C} \end{bmatrix} \\
&= \mathbf{A}\otimes\mathbf{B}+\mathbf{A}\otimes\mathbf{C}
\end{aligned}$$

(iii) Kohta (iii) todistetaan samalla tavalla kuin kohta (ii).

(iv) Kaava seuraa samaan tapaan kuin kohta (i) reaalilukujen kertolaskun liitäntälaista.

(v) Oletetaan, että $\mathbf{A} = [a_{ij}]$ on $m\times n$-, $\mathbf{B}$ on $p\times q$-, $\mathbf{C} = [c_{kl}]$ on $n\times r$- ja $\mathbf{D}$ on $q\times s$-matriisi. Ositettujen matriisien laskusäännöistä seuraa, että

$$
\begin{aligned}
(\mathbf{A}\otimes\mathbf{B})(\mathbf{C}\otimes\mathbf{D}) &= \begin{bmatrix} a_{11}\mathbf{B} & \cdots & a_{1n}\mathbf{B} \\ \vdots & & \vdots \\ a_{m1}\mathbf{B} & \cdots & a_{mn}\mathbf{B} \end{bmatrix} \begin{bmatrix} c_{11}\mathbf{D} & \cdots & c_{1r}\mathbf{D} \\ \vdots & & \vdots \\ c_{n1}\mathbf{D} & \cdots & c_{nr}\mathbf{D} \end{bmatrix} \\
&= \begin{bmatrix} \sum_{j=1}^{n} a_{1j}c_{j1}\mathbf{BD} & \cdots & \sum_{j=1}^{n} a_{1j}c_{jr}\mathbf{BD} \\ \vdots & & \vdots \\ \sum_{j=1}^{n} a_{mj}c_{j1}\mathbf{BD} & \cdots & \sum_{j=1}^{n} a_{mj}c_{jr}\mathbf{BD} \end{bmatrix} \\
&= \begin{bmatrix} [\mathbf{AC}]_{11}\mathbf{BD} & \cdots & [\mathbf{AC}]_{1r}\mathbf{BD} \\ \vdots & & \vdots \\ [\mathbf{AC}]_{m1}\mathbf{BD} & \cdots & [\mathbf{AC}]_{mr}\mathbf{BD} \end{bmatrix} \\
&= (\mathbf{AC})\otimes(\mathbf{BD})
\end{aligned}
$$

(vi) Oletetaan, että $\mathbf{A} = [a_{ij}]$ on $m\times n$-matriisi ja $\mathbf{B}$ on $p\times q$-matriisi. Ositettujen matriisien laskusäännöistä seuraa, että

$$
\begin{aligned}
(\mathbf{A}\otimes\mathbf{B})' &= \begin{bmatrix} a_{11}\mathbf{B} & \cdots & a_{1n}\mathbf{B} \\ \vdots & & \vdots \\ a_{m1}\mathbf{B} & \cdots & a_{mn}\mathbf{B} \end{bmatrix}' \\
&= \begin{bmatrix} a_{11}\mathbf{B}' & \cdots & a_{m1}\mathbf{B}' \\ \vdots & & \vdots \\ a_{1n}\mathbf{B}' & \cdots & a_{mn}\mathbf{B}' \end{bmatrix} \\
&= \mathbf{A}'\otimes\mathbf{B}'
\end{aligned}
$$

(vii) Oletetaan, että $\mathbf{A} = [a_{ij}]$ on $m\times m$-matriisi ja $\mathbf{B} = [b_{kl}]$ on $p\times p$-matriisi. Tällöin

$$
\mathrm{tr}(\mathbf{A}\otimes\mathbf{B}) = \sum_{i=1}^{m}\sum_{j=1}^{p} a_{ii}b_{jj} = \sum_{i=1}^{n} a_{ii}\sum_{j=1}^{p} b_{jj} = \sum_{i=1}^{n} a_{ii}\,\mathrm{tr}(\mathbf{B}) = \mathrm{tr}(\mathbf{A})\,\mathrm{tr}(\mathbf{B})
$$

(viii) Kohdasta (v) seuraa, että

$$
(\mathbf{A}\otimes\mathbf{B})(\mathbf{A}^{-1}\otimes\mathbf{B}^{-1}) = (\mathbf{AA}^{-1})\otimes(\mathbf{BB}^{-1}) = \mathbf{I}\otimes\mathbf{I} = \mathbf{I}
$$

ja

$$
(\mathbf{A}^{-1}\otimes\mathbf{B}^{-1})(\mathbf{A}\otimes\mathbf{B}) = (\mathbf{A}^{-1}\mathbf{A})\otimes(\mathbf{B}^{-1}\mathbf{B}) = \mathbf{I}\otimes\mathbf{I} = \mathbf{I}
$$

Siten

$$
(\mathbf{A}\otimes\mathbf{B})^{-1} = \mathbf{A}^{-1}\otimes\mathbf{B}^{-1}
$$

(ix) Oletetaan, että $\mathbf{A} = [a_{ij}]$ on $m\times m$-matriisi ja $\mathbf{B}$ on $p\times p$-matriisi. Oletetaan lisäksi, että matriisit $\mathbf{A}$ ja $\mathbf{B}$ ovat molemmat epäsingulaarisia.

Olkoon $\mathbf{A}_i$ $(m-i)\times(m-i)$-matriisi, joka saadaan matriisista $\mathbf{A}$ poistamalla sen i ensimmäistä riviä ja saraketta. Olkoon lisäksi $\mathbf{a}_{i\cdot}$ $(m-i)$-vektori, joka saadaan matriisin $\mathbf{A}$ i. rivivektorista poistamalla sen i ensimmäistä alkiota ja $\mathbf{a}_{\cdot i}$ $(m-i)$-vektori, joka saadaan poistamalla matriisin $\mathbf{A}$ i. sarakevektorista sen i ensimmäistä alkiota.

Tehdään matriisille **A** seuraava ositus:

$$\mathbf{A} = \begin{bmatrix} a_{11} & \mathbf{a}'_{1\bullet} \\ \mathbf{a}_{\bullet 1} & \mathbf{A}_1 \end{bmatrix}$$

Tällöin

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} a_{11}\mathbf{B} & \mathbf{a}'_{1\bullet} \otimes \mathbf{B} \\ \mathbf{a}_{\bullet 1} \otimes \mathbf{B} & \mathbf{A}_1 \otimes \mathbf{B} \end{bmatrix}$$

Ositettujen matriisien laskusääntöjen nojalla (ks. lausetta 2.6.1.)

$$|\mathbf{A} \otimes \mathbf{B}| = |a_{11}\mathbf{B} - (\mathbf{a}_{1\bullet} \otimes \mathbf{B})(\mathbf{A}_1 \otimes \mathbf{B})^{-1}(\mathbf{a}_{\bullet 1} \otimes \mathbf{B})||\mathbf{A}_1 \otimes \mathbf{B}|$$

Koska **A** on oletettu epäsingulaariseksi, voimme olettaa yleisyyden liikaa kärsimättä, että matriisit $\mathbf{A}_i$, $i = 1, 2, \ldots , (m–1)$ ovat epäsingulaarisia.

Siten kohdista (v) ja (viii) seuraa, että

$$(\mathbf{a}_{1\bullet} \otimes \mathbf{B})(\mathbf{A}_1 \otimes \mathbf{B})^{-1}(\mathbf{a}_{\bullet 1} \otimes \mathbf{B}) = (\mathbf{a}_{1\bullet}\mathbf{A}_1^{-1}\mathbf{a}_{\bullet 1}) \otimes \mathbf{B}$$

jossa $\mathbf{a}_{1\bullet}\mathbf{A}_1^{-1}\mathbf{a}_{\bullet 1}$ on skalaari, jolloin

$$\begin{aligned} &|a_{11}\mathbf{B} - (\mathbf{a}_{1\bullet} \otimes \mathbf{B})(\mathbf{A}_1 \otimes \mathbf{B})^{-1}(\mathbf{a}_{\bullet 1} \otimes \mathbf{B})| \\ &\qquad = |(a_{11} - \mathbf{a}_{1\bullet}\mathbf{A}_1^{-1}\mathbf{a}_{\bullet 1}) \otimes \mathbf{B}| \\ &\qquad = (a_{11} - \mathbf{a}_{1\bullet}\mathbf{A}_1^{-1}\mathbf{a}_{\bullet 1})^p |\mathbf{B}| \end{aligned}$$

Ositettujen matriisien laskusääntöjen nojalla (ks. lausetta 2.6.1.)

$$|\mathbf{A}| = (a_{11} - \mathbf{a}_{1\bullet}\mathbf{A}_1^{-1}\mathbf{a}_{\bullet 1})|\mathbf{A}_1|$$

ja siten

$$(a_{11} - \mathbf{a}_{1\bullet}\mathbf{A}_1^{-1}\mathbf{a}_{\bullet 1}) = |\mathbf{A}||\mathbf{A}_1|^{-1}$$

Siten olemme saaneet matriisin determinantille $|\mathbf{A} \otimes \mathbf{B}|$ lausekkeen

(1) $$|\mathbf{A} \otimes \mathbf{B}| = |\mathbf{A}|^p |\mathbf{A}_1|^{-p} |\mathbf{B}||\mathbf{A}_1 \otimes \mathbf{B}|$$

Soveltamalla kuvattua operaatiota determinanttiin $|\mathbf{A}_1 \otimes \mathbf{B}|$ saadaan lauseke

(2) $$|\mathbf{A}_1 \otimes \mathbf{B}| = |\mathbf{A}_1|^p |\mathbf{A}_2|^{-p} |\mathbf{B}||\mathbf{A}_2 \otimes \mathbf{B}|$$

Yhdistämällä lauseke (2) lausekkeeseen (1) saadaan

$$|\mathbf{A} \otimes \mathbf{B}| = |\mathbf{A}|^p |\mathbf{A}_2|^{-p} |\mathbf{B}|^2 |\mathbf{A}_2 \otimes \mathbf{B}|$$

Jatkamalla kuvattua operaatiota vielä $(m–2)$ kertaa saamme kaavan

$$|\mathbf{A} \otimes \mathbf{B}| = |\mathbf{A}|^p |\mathbf{A}_{m-1}|^{-p} |\mathbf{B}|^{m-1} |\mathbf{A}_{m-1} \otimes \mathbf{B}|$$

Koska

$$\mathbf{A}_{m-1} = a_{mm}$$

niin

$$|\mathbf{A}_{m-1} \otimes \mathbf{B}| = |a_{mm}\mathbf{B}| = a_{mm}^{p}\mathbf{B}$$

Koska lisäksi

$$|\mathbf{A}_{m-1}| = a_{mm}$$

niin saamme lopulta haluamamme tuloksen:

$$|\mathbf{A} \otimes \mathbf{B}| = |\mathbf{A}|^{p}\,|\mathbf{B}|^{m}$$

■