Vorläufige Fassung, April 2008

Singulärwert-Zerlegung und Methode der kleinsten Quadrate

Leo Knüsel, Institut für Statistik, Universität München

1. Einführung und Überblick

Die Methode der kleinsten Quadrate ist in der Regressionsrechnung ein wichtiges Instrument zur Bestimmung von optimalen linearen Schätzfunktionen. Die Singulärwert-Zerlegung einer beliebigen Daten-Matrix $\mathbf{X} = (n \times p)$ ($n =$ Anzahl der Beobachtungen, $p =$ Anzahl der Variablen) erlaubt es, die Kleinste-Quadrat-Schätzungen zu finden ohne Differentiation, ohne das Lösung von Normalgleichungen und ohne Voraussetzugen über den Rang der Matrix *X*. Auch im Falle von Multikollinearität liefert diese Methode die einfachen natürlichen Lösungen. Überdies ist die numerische Bestimmung der Singulärwert-Zerlegung numerisch stabiler als die Bestimmung der Eigenwert-Zerlegung (Spektraldarstellung) der symmetrischen Matrix $\mathbf{X}^\top \mathbf{X}$ oder als die Bestimmung der inversen Matrix $(\mathbf{X}^\top \mathbf{X})^{-1}$. Die Ergebnisse, die in meinem Vortrag vorgestellt werden sollen, sind vor allem in der numerischen Fachliteratur entwickelt worden (vgl. Golub, 1965; Hammarling, 1985; Stör, 1999), sie haben aber in unseren Statistik-Lehrbüchern noch kaum ihren Niederschlag gefunden.

Inhaltsübersicht

2. Lineares Modell und Singulärwert-Zerlegung in der deskriptiven Statistik

Wir betrachten das lineare Modell (bzw. den linearen Ansatz) der beschreibenden Statistik in Matrix-Notation

(1) $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$,

wobei

$\mathbf{X} = (x_{ij}) = (n \times p) =$ Datenmatrix (Design-Matrix),
($n =$ Anzahl der Beobachtungen, $p =$ Anzahl der Variablen, $n > p$),

$\mathbf{y} = (n \times 1) =$ Beobachtungen der abhängigen Variablen ,

$\boldsymbol{\beta} = (p \times 1) =$ unbekannter Parametervektor .

Es sei

(2) $\mathbf{e} = \mathbf{y} - \mathbf{X}\boldsymbol{\beta}$, d.h. $e_i = y_i - \sum_{j=1}^{p} x_{ij}\beta_j$.

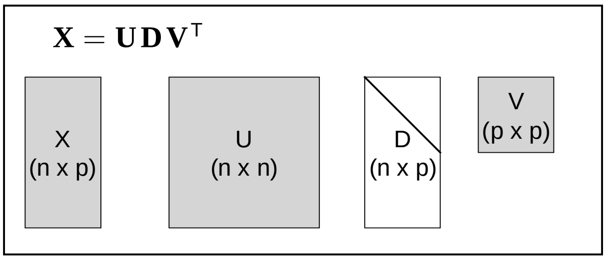
Bestimmung von $\boldsymbol{\beta}$ mit KQ-Methode (Methode der kleinsten Quadrate):

(3) $\boldsymbol{\beta}$ so wählen, dass $\sum_{i=1}^{n} e_i^2 = \mathbf{e}^\mathsf{T}\mathbf{e} = \|\mathbf{X}\boldsymbol{\beta} - \mathbf{y}\|^2 = \min$

Lösung des KQ-Problems mit Singulärwert-Zerlegung

Die Marix $\mathbf{X}$ kann geschrieben werden als

(4) $\mathbf{X} = \mathbf{U}\mathbf{D}\mathbf{V}^\mathsf{T}$ (Singulärwert-Zerlegung),



wobei

$\mathbf{U} = (n \times n)$ orthogonal,

$\mathbf{V} = (p \times p)$ orthogonal,

$\mathbf{D} = (n \times p) = diag(d_1, \ldots, d_p)$

$d_1 \geq d_2 \geq \cdots \geq d_r > 0,\ d_{r+1} = \cdots = d_p = 0$

$d_1, \ldots, d_p$ heißen *singuläre Werte* von $\mathbf{X}$, $r = rk(\mathbf{X}) =$ Rang von $\mathbf{X}$

Nun gilt

(5) $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e} = \mathbf{U}\mathbf{D}\mathbf{V}^\mathsf{T}\boldsymbol{\beta} + \mathbf{e}$.

Wenn wir (5) von links mit $\mathbf{U}^\mathsf{T}$ multiplizieren, erhalten wir

(6) $\underbrace{\mathbf{U}^\mathsf{T}\mathbf{y}}_{=\tilde{\mathbf{y}}} = \underbrace{\mathbf{U}^\mathsf{T}\mathbf{U}}_{=\mathbf{I}_n}\mathbf{D}\underbrace{\mathbf{V}^\mathsf{T}\boldsymbol{\beta}}_{=\tilde{\boldsymbol{\beta}}} + \underbrace{\mathbf{U}^\mathsf{T}\mathbf{e}}_{=\tilde{\mathbf{e}}}$

d.h.

(7) $\tilde{\mathbf{y}} = \mathbf{D}\tilde{\boldsymbol{\beta}} + \tilde{\mathbf{e}}$

wobei

$\tilde{\mathbf{y}} = \mathbf{U}^\mathsf{T}\mathbf{y}$, $\tilde{\boldsymbol{\beta}} = \mathbf{V}^\mathsf{T}\boldsymbol{\beta}$, $\tilde{\mathbf{e}} = \mathbf{U}^\mathsf{T}\mathbf{e}$.

Es gilt

$\tilde{\mathbf{e}}^\mathsf{T}\tilde{\mathbf{e}} = \mathbf{e}^\mathsf{T}\underbrace{\mathbf{U}\mathbf{U}^\mathsf{T}}_{=\mathbf{I}_n}\mathbf{e} = \mathbf{e}^\mathsf{T}\mathbf{e}$,

d.h.

$$\sum_{i=1}^{n} \tilde{e}_i^2 = \sum_{i=1}^{n} e_i^2 .$$

Das KQ-Problem lautet jetzt

$$\tilde{\boldsymbol{\beta}} \ \ (= \mathbf{V}^{\mathsf{T}}\boldsymbol{\beta}) \text{ in (7) so wählen, dass } \sum_{i=1}^{n} \tilde{e}_i^2 = \tilde{\mathbf{e}}^{\mathsf{T}}\tilde{\mathbf{e}} = \min .$$

Gemäß (7) gilt $\tilde{\mathbf{e}} = \tilde{\mathbf{y}} - \mathbf{D}\tilde{\boldsymbol{\beta}}$, d.h.

$$\begin{aligned}
&\tilde{e}_1 = \tilde{y}_1 - d_1\tilde{\beta}_1 \\
&\dots \\
&\tilde{e}_r = \tilde{y}_r - d_r\tilde{\beta}_r \\
&\tilde{e}_{r+1} = \tilde{y}_{r+1} - d_{r+1}\tilde{\beta}_{r+1} = \tilde{y}_{r+1} \quad \text{(unabhängig von } \tilde{\beta}_{r+1}) \\
&\dots \\
&\tilde{e}_p = \tilde{y}_p - d_p\tilde{\beta}_p = \tilde{y}_p \qquad\qquad \text{(unabhängig von } \tilde{\beta}_p) \\
&\tilde{e}_{p+1} = \tilde{y}_{p+1} \\
&\dots \\
&\tilde{e}_n = \tilde{y}_n .
\end{aligned}$$

KQ-Methode:

$\tilde{\beta}_1,\dots,\tilde{\beta}_p$ so wählen, dass

$$\sum_{i=1}^{n} \tilde{e}_i^2 = \min$$

Offensichtlich gilt

Quadratsumme $\sum \tilde{e}_i^2$ nur abhängig von $\tilde{\beta}_1,\dots,\tilde{\beta}_r$,

$\tilde{\beta}_{r+1},\dots,\tilde{\beta}_p$ ohne Einfluss auf Quadratsumme.

Quadratsumme $\sum \tilde{e}_i^2$ $(= \sum e_i^2)$ wird minimal falls

$\tilde{\beta}_1,\dots,\tilde{\beta}_r$ so, dass $\tilde{e}_1 = \cdots = \tilde{e}_r = 0$

$\tilde{\beta}_{r+1},\dots,\tilde{\beta}_p$ beliebig (ohne Einfluss auf Quadratsumme!)

Somit ergibt sich die KQ-Lösung

$$\tilde{\beta}_i = \begin{cases} \tilde{y}_i / d_i & \text{für } i = 1,\dots,r \\ \text{beliebig} & \text{für } i = r+1,\dots,p. \end{cases}$$

Minimale Quadratsumme:

$$Q_{\min} = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} \tilde{e}_i^2 = \sum_{i=r+1}^{n} \tilde{e}_i^2 = \sum_{i=r+1}^{n} \tilde{y}_i^2 .$$

Zurückspielen auf Originalgrößen:

Da $\tilde{\boldsymbol{\beta}} = \mathbf{V}^{\mathsf{T}}\boldsymbol{\beta}$ so gilt $\boldsymbol{\beta} = \mathbf{V}\tilde{\boldsymbol{\beta}}$, wobei $\tilde{\beta}_1, \tilde{\beta}_2, \dots, \tilde{\beta}_r$ eindeutig festgelegt (identifizierbar) und $\tilde{\beta}_{r+1},\dots,\tilde{\beta}_p$ beliebig wählbar (Minimum eindeutig, Minimalstelle nicht eindeutig).

Gesamtheit aller Lösungsvektoren $\boldsymbol{\beta}$: $(p-r)$-dimensionaler Raum.

Beachte:

- Keine Voraussetzungen, Rang r von $\mathbf{X}$ kann kleiner als p sein (Multikollinearität);
- Keine Differentiation zur Bestimmung der Extremalstelle, keine Normalgleichungen;

Vorgehen ähnlich wie bei *Minimumseigenschaft des arithmetischen Mittels*:

$x_1,\dots,x_n$ beliebige reelle Zahlen;

a so wählen, dass $\sum(x_i - a)^2 = \min$;

Satz von Steiner: $\sum(x_i - a)^2 = \underbrace{\sum(x_i - \bar{x})^2}_{(1)} + \underbrace{n(\bar{x} - a)^2}_{(2)}$

(1) unabhängig von a;

(2) minimal für $a = \bar{x}$

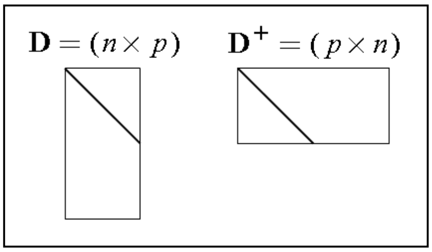
3. Lineares Modell im regulären Fall

Wir setzen voraus, dass die Datenmatrix $\mathbf{X} = (n \times p)$ vollen Rang hat, d.h. dass $rk(\mathbf{X}) = p$. Dann gilt

$$\mathbf{D} = (n \times p) = diag(d_1,\cdots,d_p) \text{ mit } d_1 \ge d_2 \ge \cdots \ge d_p > 0$$

Wir setzen

$$\mathbf{D}^+ = (p \times n) = diag\left(\frac{1}{d_1},\cdots,\frac{1}{d_p}\right).$$



Dann gilt

$$\mathbf{D}^+\mathbf{D} = (p \times p) = \mathbf{I}_p$$

$$\mathbf{D}\mathbf{D}^+ = (n \times n) = \left(\begin{array}{c|c} \mathbf{I}_p & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} \end{array}\right)$$

$\mathbf{D}^+$ heißt *verallgemeinerte Inverse* zu $\mathbf{D}$

Wir betrachten das transfomierte Modell Modell (7)

$$\tilde{\mathbf{y}} = \mathbf{D}\tilde{\boldsymbol{\beta}} + \tilde{\mathbf{e}}\,.$$

Es gilt jetzt $\tilde{\mathbf{e}} = \tilde{\mathbf{y}} - \mathbf{D}\tilde{\boldsymbol{\beta}}$, d.h.

$$\tilde{e}_i = \begin{cases} \tilde{y}_i - d_i\tilde{\beta}_i & \text{für } i = 1,\dots,p \\ \tilde{y}_i & \text{für } i = p+1,\dots,n. \end{cases}$$

KQ-Methode:

$\tilde{\beta}_1,\dots,\tilde{\beta}_p$ so wählen, dass

$$\sum_{i=1}^{n} \tilde{e}_i^2 = \min$$

Quadratsumme $\sum\tilde{e}_i^2$ $(= \sum e_i^2)$ wird minimal falls

$$\tilde{\beta}_1,\dots,\tilde{\beta}_p \text{ so, dass } \tilde{e}_1 = \cdots = \tilde{e}_p = 0\,,$$

d.h.

$$\tilde{\beta}_j = \tilde{y}_j / d_j\,,\ \ j = 1,\dots,p$$

In Matrix-Notation:

$$\tilde{\boldsymbol{\beta}} = \mathbf{D}^+\tilde{\mathbf{y}}$$

und da

$$\tilde{\boldsymbol{\beta}} = \mathbf{V}^{\mathsf{T}}\boldsymbol{\beta} \quad \text{und} \quad \tilde{\mathbf{y}} = \mathbf{U}^{\mathsf{T}}\mathbf{y}$$

so erhalten wir

$$\underbrace{\mathbf{V}^{\mathsf{T}}\boldsymbol{\beta}}_{=\tilde{\boldsymbol{\beta}}} = \mathbf{D}^+ \underbrace{\mathbf{U}^{\mathsf{T}}\mathbf{y}}_{=\tilde{\mathbf{y}}}$$

d.h.

(8) $$\boxed{\boldsymbol{\beta} = \mathbf{V}\,\mathbf{D}^{+}\mathbf{U}^{\top}\mathbf{y} = \mathbf{X}^{+}\mathbf{y}}$$

wobei

$\mathbf{X} = \mathbf{UDV}^{\top}$ = ursprüngliche Datenmatrix (Design-Matrix),

$\mathbf{X}^{+} = \mathbf{VD}^{+}\mathbf{U}^{\top}$ = verallgemeinerte Inverse zu $\mathbf{X}$ (Moore-Penrose-Inverse).

Resultat:

Modell: $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$

KQ-Lösung: $\boldsymbol{\beta} = \mathbf{X}^{+}\mathbf{y}$.

Klassisches Lösung:

Lineares Modell:

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

KQ-Methode:

$\boldsymbol{\beta}$ so, dass $\sum e_i^2 = \mathbf{e}^{\top}\mathbf{e} = \min$

Notwendige Bedingung für Extremalstelle: Normalgleichungen

$$\mathbf{X}^{\top}\mathbf{X}\boldsymbol{\beta} = \mathbf{X}^{\top}\mathbf{y}$$

Voraussetzung: $rk(\mathbf{X}) = p$, d.h. $\mathbf{X}^{\top}\mathbf{X}$ regulär

Dann gilt

(9) $$\boxed{\boldsymbol{\beta} = (\mathbf{X}^{\top}\mathbf{X})^{-1}\mathbf{X}^{\top}\mathbf{y}}$$

Sind die beiden Lösungen identisch?

Singulärwert-Zerlegung von $\mathbf{X}$:

$$\mathbf{X} = \mathbf{UDV}^{\top}$$

$$\mathbf{X}^{\top}\mathbf{X} = \underbrace{\mathbf{VD}^{\top}\mathbf{U}^{\top}}_{=\mathbf{X}^{\top}}\ \underbrace{\mathbf{UDV}^{\top}}_{=\mathbf{X}} = \mathbf{VD}^{\top}\ \underbrace{\mathbf{U}^{\top}\mathbf{U}}_{=\mathbf{I}_n}\mathbf{DV}^{\top} = \mathbf{VD}^{\top}\mathbf{DV}^{\top}$$

$$\mathbf{D}^{\top}\mathbf{D} = (p \times p) = diag(d_1^2, \ldots, d_p^2) = \boldsymbol{\Lambda},\ \lambda_i = d_i^2 > 0$$

$\mathbf{X}^{\top}\mathbf{X} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^{\top}$ (Spektralzerlegung, Eigenwertdarstellung von $\mathbf{X}^{\top}\mathbf{X}$)

$$(\mathbf{X}^{\top}\mathbf{X})^{-1} = (\mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^{\top})^{-1} = \mathbf{V}\boldsymbol{\Lambda}^{-1}\mathbf{V}^{\top}$$

$$(\mathbf{X}^{\top}\mathbf{X})^{-1}\mathbf{X}^{\top} = \mathbf{V}\boldsymbol{\Lambda}^{-1}\ \underbrace{\mathbf{V}^{\top}\mathbf{V}}_{=\mathbf{I}_p}\mathbf{D}^{\top}\mathbf{U}^{\top} = \mathbf{V}\ \underbrace{\boldsymbol{\Lambda}^{-1}\mathbf{D}^{\top}}_{=\mathbf{D}^{+}}\mathbf{U}^{\top} = \mathbf{VD}^{+}\mathbf{U}^{\top} = \mathbf{X}^{+}$$

Somit gilt

$$\boldsymbol{\beta} = (\mathbf{X}^{\top}\mathbf{X})^{-1}\mathbf{X}^{\top}\mathbf{y} = \mathbf{X}^{+}\mathbf{y}\ .$$

Falls $rk(\mathbf{X}) = p$, so ist die klassische Lösung äquivalent zur SVD-Lösung.

4. Multikollinearität im lineares Modell

Vorher: Regulärer Fall, $rk(\mathbf{X}) = p$, $\mathbf{X}^\top\mathbf{X} = (p \times p)$ regulär, $(\mathbf{X}^\top\mathbf{X})^{-1}$ existiert .

Jetzt: $rk(\mathbf{X}) = r < p$, $(\mathbf{X}^\top\mathbf{X})^{-1}$ existiert nicht, Multikollinearität.

Lineares Modell:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}\text{, wobei } \mathbf{X} = (n \times p)\text{, } \mathbf{y} = (n \times 1) \text{ und } \boldsymbol{\beta} = (p \times 1)$$

Gesucht:

$$\boldsymbol{\beta} \text{ so, dass } \sum_{i=1}^{n} e_i^2 = \mathbf{e}^\top\mathbf{e} = \left\|\mathbf{X}\boldsymbol{\beta} - \mathbf{y}\right\|^2 = \min$$

Lösung:

Singulärwert-Zerlegung von $\mathbf{X}$

$$\mathbf{X} = \mathbf{UDV}^\top,$$

wobei

$$\mathbf{D} = (n \times p) = diag(d_1, \dots, d_p) \text{ mit } d_1 \ge d_2 \ge \dots \ge d_r > 0,\ d_{r+1} = \dots = d_p = 0$$

Transformation des Modell:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e} = \mathbf{UDV}^\top\boldsymbol{\beta} + \mathbf{e}$$

$$\Rightarrow \mathbf{U}^\top\mathbf{y} = \mathbf{DV}^\top\boldsymbol{\beta} + \mathbf{U}^\top\mathbf{e} \qquad \text{(Multiplikation mit } \mathbf{U}^\top\text{)}$$

$$\Rightarrow \tilde{\mathbf{y}} = \mathbf{D}\tilde{\boldsymbol{\beta}} + \tilde{\mathbf{e}},\ \text{ wobei } \tilde{\mathbf{y}} = \mathbf{U}^\top\mathbf{y},\ \ \tilde{\mathbf{e}} = \mathbf{U}^\top\mathbf{e},\ \ \tilde{\boldsymbol{\beta}} = \mathbf{V}^\top\boldsymbol{\beta}.$$

Es gilt

$$\tilde{\mathbf{e}}^\top\tilde{\mathbf{e}} = \mathbf{e}^\top\mathbf{U}^\top\mathbf{U}\,\mathbf{e} = \mathbf{e}^\top\mathbf{e}$$

KQ-Problem

$$\tilde{\boldsymbol{\beta}}\ (= \mathbf{V}^\top\boldsymbol{\beta}) \text{ so wählen, dass } \tilde{\mathbf{e}}^\top\tilde{\mathbf{e}} = \min.$$

Nun gilt $\tilde{\mathbf{e}} = \tilde{\mathbf{y}} - \mathbf{D}\tilde{\boldsymbol{\beta}}$, d.h.

$$\tilde{e}_i = \begin{cases} \tilde{y}_i - d_i\tilde{\beta}_i & \text{für } i = 1, \dots, r \\ \tilde{y}_i & \text{für } i = r+1, \dots, n. \end{cases}$$

Quadratsumme $\sum \tilde{e}_i^2$ nur abhängig von $\tilde{\beta}_1, \dots, \tilde{\beta}_r$

$\tilde{\beta}_{r+1}, \dots, \tilde{\beta}_p$ ohne Einfluss auf Quadratsumme

$\sum \tilde{e}_i^2\ (= \sum e_i^2)$ wird minimal falls

$\tilde{\beta}_1, \dots, \tilde{\beta}_r$ so, dass $\tilde{e}_1 = \dots = \tilde{e}_r = 0$

$\tilde{\beta}_{r+1}, \dots, \tilde{\beta}_p$ beliebig (ohne Einfluss auf Quadratsumme!)

d.h.

$$\tilde{\beta}_1 = \tilde{y}_1/d_1,\ \tilde{\beta}_2 = \tilde{y}_2/d_2, \cdots, \tilde{\beta}_r = \tilde{y}_r/d_r$$

$$\tilde{\beta}_{r+1}, \cdots, \tilde{\beta}_p \ \text{ beliebig}$$

Da $\tilde{\boldsymbol{\beta}} = \mathbf{V}^\top\boldsymbol{\beta}$ so gilt $\boldsymbol{\beta} = \mathbf{V}\tilde{\boldsymbol{\beta}}$,wobei $\tilde{\beta}_1, \tilde{\beta}_2, \dots, \tilde{\beta}_r$ eindeutig festgelegt (identifizierbar) und $\tilde{\beta}_{r+1}, \dots, \tilde{\beta}_p$ beliebig wählbar (Minimum eindeutig, Minimalstelle nicht eindeutig).

Gesamtheit aller Lösungsvektoren $\boldsymbol{\beta}$: $(p - r)$ -dimensionaler Raum.

Zur Interpretation:
Es gilt

$$\mathbf{X} = \mathbf{U}\mathbf{D}\mathbf{V}^\mathsf{T} \Leftrightarrow \mathbf{X}\mathbf{V} = \mathbf{U}\mathbf{D}$$

Nun ist

$$\mathbf{U} = (n \times n) = (\mathbf{u}_1, \ldots, \mathbf{u}_n)$$
$$\mathbf{U}\mathbf{D} = (n \times p) = (d_1\mathbf{u}_1, \ldots, d_r\mathbf{u}_r, \mathbf{0}, \ldots, \mathbf{0})$$
$$\mathbf{V} = (p \times p) = (\mathbf{v}_1, \ldots, \mathbf{v}_p)$$
$$\mathbf{X}\mathbf{V} = (n \times p) = (\mathbf{X}\mathbf{v}_1, \ldots, \mathbf{X}\mathbf{v}_p)$$

und da $\mathbf{X}\mathbf{V} = \mathbf{U}\mathbf{D}$, so gilt

$$\mathbf{X}\mathbf{v}_{r+1} = \cdots = \mathbf{X}\mathbf{v}_p = 0\,.$$

Es ist

$\mathbf{X}\mathbf{v}_j$ = Linearkombination der Spalten von $\mathbf{X}$ mit Koeffizientenvektor $\mathbf{v}_j$;

Die Information $\mathbf{X}\mathbf{v}_j = 0$ kann wertvoll sein (unerwarteter Zusammenhang zwischen erklärenden Variablen!). Eine Spalte von $\mathbf{X}$ und der zugehörige Parameter sind eliminierbar!

Beispiel:

$\mathbf{X} = (n \times p) = (\mathbf{x}_1, \ldots, \mathbf{x}_p)$

$\mathbf{X}\boldsymbol{\beta} = \beta_1\mathbf{x}_1 + \cdots + \beta_p\mathbf{x}_p$

Annahme: 1. und 2. Spalte von $\mathbf{X}$ identisch, d.h. $\mathbf{x}_1 = \mathbf{x}_2$

dann gilt $\beta_1\mathbf{x}_1 + \beta_2\mathbf{x}_2 = (\beta_1 + \beta_2)\mathbf{x}_1$

$\tilde{\beta}_1 = \beta_1 + \beta_2$ ist eindeutig festgelegt, identifizierbar,

$\tilde{\beta}_2 = \beta_1 - \beta_2$ nicht identifizierbar, kann bei gegebenem $\tilde{\beta}_1$ beliebige Werte annehmen, ohne dass sich das Modell ändert!

Zusätzliche Forderung an Parametervektor, damit Lösung eindeutig wird:

$\tilde{\beta}_1, \ldots, \tilde{\beta}_r$ identifizierbar;

$\tilde{\beta}_{r+1}, \ldots, \tilde{\beta}_p$ beliebig wählbar (nicht identifizierbar), ohne Einfluß auf Quadratsumme.

Wähle $\tilde{\beta}_{r+1} = \cdots = \tilde{\beta}_p = 0$;

dann gilt $\sum_{i=1}^{p} \tilde{\beta}_i^2 = \sum_{i=1}^{r} \tilde{\beta}_i^2 = \min$ (Parametervektor mit minimaler Länge).

Da $\tilde{\boldsymbol{\beta}} = \mathbf{V}^\mathsf{T}\boldsymbol{\beta}$ so gilt $\boldsymbol{\beta} = \mathbf{V}\tilde{\boldsymbol{\beta}}$ und somit

$$\boldsymbol{\beta}^\mathsf{T}\boldsymbol{\beta} = \tilde{\boldsymbol{\beta}}^\mathsf{T} \underbrace{\mathbf{V}^\mathsf{T}\mathbf{V}}_{=\mathbf{I}_p} \tilde{\boldsymbol{\beta}} = \tilde{\boldsymbol{\beta}}^\mathsf{T}\tilde{\boldsymbol{\beta}}$$

d.h.

$$\tilde{\beta}_{r+1} = \cdots = \tilde{\beta}_p = 0 \quad \Leftrightarrow \quad \sum_{i=1}^{p} \beta_i^2 = \sum_{i=1}^{p} \tilde{\beta}_i^2 = \sum_{i=1}^{r} \tilde{\beta}_i^2 = \min$$

Zusatzforderung an Parametervektor, um Eindeutigkeit zu erreichen.

Zusammenfassung: KQ-Methode mit Zusatzforderung:

$\boldsymbol{\beta}$ so wählen, dass

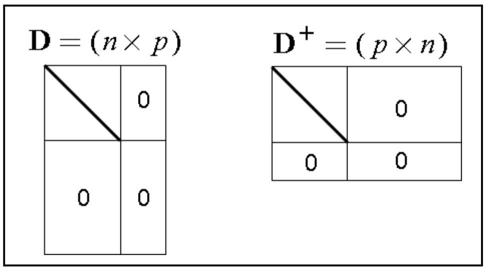
1) $\mathbf{e}^{\mathsf{T}}\mathbf{e} = \min$
2) $\boldsymbol{\beta}^{\mathsf{T}}\boldsymbol{\beta} = \min$

Dann sind alle Parameter eindeutig festgelegt (identifizierbar).

Lösung in Matrix-Notation

$$\mathbf{D} = (n \times p) = diag(d_1, \ldots, d_r, 0, \ldots, 0)$$

$$\mathbf{D}^{+} = (p \times n) = diag(\frac{1}{d_1}, \ldots, \frac{1}{d_r}, 0, \ldots, 0)$$



$\mathbf{D}^{+}$ ist die *verallgemeinerte Inverse* zu $\mathbf{D}$. Dann gilt

$$\tilde{\boldsymbol{\beta}} = \mathbf{D}^{+}\tilde{\mathbf{y}}$$

und da $\tilde{\boldsymbol{\beta}} = \mathbf{V}^{\mathsf{T}}\boldsymbol{\beta}$ und $\tilde{\mathbf{y}} = \mathbf{U}^{\mathsf{T}}\mathbf{y}$ so erhalten wir

$$\mathbf{V}^{\mathsf{T}}\boldsymbol{\beta} = \mathbf{D}^{+}\mathbf{U}^{\mathsf{T}}\mathbf{y} \quad \text{d.h.} \quad \boldsymbol{\beta} = \mathbf{V}\,\mathbf{D}^{+}\mathbf{U}^{\mathsf{T}}\mathbf{y} = \mathbf{X}^{+}\mathbf{y}\text{, und es gilt } \sum \beta_i^2 = \sum \tilde{\beta}_i^2 = \min$$

5. Stochastisches Modell

Es sei

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

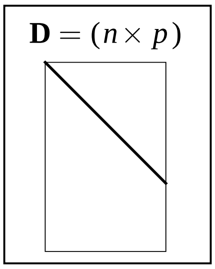
wobei

$\mathbf{X} = (n \times p)$	gegebene Datenmatrix (Design-Matrix)
$\boldsymbol{\beta} = (p \times 1)$	unbekannter Parametervektor
$\mathbf{y} = (n \times 1)$	Vektor von beobachtbaren Zufallsgrößen
$\boldsymbol{\varepsilon} = (n \times 1)$	Vektor von nicht-beobachtbaren Zufallsgrößen (Fehlervariablen)

$$E(\boldsymbol{\varepsilon}) = \mathbf{0}, \ E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\mathsf{T}}) = \sigma^2\mathbf{I}_n$$

KQ-Methode:

$$\boldsymbol{\beta} = \hat{\boldsymbol{\beta}} \text{ so, dass } \mathbf{e}^{\mathsf{T}}\mathbf{e} = \sum e_i^2 = \min, \text{ wobei } \mathbf{e} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}$$



Singulärwert-Zerlegung von $\mathbf{X}$: $\mathbf{X} = \mathbf{U}\mathbf{D}\mathbf{V}^{\mathsf{T}}$

Es gilt

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} = \mathbf{U}\mathbf{D}\mathbf{V}^{\mathsf{T}}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

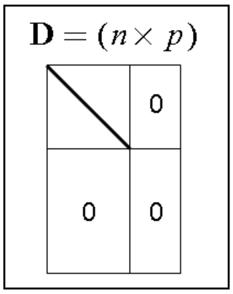
$$\Rightarrow \underbrace{\mathbf{U}^{\mathsf{T}}\mathbf{y}}_{=\tilde{\mathbf{y}}} = \mathbf{D}\underbrace{\mathbf{V}^{\mathsf{T}}\boldsymbol{\beta}}_{=\tilde{\boldsymbol{\beta}}} + \underbrace{\mathbf{U}^{\mathsf{T}}\boldsymbol{\varepsilon}}_{=\tilde{\boldsymbol{\varepsilon}}} \quad \text{(Multiplikation mit } \mathbf{U}^{\mathsf{T}}\text{)}$$

$$\Rightarrow \tilde{\mathbf{y}} = \mathbf{D}\tilde{\boldsymbol{\beta}} + \tilde{\boldsymbol{\varepsilon}},$$

und weiter

$$E(\tilde{\boldsymbol{\varepsilon}}) = \mathbf{U}^{\mathsf{T}}E(\boldsymbol{\varepsilon}) = \mathbf{0}, \ E(\tilde{\boldsymbol{\varepsilon}}\tilde{\boldsymbol{\varepsilon}}^{\mathsf{T}}) = \mathbf{U}^{\mathsf{T}}E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\mathsf{T}})\mathbf{U} = \sigma^2\mathbf{I}_n$$

Wir erhalten das einfache Modell (in neuer Notation, ohne Schlangen)



$$\text{(10)} \quad \begin{aligned} &\mathbf{y} = \mathbf{D}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \\ &E(\boldsymbol{\varepsilon}) = \mathbf{0},\ E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^\top) = \sigma^2 \mathbf{I}_n \end{aligned}$$

Annahme: $rk(\mathbf{X}) = r \le p$.

Dann gilt

$$\mathbf{D} = (n \times p) = \left(\begin{array}{c|c} \mathbf{D}_r & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} \end{array}\right), \text{ wobei } \mathbf{D}_r = (r \times r) = diag(d_1, \ldots, d_r),$$

$$\boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix},\ \boldsymbol{\beta}_1^\top = (\beta_1, \ldots, \beta_r), \boldsymbol{\beta}_2^\top = (\beta_{r+1}, \ldots, \beta_p),$$

$$\mathbf{y} = \mathbf{D}\boldsymbol{\beta} + \boldsymbol{\varepsilon} = \mathbf{D}_r \boldsymbol{\beta}_1 + \boldsymbol{\varepsilon}.$$

Nur die r Komponenten von $\boldsymbol{\beta}_1$ treten im Modell auf, die $p-r$ Komponenten von $\boldsymbol{\beta}_2$ können beliebig gewählt werden (Pseudoparameter, nicht identifizierbar).

Sinnvolle Forderung: $\beta_{r+1} = \cdots = \beta_p = 0$ (Pseudoparameter gleich null setzen)

Frage:

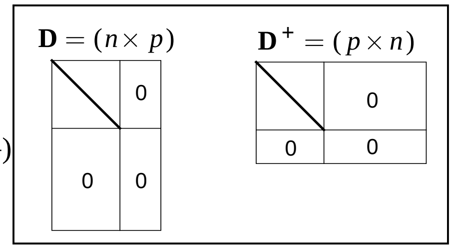
Schätzfunktionen $\hat{\boldsymbol{\beta}}$?

Varianz-Kovarianz-Matrix der Schätzfunktionen?

KQ-Methode mit Zusatzforderung:

$$\hat{\beta}_1 = y_1/d_1, \ldots, \hat{\beta}_r = y_r/d_r,$$
$$\hat{\beta}_{r+1} = \cdots = \hat{\beta}_p = 0.$$

In Matrix-Notation



$$\mathbf{D} = (n \times p) = \left(\begin{array}{c|c} \mathbf{D}_r & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} \end{array}\right), \text{ wobei } \mathbf{D}_r = (r \times r) = diag(d_1, \ldots, d_r)$$

$$\mathbf{D}^+ = (p \times r) = \left(\begin{array}{c|c} \mathbf{D}_r^{-1} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} \end{array}\right), \text{ wobei } \mathbf{D}_r^{-1} = (r \times r) = diag(\frac{1}{d_1}, \ldots, \frac{1}{d_r})$$

$$\boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix},\ \boldsymbol{\beta}_1 = (r \times 1) \text{ identifizierbar}, \boldsymbol{\beta}_2 \text{ nicht identifizierbar}$$

$$\mathbf{y} = (n \times 1) = \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{pmatrix},\ \mathbf{y}_1 = (r \times 1).$$

KQ-Methode:

$$\hat{\boldsymbol{\beta}}_1 = \mathbf{D}_r^{-1}\mathbf{y}_1,\ \hat{\boldsymbol{\beta}}_2 = \mathbf{0}, \text{ d.h. } \hat{\boldsymbol{\beta}} = \mathbf{D}^+\mathbf{y}$$

Nun gilt

$$E(\mathbf{y}) = \mathbf{D}\boldsymbol{\beta} \quad \Rightarrow \quad E(\mathbf{y}_1) = \mathbf{D}_r\boldsymbol{\beta}_1,\ E(\mathbf{y}_2) = \mathbf{0}$$

$$\Rightarrow \quad E(\hat{\boldsymbol{\beta}}) = E(\mathbf{D}^+\mathbf{y}) = \mathbf{D}^+ E(\mathbf{y}) = \mathbf{D}^+\mathbf{D}\boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{\beta}_1 \\ \mathbf{0} \end{pmatrix}$$

d.h. die Schätzung $\hat{\boldsymbol{\beta}} = \mathbf{D}^+\mathbf{y}$ ist erwartungstreu. Weiter gilt

$$\text{var}(\hat{\boldsymbol{\beta}}_1) = \mathbf{D}_r^{-1} \underbrace{\text{var}(\mathbf{y}_1)}_{=\sigma^2 \mathbf{I}_r} \mathbf{D}_r^{-1} = \sigma^2 \mathbf{D}_r^{-2}, \quad \text{var}(\hat{\boldsymbol{\beta}}_2) = (\mathbf{0}), \quad \text{cov}(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2) = 0$$

d.h.

$$\operatorname{var}(\hat{\boldsymbol\beta}) = \mathbf{D}^+ \operatorname{var}(\mathbf{y})\,(\mathbf{D}^+)^\mathsf{T} = \sigma^2 \mathbf{D}^+ (\mathbf{D}^+)^\mathsf{T} = \sigma^2 \left(\begin{array}{c|c} \mathbf{D}_r^{-2} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} \end{array}\right) = (p \times p)\,.$$

Da die Schätzung $\hat{\boldsymbol\beta}_1$ die beste unter den lineare erwartungstreuen Schätzungen ist (BLUE, Gauss-Markoff), und da wir $\boldsymbol\beta_2 = \hat{\boldsymbol\beta}_2 = \mathbf{0}$ gesetzt haben, so besitzt auch $\hat{\boldsymbol\beta} = \mathbf{D}^+\mathbf{y}$ die BLUE-Eigenschaften.

Weiter gilt für $\mathbf{e} = \mathbf{y} - \mathbf{D}\hat{\boldsymbol\beta}$

$$\mathbf{e}^\mathsf{T}\mathbf{e} = \sum_{i=1}^{n} e_i^2 = \sum_{i=r+1}^{n} y_i^2 = \mathbf{y}_2^\mathsf{T}\mathbf{y}_2,\ \ E(\mathbf{y}_2) = 0,\ \ \operatorname{var}(\mathbf{y}_2) = E(\mathbf{y}_2\mathbf{y}_2^\mathsf{T}) = \sigma^2 \mathbf{I}_{n-r}$$

$$\Rightarrow\ \ E(\mathbf{e}^\mathsf{T}\mathbf{e}) = (n-r)\sigma^2$$

$$\Rightarrow\ \ \hat\sigma^2 = \frac{1}{n-r}\sum_{i=r+1}^{n} y_i^2 \ \ \text{ist erwartungstreuer Schätzer für } \sigma^2$$

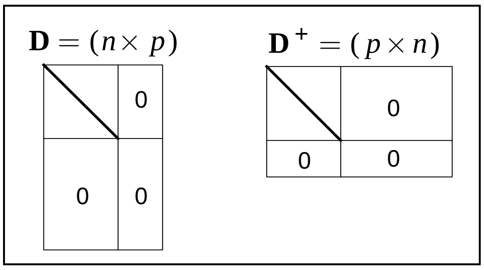
Zusammenfassung zum einfachen Modell:

Modell:

$$\mathbf{D} = (n \times p) = \left(\begin{array}{c|c} \mathbf{D}_r & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} \end{array}\right),\ \text{wobei}\ \mathbf{D}_r = (r \times r) = diag(d_1,\ldots,d_r),$$

$$\boldsymbol\beta = \begin{pmatrix} \boldsymbol\beta_1 \\ \boldsymbol\beta_2 \end{pmatrix},\ \boldsymbol\beta_1^\mathsf{T} = (\beta_1,\ldots,\beta_r), \boldsymbol\beta_2^\mathsf{T} = (\beta_{r+1},\ldots,\beta_p),$$

$$\mathbf{y} = \mathbf{D}\boldsymbol\beta + \boldsymbol\varepsilon = \mathbf{D}_r\boldsymbol\beta_1 + \boldsymbol\varepsilon,\ \ E(\boldsymbol\varepsilon) = 0,\ \ E(\boldsymbol\varepsilon\boldsymbol\varepsilon^\mathsf{T}) = \sigma^2\mathbf{I}_n\,.$$



Pseudoparameter gleich null setzen: $\beta_{r+1} = \cdots = \beta_p = 0$

KQ-Methode:

$$\hat{\boldsymbol\beta} = \begin{pmatrix} \hat{\boldsymbol\beta}_1 \\ \hat{\boldsymbol\beta}_2 \end{pmatrix} = \mathbf{D}^+\mathbf{y} = \begin{pmatrix} \mathbf{D}_r^{-1}\mathbf{y}_1 \\ \mathbf{0} \end{pmatrix}$$

Es gilt:

$$E(\hat{\boldsymbol\beta}) = \boldsymbol\beta,\ \ \operatorname{var}(\hat{\boldsymbol\beta}) = \sigma^2\mathbf{D}^+(\mathbf{D}^+)^\mathsf{T} = \sigma^2\left(\begin{array}{c|c} \mathbf{D}_r^{-2} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} \end{array}\right)$$

Zusammenfassung zum Originalmodell:

Originalmodell:

(11) $\quad \mathbf{y} = \mathbf{X}\boldsymbol\beta + \boldsymbol\varepsilon,\ E(\boldsymbol\varepsilon) = 0,\ E(\boldsymbol\varepsilon\boldsymbol\varepsilon^\mathsf{T}) = \sigma^2\mathbf{I}_n$

Mit der Singulärwert-Zerlegung $\mathbf{X} = \mathbf{U}\mathbf{D}\mathbf{V}^\mathsf{T}$ erhalten wir aus (11) das einfache Modell

(12) $\quad \tilde{\mathbf{y}} = \mathbf{D}\tilde{\boldsymbol\beta} + \tilde{\boldsymbol\varepsilon}$ wobei $\tilde{\mathbf{y}} = \mathbf{U}^\mathsf{T}\mathbf{y},\ \tilde{\boldsymbol\varepsilon} = \mathbf{U}^\mathsf{T}\boldsymbol\varepsilon,\ \tilde{\boldsymbol\beta} = \mathbf{V}^\mathsf{T}\boldsymbol\beta,\ \ E(\tilde{\boldsymbol\varepsilon}) = 0,\ E(\tilde{\boldsymbol\varepsilon}\tilde{\boldsymbol\varepsilon}^\mathsf{T}) = \sigma^2\mathbf{I}_n\,.$

Identifizierbarkeit der Parameter:

Falls $rk(\mathbf{X}) = rk(\mathbf{D}) < p$, so sind die Parameter $\tilde\beta_{r+1},\ldots,\tilde\beta_p$ Pseudoparameter, die im einfachen Modell (12) gar nicht auftreten; wir setzen diese Parameter gleich null. Wir verlangen also, dass die freien Modellparameter so festgelegt werden, dass $\tilde{\boldsymbol\beta}^\mathsf{T}\tilde{\boldsymbol\beta} = \min$. Daraus ergibt $\tilde\beta_{r+1} = \ldots = \tilde\beta_p = 0$. Da $\tilde{\boldsymbol\beta}^\mathsf{T}\tilde{\boldsymbol\beta} = \boldsymbol\beta^\mathsf{T}\boldsymbol\beta$, so lautet die entsprechende Forderung im Originalmodell: Parameter so festlegen, dass $\boldsymbol\beta^\mathsf{T}\boldsymbol\beta = \min$ (Parametervektor mit minimaler Länge). Mit dieser Zusatzforderung werden alle Modellparameter eindeutig festgelegt auf $\boldsymbol\beta = \mathbf{V}\tilde{\boldsymbol\beta}$.

KQ-Methode: $\tilde{\boldsymbol{\beta}} = \hat{\tilde{\boldsymbol{\beta}}}$ so wählen, dass

1) $\tilde{\mathbf{e}}^{\mathsf{T}}\tilde{\mathbf{e}} = \min$, wobei $\tilde{\mathbf{e}} = \tilde{\mathbf{y}} - \mathbf{X}\hat{\tilde{\boldsymbol{\beta}}}$
2) $\hat{\tilde{\boldsymbol{\beta}}}^{\mathsf{T}}\hat{\tilde{\boldsymbol{\beta}}} = \min$, d.h. $\hat{\tilde{\beta}}_{r+1} = \ldots = \hat{\tilde{\beta}}_p = 0$

Dann sind die Schätzfunktionen eindeutig festgelegt (identifizierbar): $\hat{\tilde{\boldsymbol{\beta}}} = \mathbf{D}^{+}\tilde{\mathbf{y}}$

Nun wollen wir die Ergebnisse beim einfachen Modell zurückspielen auf das Originalmodell.

KQ-Methode im Originalmodell: $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}$ so wählen, dass

1) $\mathbf{e}^{\mathsf{T}}\mathbf{e} = \min$, wobei $\mathbf{e} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}$
2) $\hat{\boldsymbol{\beta}}^{\mathsf{T}}\hat{\boldsymbol{\beta}} = \min$, d.h. Schätzvektor mit minimaler Länge

Da $\mathbf{e}^{\mathsf{T}}\mathbf{e} = \tilde{\mathbf{e}}^{\mathsf{T}}\tilde{\mathbf{e}}$, $\boldsymbol{\beta}^{\mathsf{T}}\boldsymbol{\beta} = \tilde{\boldsymbol{\beta}}^{\mathsf{T}}\tilde{\boldsymbol{\beta}}$
so erhalten wir

$$\hat{\boldsymbol{\beta}} = \mathbf{V}\hat{\tilde{\boldsymbol{\beta}}} = \mathbf{V}\mathbf{D}^{+}\tilde{\mathbf{y}} = \mathbf{V}\mathbf{D}^{+}\mathbf{U}^{\mathsf{T}}\mathbf{y} = \mathbf{X}^{+}\mathbf{y}\,.$$

Die Singulärwert-Zerlegung $\mathbf{X} = \mathbf{U}\mathbf{D}\mathbf{V}^{\mathsf{T}}$ liefert also die natürlichen Transfomationen:

$\tilde{\mathbf{y}} = \mathbf{U}^{\mathsf{T}}\mathbf{y}$, dann gilt $\operatorname{var}(\tilde{\mathbf{y}}) = \operatorname{var}(\mathbf{y}) = \sigma^2\mathbf{I}_n$;

$\tilde{\boldsymbol{\varepsilon}} = \mathbf{U}^{\mathsf{T}}\boldsymbol{\varepsilon}$, dann gilt $\operatorname{var}(\tilde{\boldsymbol{\varepsilon}}) = \operatorname{var}(\boldsymbol{\varepsilon}) = \sigma^2\mathbf{I}_n$;

$\tilde{\boldsymbol{\beta}} = \mathbf{V}^{\mathsf{T}}\boldsymbol{\beta}$, dann gilt $\boldsymbol{\beta}^{\mathsf{T}}\boldsymbol{\beta} = \tilde{\boldsymbol{\beta}}^{\mathsf{T}}\tilde{\boldsymbol{\beta}}$ (minimale Länge bewirkt Identifizierbarkeit)

$\hat{\boldsymbol{\beta}} = \mathbf{V}\hat{\tilde{\boldsymbol{\beta}}}$, dann gilt: $\operatorname{var}(\hat{\boldsymbol{\beta}}) = \mathbf{V}\operatorname{var}(\hat{\tilde{\boldsymbol{\beta}}})\,\mathbf{V}^{\mathsf{T}} = \sigma^2\mathbf{V}\left(\begin{array}{c|c}\mathbf{D}_r^{-2} & \mathbf{0}\\ \hline \mathbf{0} & \mathbf{0}\end{array}\right)\mathbf{V}^{\mathsf{T}}$.

6. Singulärwert-Zerlegung und verallgemeinerte Inverse

Definition der verallgemeinerten Inversen (vgl. Toutenburg, 2003, S. 504)
Es sei $\mathbf{A} = (m \times n)$ eine beliebige reelle Matrix. Eine Matrix $\mathbf{A}^{+} = (n \times m)$ mit

1) $\mathbf{A}\mathbf{A}^{+}\mathbf{A} = \mathbf{A}$
2) $\mathbf{A}^{+}\mathbf{A}\mathbf{A}^{+} = \mathbf{A}^{+}$
3) $(\mathbf{A}\mathbf{A}^{+})^{\mathsf{T}} = \mathbf{A}\mathbf{A}^{+} = (m \times m)$ (Symmetrie von $\mathbf{A}\mathbf{A}^{+}$)
4) $(\mathbf{A}^{+}\mathbf{A})^{\mathsf{T}} = \mathbf{A}^{+}\mathbf{A} = (n \times n)$ (Symmetrie von $\mathbf{A}^{+}\mathbf{A}$)

heißt *verallgemeinerte Inverse (Moore-Penrose-Inverse, Pseudo-Inverse)* von **A**.

Satz: Zu jeder reellen Matrix $\mathbf{A} = (m \times n)$ existiert genau eine verallgemeinerte Inverse $\mathbf{A}^{+}$.

Beweis der Eindeutigkeit (vgl. Stör, 1999, S. 248):
Es seien **U** und **V** zwei Exemplare von verallgemeinerten Inversen $\mathbf{A}^{+}$ mit den Eigenschaften 1) bis 4). Wir wollen zeigen, dass die beiden Matrizen identisch sein müssen. Es gilt

$$\begin{aligned}
\mathbf{U} &= \underbrace{\mathbf{U}\mathbf{A}\mathbf{U}}_{\text{wegen 2)}} = \mathbf{U}(\underbrace{\mathbf{A}\mathbf{V}\mathbf{A}}_{=\mathbf{A}})\mathbf{U} = \mathbf{U}(\underbrace{\mathbf{A}\mathbf{V}\mathbf{A}}_{=\mathbf{A}})\mathbf{V}(\underbrace{\mathbf{A}\mathbf{V}\mathbf{A}}_{=\mathbf{A}})\mathbf{U}\\
&= \underbrace{(\mathbf{U}\mathbf{A})^{\mathsf{T}}}_{=\mathbf{U}\mathbf{A}}\underbrace{(\mathbf{V}\mathbf{A})^{\mathsf{T}}}_{=\mathbf{V}\mathbf{A}}\mathbf{V}\underbrace{(\mathbf{A}\mathbf{V})^{\mathsf{T}}}_{=\mathbf{A}\mathbf{V}}\underbrace{(\mathbf{A}\mathbf{U})^{\mathsf{T}}}_{=\mathbf{A}\mathbf{U}}\\
&= \underbrace{(\mathbf{A}^{\mathsf{T}}\mathbf{U}^{\mathsf{T}}\mathbf{A}^{\mathsf{T}})}_{=\mathbf{A}^{\mathsf{T}}}\mathbf{V}^{\mathsf{T}}\mathbf{V}\,\mathbf{V}^{\mathsf{T}}\underbrace{(\mathbf{A}^{\mathsf{T}}\mathbf{U}^{\mathsf{T}}\mathbf{A}^{\mathsf{T}})}_{=\mathbf{A}^{\mathsf{T}}}\\
&= \underbrace{\mathbf{A}^{\mathsf{T}}\mathbf{V}^{\mathsf{T}}}_{=\mathbf{V}\mathbf{A}}\mathbf{V}\underbrace{\mathbf{V}^{\mathsf{T}}\mathbf{A}^{\mathsf{T}}}_{=\mathbf{A}\mathbf{V}} = \mathbf{V}\underbrace{\mathbf{A}\mathbf{V}\mathbf{A}}_{=\mathbf{A}}\mathbf{V} = \mathbf{V}\mathbf{A}\mathbf{V} = \mathbf{V}
\end{aligned}$$

Somit ist die die Moore-Penrose-Inverse eindeutig bestimmt.

Zusammenhang mit Singulärtwert-Zerlegung:

Die Singulärwert-Zerlegung von **A** hat die Form $\mathbf{A} = \mathbf{UDV}^{\mathsf{T}}$, wobei $\mathbf{U}, \mathbf{V}$ orthogonal und $\mathbf{D} = (m \times n) = diag(d_1, \ldots, d_p)$ mit $d_1 \geq d_2 \geq \cdots \geq d_r > 0,\, d_{r+1} = \cdots = d_p = 0$. Es sei

$$\mathbf{D}^{+} = (n \times m) = diag\left(\frac{1}{d_1}, \ldots, \frac{1}{d_r}, 0, \ldots. 0\right),$$

$$\mathbf{A}^{+} = \mathbf{V}\,\mathbf{D}^{+}\mathbf{U}^{\mathsf{T}}.$$

Dann ist $\mathbf{D}^{+}$ die verallgemeinerte Inverse zu **D** und $\mathbf{A}^{+}$ die verallgemeinerte Inverse zu **A**.

Beweis: Es gilt

$$\mathbf{DD}^{+} = (m \times m) = \left(\begin{array}{c|c} \mathbf{I}_r & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} \end{array}\right) \quad \text{und} \quad \mathbf{D}^{+}\mathbf{D} = (n \times n) = \left(\begin{array}{c|c} \mathbf{I}_r & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} \end{array}\right)$$

und daraus folgen sofort die Eigenschaften 1) bis 4) für **D** und $\mathbf{D}^{+}$. Weiter gilt

1) $\mathbf{A}\,\mathbf{A}^{+}\mathbf{A} = \underbrace{\mathbf{UDV}^{\mathsf{T}}}_{=\mathbf{A}}\underbrace{\mathbf{VD}^{+}\mathbf{U}^{\mathsf{T}}}_{=\mathbf{A}^{+}}\underbrace{\mathbf{UDV}^{\mathsf{T}}}_{=\mathbf{A}} = \mathbf{U}\underbrace{\mathbf{DD}^{+}\mathbf{D}}_{=\mathbf{D}}\mathbf{V}^{\mathsf{T}} = \mathbf{A}$

2) $\mathbf{A}^{+}\mathbf{A}\,\mathbf{A}^{+} = \underbrace{\mathbf{VD}^{+}\mathbf{U}^{\mathsf{T}}}_{=\mathbf{A}^{+}}\underbrace{\mathbf{UDV}^{\mathsf{T}}}_{=\mathbf{A}}\underbrace{\mathbf{VD}^{+}\mathbf{U}^{\mathsf{T}}}_{=\mathbf{A}^{+}} = \mathbf{V}\underbrace{\mathbf{D}^{+}\mathbf{DD}^{+}}_{=\mathbf{D}^{+}}\mathbf{U}^{\mathsf{T}} = \mathbf{A}^{+}$

3) $\mathbf{A}\,\mathbf{A}^{+} = \underbrace{\mathbf{UDV}^{\mathsf{T}}}_{=\mathbf{A}}\underbrace{\mathbf{VD}^{+}\mathbf{U}^{\mathsf{T}}}_{=\mathbf{A}^{+}} = \mathbf{UDD}^{+}\mathbf{U}^{\mathsf{T}}$ ist offensichtlich symmetrisch

4) $\mathbf{A}^{+}\mathbf{A} = \underbrace{\mathbf{VD}^{+}\mathbf{U}^{\mathsf{T}}}_{=\mathbf{A}^{+}}\underbrace{\mathbf{UDV}^{\mathsf{T}}}_{=\mathbf{A}} = \mathbf{VD}^{+}\mathbf{DV}^{\mathsf{T}}$ ist offensichtlich symmetrisch

und somit ist $\mathbf{A}^{+}$ die verallgemeinerte Inverse zu $\mathbf{A}$. Die Singulärwert-Zerlegung liefert also eine einfache und übersichtliche Darstellung der Moore-Penrose-Inversen!

Literatur

Golub, G. H. (1965). Numerical Methods for Solving Linear Least Squares Problems. Numerische Mathematik 7, 206-216.

Golub, G. H. and Kahan, W. (1965).Calculating the Singular Values and Pseudoinverse of a Matrix. SIAM Journal of Numerical Analysis 2, 205-224.

Golub, Gene H. and van Loan, Charles F. (1996). Matrix Computations. The Johns Hopkins University Press, Baltimore and London.

Hammarling, Sven (1985). The Singular Value Decomposition in Multivariate Statistics. ACM Signum Newsletter 20, 2-25.

Stoer, Josef (1999). Numerische Mathematik. Springer, Berlin.

Toutenburg, Helge (2003). Lineare Modelle. Theorie und Anwendungen. Physica-Verlag, Heidelberg.