

Matrixalgebra

mit einer Einführung in lineare Modelle

Stefan Lang

Institut für Statistik

Ludwigstrasse 33

email: lang@stat.uni-muenchen.de

25. August 2004

Vielen Dank an

Christiane Belitz,
Manuela Hummel und
Sebastian Kaiser

für die große Hilfe bei der Erstellung des Skripts.

Vielen Dank an

Prof. Ludwig Fahrmeir und
Prof. Gerhard Tutz

für zahlreiche Verbesserungsvorschläge.

Die Kapitel 9-11 sind durch die grandiose Vorlesung
„lineare Modelle“ von Prof. Franz Ferschl beeinflusst.

Inhaltsverzeichnis

1. Einführung	1
1.1 Vektoren im \mathbf{R}^n	1
1.2 Reelle Matrizen	7
1.3 Matrixmultiplikation	13
1.4 Reduktion auf Dreiecksform und Diagonalform	21
2. Vektorräume	33
2.1 Komplexe Zahlen	33
2.2 Definition von Vektorräumen und Beispiele	37
2.3 Unterräume	39
2.4 Basis und Dimension	43
2.5 Lineare Abbildungen zwischen Vektorräumen	53
2.6 Euklidische Vektorräume	59
3. Der Rang einer Matrix	69
3.1 Definition und Eigenschaften des Rangs	69
3.2 Inverse einer Matrix	73
3.3 Praktische Bestimmung des Rangs einer Matrix	75
3.4 Vollrang Zerlegung einer Matrix	78
4. Determinante und Spur	83
4.1 Permutationen	83

4.2	Determinante einer Matrix	84
4.3	Die Spur einer Matrix	93
5.	Lineare Gleichungssysteme	95
5.1	Definition und allgemeine Eigenschaften	95
5.2	Lösen von linearen Gleichungssystemen	97
5.3	Verallgemeinerte Inverse	103
6.	Eigenwerte und Eigenvektoren	109
6.1	Definition und allgemeine Eigenschaften	109
6.2	Ähnliche Matrizen	115
6.3	Eigenwerte symmetrischer Matrizen	116
7.	Quadratische Formen und definite Matrizen	121
7.1	Definition und allgemeine Eigenschaften	121
7.2	Choleskyzerlegung	126
8.	Differenziation von Matrizen	137
8.1	Differenziation nach einem Skalar	137
8.2	Differenziation einer Matrixfunktion nach der Matrix	138
9.	Die multivariate Normalverteilung	139
9.1	Die univariate Normalverteilung	139
9.2	Die multivariate Normalverteilung	143
9.2.1	Marginalverteilungen, Unkorreliertheit, Unabhängigkeit	146
9.3	Testverteilungen: χ^2 – t – und F –Verteilung	147
9.3.1	Die χ^2 –Verteilung	147
9.3.2	Die t –Verteilung	152
9.3.3	Die F –Verteilung	153

10. Das klassische lineare Regressionsmodell	155
10.1 Modelldefinition	155
10.2 Die Methode der kleinsten Quadrate	158
10.3 Die KQ-Methode unter linearen Nebenbedingungen	162
10.4 Geschätzte Werte, Residuen, Streuungszerlegung	166
10.4.1 Geschätzte Werte	166
10.4.2 Residuen	166
10.4.3 Geometrische Eigenschaften der KQ-Schätzung	167
10.4.4 Streuungszerlegung	169
10.5 Eigenschaften des KQ-Schätzers	172
10.5.1 Gütekriterien	173
10.5.2 Statistische Eigenschaften linearer Schätzer	174
10.5.3 Eine Schätzung für die Varianz σ^2	178
10.5.4 Vergleich des unrestringierten mit dem KQ-Schätzer unter linearen Nebenbedingungen	180
11. Klassische Normalregression	183
11.1 Maximum Likelihood-Schätzung	183
11.2 Verteilungstheorie im klassischen linearen Modell	185
11.3 Tests für allgemeine lineare Hypothesen	186
11.3.1 Bestimmung von SSE_H und ΔSSE	187
11.3.2 Stochastische Eigenschaften der Differenz ΔSSE	189
11.3.3 Einige spezielle Testprobleme	192
11.3.4 Konfidenzintervalle und Konfidenzellipsoide	194
A. Körper	197
Literaturverzeichnis	199

Index 201

1

Einführung

In diesem Einführungskapitel beschäftigen wir uns zunächst mit n -Tupeln

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

reeller Zahlen x_1, \dots, x_n . Für die Menge der reellen n -Tupel kann eine Addition und eine (skalare) Multiplikation definiert werden. Mit diesen Definitionen wird die Menge der n -Tupel zu einem sogenannten Vektorraum. Die Theorie allgemeiner Vektorräume ist Gegenstand von Kapitel 2. In den Abschnitten 1.2 und 1.3 führen wir dann den für dieses Skript zentralen Begriff der Matrix ein und definieren die Addition, skalare Multiplikation und Multiplikation von Matrizen. Da Matrizen eng mit sogenannten linearen Gleichungssystemen verknüpft sind und die Lösung linearer Gleichungssysteme im Laufe des Skriptes immer wieder benötigt wird, soll bereits in diesem ersten Kapitel in Abschnitt 1.4 der Zusammenhang zwischen Matrizen und linearen Gleichungssystemen hergestellt werden. Dabei können aber noch nicht alle Details bei der Lösung linearer Gleichungssysteme behandelt werden. Eine ausführlichere Diskussion linearer Gleichungssysteme erfolgt dann in Kapitel 5 und in Abschnitt 7.2 von Kapitel 7.

1.1 Vektoren im \mathbb{R}^n

In der Statistik (und in vielen anderen Wissenschaften) ist es häufig zweckmäßig eine Menge von (reellen) Zahlen x_1, \dots, x_n , zu einem geordneten n -Tupel

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \tag{1.1}$$

zusammenzufassen. Beispielsweise könnte es sich bei den Zahlen x_1, \dots, x_n um eine Stichprobe von n Personen aus einer größeren Grundgesamtheit handeln und bei den Werten

x_i , $i = 1, \dots, n$, um die gemessene Körpergröße, das Gewicht, Einkommen etc. der i -ten Person. In der folgenden Definition definieren wir für n -Tupel der Form (1.1) zwei Verknüpfungen, die Addition von Vektoren und die skalare Multiplikation eines Vektors mit einer reellen Zahl (dem sogenannten Skalar). Wir bezeichnen geordnete n -Tupel der Form (1.1) versehen mit Addition und skalarer Multiplikation als Vektoren im \mathbb{R}^n .

Definition 1.1 (n -dimensionaler Vektorraum über \mathbb{R})

Die Menge aller n -Tupel

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

reeller Zahlen x_1, \dots, x_n versehen mit den beiden unten definierten Verknüpfungen wird n -dimensionaler Vektorraum über \mathbb{R} , kurz \mathbb{R}^n genannt. Die Zahlen x_1, \dots, x_n heißen auch Skalare. Wir definieren für Vektoren $x \in \mathbb{R}^n, y \in \mathbb{R}^n$ und dem Skalar $\lambda \in \mathbb{R}$ folgende Operationen:

(i) (Vektoraddition)

$$x + y = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{pmatrix}$$

(ii) (Multiplikation mit einem Skalar)

$$\lambda \cdot x = \lambda \cdot \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} \lambda \cdot x_1 \\ \lambda \cdot x_2 \\ \vdots \\ \lambda \cdot x_n \end{pmatrix}$$

Den Nullvektor

$$\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

bezeichnen wir im Folgenden mit $\mathbf{0}$ und den Einsvektor

$$\begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$$

mit **1**.

Bemerkung:

Vektoren des \mathbb{R}^n werden in der Literatur sowohl spaltenweise als auch zeilenweise dargestellt. Eine zeilenweise Darstellung von

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

ist gegeben durch

$$x = (x_1, \dots, x_n).$$

Im Zusammenhang mit Matrizen kann die gleichzeitige spalten- und zeilenweise Darstellung von Vektoren des \mathbb{R}^n besonders für Anfänger verwirrend sein. Fasst man nämlich Vektoren des \mathbb{R}^n als spezielle Matrizen auf (vgl. Abschnitt 1.2), dann identifiziert die zeilenweise Darstellung eines Vektors eine andere Matrix als die spaltenweise Darstellung. Wenn im Folgenden von Vektoren des \mathbb{R}^n die Rede ist, gehen wir daher stets von der *spaltenweise* Darstellung aus. Um trotzdem Vektoren des \mathbb{R}^n platzsparend auch zeilenweise darstellen zu können, schreiben wir im Folgenden gelegentlich $x = (x_1, \dots, x_n)'$ und meinen damit den Vektor

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}.$$

▽

Vektoren im \mathbb{R}^2 und die Vektoraddition bzw. Multiplikation mit einem Skalar können geometrisch veranschaulicht werden. Wir können den Vektor $x = (x_1, x_2)'$ in einem kartesischen Koordinatensystem als Pfeil vom Ursprung (Punkt $(0,0)'$) zu den Koordinaten $(x_1, x_2)'$ darstellen (Abbildung 1.1). Die Addition zweier Vektoren $x = (x_1, x_2)'$ und $y = (y_1, y_2)'$ ergibt sich als die Diagonale des von x und y aufgespannten Parallelogramms (Abbildung 1.2). Das Produkt eines Vektors $x = (x_1, x_2)'$ mit einem Skalar $\lambda \in \mathbb{R}$ bedeutet eine Streckung (falls $|\lambda| > 1$) bzw. Stauchung ($|\lambda| < 1$) des Vektors x . Falls $\lambda > 0$ bleibt die Richtung erhalten, im Falle $\lambda < 0$ ändert sich die Richtung des Vektors (Abbildung 1.3).

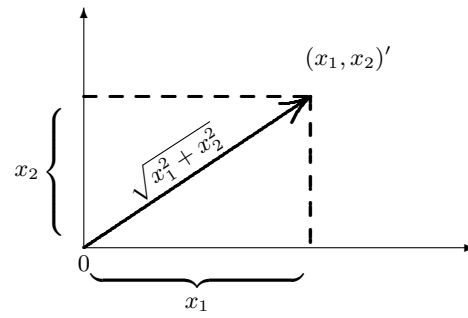


Abbildung 1.1. Geometrische Veranschaulichung eines Vektors im \mathbb{R}^2 .

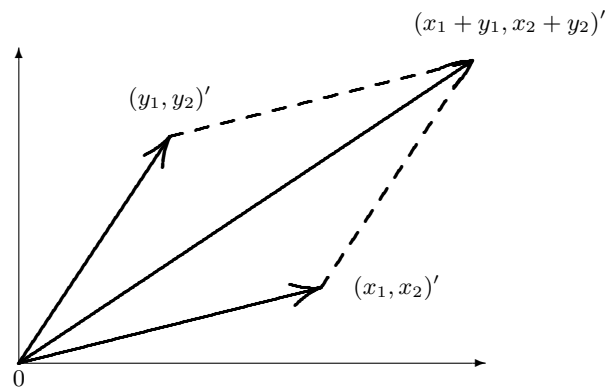


Abbildung 1.2. Geometrische Veranschaulichung der Vektoraddition im \mathbb{R}^2 .

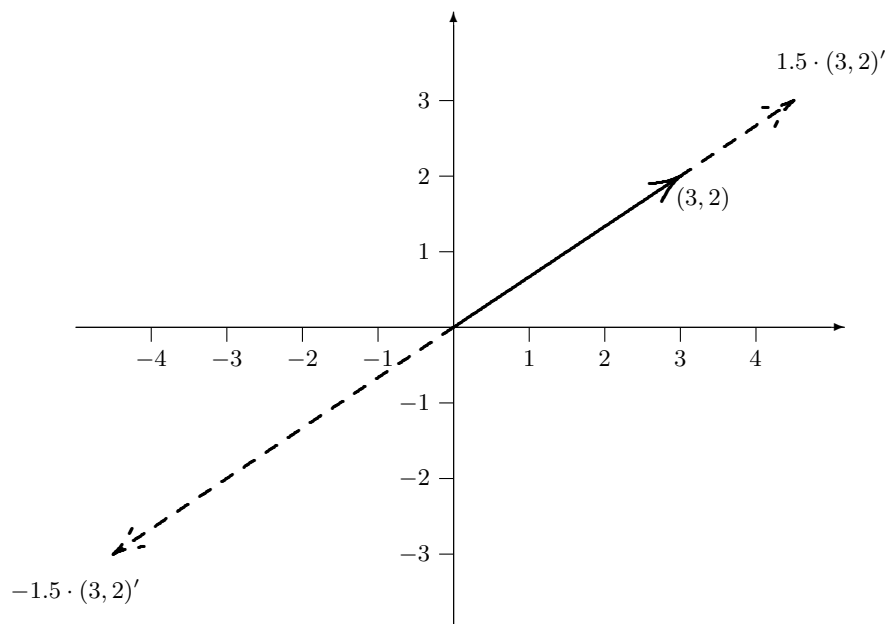


Abbildung 1.3. Veranschaulichung der Multiplikation mit einem Skalar

Für Vektoren in \mathbb{R}^n gelten folgende einfache Rechenregeln:

Satz 1.1 (Rechenregeln für Vektoren im \mathbb{R}^n)

Für beliebige Vektoren $x, y, z \in \mathbb{R}^n$ und Skalare $\lambda, \mu \in \mathbb{R}^n$ gilt:

1. Assoziativgesetz für die Addition: $x + (y + z) = (x + y) + z$
2. Kommutativgesetz: $x + y = y + x$
3. $x + \mathbf{0} = x$
4. $x + (-x) = \mathbf{0}$
5. Distributivgesetze für die skalare Multiplikation: $(\lambda + \mu)x = \lambda x + \mu x$ bzw. $\lambda(x + y) = \lambda x + \lambda y$
6. Assoziativgesetz für die skalare Multiplikation: $(\lambda\mu)x = \lambda(\mu x)$
7. $1 \cdot x = x$

Beweis:

Einfaches Nachrechnen. □

Wir werden im nächsten Kapitel sehen, dass die Vektoren im \mathbb{R}^n nicht die einzigen Strukturen sind, für die obige Rechengesetze gelten. Vielmehr kann man für eine Vielzahl von Mengensystemen eine Vektoraddition und eine skalare Multiplikation derart definieren, dass obige fundamentale Rechenregeln gelten.

Definition 1.2 (Skalarprodukt)

Das Skalarprodukt oder inneres Produkt $\langle x, y \rangle$ der Vektoren $x, y \in \mathbb{R}^n$ ist definiert als

$$\langle x, y \rangle = x_1 \cdot y_1 + x_2 \cdot y_2 + \cdots + x_n \cdot y_n.$$

Zwei Vektoren heißen orthogonal, wenn

$$\langle x, y \rangle = 0$$

gilt.

In Kapitel 2.6 werden wir noch eine Verallgemeinerung des Standardskalarprodukts kennenlernen. Im \mathbb{R}^2 läßt sich die Orthogonalität zweier Vektoren wieder geometrisch veranschaulichen. Sind nämlich zwei Vektoren zueinander orthogonal, so stehen sie senkrecht aufeinander (Abbildung 1.4).

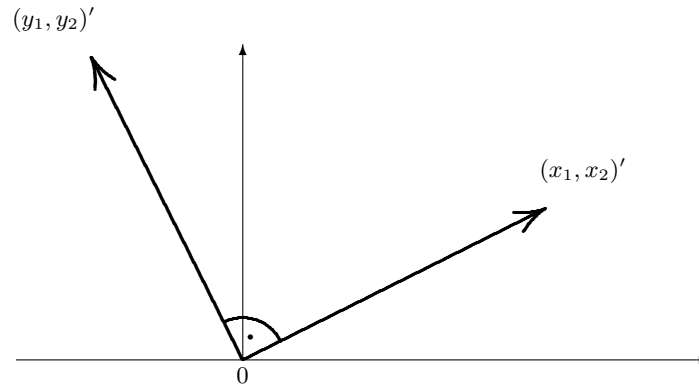


Abbildung 1.4. Beispiel für zwei Vektoren x und y mit $\langle x, y \rangle = 0$.

Beispiel 1.1

Wir betrachten die Vektoren $x = (1, 2, 3)'$, $y = (2, -1, 2)'$ und $z = (-1, 0, \frac{1}{3})'$ des \mathbb{R}^3 . Es gilt

$$\langle x, y \rangle = 1 \cdot 2 + 2 \cdot (-1) + 3 \cdot 2 = 6$$

und

$$\langle x, z \rangle = 1 \cdot (-1) + 2 \cdot 0 + 3 \cdot \frac{1}{3} = 0.$$

Die Vektoren x und z sind also zueinander orthogonal.

△

Bemerkung:

Der Raum \mathbb{R}^n versehen mit der Vektoraddition, der skalaren Multiplikation und dem Skalarprodukt heißt euklidischer Raum.

▽

Definition 1.3 (Abstand und Länge)

Gegeben seien die Vektoren x und y im \mathbb{R}^n . Der (euklidische) Abstand $d(x, y)$ zwischen den Punkten x und y ist definiert als

$$\begin{aligned} d(x, y) &= \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots + (x_n - y_n)^2} \\ &= \sqrt{\langle x - y, x - y \rangle}. \end{aligned}$$

Die (euklidische) Länge $\|x\|$ eines Vektors $x \in \mathbb{R}^n$ ist definiert als

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

Der Abstand zweier Vektoren x und y im \mathbb{R}^2 ist in Abbildung 1.5 veranschaulicht. Die Länge eines Vektors x im \mathbb{R}^2 ist in Abbildung 1.6 geometrisch veranschaulicht. Wir werden in Kapitel 2.6 Verallgemeinerungen des Abstands und der Länge eines Vektors kennenlernen.

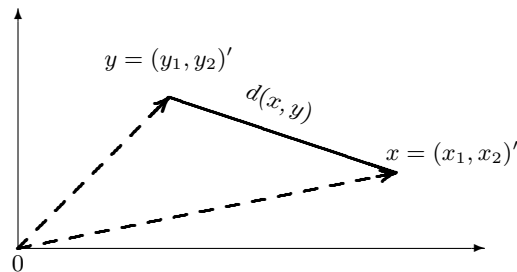


Abbildung 1.5. Veranschaulichung des euklidischen Abstands zwischen zwei Vektoren x und y im \mathbb{R}^2

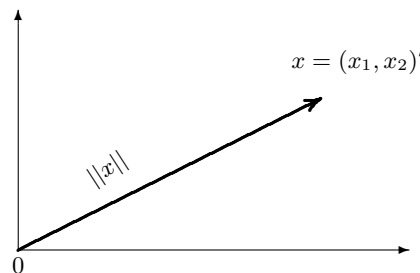


Abbildung 1.6. Veranschaulichung der Länge $\|x\|$ eines Vektors x im \mathbb{R}^2 .

1.2 Reelle Matrizen

In der Statistik interessiert man sich in der Regel nicht nur für ein Merkmal einer Person oder Untersuchungseinheit, sondern gleichzeitig für mehrere Merkmale (etwa das Alter, das Gewicht, usw. einer Person). In diesem Fall erweist es sich als zweckmäßig die Merkmalsausprägungen in einem geordneten rechteckigen Schema anzuordnen. Dieses Schema besteht dann aus $m = \text{Anzahl der Untersuchungseinheiten}$ Zeilen und $n = \text{Anzahl der untersuchten Merkmale}$ Spalten. Dies führt zu folgender Definition:

Definition 1.4 (reelle Matrix)

Ein nach m Zeilen und n Spalten geordnetes Schema \mathbf{A} von mn Elementen $a_{ij} \in \mathbb{R}$

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

heißt *reelle Matrix* von der Ordnung $m \times n$ oder kurz $m \times n$ Matrix. Kurzschreibweise:

$$\mathbf{A} = (a_{ij}), i = 1, \dots, m, j = 1, \dots, n.$$

Die Zeilen von \mathbf{A} können dabei als Vektoren des \mathbf{R}^n (sog. Zeilenvektoren) und die Spalten als Vektoren des \mathbf{R}^m (sog. Spaltenvektoren) angesehen werden. Dabei wird der j -te Zeilenvektor von \mathbf{A} mit $a^j = (a_{j1}, \dots, a_{jn})$ und der j -te Spaltenvektor mit $a_j = (a_{1j}, \dots, a_{nj})'$ bezeichnet. Zwei $m \times n$ Matrizen $\mathbf{A} = (a_{ij})$ und $\mathbf{B} = (b_{ij})$ sind genau dann gleich, wenn für alle i, j gilt: $a_{ij} = b_{ij}$.

Beispiel 1.2

Nach dem Gesetz zur Regelung der Miethöhe kann der Vermieter die Zustimmung zu einer Erhöhung des Mietzinses verlangen, wenn „der Mietzins die üblichen Entgelte nicht übersteigt, die in der Gemeinde für nicht preisgebundenen Wohnraum vergleichbarer Art, Größe, Ausstattung, Beschaffenheit und Lage in den letzten vier Jahren vereinbart oder Erhöhungen geändert worden sind“.

Zur Feststellung der „üblichen Entgelte“ erstellen die meisten Städte und viele Gemeinden sogenannte Mietspiegel. Diese ermöglichen die Berechnung der „durchschnittlichen“ Miete, die pro Quadratmeter und Monat für eine Wohnung mit einer bestimmten Wohnfläche (in Quadratmeter), einem Baualter (in Jahren) und Merkmalen, welche die Ausstattung der Wohnung, den Haustyp und die Lage der Wohnung, den Haustyp und die Lage der Wohnung in der Gemeinde charakterisieren, bezahlt wird.

Da in größeren Städten wie München eine Erfassung aller Mietpreise schon aus Zeit- und Kostengründen nicht möglich ist, werden Daten zu Miethöhen und zugehörigen Merkmalen über eine repräsentative Stichprobe gesammelt.

Folgende Merkmale werden unter anderen erhoben:

Y	Nettomiete der Wohnung
X_1	Wohnflächen
X_2	Baualter
X_4	gehobene Küchenausstattung (1 = ja, 0 = nein)
X_5	gehobener Neubau (1 = ja, 0 = nein)
X_3	geographische Lage

Die erhobenen Merkmale werden zweckmäßigerweise in einer Matrix \mathbf{A} abgelegt, deren erste zehn Zeilen folgende Gestalt besitzt:

235.9	35	39	0	0	1112
852.1	104	39	0	0	1112
693.7	29	71	0	0	2114
551.7	39	72	0	0	2148
1574.1	97	85	0	0	2222
941.5	62	62	0	0	2222
631.2	31	65	0	0	2211
723.4	61	57.5	0	0	2142
728.7	72	78	0	0	2143
1017.3	75	68	0	0	2142

In der ersten Spalte dieser Matrix sind die beobachteten Nettomieten zu finden, in der zweiten Spalte die Wohnfläche usw. Die Zeilen der Matrix beinhalten jeweils die erhobenen Merkmalsausprägungen einer bestimmten Wohnung. Die 1. Zeile besagt beispielsweise, dass die erste Wohnung eine Nettomiete von 235.9 DM, eine Wohnfläche von 35 qm usw. besitzt.

△

Definition 1.5 (transponierte Matrix)

Sei $\mathbf{A} = (a_{ij})$ eine $m \times n$ Matrix. Dann ist die transponierte Matrix \mathbf{A}' definiert als diejenige Matrix, die man durch das Vertauschen der Zeilen und Spalten von \mathbf{A} erhält, d.h.

$$\mathbf{A}' = \begin{pmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ \vdots & & & \vdots \\ \vdots & & & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{mn} \end{pmatrix}$$

Sie ist also von der Ordnung $n \times m$.

Beispiel 1.3

Betrachte die 3×4 Matrix

$$\mathbf{A} = \begin{pmatrix} 2 & 4 & 1 & 6 \\ 1 & 0 & 3 & 2 \\ 9 & 3 & 4 & 3 \end{pmatrix}.$$

Die transponierte von \mathbf{A} ist gegeben durch die 4×3 Matrix

$$\mathbf{A}' = \begin{pmatrix} 2 & 1 & 9 \\ 4 & 0 & 3 \\ 1 & 3 & 4 \\ 6 & 2 & 3 \end{pmatrix}.$$

△

Wir definieren im Folgenden noch einige spezielle Matrizen, die immer wieder auftauchen werden.

Definition 1.6 (quadratische Matrix)

Eine Matrix \mathbf{A} heißt quadratisch, falls sie von der Ordnung $n \times n$ ist. Die Diagonale, welche aus den Elementen a_{11}, \dots, a_{nn} besteht, heißt Hauptdiagonale.

Eine wichtige quadratische Matrix ist die sogenannte *Einheitsmatrix* \mathbf{I}_n , deren Einträge auf der Hauptdiagonalen sämtlich gleich Eins und ober bzw. unterhalb der Hauptdiagonalen Null sind, d.h.

$$\mathbf{I}_n = \begin{pmatrix} 1 & 0 & \dots & 0 \\ \vdots & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ 0 & \dots & \dots & 1 \end{pmatrix}.$$

Weitere spezielle quadratische Matrizen werden in den folgenden Definitionen angegeben:

Definition 1.7 (Diagonalmatrix)

Eine quadratische Matrix \mathbf{D} heißt Diagonalmatrix, wenn ihre Einträge unter- und oberhalb der Hauptdiagonalen Null sind. \mathbf{D} hat also folgende Gestalt:

$$\mathbf{D} = \begin{pmatrix} d_1 & 0 & \dots & 0 \\ \vdots & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ 0 & \dots & \dots & d_n \end{pmatrix}$$

Schreibweise: $\mathbf{D} = \text{diag}(d_1, \dots, d_n)$

Um eine spezielle Diagonalmatrix handelt es sich beispielsweise bei der Einheitsmatrix.

Definition 1.8 (symmetrische Matrix)

Eine quadratische Matrix \mathbf{A} heißt symmetrisch, wenn gilt: $\mathbf{A} = \mathbf{A}'$.

Offenbar ist jede Diagonalmatrix, also auch die Einheitsmatrix, eine symmetrische Matrix.

Beispiel 1.4

Ein Beispiel für eine symmetrische Matrix ist gegeben durch

$$\mathbf{A} = \begin{pmatrix} 2 & 3 & 1 & 8 \\ 3 & 2 & 7 & 5 \\ 1 & 7 & 6 & 6 \\ 8 & 5 & 6 & 0 \end{pmatrix}.$$

In manchen Situationen ist es nützlich eine Matrix \mathbf{A} in Teilmatrizen (auch Submatrizen) \mathbf{A}_{ij} zu partitionieren:

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} & \cdots & \mathbf{A}_{1c} \\ \vdots & & & \vdots \\ \vdots & & & \vdots \\ \mathbf{A}_{r1} & \mathbf{A}_{r2} & \cdots & \mathbf{A}_{rc} \end{pmatrix} = (\mathbf{A}_{ij})$$

Dabei haben die Teilmatrizen $\mathbf{A}_{i1}, \dots, \mathbf{A}_{ic}$, $i = 1, \dots, r$ jeweils gleiche Zeilenzahl und die Teilmatrizen $\mathbf{A}_{1j}, \dots, \mathbf{A}_{rj}$, $j = 1, \dots, c$ gleiche Spaltenzahl.

Es ist leicht einzusehen, dass die Transponierte einer partitionierten Matrix die transponierte Matrix der Transponierten der Teilmatrizen ist, d.h.

$$\mathbf{A}' = \begin{pmatrix} \mathbf{A}'_{11} & \mathbf{A}'_{21} & \cdots & \mathbf{A}'_{r1} \\ \vdots & & & \vdots \\ \vdots & & & \vdots \\ \mathbf{A}'_{1c} & \mathbf{A}'_{2c} & \cdots & \mathbf{A}'_{rc} \end{pmatrix}.$$

Beispiel 1.5

Betrachte die Matrix

$$\mathbf{A} = \left(\begin{array}{cc|cc} 1 & 2 & -1 & 3 \\ 2 & -2 & 1 & 0 \\ \hline 1 & -2 & 3 & 4 \\ -2 & 4 & 5 & 1 \end{array} \right)$$

Definiert man die Matrizen

$$\mathbf{A}_{11} = \begin{pmatrix} 1 & 2 \\ 2 & -2 \end{pmatrix}, \quad \mathbf{A}_{12} = \begin{pmatrix} -1 & 3 \\ 1 & 0 \end{pmatrix}$$

$$\mathbf{A}_{21} = \begin{pmatrix} 1 & -2 \\ -2 & 4 \end{pmatrix}, \quad \mathbf{A}_{22} = \begin{pmatrix} 3 & 4 \\ 5 & 1 \end{pmatrix},$$

so gilt

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix}.$$

Die transponierte Matrix \mathbf{A}' der Matrix \mathbf{A} ist gegeben durch

$$\mathbf{A}' = \begin{pmatrix} 1 & 2 & 1 & -2 \\ 2 & -2 & -2 & 4 \\ -1 & 1 & 3 & 5 \\ 3 & 0 & 4 & 1 \end{pmatrix}.$$

Wir definieren jetzt ähnlich wie für Vektoren des \mathbf{R}^n die Addition zweier Matrizen und die skalare Multiplikation eines Skalars mit einer Matrix.

Definition 1.9 (Summe und skalare Multiplikation von Matrizen)

Die Summe $\mathbf{A} + \mathbf{B}$ zweier $m \times n$ Matrizen $\mathbf{A} = (a_{ij})$ und $\mathbf{B} = (b_{ij})$ ist definiert als:

$$\mathbf{A} + \mathbf{B} := (a_{ij} + b_{ij}).$$

Die Multiplikation von \mathbf{A} mit einem Skalar $\lambda \in \mathbf{R}$ ist definiert als

$$\lambda \mathbf{A} := (\lambda a_{ij}).$$

Beispiel 1.6

Betrachte die Matrizen

$$\mathbf{A} = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 5 & 2 \\ 1 & 2 & 2 \end{pmatrix} \quad \text{und} \quad \mathbf{B} = \begin{pmatrix} 1 & 4 & 2 \\ 3 & 1 & 0 \\ -1 & 2 & -4 \end{pmatrix}.$$

Dann gilt für die Summe von \mathbf{A} und \mathbf{B} :

$$\mathbf{A} + \mathbf{B} = \begin{pmatrix} 1+1 & 2+4 & 3+2 \\ 3+3 & 5+1 & 2+0 \\ 1-1 & 2+2 & 2-4 \end{pmatrix} = \begin{pmatrix} 2 & 6 & 5 \\ 6 & 6 & 2 \\ 0 & 4 & -2 \end{pmatrix}.$$

△

Wir stellen im Folgenden einige fundamentale Rechenregeln für Matrizen zusammen. Bei den ersten 7 Rechenregeln handelt es sich dabei um dieselben Regeln, die bereits in Satz 1.1 für Vektoren des \mathbf{R}^n behandelt wurden. Wie bereits erwähnt, werden wir im nächsten Kapitel weitere Strukturen mit analogen Eigenschaften kennenlernen.

Satz 1.2 (Rechenregeln)

Für beliebige $m \times n$ Matrizen $\mathbf{A}, \mathbf{B}, \mathbf{C}$ und beliebige Skalare $r, k \in \mathbf{R}$ gilt:

1. Assoziativgesetz für die Addition: $\mathbf{A} + (\mathbf{B} + \mathbf{C}) = (\mathbf{A} + \mathbf{B}) + \mathbf{C}$
2. Kommutativgesetz: $\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$
3. $\mathbf{A} + \mathbf{0} = \mathbf{A}$, wobei die Nullmatrix $\mathbf{0}$ diejenige Matrix ist, deren sämtliche Einträge gleich Null sind.

$$4. \mathbf{A} + (-\mathbf{A}) = \mathbf{0}$$

$$5. \text{Distributivgesetze für die skalare Multiplikation: } (k+r)\mathbf{A} = k\mathbf{A} + r\mathbf{A} \text{ bzw. } k(\mathbf{A} + \mathbf{B}) = k\mathbf{A} + k\mathbf{B}$$

$$6. \text{Assoziativgesetz für die skalare Multiplikation: } (kr)\mathbf{A} = k(r\mathbf{A})$$

$$7. 1 \cdot \mathbf{A} = \mathbf{A}$$

$$8. 0 \cdot \mathbf{A} = \mathbf{0}.$$

$$9. (k\mathbf{A})' = k\mathbf{A}'$$

$$10. (\mathbf{A} + \mathbf{B})' = \mathbf{A}' + \mathbf{B}'$$

Beweis:

Die Regeln folgen unmittelbar aus der Definition für die Addition von Matrizen und der Definition der skalaren Multiplikation bzw. der Definition der transponierten Matrix. \square

1.3 Matrixmultiplikation

Definition 1.10 (Matrixmultiplikation)

Das Produkt der $m \times n$ Matrix $\mathbf{A} = (a_{ij})$ mit der $n \times p$ Matrix $\mathbf{B} = (b_{ij})$ ist die $m \times p$ Matrix

$$\mathbf{AB} = \mathbf{C} = (c_{ik}) \quad \text{mit} \quad c_{ik} = \sum_{j=1}^n a_{ij}b_{jk}.$$

Ausführlich erhalten wir demnach

$$\mathbf{A} \cdot \mathbf{B} = \begin{pmatrix} \sum_{j=1}^n a_{1j}b_{j1} & \sum_{j=1}^n a_{1j}b_{j2} & \cdots & \sum_{j=1}^n a_{1j}b_{jp} \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ \sum_{j=1}^n a_{mj}b_{j1} & \sum_{j=1}^n a_{mj}b_{j2} & \cdots & \sum_{j=1}^n a_{mj}b_{jp} \end{pmatrix}.$$

Man beachte, dass zwei Matrizen \mathbf{A} und \mathbf{B} nur dann multiplizierbar sind, wenn die Anzahl der Spalten von \mathbf{A} gleich der Anzahl der Zeilen von \mathbf{B} ist.

Beispiel 1.7

Betrachte die Matrizen

$$\mathbf{A} = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \quad \text{und} \quad \mathbf{B} = \begin{pmatrix} -1 & 2 \\ 1 & 2 \end{pmatrix}.$$

Dann erhalten wir für das Produkt

$$\mathbf{A} \cdot \mathbf{B} = \begin{pmatrix} -1 \cdot 1 + 2 \cdot 1 & 1 \cdot 2 + 2 \cdot 2 \\ -1 \cdot 3 + 4 \cdot 1 & 3 \cdot 2 + 4 \cdot 2 \end{pmatrix} = \begin{pmatrix} 1 & 6 \\ 1 & 14 \end{pmatrix}.$$

Man beachte, dass die Matrixmultiplikation nicht kommutativ ist, d.h.

$$\mathbf{B} \cdot \mathbf{A} = \begin{pmatrix} -1 \cdot 1 + 2 \cdot 3 & -1 \cdot 2 + 2 \cdot 4 \\ 1 \cdot 1 + 2 \cdot 3 & 1 \cdot 2 + 2 \cdot 4 \end{pmatrix} = \begin{pmatrix} 5 & 6 \\ 7 & 10 \end{pmatrix} \neq \mathbf{A} \cdot \mathbf{B}.$$

△

Beispiel 1.8

Falls $a \in \mathbf{R}$ und $b \in \mathbf{R}$ zwei Skalare sind, ist bekannt, dass

$$a \cdot b = 0$$

genau dann gilt, wenn entweder $a = 0$ oder $b = 0$ ist. Diese Tatsache wird auch in vielen Beweisen verwendet. Wir zeigen im Folgenden in einem Gegenbeispiel dass für Matrixprodukte aus

$$\mathbf{A} \cdot \mathbf{B} = \mathbf{0}$$

keineswegs folgt, dass \mathbf{A} oder \mathbf{B} Nullmatrizen sein müssen. Wir betrachten dazu die Matrizen

$$\mathbf{A} = \begin{pmatrix} 2 & 4 & 16 \\ 1 & -3 & -7 \\ -2 & 2 & 2 \end{pmatrix}$$

und

$$\mathbf{B} = \begin{pmatrix} -2 & -4 & -8 \\ -3 & -6 & -12 \\ 1 & 2 & 4 \end{pmatrix}.$$

Für das Produkt $\mathbf{A} \cdot \mathbf{B}$ erhalten wir

$$\begin{pmatrix} 2 & 4 & 16 \\ 1 & -3 & -7 \\ -2 & 2 & 2 \end{pmatrix} \begin{pmatrix} -2 & -4 & -8 \\ -3 & -6 & -12 \\ 1 & 2 & 4 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Das Produkt der beiden Matrizen ist also die Nullmatrix, obwohl es sich bei keinem der beiden Faktoren um die Nullmatrix handelt.

△

Beispiel 1.9

Um ein spezielles Matrizenprodukt handelt es sich beim sogenannten dyadischen Produkt xy' , das durch Multiplikation eines $m \times 1$ Spaltenvektors mit einem $1 \times n$ Zeilenvektor entsteht.

△

Beispiel 1.10

Mit Hilfe der Matrixmultiplikation lassen sich auch einige Summen darstellen. Seien $x, y \in \mathbb{R}^n$ und sei $\mathbf{1}$ der $n \times 1$ Einsvektor, dessen Einträge sämtlich aus Einsen bestehen. Dann gilt:

1. $\sum_{i=1}^n x_i = \mathbf{1}'x = x'\mathbf{1}$
2. $\sum_{i=1}^n x_i y_i = x'y = y'x$
3. $\sum_{i=1}^n x_i^2 = x'x$

Damit lassen sich das arithmetische Mittel \bar{x} und die Varianz s^2 der Zahlen $x = (x_1, x_2, \dots, x_n)'$ wie folgt in Matrixschreibweise darstellen:

1. $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{n} \mathbf{1}'x = w'x$,
wobei $w = \frac{1}{n} \mathbf{1} = (\frac{1}{n}, \dots, \frac{1}{n})'$.
2. $s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n} (x - \bar{\mathbf{x}})'(x - \bar{\mathbf{x}})$,
wobei $\bar{\mathbf{x}} = (\bar{x}, \dots, \bar{x})'$.

△

Sind die Matrizen \mathbf{A} und \mathbf{B} partitioniert in Teilmatrizen \mathbf{A}_{ij} , $i = 1, \dots, r$, $j = 1, \dots, c$, und \mathbf{B}_{lk} , $l = 1, \dots, c$, $k = 1, \dots, d$, dann ergibt sich das Produkt der beiden Matrizen zu

$$\mathbf{AB} = \begin{pmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} & \cdots & \mathbf{C}_{1d} \\ \vdots & & & \vdots \\ \vdots & & & \vdots \\ \mathbf{C}_{r1} & \mathbf{C}_{r2} & \cdots & \mathbf{C}_{rd} \end{pmatrix},$$

mit

$$\mathbf{C}_{ik} = \sum_{j=1}^c \mathbf{A}_{ij} \mathbf{B}_{jk} \quad i = 1, \dots, r \quad k = 1, \dots, d.$$

Partitionierte Matrizen können also nur in partitionierter Form multipliziert werden, wenn die entsprechenden Teilmatrizen die für die Multiplikation passende Ordnung besitzen.

Beispiel 1.11

Wir betrachten die im Beispiel 1.5 definierte Matrix \mathbf{A} , die in die 4 Teilmatrizen \mathbf{A}_{11} , \mathbf{A}_{12} , \mathbf{A}_{21} und \mathbf{A}_{22} partitioniert wurde. Weiterhin betrachten wir die Matrix

$$\mathbf{B} = \begin{pmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{21} & \mathbf{B}_{22} \end{pmatrix}$$

mit

$$\mathbf{B}_{11} = \begin{pmatrix} 2 & 3 \\ -1 & 4 \end{pmatrix}, \quad \mathbf{B}_{12} = \begin{pmatrix} 2 & 4 \\ 1 & 1 \end{pmatrix}$$

$$\mathbf{B}_{21} = \begin{pmatrix} -1 & -2 \\ 0 & 0 \end{pmatrix}, \quad \mathbf{B}_{22} = \begin{pmatrix} 1 & 2 \\ 0 & -2 \end{pmatrix}.$$

Dann können wir das Produkt $\mathbf{A} \cdot \mathbf{B}$ in partitionierter Form berechnen, d.h.

$$\mathbf{A} \cdot \mathbf{B} = \begin{pmatrix} \mathbf{A}_{11}\mathbf{B}_{11} + \mathbf{A}_{12}\mathbf{B}_{21} & \mathbf{A}_{11}\mathbf{B}_{12} + \mathbf{A}_{12}\mathbf{B}_{22} \\ \mathbf{A}_{21}\mathbf{B}_{11} + \mathbf{A}_{22}\mathbf{B}_{21} & \mathbf{A}_{21}\mathbf{B}_{12} + \mathbf{A}_{22}\mathbf{B}_{22} \end{pmatrix}.$$

Es gilt

$$\mathbf{A}_{11}\mathbf{B}_{11} + \mathbf{A}_{12}\mathbf{B}_{21} = \begin{pmatrix} 1 & 13 \\ 5 & -4 \end{pmatrix}$$

$$\mathbf{A}_{11}\mathbf{B}_{12} + \mathbf{A}_{12}\mathbf{B}_{22} = \begin{pmatrix} 3 & -2 \\ 3 & 8 \end{pmatrix}$$

$$\mathbf{A}_{21}\mathbf{B}_{11} + \mathbf{A}_{22}\mathbf{B}_{21} = \begin{pmatrix} 1 & -11 \\ 13 & 0 \end{pmatrix}$$

$$\mathbf{A}_{21}\mathbf{B}_{12} + \mathbf{A}_{22}\mathbf{B}_{22} = \begin{pmatrix} 3 & 0 \\ 5 & 4 \end{pmatrix}$$

und wir erhalten

$$\mathbf{A} \cdot \mathbf{B} = \begin{pmatrix} 1 & 13 & 3 & -2 \\ 5 & -14 & 3 & 8 \\ 1 & -11 & 3 & 0 \\ -13 & 0 & 5 & 4 \end{pmatrix}.$$

Für die Matrixmultiplikation gelten folgende Rechenregeln:

Satz 1.3 (Rechenregeln für die Matrixmultiplikation)

Für Matrizen $\mathbf{A}, \mathbf{B}, \mathbf{C}$ passender Ordnung gilt:

1. $\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC}$
2. $(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC})$
3. $(\mathbf{AB})' = \mathbf{B}'\mathbf{A}'$
4. $\mathbf{A}\mathbf{I}_n = \mathbf{A}$ bzw. $\mathbf{I}_n\mathbf{A} = \mathbf{A}$

Beweis:

zu 1) : Sei $\mathbf{A} = (a_{ij})$ eine $m \times n$ Matrix, $\mathbf{B} = (b_{jk})$ eine $n \times p$ Matrix und $\mathbf{C} = (c_{jk})$ ebenfalls von der Ordnung $n \times p$. Dann gilt mit $\mathbf{B} + \mathbf{C} =: \mathbf{D} = (b_{jk} + c_{jk}) = (d_{jk})$:

$$\begin{aligned} \mathbf{A}(\mathbf{B} + \mathbf{C}) &= \mathbf{AD} = \left(\sum_{j=1}^n a_{ij}d_{jk} \right) = \left(\sum_{j=1}^n a_{ij}(b_{jk} + c_{jk}) \right) \\ &= \left(\sum_{j=1}^n a_{ij}b_{jk} + \sum_{j=1}^n a_{ij}c_{jk} \right) = \mathbf{AB} + \mathbf{AC} \end{aligned}$$

zu 2) : Sei $\mathbf{A} = (a_{ij})$ von der Ordnung $m \times n$, $\mathbf{B} = (b_{jk})$ von der Ordnung $n \times p$ und $\mathbf{C} = (c_{ks})$ von der Ordnung $p \times q$. Dann ist \mathbf{AB} eine $m \times p$ Matrix und \mathbf{BC} eine $n \times q$ Matrix und es gilt

$$\mathbf{AB} =: \mathbf{D} = (d_{ik}) = \left(\sum_{j=1}^n a_{ij}b_{jk} \right)$$

und

$$\mathbf{BC} =: \mathbf{E} = (e_{js}) = \left(\sum_{k=1}^p b_{jk}c_{ks} \right).$$

Damit folgt:

$$\begin{aligned} (\mathbf{AB})\mathbf{C} &= (f_{is}) = \left(\sum_{k=1}^p d_{ik}c_{ks} \right) = \left(\sum_{k=1}^p \left(\sum_{j=1}^n a_{ij}b_{jk} \right) c_{ks} \right) \\ &= \left(\sum_{j=1}^n a_{ij} \left(\sum_{k=1}^p b_{jk}c_{ks} \right) \right) = \mathbf{A}(\mathbf{BC}) \end{aligned}$$

zu 3) : Sei $\mathbf{A} = (a_{ij})$ von der Ordnung $m \times n$ und $\mathbf{B} = (b_{jk})$ von der Ordnung $n \times p$. Für das Element in der i -ten Zeile und j -ten Spalte von \mathbf{AB} gilt

$$a_{i1}b_{1j} + a_{i2}b_{2j} + \dots + a_{in}b_{nj}. \quad (1.2)$$

Das Element (1.2) ist auch das Element in der j -ten Zeile und i -ten Spalte von $(\mathbf{AB})'$. Andererseits ist die j -te Zeile von \mathbf{B}' gegeben durch

$$(b_{1j}, b_{2j}, \dots, b_{nj}) \quad (1.3)$$

und die i -te Spalte von \mathbf{A}' ist gegeben durch

$$\begin{pmatrix} a_{i1} \\ a_{i2} \\ \vdots \\ a_{in} \end{pmatrix}. \quad (1.4)$$

Also ist das Element in der j -ten Zeile und i -ten Spalte von $\mathbf{B}'\mathbf{A}'$ das Produkt von (1.3) und (1.4), also (1.2), woraus die Behauptung folgt.

zu 4): Die Behauptung ist offensichtlich. □

Der folgende Satz erweist sich in Kapitel 10 als nützlich:

Satz 1.4 (Kürzungsregel)

Sei $\mathbf{X} \neq 0$ eine $m \times n$ Matrix und \mathbf{A} eine $n \times n$ Matrix. Dann folgt aus

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

die Beziehung

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

Beweis

Wegen $\mathbf{X}'\mathbf{X}\mathbf{A}\mathbf{X}'\mathbf{X} - \mathbf{X}'\mathbf{X} = 0$ folgt

$$\begin{aligned} 0 &= (\mathbf{X}'\mathbf{X}\mathbf{A} - \mathbf{I})(\mathbf{X}'\mathbf{X}\mathbf{A}\mathbf{X}'\mathbf{X} - \mathbf{X}'\mathbf{X}) \\ &= (\mathbf{X}'\mathbf{X}\mathbf{A} - \mathbf{I})\mathbf{X}'(\mathbf{X}\mathbf{A}\mathbf{X}'\mathbf{X} - \mathbf{X}) \\ &= (\mathbf{X}'\mathbf{X}\mathbf{A}\mathbf{X}' - \mathbf{X}')(\mathbf{X}\mathbf{A}\mathbf{X}'\mathbf{X} - \mathbf{X}) \\ &= y'y, \end{aligned}$$

wobei $y = \mathbf{X}\mathbf{A}\mathbf{X}'\mathbf{X} - \mathbf{X}$. Es folgt $y = 0$, und damit die Behauptung. □

Im Folgenden definieren wir noch einige spezielle Matrizen, die im weiteren Verlauf dieses Skriptes gelegentlich eine Rolle spielen:

Definition 1.11 (orthogonale Matrix)

Eine quadratische Matrix \mathbf{A} heißt orthogonal, wenn $\mathbf{A}\mathbf{A}' = \mathbf{A}'\mathbf{A} = \mathbf{I}$ gilt.

Orthogonale Matrizen besitzen folgende Eigenschaften:

Satz 1.5 (Eigenschaften orthogonaler Matrizen)

Sei \mathbf{A} eine orthogonale Matrix. Dann gilt:

1. Die Zeilenvektoren bzw. die Spaltenvektoren bilden ein Orthonormalsystem. (Eine Menge von Vektoren bilden ein Orthonormalsystem, wenn jeder Vektor die Länge Eins hat und wenn je zwei Vektoren orthogonal sind.)
2. $\mathbf{A}\mathbf{B}$ ist orthogonal, wenn \mathbf{A} und \mathbf{B} orthogonal sind.

Beweis:

zu 1) : Unter Verwendung der Zeilenvektoren von \mathbf{A} gilt:

$$\mathbf{A}\mathbf{A}' = \begin{pmatrix} a^1 \\ \vdots \\ a^n \end{pmatrix} ((a^1)' \dots (a^n)') = \begin{pmatrix} a^1(a^1)' & \dots & \dots & a^1(a^n)' \\ \vdots & & & \vdots \\ \vdots & & & \vdots \\ a^n(a^1)' & \dots & \dots & a^n(a^n)' \end{pmatrix}$$

Nun folgt aus $\mathbf{A}\mathbf{A}' = \mathbf{I} = ((e^1)' \dots (e^n)')$ die Behauptung für die Zeilenvektoren. Analog beweist man die Behauptung für die Spaltenvektoren, indem man $\mathbf{A}'\mathbf{A}$ in Abhängigkeit der Spaltenvektoren von \mathbf{A} darstellt und $\mathbf{A}'\mathbf{A} = \mathbf{I}$ beachtet.

zu 2) : Es gilt:

$$\mathbf{A}\mathbf{B}(\mathbf{A}\mathbf{B})' = \mathbf{A}\mathbf{B}\mathbf{B}'\mathbf{A}' = \mathbf{A}\mathbf{I}\mathbf{A}' = \mathbf{A}\mathbf{A}' = \mathbf{I}$$

□

Definition 1.12 (idempotente Matrix)

Eine quadratische Matrix \mathbf{A} heißt idempotent, wenn gilt: $\mathbf{A}\mathbf{A} = \mathbf{A}^2 = \mathbf{A}$.

Beispiel 1.12 (eine spezielle idempotente Matrix)

Die spezielle $n \times n$ Matrix

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

spielt eine wichtige Rolle in der Statistik. Die $n \times n$ Matrix $\mathbf{1}\mathbf{1}'$ besteht sämtlich aus Einsen, so daß \mathbf{C} auf der Hauptdiagonalen die Werte $1 - \frac{1}{n}$ stehen hat, während ansonsten der Wert $-\frac{1}{n}$ steht. \mathbf{C} ist offensichtlich symmetrisch und idempotent. Weiter gilt

1. Multiplikation von \mathbf{C} mit einem beliebigen $n \times 1$ Vektor a ergibt

$$\mathbf{C}a = \begin{pmatrix} a_1 - \bar{a} \\ \vdots \\ a_n - \bar{a} \end{pmatrix},$$

d.h. man erhält den mittelwertszentrierten Vektor von a .

2. Multiplikation von \mathbf{C} mit einer $n \times m$ Matrix \mathbf{A} liefert

$$\mathbf{C}\mathbf{A} = \begin{pmatrix} a_{11} - \bar{a}_1 & \cdots & a_{1m} - \bar{a}_m \\ \vdots & & \vdots \\ a_{n1} - \bar{a}_1 & \cdots & a_{nm} - \bar{a}_m \end{pmatrix},$$

wobei $\bar{a}_1, \dots, \bar{a}_m$ die Mittelwerte der Spalten von A sind.

3. $\mathbf{C}\mathbf{1} = \mathbf{0}$

4. $\mathbf{1}'\mathbf{C} = \mathbf{0}'$

5. $\mathbf{1}\mathbf{1}'\mathbf{C} = \mathbf{C}\mathbf{1}\mathbf{1}' = \mathbf{0}$

6. $\sum_{i=1}^n (x_i - \bar{x})^2 = x'\mathbf{C}x$ wobei $x = (x_1, \dots, x_n)'$.

Beweis:

zu 1) - 5): Der Beweis der ersten 5 Eigenschaften ist relativ einfach. Wir beweisen die erste Eigenschaft:

$$\mathbf{C}a = \left(\mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}'\right)a = a - \frac{1}{n}\mathbf{1}\mathbf{1}'a = a - \bar{a}\mathbf{1} = \begin{pmatrix} a_1 - \bar{a} \\ \vdots \\ a_n - \bar{a} \end{pmatrix}.$$

Dabei wurde $\mathbf{1}'a = \sum_{i=1}^n a_i$ benutzt. Die Eigenschaften 2) - 5) ergeben sich als Spezialfall.

zu 6): Es gilt:

$$\begin{aligned} x'\mathbf{C}x &= x'\left(\mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}'\right)x = x'x - x'\frac{1}{n}\mathbf{1}\mathbf{1}'x = \sum_{i=1}^n x_i^2 - \frac{1}{n} \sum_{i=1}^n x_i \sum_{i=1}^n x_i \\ &= \sum_{i=1}^n x_i^2 - n\bar{x}^2 = \sum_{i=1}^n (x_i - \bar{x})^2 \end{aligned}$$

Eine Anwendung der Matrix \mathbf{C} und ihrer Eigenschaften findet man in Kapitel 9 beim Beweis von Satz 9.14 sowie in Kapitel 10 bei der Herleitung der Streuungszersetzung im Abschnitt 10.4.4.

△

Satz 1.6 (Eigenschaften idempotenter Matrizen)

Für idempotente Matrizen \mathbf{A} und \mathbf{B} gilt:

1. $\mathbf{AB} = \mathbf{BA} \implies \mathbf{AB}$ idempotent.
2. $\mathbf{I} - \mathbf{A}$ ist idempotent.
3. $\mathbf{A}(\mathbf{I} - \mathbf{A}) = (\mathbf{I} - \mathbf{A})\mathbf{A} = \mathbf{0}$.

Beweis:

zu 1) : Es gilt

$$\mathbf{ABAB} = \mathbf{ABBA} = \mathbf{ABA} = \mathbf{AAB} = \mathbf{AB}.$$

zu 2) : Es gilt

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

Die Aussage 3) folgt direkt aus der Definition idempotenter Matrizen. □

1.4 Reduktion auf Dreiecksform und Diagonalform

Um einige charakteristische Matrixgrößen, wie den Rang (Kapitel 3) oder die Determinante (Kapitel 4) einer Matrix berechnen zu können, ist es meist günstig die Matrix auf sogenannte Dreiecksform zu reduzieren. Im nächsten Kapitel erweist sich die Reduktion auf Dreiecksform als vorteilhaft beim Nachweis der linearen (Un)-abhängigkeit von Vektoren. Darüberhinaus benötigt man die Reduktion einer Matrix auf Dreiecksform vor allem zum Lösen linearer Gleichungssysteme (Kapitel 5).

Definition 1.13 (Dreiecksform einer Matrix)

Eine $m \times n$ Matrix $\mathbf{A} \neq \mathbf{0}$ liegt in Dreiecksform vor, wenn sämtliche Elemente unterhalb der Hauptdiagonalen a_{11}, a_{22}, \dots Null sind und die ersten r , $r \geq 1$, Elemente auf der Hauptdiagonalen ungleich Null sind. Sie hat also folgende Gestalt:

$$\mathbf{A} = \begin{pmatrix} a_{11} & & & & & & \\ 0 & a_{22} & & & & & \\ \vdots & & \ddots & & & & \\ 0 & 0 & \dots & a_{rr} & & & \\ 0 & 0 & \dots & 0 & 0 & \dots & 0 \\ \vdots & & & & & \ddots & \vdots \\ 0 & \dots & & & \dots & 0 & \end{pmatrix}$$

Beispiel 1.13

Die Matrix

$$\mathbf{A} = \begin{pmatrix} 1 & 3 & 3 & 2 \\ 0 & 4 & 3 & 8 \\ 0 & 0 & 3 & 1 \\ 0 & 0 & 0 & 4 \end{pmatrix}$$

befindet sich genauso wie die Matrix

$$\mathbf{C} = \begin{pmatrix} 1 & 3 & 3 & 2 \\ 0 & 4 & 3 & 8 \\ 0 & 0 & 3 & 1 \\ 0 & 0 & 0 & 4 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

in Dreiecksform. Die Matrix

$$\mathbf{C} = \begin{pmatrix} 1 & 3 & 3 & 2 \\ 0 & 4 & 3 & 8 \\ 0 & 0 & 3 & 1 \\ 6 & 0 & 0 & 4 \end{pmatrix}$$

liegt hingegen nicht in Dreiecksform vor.

△

Da die Dreiecksform einer Matrix vor allem beim Lösen linearer Gleichungssysteme eine zentrale Rolle spielt, wollen wir im Folgenden definieren was wir darunter verstehen:

Definition 1.14 (Lineares Gleichungssystem)

Unter einem linearen Gleichungssystem mit Unbekannten $x_1, \dots, x_n \in \mathbb{R}$ versteht man ein System von m Gleichungen der Form

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= c_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= c_2 \\ \vdots & \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= c_m \end{aligned}$$

wobei die Skalare $a_{ij}, c_i \in \mathbb{R}$ bekannte Koeffizienten sind. Fasst man die Skalare a_{ij} , $i = 1, \dots, m$, $j = 1, \dots, n$, zur $m \times n$ Matrix \mathbf{A} und x_i und c_i zu den $n \times 1$ bzw. $m \times 1$ Spaltenvektoren x und c zusammen so lässt sich ein lineares Gleichungssystem durch

$$\mathbf{A}x = c$$

in Matrixnotation schreiben.

Beispiel 1.14

Das Gleichungssystem

$$\begin{aligned} 2x_1 + 1x_2 - 2x_3 &= 10 \\ 6x_1 + 4x_2 + 4x_3 &= 2 \\ 5x_1 + 4x_2 + 3x_3 &= 4 \end{aligned}$$

besteht aus 3 Gleichungen mit 3 Unbekannten. In Matrixnotation erhalten wir

$$\begin{pmatrix} 2 & 1 & -2 \\ 6 & 4 & 4 \\ 5 & 4 & 3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 10 \\ 2 \\ 4 \end{pmatrix}.$$

Dieses sehr einfache Gleichungssystem ist noch ohne spezielle Lösungstechniken lösbar. Man löse eine Gleichung nach einer Unbekannten auf und setze in die beiden anderen Gleichungen ein, usw. In der Statistik und in vielen anderen Anwendungsgebieten treten aber Gleichungssysteme mit hunderten oder gar tausenden von Gleichungen und Unbekannten auf. Hier ist man darauf angewiesen automatisierte Lösungstechniken zur Verfügung zu haben, die auch im Computer programmiert werden können.

△

Beispiel 1.15 (Dreiecksform und lineare Gleichungssysteme)

Wir demonstrieren anhand eines einfachen Beispiels, warum Matrizen in Dreiecksform eine wichtige Rolle bei der Lösung linearer Gleichungssysteme spielen. Betrachte das Gleichungssystem

$$\begin{pmatrix} 2 & 1 & -2 \\ 0 & 1 & 10 \\ 0 & 0 & -7 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 10 \\ -28 \\ 21 \end{pmatrix}.$$

Offensichtlich befindet sich die Koeffizientenmatrix in Dreiecksform und wir erkennen unmittelbar den entscheidenden Vorteil. Da die Koeffizientenmatrix sich in Dreiecksform befindet, können wir (fast) ohne weitere Umformungen die Lösungen “ablesen”. Wir beginnen bei der Berechnung von x_3 und erhalten unmittelbar

$$x_3 = -21/7 = -3.$$

Weiter erhalten wir (unter Verwendung der soeben erhaltenen Lösung für x_3)

$$x_2 = (-28 - 10x_3)/1 = -28 + 10 \cdot 3 = 2.$$

Zuletzt ergibt sich

$$x_1 = (10 + 2x_3 - 1x_2)/2 = (10 - 2 \cdot 3 - 1 \cdot 2)/2 = 1.$$

Beispiel 1.16 (Fertigungsprozess eines Produktes)

Als Anwendungsbeispiel für lineare Gleichungssysteme betrachten wir den Fertigungsprozess eines Endproduktes D. Wir nehmen an, dass zur Fertigung von D die Zwischenprodukte A, B, und C benötigt werden, die wiederum aus mehreren Teilen zusammengesetzt sein können. Eine solche Situation kann man in dem in Abbildung 1.7 dargestellten Graphen veranschaulichen. In diesem Fall wird das Endprodukt D aus je einem Teil des Produktes A, drei Teilen von B und vier Teilen von C hergestellt. Zur Fertigung des Zwischenproduktes B wird ein Teil von A benötigt, zur Fertigung von C sind je zwei Teile von A und B nötig.

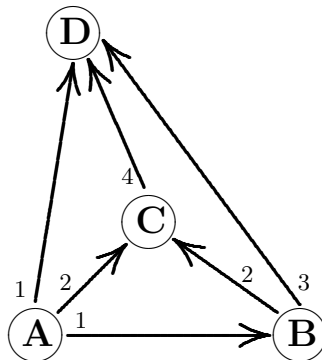


Abbildung 1.7. Graphische Veranschaulichung des Fertigungsprozesses eines Produktes D.

Es stellt sich die Frage, wie groß der Gesamtbedarf aller Produktionselemente bei Herstellung einer gewissen Anzahl von Endprodukten ist. Wir können diese Fragestellung in ein System von vier Gleichungen übersetzen. Dazu definieren wir den Vektor $x = (x_1, x_2, x_3, x_4)'$, der angibt wie viele Teile der Produkte A, B, C und D produziert werden müssen. Wir assoziieren x_1 mit dem Produkt A, x_2 mit dem Produkt B usw.. Angenommen wir wollen 25 Stück des Endproduktes D produzieren, dann gilt $x_4 = 25$. Das Zwischenprodukt A wird jeweils einmal zur Produktion eines Endproduktes D und eines Zwischenproduktes B und zweimal zur Produktion von C benötigt. Es muss also

$$x_1 = 1 \cdot x_2 + 2 \cdot x_3 + 1 \cdot x_4$$

gelten. Das Zwischenprodukt B wird zweimal zur Produktion eines Teils von C und dreimal zur Produktion von D benötigt. Wir erhalten also

$$x_2 = 2 \cdot x_3 + 3 \cdot x_4.$$

Schließlich benötigen wir 4 Teile von C zur Produktion eines Endproduktes D, woraus die Gleichung

$$x_3 = 4 \cdot x_4$$

folgt. Zusammenfassend ergibt sich ein System von vier Gleichungen:

$$\begin{aligned} 1 \cdot x_1 - 1 \cdot x_2 - 2 \cdot x_3 - 1 \cdot x_4 &= 0 \\ 0 \cdot x_1 + 1 \cdot x_2 - 2 \cdot x_3 - 3 \cdot x_4 &= 0 \\ 0 \cdot x_1 + 0 \cdot x_2 + 1 \cdot x_3 - 4 \cdot x_4 &= 0 \\ 0 \cdot x_1 + 0 \cdot x_2 + 0 \cdot x_3 + 1 \cdot x_4 &= 25 \end{aligned}$$

Hierbei handelt es sich wieder um ein Beispiel für ein lineares Gleichungssystem in Dreiecksform. Wir können das Gleichungssystem kompakt in Matrixschreibweise darstellen als

$$\begin{pmatrix} 1 & -1 & -2 & -1 \\ 0 & 1 & -2 & -3 \\ 0 & 0 & 1 & -4 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 25 \end{pmatrix}.$$

Mit der Koeffizientenmatrix

$$\mathbf{A} = \begin{pmatrix} 1 & -1 & -2 & -1 \\ 0 & 1 & -2 & -3 \\ 0 & 0 & 1 & -4 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

und dem Ergebnisvektor

$$c = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 25 \end{pmatrix}$$

erhalten wir

$$\mathbf{A}x = c.$$

Die Lösung dieses Gleichungssystems stellt sich wieder als vergleichsweise einfach dar, weil die Koeffizientenmatrix \mathbf{A} in Dreiecksform vorliegt. Durch die spezielle Form der Matrix können wir die Lösungen mehr oder weniger „ablesen“. Wir beginnen mit x_4 und erhalten $x_4 = 25$. Einsetzen von x_4 in die dritte Gleichung liefert $x_3 - 4 \cdot 25 = 0$, also $x_3 = 100$. Anschließend fahren wir fort mit x_2 und berechnen zuletzt x_1 . Als Lösungsvektor erhalten wir

$$x = \begin{pmatrix} 500 \\ 275 \\ 100 \\ 25 \end{pmatrix}.$$

Um das Produkt D in 25 facher Ausfertigung herzustellen, braucht man also 500 Stück von Produkt A, 275 Stück von B, sowie 100 Stück von C.

Läge die Koeffizientenmatrix \mathbf{A} nicht in Dreiecksform vor, dann könnte die Lösung nicht so leicht berechnet werden wie in diesem Beispiel. Es ist also wünschenswert, dass die Koeffizientenmatrix in Dreiecksform vorliegt. Tatsächlich ist ein möglicher allgemeiner Ansatz zur Lösung linearer Gleichungssysteme dadurch gegeben, dass die Koeffizientenmatrix durch bestimmte Matrixoperationen in Dreiecksform transformiert wird, so dass anschließend die Lösung abgelesen (bzw. leicht berechnet) werden kann.

Wir können den Produktionsprozess aus Abbildung 1.7 auch noch auf andere Art und Weise ableiten. Wir definieren die Matrix

$$\mathbf{D} = \begin{pmatrix} 0 & 1 & 2 & 1 \\ 0 & 0 & 2 & 3 \\ 0 & 0 & 0 & 4 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

die sich wie folgt interpretieren lässt: In jeder Zeile kann man ablesen, wieviele Teile eines Produktes man benötigt, um das Produkt der jeweiligen Spalte herzustellen. Beispielsweise benötigt man zur Herstellung des Produktes C (3. Spalte) 2 Teile des Produktes B (2. Zeile). In den Spalten lässt sich ablesen, wieviele Stücke aller anderen Produkte gebraucht werden, um das jeweilige Produkt zusammenzusetzen. Zur Herstellung von D (4. Spalte) benötigt man also 1 Teil von A (1. Zeile), 3 Teile von B (2. Zeile), 4 Teile von C (3. Zeile) und 0 Teile von D (4. Zeile). Die Matrix \mathbf{A} ist dann gegeben durch $\mathbf{A} = \mathbf{I} - \mathbf{D}$.

Wie wir in Kapitel 5.3 sehen werden, kann die Lösung auch in Abhängigkeit einer sogenannten Inversen \mathbf{A}^{-1} von \mathbf{A} geschrieben werden. Es gilt

$$x = \mathbf{A}^{-1}c = \begin{pmatrix} 1 & 1 & 4 & 20 \\ 0 & 1 & 2 & 11 \\ 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 0 \\ 25 \end{pmatrix} = \begin{pmatrix} 500 \\ 275 \\ 100 \\ 25 \end{pmatrix}.$$

Die Darstellung des Fertigungsprozesses durch Matrizen hat den Vorteil, dass der Bedarfsvektor b beliebig verändert werden kann, ohne dass jedesmal eine neue Berechnung angestellt werden muss. Wenn z.B. zusätzlich noch drei Stück von Produkt C benötigt werden, erhalten wir

$$x = \begin{pmatrix} 1 & 1 & 4 & 20 \\ 0 & 1 & 2 & 11 \\ 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 3 \\ 25 \end{pmatrix} = \begin{pmatrix} 512 \\ 281 \\ 103 \\ 25 \end{pmatrix}.$$

Die benötigte Stückzahl von Teil C erhöht sich um drei, die Stückzahl von Teil B demnach um sechs Stück usw..

Beispiel 1.17 (Ein inkonsistentes Gleichungssystem)

Betrachte das folgende lineare Gleichungssystem:

$$\begin{pmatrix} 2 & 3 & -2 \\ 0 & -\frac{7}{2} & 4 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 5 \\ -\frac{1}{2} \\ -8 \end{pmatrix}.$$

Offensichtlich befindet sich die Koeffizientenmatrix wieder in Dreiecksform. Anhand dieses Beispiels erkennen wir einen weiteren Vorteil eines Gleichungssystems in Dreiecksform. Offensichtlich können wir mit einem Blick ablesen, ob das Gleichungssystem lösbar ist oder nicht. Im vorliegenden Fall handelt es sich um ein unlösbares Gleichungssystem, weil die dritte Gleichung

$$0 \cdot x_1 + 0 \cdot x_2 + 0 \cdot x_3 = -8$$

keine Lösungen besitzt. Ein unlösbares Gleichungssystem heißt auch inkonsistent.

△

Wir können in diesem Einführungskapitel noch nicht auf alle Details bei der Lösung linearer Gleichungssysteme eingehen. Eine ausführliche Darstellung erfolgt in Kapitel 5 und in Teilen von Kapitel 7. Trotzdem kann die grundlegende Vorgehensweise bereits jetzt skizziert werden. Bei der Lösung eines beliebigen Gleichungssystems kann in etwa wie folgt vorgegangen werden:

- Reduziere die Koeffizientenmatrix \mathbf{A} eines linearen Gleichungssystems durch noch zu präzisierende Zeilen- und Spaltenoperationen auf Dreiecksform. Wenn die dazu nötigen Operationen auch auf den Ergebnisvektor c angewendet werden, dann lässt sich (in Kapitel 5) zeigen, dass die Lösungen des Gleichungssystems durch diese Operationen *unverändert* bleiben. Ein Beispiel haben wir bereits kennengelernt. Das Gleichungssystem in Beispiel 1.15 ist nämlich durch Zeilen- und Spaltenoperationen aus dem Gleichungssystem in Beispiel 1.14 hervorgegangen.
- In einem zweiten Schritt können dann die Lösungen aus dem System in Dreiecksform “abgelesen” werden. Dabei geht man völlig analog zu Beispiel 1.15 vor. In einigen Fällen stellt sich heraus, dass das Gleichungssystem nicht lösbar ist. Dies kann aber auch leicht abgelesen werden, wenn sich die Koeffizientenmatrix in Dreiecksform befindet, vergleiche Beispiel 1.17.

Wir wollen im Folgenden einen Algorithmus zur Reduzierung (Umformung) einer Matrix auf Dreiecksform vorstellen. Um eine Matrix auf Dreiecksform zu reduzieren werden sogenannte *elementare Matrixoperationen* benötigt.

Elementare Matrixoperationen sind

1. das Vertauschen der i -ten und j -ten Zeile (Spalte) einer Matrix \mathbf{A} ,
2. die Multiplikation der i -ten Zeile (Spalte) mit einem Skalar λ ,
3. die Addition des λ -fachen der i -ten Zeile (Spalte) zur j -ten Zeile (Spalte).

Mit den soeben definierten Matrixoperationen lässt sich nun folgender Algorithmus zur Reduktion einer Matrix auf Dreiecksgestalt angeben:

Algorithmus 1.1 (zur Reduzierung auf Dreiecksgestalt)

Gegeben sei die $m \times n$ Matrix \mathbf{A} mit $\mathbf{A} \neq \mathbf{0}$. Diese lässt sich gemäß dem folgenden Algorithmus auf Dreiecksform reduzieren:

1. Setze $i = 1$.
2. Sind alle Zeilen $i + 1, \dots, m$ Null, dann Abbruch des Verfahrens. Die Matrix befindet sich in Dreiecksgestalt.
3. Ist das Element a_{ii} ungleich Null, dann fahre fort mit 4. Ansonsten suche eine Zeile k ($k > i$), in der das Element a_{ki} ungleich Null ist und vertausche die Zeilen i und k . Kann keine solche Zeile gefunden werden, dann suche eine Spalte k ($k > i$), in der mindestens eines der Elemente a_{ik}, \dots, a_{mk} ungleich Null ist (hier: a_{rk}) und vertausche die Spalten i und k . Sodann vertausche die Zeilen i und r .
4. Addiere für $j = i + 1, \dots, m$ zur j -ten Zeile das $-\frac{a_{ji}}{a_{ii}}$ fache der i -ten Zeile.
5. Setze $i = i + 1$. Für $i = m$ Abbruch des Verfahrens. Die Matrix befindet sich in diesem Fall in Dreiecksform. Ansonsten Rücksprung auf 2.

Wie aus dem Algorithmus ersichtlich ist, kann jede von Null verschiedene Matrix in Dreiecksform gebracht werden.

Beispiel 1.18

Gegeben sei die Matrix

$$\mathbf{A} = \begin{pmatrix} 2 & 3 & 1 \\ 1 & 1 & 1 \\ 3 & 5 & 1 \end{pmatrix}.$$

Wir bringen die Matrix durch folgende Schritte auf Dreiecksform:

1. Schritt: ($i = 1, j = 2$)

Da $a_{11} = 2 \neq 0$, addieren wir zur 2. Zeile das $-\frac{a_{21}}{a_{11}} = -\frac{1}{2}$ fache der 1. Zeile. Wir erhalten die Matrix

$$\mathbf{A}^{(1)} := \begin{pmatrix} 2 & 3 & 1 \\ 1 - \frac{1}{2} \cdot 2 & 1 - \frac{1}{2} \cdot 3 & 1 - \frac{1}{2} \cdot 1 \\ 3 & 5 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 3 & 1 \\ 0 & -\frac{1}{2} & \frac{1}{2} \\ 3 & 5 & 1 \end{pmatrix}.$$

2. Schritt: ($i = 1, j = 3$)

Wir addieren zur 3. Zeile das $-\frac{a_{31}}{a_{11}} = -\frac{3}{2}$ fache der 1. Zeile. Wir erhalten

$$\mathbf{A}^{(2)} := \begin{pmatrix} 2 & 3 & 1 \\ 0 & -\frac{1}{2} & \frac{1}{2} \\ 3 - \frac{3}{2} \cdot 2 & 5 - \frac{3}{2} \cdot 3 & 1 - \frac{3}{2} \cdot 1 \end{pmatrix} = \begin{pmatrix} 2 & 3 & 1 \\ 0 & -\frac{1}{2} & \frac{1}{2} \\ 0 & \frac{1}{2} & -\frac{1}{2} \end{pmatrix}.$$

3. Schritt: ($i = 2, j = 3$)

Da in $\mathbf{A}^{(2)}$ das Element $a_{22} = -\frac{1}{2} \neq 0$, addieren wir zur 3. Zeile das $-\frac{a_{32}}{a_{22}} = 1$ fache der 2. Zeile und erhalten

$$\mathbf{A}^{(3)} := \begin{pmatrix} 2 & 3 & 1 \\ 0 & -\frac{1}{2} & \frac{1}{2} \\ 0 & \frac{1}{2} - 1 \cdot \frac{1}{2} & -\frac{1}{2} + 1 \cdot \frac{1}{2} \end{pmatrix} = \begin{pmatrix} 2 & 3 & 1 \\ 0 & -\frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 0 \end{pmatrix}.$$

Damit befindet sich die Matrix $\mathbf{A}^{(3)}$ in Dreiecksform. Es ist aber wichtig sich klarzumachen, dass die aus der Matrix \mathbf{A} hervorgegangene Matrix $\mathbf{A}^{(3)}$ *nicht* gleich \mathbf{A} oder irgendwie äquivalent ist.

△

Durch elementare Zeilen- und Spaltenoperationen kann jede Matrix sogar auf Diagonalform gebracht werden. Dabei spricht man von einer Matrix in Diagonalform, wenn alle Elemente unter- und oberhalb der Hauptdiagonalen Null sind und die ersten $r \geq 1$ Elemente der Hauptdiagonalen von Null verschieden sind. Das heißt eine Matrix \mathbf{D} in Diagonalform hat folgende Gestalt

$$\mathbf{D} = \begin{pmatrix} \mathbf{D}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix},$$

wobei \mathbf{D}_r eine $r \times r$ Diagonalmatrix ist und die Nullmatrizen Matrizen passender Ordnung sind. Der folgende Algorithmus liefert eine Reduzierung auf Diagonalform:

Algorithmus 1.2 (zur Reduzierung auf Diagonalform)

Sei $\mathbf{A} \neq \mathbf{0}$ eine $m \times n$ Matrix. Diese lässt sich gemäß dem folgenden Algorithmus auf Diagonalform \mathbf{D} bringen:

1. Setze $i = 1$.

2. Ist die Submatrix

$$\mathbf{A}_i = \begin{pmatrix} a_{ii} & \dots & \dots & a_{in} \\ \vdots & & & \vdots \\ \vdots & & & \vdots \\ a_{mi} & \dots & \dots & a_{mn} \end{pmatrix}$$

die Nullmatrix, dann beende den Algorithmus. Die Matrix befindet sich bereits in Diagonalform. Ansonsten fahre fort mit 3.

3. Ist das Element a_{ii} ungleich Null, dann fahre fort mit 4. Ansonsten suche ein Element $a_{rk} \neq 0$ aus \mathbf{A}_i und vertausche die Zeilen r und i und die Spalten k und i .

4. Falls $i < m$, addiere für $j = i + 1, \dots, m$ zur j -ten Zeile das $-\frac{a_{ji}}{a_{ii}}$ fache der i -ten Zeile.

5. Falls $i < n$, addiere für $j = i + 1, \dots, n$ zur j -ten Spalte das $-\frac{a_{ij}}{a_{ii}}$ fache der i -ten Spalte.

6. Setze $i = i + 1$. Für $i = m + 1$ ist der Algorithmus beendet, ansonsten Rücksprung auf 2.

Beispiel 1.19

Gegeben sei wieder die Matrix

$$\mathbf{A} = \begin{pmatrix} 2 & 3 & 1 \\ 1 & 1 & 1 \\ 3 & 5 & 1 \end{pmatrix}.$$

Wir bringen \mathbf{A} durch folgende Schritte auf Diagonalform:

1. Schritt: ($i=1, j=2$)

Da $a_{11} = 2 \neq 0$, addieren wir zur 2. Zeile das $-\frac{1}{2}$ fache der 1. Zeile und erhalten

$$\mathbf{A}^{(1)} := \begin{pmatrix} 2 & 3 & 1 \\ 1 - \frac{1}{2} \cdot 2 & 1 - \frac{1}{2} \cdot 3 & 1 - \frac{1}{2} \cdot 1 \\ 3 & 5 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 3 & 1 \\ 0 & -\frac{1}{2} & \frac{1}{2} \\ 3 & 5 & 1 \end{pmatrix}$$

2. Schritt: ($i=1, j=3$)

Wir addieren zur 3. Zeile das $-\frac{3}{2}$ fache der 1. Zeile:

$$\mathbf{A}^{(2)} := \begin{pmatrix} 2 & 3 & 1 \\ 0 & -\frac{1}{2} & \frac{1}{2} \\ 3 - \frac{3}{2} \cdot 2 & 5 - \frac{3}{2} \cdot 3 & 1 - \frac{3}{2} \cdot 1 \end{pmatrix} = \begin{pmatrix} 2 & 3 & 1 \\ 0 & -\frac{1}{2} & \frac{1}{2} \\ 0 & \frac{1}{2} & -\frac{1}{2} \end{pmatrix}$$

3. Schritt: (i=1, j=2)

Addiere zur 2. Spalte das $-\frac{3}{2}$ fache der 1. Spalte:

$$\mathbf{A}^{(3)} := \begin{pmatrix} 2 & 3 - \frac{3}{2} \cdot 2 & 1 \\ 0 & -\frac{1}{2} - \frac{3}{2} \cdot 0 & \frac{1}{2} \\ 0 & \frac{1}{2} - \frac{3}{2} \cdot 0 & -\frac{1}{2} \end{pmatrix} = \begin{pmatrix} 2 & 0 & 1 \\ 0 & -\frac{1}{2} & \frac{1}{2} \\ 0 & \frac{1}{2} & -\frac{1}{2} \end{pmatrix}$$

4. Schritt: (i=2, j=3)

Addiere zur 3. Spalte das $-\frac{1}{2}$ fache der 1. Spalte:

$$\mathbf{A}^{(4)} := \begin{pmatrix} 2 & 0 & 1 - \frac{1}{2} \cdot 2 \\ 0 & -\frac{1}{2} & \frac{1}{2} - \frac{1}{2} \cdot 0 \\ 0 & \frac{1}{2} & -\frac{1}{2} - \frac{1}{2} \cdot 0 \end{pmatrix} = \begin{pmatrix} 2 & 0 & 0 \\ 0 & -\frac{1}{2} & \frac{1}{2} \\ 0 & \frac{1}{2} & -\frac{1}{2} \end{pmatrix}$$

5. Schritt: (i=2, j=3)

Addiere zur 3. Zeile das 1 fache der 2. Zeile:

$$\mathbf{A}^{(5)} := \begin{pmatrix} 2 & 0 & 0 \\ 0 & -\frac{1}{2} & \frac{1}{2} \\ 0 & \frac{1}{2} - 1 \cdot \frac{1}{2} & -\frac{1}{2} + 1 \cdot \frac{1}{2} \end{pmatrix} = \begin{pmatrix} 2 & 0 & 0 \\ 0 & -\frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 0 \end{pmatrix}$$

6. Schritt: (i=2, j=3)

Addiere zur 3. Spalte das 1 fache der 2. Spalte:

$$\mathbf{A}^{(6)} := \begin{pmatrix} 2 & 0 & 0 \\ 0 & -\frac{1}{2} & \frac{1}{2} - \frac{1}{2} \\ 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 2 & 0 & 0 \\ 0 & -\frac{1}{2} & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

△

Die Reduzierung einer Matrix auf Diagonalform wird bei der Herleitung der wichtigen Vollrangzerlegung einer Matrix in Kapitel 3.4 eine wichtige Rolle spielen.

2

Vektorräume

In Kapitel 1 haben wir die Menge der reellen n -Tupel versehen mit der Vektoraddition und der skalaren Multiplikation als den n -dimensionalen Vektorraum über \mathbb{R} (kurz \mathbb{R}^n) definiert. Tatsächlich ist der \mathbb{R}^n lediglich ein (wichtiger) Spezialfall allgemeiner Vektorräume. Die folgende Definition eines Vektorraums verwendet die algebraische Struktur des *Körpers*. Eine allgemeine Definition eines Körpers K findet man im Anhang. Für die Zwecke dieses Skripts genügt es jedoch sich $K = \mathbb{R}$, den Körper der reellen Zahlen oder $K = \mathbb{C}$, den Körper der komplexen Zahlen vorzustellen. Da komplexe Zahlen dem einen oder anderen unbekannt sein mögen, beginnen wir dieses Kapitel mit einer kurzen Einführung in die komplexen Zahlen. Anschließend definieren wir allgemein den Begriff des Vektorraums (Abschnitt 2.2). Teilmengen von Vektorräumen sind unter bestimmten Umständen selbst wieder Vektorräume und werden als Unterräume bezeichnet, vgl. Abschnitt 2.3. Häufig ist es möglich, sämtliche Elemente eines Vektorraums mit Hilfe einer Menge von sogenannten Basisvektoren darzustellen. Abschnitt 2.4 befasst sich daher mit der Darstellung von Vektoren durch eine möglichst kleine Menge von Basisvektoren. In Abschnitt 2.5 untersuchen wir dann Beziehungen zwischen Vektorräumen und zwar sogenannte lineare Abbildungen und stellen den Zusammenhang mit Matrizen her. Schließlich verallgemeinern wir in Abschnitt 2.6 die bereits in Kapitel 1 eingeführten Begriffe der Länge und des Abstands von Vektoren.

2.1 Komplexe Zahlen

Definition 2.1 (Körper der komplexen Zahlen)

Eine komplexe Zahl x ist ein geordnetes Paar $x = (x_1, x_2)$ reeller Zahlen. Die Menge aller komplexen Zahlen wird mit \mathbb{C} bezeichnet. Zwei komplexe Zahlen $x = (x_1, x_2)$ und $y = (y_1, y_2)$ heißen gleich, wenn $x_1 = y_1$ und $x_2 = y_2$ gilt. Die Addition und Multiplikation komplexer Zahlen ist wie folgt definiert:

1. $(x_1, x_2) + (y_1, y_2) = (x_1 + y_1, x_2 + y_2)$
2. $(x_1, x_2) \cdot (y_1, y_2) = (x_1y_1 - x_2y_2, x_1y_2 + x_2y_1)$

Bemerkungen:

- Die Menge der komplexen Zahlen bildet einen Körper (siehe Anhang A). Das Nullelement ist die Zahl $(0, 0)$, das Einselement die Zahl $(1, 0)$.
- Da $(x_1, 0) + (y_1, 0) = (x_1 + y_1, 0)$ und $(x_1, 0)(y_1, 0) = (x_1 y_1, 0)$ ist, kann man eine reelle Zahl x mit der komplexen Zahl $(x, 0)$ identifizieren, so dass \mathbf{R} eine Teilmenge von \mathbf{C} ist.
- Eine wichtige Rolle spielt die komplexe Zahl $i = (0, 1)$. Für i gilt:

$$i^2 = i \cdot i = (0, 1)(0, 1) = (-1, 0) = -1$$

Darüberhinaus lässt sich jede komplexe Zahl $x = (x_1, x_2)$ darstellen als

$$x = (x_1, x_2) = (x_1, 0) + (x_2, 0) \cdot (0, 1) = x_1 + x_2 \cdot i.$$

Dies ist die übliche Darstellung komplexer Zahlen. In dieser Darstellung kann man in der für reelle Zahlen gewohnten Art und Weise rechnen. x_1 heißt *Realteil* von x und x_2 heißt *Imaginärteil* von x .

- Die komplexe Zahl

$$\bar{x} = x_1 - x_2 \cdot i$$

heißt die zur komplexen Zahl

$$x = x_1 + x_2 \cdot i$$

konjugiert komplexe Zahl. Es gilt

$$x \cdot \bar{x} = x_1^2 + x_2^2.$$

▽

Beispiel 2.1

Gegeben seien die komplexen Zahlen $(3, 2) = 3 + 2i$ und $(2, 1) = 2 + 1i$. Wir berechnen die Summe und das Produkt der beiden Zahlen:

$$(3, 2) + (2, 1) = 3 + 2i + 2 + 1i = 5 + 3i = (5, 3)$$

$$\begin{aligned} (3, 2) \cdot (2, 1) &= (3 + 2i)(2 + 1i) = 6 + 3i + 4i + 2i^2 \\ &= 6 + 7i - 2 = 4 + 7i = (4, 7). \end{aligned}$$

Betrachte weiterhin den Ausdruck

$$\frac{1}{3 + 2i}.$$

Wir vereinfachen wie folgt:

$$\begin{aligned}\frac{1}{3+2i} &= \frac{3-2i}{(3+2i)(3-2i)} = \frac{3-2i}{9+4} = \frac{3}{13} - \frac{2}{13}i \\ &= \left(\frac{3}{13}, \frac{2}{13}\right).\end{aligned}$$

△

Beispiel 2.2

Wir betrachten allgemein die quadratische Gleichung

$$x^2 + p = 0$$

Im Falle $p < 0$ erhalten wir die beiden reellen Lösungen

$$x_1 = \sqrt{-p}$$

und

$$x_2 = -\sqrt{-p}.$$

Im Falle $p > 0$ existieren keine reellwertigen Lösungen. Allerdings existieren die beiden komplexen Lösungen

$$x_1 = \sqrt{p} \cdot i$$

und

$$x_2 = -\sqrt{p} \cdot i.$$

Denn

$$x_1^2 + p = (\sqrt{p} \cdot i)^2 + p = p \cdot i^2 + p = p \cdot (-1) + p = 0$$

und

$$x_2^2 + p = (-\sqrt{p} \cdot i)^2 + p = -p + p = 0.$$

△

Definition 2.2 (Betrag einer komplexen Zahl)

Der Betrag einer komplexen Zahl $x = (x_1, x_2) = x_1 + x_2 \cdot i$ ist definiert als

$$|x| = \sqrt{x_1^2 + x_2^2}.$$

Beispiel 2.3

Der Betrag der Zahl $x = 4 + 3i$ ist

$$|x| = \sqrt{4^2 + 3^2} = \sqrt{16 + 9} = 5.$$

△

Definition 2.3 (Komplexe Matrizen)

Ein nach m Zeilen und n Spalten geordnetes Schema \mathbf{A} von mn Elementen $a_{ij} \in \mathbb{C}$ heißt komplexe Matrix. Addition, skalare Multiplikation und Multiplikation sind analog zu reellen Matrizen definiert. Auch lassen sich die Sätze 1.2 und 1.3 analog übertragen.

Beispiel 2.4 (Rechnen mit komplexen Matrizen)

Wir betrachten die komplexen Matrizen

$$\mathbf{A} = \begin{pmatrix} 2 + 2i & 3 - i \\ 1 + i & -1 + 2i \end{pmatrix}$$

und

$$\mathbf{B} = \begin{pmatrix} 1 + i & 2 - i \\ -2 + 2i & 3 \end{pmatrix}.$$

Wir erhalten

$$\begin{aligned} \mathbf{A} + \mathbf{B} &= \begin{pmatrix} 2 + 2i + 1 + i & 3 - i + 2 - i \\ 1 + i - 2 + 2i & -1 + 2i + 3 \end{pmatrix} \\ &= \begin{pmatrix} 3 + 3i & 5 - 2i \\ -1 + 3i & 2 + 2i \end{pmatrix}, \end{aligned}$$

$$\begin{aligned} \mathbf{A} \cdot \mathbf{B} &= \begin{pmatrix} (2 + 2i)(1 + i) + (3 - i)(-2 + 2i) & (2 + 2i)(2 - i) + (3 - i)3 \\ (1 + i)(1 + i) + (-1 + 2i)(-2 + 2i) & (1 + i)(2 - i) + (-1 + 2i)3 \end{pmatrix} \\ &= \begin{pmatrix} 4i - 4 + 8i & 6 + 2i + 9 - 3i \\ 2i - 2 - 6i & 3 + i - 3 + 6i \end{pmatrix} \\ &= \begin{pmatrix} -4 + 12i & 15 - i \\ -2 - 4i & 7i \end{pmatrix} \end{aligned}$$

und

$$\begin{aligned} \mathbf{B} \cdot \mathbf{A} &= \begin{pmatrix} (1 + i)(2 + 2i) + (2 - i)(1 + i) & (1 + i)(3 - i) + (2 - i)(-1 + 2i) \\ (-2 + 2i)(2 + 2i) + 3(1 + i) & (-2 + 2i)(3 - i) + 3(-1 + 2i) \end{pmatrix} \\ &= \begin{pmatrix} 4i + 3 + i & 4 + 2i + 5i \\ -8 + 3 + 3i & -4 + 8i - 3 + 6i \end{pmatrix} \\ &= \begin{pmatrix} 3 + 5i & 4 + 7i \\ -5 + 3i & -7 + 14i \end{pmatrix}. \end{aligned}$$

Auch bei komplexen Matrizen ist also das Matrixprodukt nicht kommutativ, d.h.

$$\mathbf{AB} \neq \mathbf{BA}.$$

△

2.2 Definition von Vektorräumen und Beispiele

Wir definieren jetzt allgemein was man unter einem Vektorraum versteht. Ein Beispiel für einen Vektorraum haben wir mit dem \mathbb{R}^n bereits in Kapitel 1 kennengelernt.

Definition 2.4 (Vektorraum)

Sei K ein Körper. Ein K -Vektorraum ist ein Tripel $(V, +, \cdot)$, bestehend aus einer Menge V , einer Verknüpfung (Addition)

$$\begin{aligned} + : V \times V &\mapsto V \\ (x, y) &\mapsto x + y \end{aligned}$$

und einer Verknüpfung (Multiplikation mit Skalaren)

$$\begin{aligned} \cdot : K \times V &\mapsto V \\ (a, x) &\mapsto a \cdot x \end{aligned}$$

so dass folgende Axiome erfüllt sind:

I. Axiome der Addition

1. Assoziativität: $x + (y + z) = (x + y) + z$ für alle $x, y, z \in V$.
2. Kommutativität: $x + y = y + x$ für alle $x, y \in V$.
3. Existenz der Null: Es gibt einen Vektor $0 \in V$ mit $x + 0 = x$ für alle $x \in V$.
4. Existenz des Negativen: Zu jedem $x \in V$ existiert ein Vektor $-x \in V$ mit $x + (-x) = 0$.

II. Axiome der skalaren Multiplikation

1. Distributivgesetze: $(a + b)x = ax + bx$ bzw. $a(x + y) = ax + ay$ für alle $x, y \in V$, $a, b \in K$.
2. Assoziativgesetz: $(ab)x = a(bx)$ für alle $x \in V$, $a, b \in K$.
3. Existenz der Eins: Es gibt eine Zahl $1 \in K$, so dass $1 \cdot x = x$ für alle $x \in V$.

Aus den Axiomen ergeben sich einige einfache Folgerungen:

Satz 2.1

Sei V ein K -Vektorraum. Dann gilt:

1. $a \cdot 0 = 0$ mit $a \in K$ und $0 \in V$.
2. $0 \cdot x = 0$ mit $0 \in K$, $x \in V$.
3. Ist $ax = 0$, wobei $a \in K$ und $x \in V$ seien, so folgt $a = 0$ oder $x = 0$.
4. $(-a)x = a(-x) = -ax$ für alle $a \in K$, $x \in V$.

Beispiel 2.5 (Vektorraum der n-Tupel)

Sei K ein Körper. Dann ist die Menge aller n -Tupel der Elemente von K mit Vektoraddition und skalarer Multiplikation definiert durch

$$x + y = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{pmatrix}$$

und

$$\lambda \cdot x = \lambda \cdot \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} \lambda \cdot x_1 \\ \lambda \cdot x_2 \\ \vdots \\ \lambda \cdot x_n \end{pmatrix},$$

wobei $\lambda, x_i, y_i \in K$ gilt, ein K -Vektorraum. Dieser Raum wird mit K^n bezeichnet. Die wichtigsten Spezialfälle ergeben sich für $K = \mathbf{R}$ und $K = \mathbf{C}$.

△

Beispiel 2.6 (Vektorraum der Matrizen)

Sei V die Menge der reellen $m \times n$ Matrizen und $K = \mathbf{R}$. Dann ist V für festes m und n unter Berücksichtigung der Matrixaddition und der skalaren Multiplikation wegen Satz 1.2 1)-7) ein Vektorraum über $K = \mathbf{R}$. Das Nullelement ist die Nullmatrix, deren Elemente sämtlich aus Null bestehen, das Einselement ist die Zahl Eins. Darüber hinaus stellt auch die Menge der komplexen $m \times n$ Matrizen einen Vektorraum über $K = \mathbf{C}$ dar, da für komplexe Matrizen dieselben Regeln wie in Satz 1.2 für reelle Matrizen gelten. Allgemein kann man $m \times n$ Matrizen für einen beliebigen Körper K definieren, d.h. die Elemente der Matrix sind Elemente aus K . Auch diese Menge stellt einen Vektorraum dar.

△

Beispiel 2.7 (Vektorraum der Polynome)

Sei V die Menge aller Polynome vom Grad n

$$P(t) = a_0 + a_1t + a_2t^2 + \cdots + a_nt^n$$

mit Koeffizienten $a_i \in \mathbb{R}$.

Wir definieren für

$$P_1(t) = a_0 + a_1t + a_2t^2 + \cdots + a_nt^n$$

und

$$P_2(t) = b_0 + b_1t + b_2t^2 + \cdots + b_nt^n$$

die Addition von Polynomen durch

$$P_1(t) + P_2(t) = (a_0 + b_0) + (a_1 + b_1)t + (a_2 + b_2)t^2 + \cdots + (a_n + b_n)t^n.$$

Die skalare Multiplikation mit einem Skalar $b \in \mathbb{R}$ ist definiert durch

$$bP(t) = ba_0 + ba_1t + ba_2t^2 + \cdots + ba_nt^n.$$

Mit diesen beiden Verknüpfungen ist die Menge der Polynome vom Grad n ein Vektorraum über \mathbb{R} .

△

Beispiel 2.8

Sei V die Menge aller reellwertigen Funktionen $f : \mathbb{R} \mapsto \mathbb{R}$.

Wir definieren die Summe zweier Funktionen $f, g \in V$ durch

$$(f + g)(x) = f(x) + g(x)$$

und das Produkt mit einem Skalar $k \in \mathbb{R}$ durch

$$(kf)(x) = kf(x).$$

Dann ist V ein Vektorraum über \mathbb{R} . Das Nullelement ist die Nullfunktion $0(x) = 0$.

△

2.3 Unterräume

Wir betrachten im Folgenden Teilmengen von Vektorräumen. Unter bestimmten Voraussetzungen sind diese Teilmengen selbst wieder Vektorräume.

Definition 2.5 (Unterraum)

Sei U eine Teilmenge eines Vektorraumes V . U heißt *Unterraum* des Vektorraumes V , wenn U unter Berücksichtigung der Vektoraddition und skalaren Multiplikation selbst ein Vektorraum ist.

Bemerkung:

Definitionsgemäß muss ein Untervektorraum den Nullvektor 0 enthalten, da Axiom I 3) für Vektorräume erfüllt sein muss.

▽

Der folgende Satz liefert eine Nachweismöglichkeit für Unterräume.

Satz 2.2 (Unterraum ist Vektorraum)

$U \subseteq V$ ist genau dann ein Unterraum, wenn gilt:

1. U ist nicht leer.
2. U ist abgeschlossen bzgl. der Vektoraddition, d.h. für $u_1, u_2 \in U$ gilt $u_1 + u_2 \in U$.
3. U ist abgeschlossen bzgl. der skalaren Multiplikation, d.h. für $u \in U$ gilt $k \cdot u \in U$ für jedes $k \in K$.

Beweis:

Wir nehmen zunächst an, dass 1.-3. gilt.

Es ist zu zeigen, dass dann die Vektorraumaxiome gelten. Die Axiome I1, I2, II1, II2, II3 gelten in U , da die Vektoren in U zu V gehören.

Die Existenz der Null (I3) zeigt man wie folgt: Aufgrund von 1) ist U nicht leer, sei also z.B. $u \in U$. Wegen 3) gilt

$$0 \cdot u = 0 \in U$$

und für jedes $u \in U$ gilt

$$u + 0 = u,$$

so dass also Axiom I3 gilt.

Wegen 3) gilt

$$(-1) \cdot u = -u \in U$$

und

$$u + (-u) = 0$$

wenn $u \in U$. Also gilt auch Axiom I4 und U ist ein Unterraum.

Ist umgekehrt U ein Unterraum, dann gelten 1), 2) und 3).

□

Bemerkung:

Der Satz liefert eine Nachweismöglichkeit für Unterräume. Zum Nachweis, dass eine Teilmenge U eines Vektorraumes V ein Unterraum ist, können wir wie folgt vorgehen:

- Zeige, dass $0 \in U$.
- Zeige, dass für $k_1, k_2 \in K$ und $u_1, u_2 \in U$ der Vektor $k_1u_1 + k_2u_2$ in U enthalten ist.

▽

Beispiel 2.9

Sei V ein beliebiger Vektorraum. Dann ist die Menge $\{0\}$, die nur aus dem Nullvektor besteht ein Unterraum. Die gesamte Menge V ist ebenfalls ein Unterraum.

△

Beispiel 2.10

Sei V der Vektorraum \mathbb{R}^3 . Wir zeigen, dass die Menge $U := \{(0, a, b) : a, b \in \mathbb{R}\}$ ein Unterraum des \mathbb{R}^3 ist. Offensichtlich ist $0 \in U$ (setze $a = b = 0$). Seien $(0, a_1, b_1)$ und $(0, a_2, b_2)$ zwei beliebige Vektoren aus U und $k_1, k_2 \in K$. Dann gilt

$$k_1(0, a_1, b_1) + k_2(0, a_2, b_2) = (0, k_1a_1 + k_2a_2, k_1b_1 + k_2b_2) \in U.$$

U ist also ein Unterraum.

△

Beispiel 2.11

Sei V der Vektorraum \mathbb{R}^2 . Betrachte die Menge $U := \{(y, x) : y = a + bx, a, b \in \mathbb{R}\}$. Bei U handelt es sich um eine Gerade mit Ordinatenabschnitt a und Steigung b . Für $a \neq 0$ ist U kein Unterraum, da U nicht den Nullvektor enthält. Für $a = 0$ jedoch gilt $(0, 0) \in U$. Weiter gilt für zwei Vektoren $(y_1, x_1) \in U$ und $(y_2, x_2) \in U$ und falls $a = 0$

$$\begin{aligned} k_1(y_1, x_1) + k_2(y_2, x_2) &= (k_1y_1 + k_2y_2, k_1x_1 + k_2x_2) \\ &= (k_1bx_1 + k_2bx_2, k_1x_1 + k_2x_2) \\ &= (b(k_1x_1 + k_2x_2), k_1x_1 + k_2x_2), \end{aligned}$$

wobei $k_1, k_2 \in \mathbb{R}$. Damit liegt der Punkt $k_1(y_1, x_1) + k_2(y_2, x_2)$ ebenfalls auf der Gerade mit Steigung b und U ist ein Unterraum (falls $a = 0$).

△

Beispiel 2.12

Sei V der Vektorraum der Polynome vom Grad n . Dann ist die Menge U der Polynome vom Grad p mit $p \leq n$ ein Unterraum von V .

△

Im folgenden Satz zeigen wir, dass der Durchschnitt von Unterräumen wieder ein Unterraum ist.

Satz 2.3

Der Durchschnitt beliebig vieler Unterräume ist wieder ein Unterraum.

Beweis

Seien U_1 und U_2 Unterräume des Vektorraumes V . Wegen $0 \in U_1$ und $0 \in U_2$ gilt auch $0 \in U_1 \cap U_2$. Gelte nun $w_1, w_2 \in U_1 \cap U_2$. Dann gilt $w_1, w_2 \in U_1$ und $w_1, w_2 \in U_2$. Da U_1 und U_2 Unterräume sind gilt auch

$$aw_1 + bw_2 \in U_1$$

und

$$aw_1 + bw_2 \in U_2$$

für $a, b \in K$.

Damit ist aber $aw_1 + bw_2 \in U_1 \cap U_2$ und folglich $U_1 \cap U_2$ ein Unterraum von V .

□

Definition 2.6 (Summe zweier Unterräume, direkte Summe)

Die Summe der Teilräume U_1, U_2 aus V ist die Menge

$$U_1 + U_2 := \{k_1u_1 + k_2u_2 : u_1 \in U_1, u_2 \in U_2, k_1, k_2 \in K\}.$$

Gilt darüberhinaus, dass $U_1 \cap U_2 = 0$, dann heißt die Summe direkt. U_1 und U_2 heißen dann komplementär zueinander. Schreibweise: $U_1 \oplus U_2$

Bemerkung:

- Es ist sofort ersichtlich, dass die Summe zweier Unterräume wieder ein Unterraum ist.
- Handelt es sich bei einem Vektorraum V um die direkte Summe zweier Unterräume U_1 und U_2 , dann läßt sich jeder Vektor $x \in V$ *eindeutig* als Summe zweier Vektoren $u_1 \in U_1$ und $u_2 \in U_2$ darstellen. Dies läßt sich folgendermaßen zeigen:
Da $V = U_1 + U_2$ existieren Vektoren u_1 und u_2 , so dass $x = u_1 + u_2$. Angenommen, es gäbe auch die Darstellung $x = u'_1 + u'_2$ mit $u'_1 \in U_1$ und $u'_2 \in U_2$. Dann gilt

$$u_1 + u_2 = u'_1 + u'_2$$

bzw.

$$u_1 - u'_1 = u'_2 - u_2.$$

Wegen $u_1 - u'_1 \in U_1$, $u'_2 - u_2 \in U_2$ und $U_1 \cap U_2 = 0$ muss $u_1 - u'_1 = 0 = u'_2 - u_2$ gelten. Folglich erhalten wir $u_1 = u'_1$ und $u_2 = u'_2$.

▽

Beispiel 2.13

Sei $V = \mathbb{R}^2$. Betrachte die Mengen $U_1 = \{(y, 0) : y \in \mathbb{R}\}$ und $U_2 = \{(0, x) : x \in \mathbb{R}\}$. Beachte, dass U_1 die Ordinate und U_2 die Abszisse in einem kartesischen Koordinatensystem sind. Offensichtlich handelt es sich bei U_1 und U_2 um Unterräume. Jeder Vektor $(y, x) \in \mathbb{R}^2$ lässt sich darstellen als Summe von Vektoren $(y, 0)$ und $(0, x)$ aus U_1 und U_2 . Da $U_1 \cap U_2 = (0, 0)$ gilt, erhalten wir $\mathbb{R}^2 = U_1 \oplus U_2$, d.h. \mathbb{R}^2 ist die direkte Summe von U_1 und U_2 .

△

2.4 Basis und Dimension

In diesem Abschnitt befassen wir uns mit der Darstellung von Vektoren eines Vektorraums. Es wird sich herausstellen, dass bei vielen Vektorräumen eine endliche Anzahl von Vektoren dieses Vektorraums (sogenannte Basisvektoren) ausreicht um alle anderen Vektoren als Linearkombination der Basisvektoren eindeutig darzustellen. Die Basisvektoren sind dabei jedoch nicht eindeutig, d.h. eine Menge von Basisvektoren kann durch andere Basisvektoren ersetzt werden. Die minimal benötigte Anzahl von Basisvektoren ist aber immer gleich und wird dann als die Dimension des Vektorraums bezeichnet.

Eine wichtige Rolle spielt im Folgenden die lineare Unabhängigkeit von Vektoren. Es wird sich herausstellen, dass eine Basis eines Vektorraums dadurch gekennzeichnet ist, dass die Basisvektoren linear unabhängig sind.

Definition 2.7 (lineare (Un)–Abhängigkeit von Vektoren)

Eine Menge von n Vektoren $x_1, x_2, \dots, x_n \in V$ heißt linear unabhängig, wenn für jede Linearkombination mit $a_1x_1 + \dots + a_nx_n = 0$ ($a_i \in K$) stets $a_1 = a_2 = \dots = a_n = 0$ gilt. Andernfalls heißen die x_1, \dots, x_n linear abhängig.

Beispiel 2.14

Wir betrachten die Vektoren

$$x_1 = \begin{pmatrix} 2 \\ 1 \\ -2 \end{pmatrix}, \quad x_2 = \begin{pmatrix} 4 \\ -3 \\ 2 \end{pmatrix}, \quad x_3 = \begin{pmatrix} 16 \\ -7 \\ 2 \end{pmatrix}.$$

Wenn wir überprüfen wollen, ob die Vektoren linear unabhängig sind, müssen wir das lineare Gleichungssystem

$$\begin{pmatrix} 2 & 4 & 16 \\ 1 & -3 & -7 \\ -2 & 2 & 2 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

lösen. Wenn sich als einzige Lösung $a_1 = a_2 = a_3 = 0$ ergibt, dann sind die Vektoren linear unabhängig, andernfalls linear abhängig. In Kapitel 1.4 wurde die Lösung linearer Gleichungssysteme bereits gestreift. Zur Lösung kann die Koeffizientenmatrix durch elementare Zeilen- und Spaltenoperationen in Dreiecksform gebracht werden und anschließend die Lösung „abgelesen“ werden. Überführung in Dreiecksform liefert (vergleiche Beispiel 5.1 in Kapitel 5)

$$\begin{pmatrix} 2 & 4 & 16 \\ 0 & -5 & -15 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

Offensichtlich gibt es neben $a_1 = a_2 = a_3 = 0$ weitere Lösungen. Wir können a_3 beliebig wählen, weil Gleichung drei immer stimmt. Mit $a_3 = 1$ erhalten wir $a_2 = -3$ und $a_1 = -2$. Die Vektoren x_1 , x_2 und x_3 sind also linear abhängig.

△

Beispiel 2.15

Wir betrachten die Vektoren aus dem \mathbf{R}^3

$$x_1 = \begin{pmatrix} 2 \\ 6 \\ 5 \end{pmatrix}, \quad x_2 = \begin{pmatrix} 1 \\ 4 \\ 4 \end{pmatrix}, \quad x_3 = \begin{pmatrix} -2 \\ 4 \\ 3 \end{pmatrix}.$$

Zum Nachweis der linearen (Un)abhängigkeit lösen wir das Gleichungssystem

$$\begin{pmatrix} 2 & 1 & -2 \\ 6 & 4 & 4 \\ 5 & 4 & 3 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

Überführen in Dreiecksform liefert das äquivalente System

$$\begin{pmatrix} 2 & 1 & -2 \\ 0 & 1 & 10 \\ 0 & 0 & -7 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

Offenbar kann das Gleichungssystem nur mit $a_1 = a_2 = a_3 = 0$ gelöst werden. Die Vektoren sind daher linear unabhängig.

△

Der folgende Satz charakterisiert die lineare Abhängigkeit von Vektoren. Sind Vektoren linear abhängig, dann lässt sich einer der Vektoren als Linearkombination (daher die Bezeichnung *linear* abhängig) einer Teilmenge der restlichen Vektoren darstellen.

Satz 2.4

Seien x_1, \dots, x_n von Null verschiedene Vektoren. Diese sind genau dann linear abhängig, wenn einer der Vektoren z.B. x_i eine Linearkombination der vorherigen ist, d.h.

$$x_i = a_1x_1 + \dots + a_{i-1}x_{i-1}.$$

Beweis

Sei x_i eine Linearkombination der vorherigen, d.h.

$$x_i = a_1x_1 + \dots + a_{i-1}x_{i-1}.$$

Dann gilt

$$a_1x_1 + \dots + a_{i-1}x_{i-1} - x_i + 0x_{i+1} + \dots + 0x_n = 0,$$

d.h. die Vektoren x_1, \dots, x_n sind linear abhängig.

Seien nun die Vektoren x_1, \dots, x_n linear abhängig. Dann existieren Skalare a_1, \dots, a_n die nicht alle null sind, so dass

$$a_1x_1 + \dots + a_nx_n = 0.$$

Sei k die größte ganze Zahl derart, dass $a_k \neq 0$ ist (d.h. auch $a_{k+1} = a_{k+2} = \dots = a_n = 0$).

Dann gilt

$$a_1x_1 + \dots + a_kx_k + 0x_{k+1} + \dots + 0x_n = 0$$

bzw.

$$a_1x_1 + \dots + a_kx_k = 0.$$

Angenommen $k = 1$, dann folgt $a_1x_1 = 0$ und wegen $a_1 \neq 0$ $x_1 = 0$. Da x_1, \dots, x_n aber ungleich Null sind, muß $k > 1$ sein und

$$x_k = -a_k^{-1}a_1x_1 - \dots - a_k^{-1}a_{k-1}x_{k-1}$$

woraus die Behauptung folgt. □

Wir befassen uns jetzt mit der Darstellung eines beliebigen Vektors x eines Vektorraumes V als Linearkombination einer (möglichst kleinen) Menge m von Vektoren x_1, x_2, \dots, x_m aus V .

Definition 2.8 (Erzeugendensystem)

Eine Menge $\{x_1, \dots, x_m\}$ heißt Erzeugendensystem eines Vektorraumes V , falls alle Vektoren $x \in V$ darstellbar sind als Linearkombination von x_1, \dots, x_m .

Die folgenden beiden Sätze 2.5 und 2.6 dienen der Charakterisierung von Erzeugendensystemen und werden anschließend bei der Definition des zentralen Begriffs der Basis und Dimension eines Vektorraums benötigt, vergleiche insbesondere Satz 2.7.

Satz 2.5

Sei $\{x_1, \dots, x_m\}$ ein Erzeugendensystem des Vektorraumes V . Dann gilt

1. Für $x \in V$ ist die Menge $\{x, x_1, \dots, x_m\}$ linear abhängig und erzeugt V .
2. Wenn x_i eine Linearkombination der vorangegangenen Vektoren ist, dann wird V durch die Menge

$$x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_m$$

erzeugt.

Beweis

zu 1): Da $\{x_1, \dots, x_m\}$ ein Erzeugendensystem ist, läßt sich x als Linearkombination von $\{x_1, \dots, x_m\}$ darstellen und die Menge $\{x, x_1, \dots, x_m\}$ ist nach Satz 2.4 linear abhängig. Da $\{x_1, \dots, x_m\}$ bereits V erzeugt, erzeugt auch $\{x, x_1, \dots, x_m\}$ V .

zu 2): Sei

$$x_i = k_1x_1 + \dots + k_{i-1}x_{i-1} \tag{2.1}$$

und sei $x \in V$. Da $\{x_1, \dots, x_m\}$ ein Erzeugendensystem ist, läßt sich x als Linearkombination von x_1, \dots, x_m darstellen, z. B.

$$x = a_1x_1 + \dots + a_mx_m.$$

Einsetzen von (2.1) liefert

$$\begin{aligned} x &= a_1x_1 + \cdots + a_{i-1}x_{i-1} + a_ik_1x_1 + \cdots + a_ik_{i-1}x_{i-1} + a_{i+1}x_{i+1} + \cdots + a_mx_m \\ &= (a_1 + a_ik_1)x_1 + \cdots + (a_{i-1} + a_ik_{i-1})x_{i-1} + a_{i+1}x_{i+1} + \cdots + a_mx_m, \end{aligned}$$

d.h. V wird durch $x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_m$ erzeugt.

□

Satz 2.6 (Austauschsatz)

Sei die Menge $\{x_1, \dots, x_n\}$ ein Erzeugendensystem eines Vektorraums V , d.h. jeder Vektor $x \in V$ lässt sich als Linearkombination von x_1, \dots, x_n darstellen. Falls die Vektoren v_1, \dots, v_m linear unabhängig sind, dann gilt $m \leq n$ und V wird durch eine Menge der Form

$$\{v_1, \dots, v_m, x_{i_1}, \dots, x_{i_{n-m}}\}$$

erzeugt, d.h. in der erzeugenden Menge $\{x_1, \dots, x_n\}$ können m Vektoren durch eine andere linear unabhängige Menge ersetzt werden.

Beweis

O.B.d.A. seien die x_i ungleich Null. Wegen Satz 2.5 1) ist die Menge

$$\{v_1, x_1, \dots, x_n\}$$

linear abhängig. Gemäß Satz 2.4 ist einer dieser Vektoren eine Linearkombination der vorherigen. Bei diesem Vektor kann es sich nicht um v_1 handeln, es muss also einer der x_i 's sein, z.B. x_j .

Wegen Satz 2.5 2) kann man dann x_j aus dem Erzeugendensystem streichen und wir erhalten als Erzeugendensystem die Menge

$$\{v_1, x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_n\}. \quad (2.2)$$

Wir wiederholen jetzt diese Argumentation für den Vektor v_2 . Da (2.2) ein Erzeugendensystem ist, ist

$$\{v_1, v_2, x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_n\}$$

linear abhängig. Einer der Vektoren ist eine Linearkombination der übrigen, wobei v_1 und v_2 nicht in Frage kommen da diese linear unabhängig sind. Sei der Vektor also z.B. x_k . Diesen kann man wieder wegen Satz 2.5 2) aus dem Erzeugendensystem streichen und wir erhalten

$$\{v_1, v_2, x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_{k-1}, x_{k+1}, \dots, x_n\}$$

Wiederholung der Argumentation liefert schließlich die Behauptung für $m \leq n$.

Wir zeigen noch, dass $m > n$ nicht möglich ist:

Nach n Schritten erhalten wir das Erzeugendensystem

$$\{v_1, \dots, v_n\}.$$

Damit würde v_{n+1} als Linearkombination der v_i 's darstellbar sein und v_1, \dots, v_n, v_{n+1} wären linear abhängig. Dies steht aber im Widerspruch zur Annahme, dass v_1, \dots, v_m ($m > n$) unabhängig sind.

□

Wir kommen jetzt zum zentralen Begriff der Basis und Dimension eines Vektorraums:

Definition 2.9 (Basis und Dimension)

Ein Vektorraum V heißt von endlicher Dimension oder n -dimensional, geschrieben $\dim(V) = n$, wenn linear unabhängige Vektoren b_1, \dots, b_n existieren, welche V erzeugen, d.h. jeder Vektor $x \in V$ lässt sich als Linearkombination der b_i darstellen. Die Menge $B := \{b_1, \dots, b_n\}$ heißt dann Basis von V .

Die Dimension eines Vektorraums ist aufgrund des folgenden Satzes wohldefiniert:

Satz 2.7

Sei V ein endlich dimensionaler Vektorraum. Dann hat jede Basis von V die gleiche Anzahl von Elementen. Außerdem ist die Darstellung eines Vektors $x \in V$ durch die Basiselemente eindeutig.

Beweis

Sei $B := \{b_1, \dots, b_n\}$ eine Basis von V und sei $\{x_1, x_2, \dots\}$ eine weitere Basis. Da V durch $\{b_1, \dots, b_n\}$ erzeugt wird, muss die Basis $\{x_1, x_2, \dots\}$ n oder weniger Vektoren enthalten, da ansonsten die x_i wegen Satz 2.6 abhängig wären. Enthält die Basis $\{x_1, x_2, \dots\}$ weniger als n Vektoren, dann wären wegen Satz 2.6 die b_i 's abhängig. Damit muss die Basis $\{x_1, x_2, \dots\}$ genau n Vektoren besitzen.

□

Satz 2.8

Sei V ein Vektorraum, der durch eine endliche Menge S erzeugt wird. Dann ist V von endlicher Dimension und eine Teilmenge von S ist eine Basis von V .

Beweis

Ist S linear unabhängig, dann ist S eine Basis und V endlich dimensional. Ist S linear abhängig, dann ist einer der Vektoren eine Linearkombination der vorherigen (Satz 2.4) und kann gemäß Satz 2.5 2) gelöscht werden. Setzt man diesen Prozeß fort, erhält man ein linear unabhängiges Erzeugendensystem und folglich eine Basis von V .

□

Satz 2.9

Sei V ein endlichdimensionaler Vektorraum mit $\dim(V) = n$. Dann gilt:

1. Eine beliebige Menge von $n + 1$ Vektoren ist linear abhängig.
2. Eine beliebige linear unabhängige Menge kann zu einer Basis erweitert werden.
3. Eine n elementige linear unabhängige Menge ist eine Basis.

Beweis

Sei $\{b_1, \dots, b_n\}$ eine Basis von V .

zu 1): Folgt aus Satz 2.6.

zu 2): Sei $\{x_1, \dots, x_m\}$ linear unabhängig. V wird gemäß Satz 2.6 durch eine Menge S der Form

$$S = \{x_1, \dots, x_m, b_{i_1}, \dots, b_{i_{n-m}}\}$$

erzeugt. Wegen Satz 2.8 ist eine Teilmenge von S eine Basis. Da S n Elemente enthält und $\dim(V) = n$ muss S eine Basis sein, d.h. die linear unabhängigen Vektoren x_1, \dots, x_m sind Teil einer Basis.

zu 3): Sei S eine linear unabhängige Menge mit n Elementen. Wegen 2) ist S Teil einer Basis. Da jede Basis wegen $\dim(V) = n$ genau n Elemente enthält, muss S eine Basis sein.

□

Beispiel 2.16 (Basis des \mathbb{R}^n)

Für den \mathbb{R}^n gilt: $\dim(\mathbb{R}^n) = n$. Eine Basis des \mathbb{R}^n ist z.B. gegeben durch die sogenannte kanonische Basis oder Standardbasis

$$E := \{e_i \in \mathbb{R}^n : e_i = (\delta_{i1}, \dots, \delta_{in})', \quad i = 1, \dots, n\}.$$

Dabei wurde das sogenannte Kroneckersymbol δ_{ij} verwendet, das wie folgt definiert ist:

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

Die kanonische Basis ist selbstverständlich nicht die einzige Basis des \mathbb{R}^n , vielmehr gibt es unendlich viele Basen. Jede linear unabhängige Menge von n Vektoren stellt eine Basis des \mathbb{R}^n dar. Beispielsweise handelt es sich bei den Vektoren

$$x_1 = \begin{pmatrix} 2 \\ 6 \\ 5 \end{pmatrix}, \quad x_2 = \begin{pmatrix} 1 \\ 4 \\ 4 \end{pmatrix}, \quad x_3 = \begin{pmatrix} -2 \\ 4 \\ 3 \end{pmatrix}.$$

aus Beispiel 2.15 um eine Basis des \mathbb{R}^3 , da diese linear unabhängig sind (vgl. Satz 2.9 3)).

Definition 2.10 (Koordinaten)

Sei $B := \{b_1, \dots, b_n\}$ die Basis eines Vektorraumes V . Dann läßt sich jeder Vektor $x \in V$ eindeutig als Linearkombination der Basisvektoren schreiben, d.h.

$$x = a_1 b_1 + \dots + a_n b_n, \quad a_i \in K.$$

Das Tupel $x_B := (a_1, \dots, a_n)'$ heißt Koordinatenvektor von x bezüglich der Basis B .

Beispiel 2.17 (Koordinaten)

Wir betrachten zunächst die Standardbasis des \mathbb{R}^3

$$e_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad e_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad e_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

Der Koordinatenvektor des Vektors $x = (1, 2, 3)'$ bezüglich der Standardbasis ist nicht überraschend $x_E = (1, 2, 3)'$, da

$$x = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = 1 \cdot \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + 2 \cdot \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + 3 \cdot \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

Eine weitere Basis $B = \{b_1, b_2, b_3\}$ des \mathbb{R}^3 ist gegeben durch

$$b_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \quad b_2 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \quad b_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

Wir stellen uns die Frage, wie der Koordinatenvektor von x bezüglich der Basis B aussieht.

Da x als Linearkombination der Basisvektoren darstellbar ist, muss

$$x = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = a_1 \cdot \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + a_2 \cdot \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} + a_3 \cdot \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

gelten. Die Koordinaten $x_B = (a_1, a_2, a_3)'$ können also einfach als Lösung eines linearen Gleichungssystems gewonnen werden. Wir erhalten $a_1 = 1$, $a_2 = 1$, $a_3 = 1$, d.h. der Vektor $x_B = (1, 1, 1)'$ ist der Koordinatenvektor von x bezüglich der Basis B .

△

Wir tragen im Folgenden noch die wichtigsten Tatsachen über die Dimension in Unterräumen zusammen:

Satz 2.10

Für die Dimension in Unterräumen gilt:

1. Sei U ein Unterraum des n dimensionalen Vektorraums V . Dann gilt $\dim(U) \leq n$. Ist $\dim(U) = n$, so gilt $U = V$.
2. Seien U_1, U_2 Unterräume des endlich dimensionalen Vektorraums V . Dann hat $U_1 + U_2$ endliche Dimension, und es gilt:

$$\dim(U_1 + U_2) = \dim(U_1) + \dim(U_2) - \dim(U_1 \cap U_2)$$

Beweis:

zu 1): Wegen $\dim(V) = n$ sind stets $n + 1$ Vektoren oder mehr linear abhängig (Satz 2.9 1)). Andererseits kann eine Basis von U nicht mehr als n Elemente besitzen, weil sie aus linear unabhängigen Vektoren gebildet wird, d.h. $\dim(U) \leq n$. Ist $\{u_1, \dots, u_n\}$ eine Basis von U , dann ist es wegen Satz 2.9 3) auch eine Basis von V und es folgt $U = V$.

zu 2): Wir stellen zunächst fest, dass $U_1 \cap U_2$ sowohl ein Unterraum von U_1 als auch von U_2 ist. Sei $\dim(U_1) = m$, $\dim(U_2) = n$ und $\dim(U_1 \cap U_2) = r$. Sei weiterhin $\{v_1, \dots, v_r\}$ eine Basis von $U_1 \cap U_2$. Aufgrund von Satz 2.9 2) kann $\{v_1, \dots, v_r\}$ zu einer Basis von U_1 erweitert werden, z. B.

$$\{v_1, \dots, v_r, u_1, \dots, u_{m-r}\} \quad (2.3)$$

Analog kann $\{v_1, \dots, v_r\}$ zu einer Basis von U_2 erweitert werden z. B.

$$\{v_1, \dots, v_r, w_1, \dots, w_{n-r}\} \quad (2.4)$$

Wir definieren

$$S := \{v_1, \dots, v_r, u_1, \dots, u_{m-r}, w_1, \dots, w_{n-r}\}$$

Da S genau $r + m - r + n - r = m + n - r$ Elemente enthält, ist die Behauptung bewiesen, wenn wir zeigen können, dass S eine Basis von $U_1 + U_2$ ist.

Zunächst ist klar, dass $U_1 + U_2$ durch die Vereinigung von (2.3) und (2.4) also von S erzeugt wird. Wir müssen also noch zeigen, dass S linear unabhängig ist. Sei

$$a_1 v_1 + \dots + a_r v_r + b_1 u_1 + \dots + b_{m-r} u_{m-r} + c_1 w_1 + \dots + c_{n-r} w_{n-r} = 0, \quad (2.5)$$

wobei a_i, b_j und c_k Skalare seien. Wir zeigen, dass $a_i = 0, b_j = 0$ und $c_k = 0$ gilt. Sei

$$v = a_1 v_1 + \dots + a_r v_r + b_1 u_1 + \dots + b_{m-r} u_{m-r}. \quad (2.6)$$

Wegen (2.5) gilt

$$v = -c_1 w_1 - \dots - c_{n-r} w_{n-r}. \quad (2.7)$$

Da v definitionsgemäß eine Linearkombination der Basisvektoren von U_1 ist, gilt $v \in U_1$. Wegen (2.7) ist v auch eine Linearkombination von Basisvektoren von U_2 , d.h. $v \in U_2$. Also gilt $v \in U_1 \cap U_2$. Da $\{v_1, \dots, v_r\}$ eine Basis von $U_1 \cap U_2$ ist, läßt sich v darstellen als

$$v = a_1 v_1 + \dots + a_r v_r.$$

Einsetzen in (2.7) und umstellen liefert

$$a_1 v_1 + \dots + a_r v_r + c_1 w_1 + \dots + c_{n-r} w_{n-r} = 0.$$

Da $\{v_1, \dots, v_r, w_1, \dots, w_{n-r}\}$ eine Basis von U_2 ist, sind die Vektoren linear unabhängig und es folgt $c_1 = c_2 = \dots = c_{n-r} = 0$. Einsetzen in (2.5) liefert

$$a_1 v_1 + \dots + a_r v_r + b_1 u_1 + \dots + b_{m-r} u_{m-r} = 0.$$

Da aber $\{v_1, \dots, v_r, u_1, \dots, u_{m-r}\}$ eine Basis von U_1 ist, folgt $a_1 = \dots = a_r = 0$ und $b_1 = \dots = b_{m-r} = 0$ und damit die Behauptung. □

Beispiel 2.18

Wir betrachten den durch die Vektoren

$$x_1 = \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix}, \quad x_2 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \quad x_3 = \begin{pmatrix} 3 \\ 5 \\ 7 \end{pmatrix}$$

aufgespannten Unterraum U_1 und den durch

$$y_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \quad y_2 = \begin{pmatrix} -1 \\ 2 \\ 2 \end{pmatrix}, \quad y_3 = \begin{pmatrix} 2 \\ -13 \\ -10 \end{pmatrix}, \quad y_4 = \begin{pmatrix} 2 \\ -1 \\ -2 \end{pmatrix}$$

aufgespannten Unterraum U_2 . Wir bestimmen zunächst die Dimension und eine Basis für U_1 und U_2 . Durch Dreieckszerlegung der entsprechenden Matrix erkennt man, dass x_1 , x_2 und x_3 linear abhängig sind. Die Vektoren x_2 und x_3 sind aber linear unabhängig, so dass $\dim(U_1) = 2$ gilt. Ausserdem bilden x_2 und x_3 eine Basis von U_1 (vergleiche auch Satz 2.9 3)). Die Vektoren y_1 , y_2 , y_3 und y_4 sind auf jeden Fall linear abhängig, da gemäß Satz 2.10 1) $\dim(U_2) \leq \dim(\mathbb{R}^3) = 3$ gilt. Wiederum durch Dreieckszerlegung erkennt man, dass $\dim(U_2) = 2$. Da die Vektoren y_1 und y_2 linear unabhängig sind, bilden diese eine Basis von U_2 .

Wir bestimmen im Folgenden noch jeweils die Dimension und eine Basis von $U_1 + U_2$ und $U_1 \cap U_2$. Für $U_1 + U_2$ gilt $\dim(U_1 + U_2) \geq 2$ und $\dim(U_1 + U_2) \leq 3$. Da die Vektoren x_2 , x_3

und y_1 linear unabhängig sind, bilden sie eine Basis von $U_1 + U_2$ und es gilt $\dim(U_1 + U_2) = 3$. Wegen $\dim(\mathbb{R}^3) = 3$ bilden die Vektoren auch eine Basis des \mathbb{R}^3 , d.h. bei $U_1 + U_2$ handelt es sich um den \mathbb{R}^3 . Aus der Dimensionsformel (Satz 2.10 2)) folgt

$$\dim(U_1 + U_2) = \dim(U_1) + \dim(U_2) - \dim(U_1 \cap U_2) = 2 + 2 - \dim(U_1 \cap U_2)$$

und damit $\dim(U_1 \cap U_2) = 1$. Zur Bestimmung einer Basis muss ein Vektor z gefunden werden, der sowohl in U_1 als auch in U_2 enthalten ist. Dies führt mit den Basisvektoren von U_1 und U_2 auf den Ansatz

$$\lambda_1 \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + \lambda_2 \begin{pmatrix} 3 \\ 5 \\ 7 \end{pmatrix} = \lambda_3 \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + \lambda_4 \begin{pmatrix} -1 \\ 2 \\ 2 \end{pmatrix}$$

bzw.

$$\begin{aligned} \lambda_1 + 3\lambda_2 - \lambda_3 + \lambda_4 &= 0 \\ \lambda_1 + 5\lambda_2 - \lambda_3 - 2\lambda_4 &= 0 \\ \lambda_1 + 7\lambda_2 - 2\lambda_4 &= 0. \end{aligned}$$

Als Lösung erhalten wir z.B.

$$\begin{aligned} \lambda_4 &= 1 \\ \lambda_3 &= -3 \\ \lambda_2 &= \frac{3}{2} \\ \lambda_1 &= -8.5. \end{aligned}$$

Damit ist

$$z = -3 \cdot \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + 1 \cdot \begin{pmatrix} -1 \\ 2 \\ 2 \end{pmatrix} = \begin{pmatrix} -4 \\ -1 \\ 2 \end{pmatrix}$$

eine Basis von $U_1 \cap U_2$.

△

2.5 Lineare Abbildungen zwischen Vektorräumen

Bisher haben wir Vektorräume gesondert betrachtet. Im Folgenden wollen wir uns mit Beziehungen zwischen zwei (endlich dimensionalen) Vektorräumen V und W befassen, genauer mit sogenannten *linearen Abbildungen*. Diese spielen eine zentrale Rolle in der linearen Algebra. Besonders wichtig ist der Zusammenhang mit Matrizen. Es stellt sich nämlich heraus, dass jeder linearen Abbildung f eine Matrix zugeordnet werden kann, die f charakterisiert. Wir beginnen mit der Definition einer linearen Abbildung.

Definition 2.11 (lineare Abbildung)

Seien V, W K -Vektorräume. Eine Abbildung $f : V \mapsto W$ heißt K -linear oder Vektorraum-Homomorphismus, wenn für alle $x, y \in V$ und $a \in K$ gilt:

1. $f(x + y) = f(x) + f(y)$
2. $f(ax) = af(x)$

Im Spezialfall $V = W$ heißt f auch linearer Operator.

Beispiel 2.19

Wir betrachten die Abbildung $f : \mathbb{R}^3 \mapsto \mathbb{R}^3$ mit

$$f(x) = (x_1 + 2x_2, x_2 + x_3, x_1 + x_2 + x_3)',$$

wobei $x = (x_1, x_2, x_3)'$. Die Abbildung ist linear, denn für $x, y \in \mathbb{R}^3$ gilt

$$\begin{aligned} f(x + y) &= (x_1 + y_1 + 2(x_2 + y_2), x_2 + y_2 + x_3 + y_3, x_1 + y_1 + x_2 + y_2 + x_3 + y_3)' \\ &= (x_1 + 2x_2, x_2 + x_3, x_1 + x_2 + x_3)' + (y_1 + 2y_2, y_2 + y_3, y_1 + y_2 + y_3)' \\ &= f(x) + f(y) \end{aligned}$$

und für $a \in \mathbb{R}$ gilt

$$f(ax) = (ax_1 + 2ax_2, ax_2 + ax_3, ax_1 + ax_2 + ax_3)' = a(x_1 + 2x_2, x_2 + x_3, x_1 + x_2 + x_3)' = af(x).$$

△

Beispiel 2.20 (Durch Matrizen induzierte lineare Abbildungen)

Sei \mathbf{A} eine $m \times n$ Matrix auf einem Körper K . Wir definieren $f : K^n \rightarrow K^m$ mit

$$f(x) = \mathbf{A}x.$$

Diese Abbildung ist linear, denn

$$\begin{aligned} f(x_1 + x_2) &= \mathbf{A}(x_1 + x_2) = \mathbf{A}x_1 + \mathbf{A}x_2 \\ &= f(x_1) + f(x_2) \end{aligned}$$

und

$$f(ax) = \mathbf{A}(ax) = a\mathbf{A}x = af(x).$$

△

Beispiel 2.20 zeigt bereits die enge Verknüpfung von Matrizen und linearen Abbildungen. Seien V und W zwei endlich dimensionale Vektorräume der Dimension $\dim(V) = n$ und

$\dim(W) = m$. Seien weiterhin $B := \{b_1, \dots, b_n\}$ und $E := \{e_1, \dots, e_m\}$ Basen der beiden Vektorräume. Wir werden sehen, dass jeder linearen Abbildung

$$f : V \rightarrow W$$

eine $m \times n$ Matrix \mathbf{A} zugeordnet werden kann, so dass

$$\mathbf{A} \cdot x_B = f_E(x),$$

wobei x_B die Koordinatendarstellung von $x \in V$ bezüglich der Basis B ist und $f_E(x)$ die Koordinatendarstellung von $f(x)$ bezüglich der Basis E ist.

Definition 2.12 (Matrixdarstellung einer linearen Abbildung)

Sei $f : V \rightarrow W$ eine lineare Abbildung und seien $B := \{b_1, \dots, b_n\}$ und $E := \{e_1, \dots, e_m\}$ Basen der beiden Vektorräume V bzw. W . Die Funktion f ausgewertet an den Basisvektoren b_j , $j = 1, \dots, n$, läßt sich in Abhängigkeit von den Basisvektoren e_i , $i = 1, \dots, m$ schreiben als

$$f(b_j) = a_{1j}e_1 + \dots + a_{mj}e_m.$$

Dann heißt die $m \times n$ Matrix

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

die Matrixdarstellung von f .

Die Definition ist durch den folgenden Satz gerechtfertigt:

Satz 2.11

Sei $f : V \rightarrow W$ eine lineare Abbildung und seien $B := \{b_1, \dots, b_n\}$ und $E := \{e_1, \dots, e_m\}$ Basen der beiden Vektorräume. Dann gilt

$$\mathbf{A} \cdot x_B = f_E(x),$$

wobei \mathbf{A} die Matrixdarstellung von f ist, x_B die Koordinatendarstellung von $x \in V$ bezüglich der Basis B ist und $f_E(x)$ die Koordinatendarstellung von $f(x) \in W$ bezüglich der Basis E .

Beweis:

Sei $x_B = (k_1, \dots, k_n)'$ die Koordinatendarstellung von x bezüglich B . Dann erhalten wir unter Ausnutzung der Linearität von f

$$\begin{aligned}
f(x) &= f\left(\sum_{j=1}^n k_j b_j\right) = \sum_{j=1}^n k_j f(b_j) \\
&= \sum_{j=1}^n k_j \sum_{i=1}^m a_{ij} e_i = \sum_{i=1}^m \left(\sum_{j=1}^n k_j a_{ij}\right) e_i \\
&= \sum_{i=1}^m (a_{i1}k_1 + \cdots + a_{in}k_n) e_i.
\end{aligned}$$

Damit ist das i -te Element der Koordinatendarstellung $f_E(x)$ von $f(x)$ bezüglich E gegeben durch

$$a_{i1}k_1 + \cdots + a_{in}k_n.$$

Dies ist zugleich die i -te Komponente des Spaltenvektors $\mathbf{A}x_B$, so dass die Behauptung folgt. □

Beispiel 2.21

Wir betrachten wieder die lineare Abbildung $f: \mathbb{R}^3 \mapsto \mathbb{R}^3$ aus Beispiel 2.19 mit

$$f(x) = (x_1 + 2x_2, x_2 + x_3, x_1 + x_2 + x_3)'$$

Wir bestimmen zunächst die Matrixdarstellung von f bezüglich der Standardbasis e_1, e_2, e_3 des \mathbb{R}^3 . Es gilt:

$$f(e_1) = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} = 1 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + 0 \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + 1 \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

$$f(e_2) = \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix} = 2 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + 1 \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + 1 \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

$$f(e_3) = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} = 0 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + 1 \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + 1 \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

Damit erhalten wir als Matrixdarstellung von f die Matrix

$$\mathbf{A} = \begin{pmatrix} 1 & 2 & 0 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}.$$

Beispielsweise berechnen wir dann für $x = (2, 1, 2)'$

$$f_E(x) = \mathbf{A} = \begin{pmatrix} 1 & 2 & 0 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} 2 \\ 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 4 \\ 3 \\ 5 \end{pmatrix}.$$

Wenn wir die Basis wechseln, ändert sich natürlich auch die Matrixdarstellung von f . Betrachte zum Beispiel die Basis B des \mathbb{R}^3 aus Beispiel 2.17 gegeben durch

$$b_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \quad b_2 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \quad b_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

Wir bestimmen im Folgenden die Matrixdarstellung von f bezüglich dieser Basis. Es gilt:

$$f(b_1) = \begin{pmatrix} 3 \\ 2 \\ 3 \end{pmatrix} = 3 \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} - 1 \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} + 1 \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

$$f(b_2) = \begin{pmatrix} 2 \\ 2 \\ 2 \end{pmatrix} = 2 \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + 0 \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} + 0 \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

$$f(b_3) = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} = 0 \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + 1 \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} + 0 \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

Als Matrixdarstellung bezüglich der Basis B erhalten wir also

$$\mathbf{A} = \begin{pmatrix} 3 & 2 & 0 \\ -1 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}.$$

Die Koordinatendarstellung x_B des Vektors $x = (2, 1, 2)'$ bezüglich B ist gegeben durch $(2, -1, 1)'$ und wir erhalten

$$f_B(x) = \mathbf{A} = \begin{pmatrix} 3 & 2 & 0 \\ -1 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} 2 \\ -1 \\ 1 \end{pmatrix} = \begin{pmatrix} 4 \\ -1 \\ 2 \end{pmatrix}$$

als Koordinatendarstellung von $f(x)$ bezüglich der Basis B .

△

Wir beenden diesen Abschnitt mit der Definition des Bilds und des Kerns einer linearen Abbildung. Beide Begriffe tauchen in der Literatur und im Folgenden immer wieder auf.

Definition 2.13 (Bild und Kern einer linearen Abbildung)

Das Bild einer linearen Abbildung $f : V \mapsto W$, geschrieben $\text{Bild}(f)$, ist die Menge der Bildpunkte von f in W :

$$\text{Bild}(f) = \{w \in W : f(x) = w, x \in V\}$$

Der Kern von f , geschrieben $\text{Ker}(f)$, ist die Menge der Elemente x in V , für die $f(x) = 0$ gilt.

Satz 2.12

Sei $f : V \mapsto W$ eine lineare Abbildung. Dann gilt:

1. Das Bild von f ist ein Unterraum von W , und der Kern von f ist ein Unterraum von V .
2. $\dim(V) = \dim(\text{Ker}(f)) + \dim(\text{Bild}(f))$

Beweis

zu 1): Wegen $f(0) = 0$ gilt $0 \in \text{Bild}(f)$. Seien nun $v, w \in \text{Bild}(f)$ und $a, b \in K$ Skalare. Es existieren Vektoren $v', w' \in V$ so dass $f(v') = v$ und $f(w') = w$ gilt. Damit gilt

$$f(av' + bw') = af(v') + bf(w') = av + bw,$$

d.h. $av + bw \in \text{Bild}(f)$ und damit $\text{Bild}(f)$ ein Unterraum von W .

Wegen $f(0) = 0$ gilt $0 \in \text{Ker}(f)$. Seien $v, w \in \text{Ker}(f)$ und $a, b \in K$. Es gilt $f(v) = 0$ und $f(w) = 0$. Damit folgt

$$f(av + bw) = af(v) + bf(w) = a0 + b0 = 0,$$

d.h. $av + bw \in \text{Ker}(f)$ und damit ist $\text{Ker}(f)$ ein Unterraum von V .

zu 2): Sei $\dim V = n$. Da $\text{Ker}(f)$ ein Unterraum von V ist, muss auch die Dimension von $\text{Ker}(f)$ endlich sein, sei also $\dim(\text{Ker}(f)) = r \leq n$. Wir müssen also zeigen, dass $\dim(\text{Bild}(f)) = n - r$.

Sei $\{k_1, \dots, k_r\}$ eine Basis von $\text{Ker}(f)$. Gemäß Satz 2.9 2), können die Basisvektoren von $\text{Ker}(f)$ zu einer Basis von V erweitert werden. Sei also $\{k_1, \dots, k_r, v_1, \dots, v_{n-r}\}$ eine Basis von V . Wir zeigen im Folgenden, dass die Menge $\{f(v_1), \dots, f(v_{n-r})\}$ eine Basis von $\text{Bild}(f)$ ist und damit $\dim(\text{Bild}(f)) = n - r$.

Sei also $b \in \text{Bild}(f)$. Dann gibt es einen Vektor $v \in V$ mit $f(v) = b$ und es existiert die Darstellung

$$v = a_1 k_1 + \dots + a_r k_r + b_1 v_1 + \dots + b_{n-r} v_{n-r}.$$

Da die Vektoren k_i zum Kern von f gehören, gilt $f(k_i) = 0$ und wir erhalten

$$\begin{aligned}
b &= f(v) = f(a_1k_1 + \cdots + a_rk_r + b_1v_1 + \cdots + b_{n-r}v_{n-r}) \\
&= a_1f(k_1) + \cdots + a_rf(k_r) + b_1f(v_1) + \cdots + b_{n-r}f(v_{n-r}) \\
&= b_1f(v_1) + \cdots + b_{n-r}f(v_{n-r}).
\end{aligned}$$

Damit erzeugen die Vektoren $f(v_1), \dots, f(v_{n-r})$ das Bild von f . Wir müssen also nur noch zeigen, dass die Vektoren linear unabhängig sind.

Sei

$$b_1f(v_1) + \cdots + b_{n-r}f(v_{n-r}) = 0.$$

Dann gilt

$$f(b_1v_1 + \cdots + b_{n-r}v_{n-r}) = 0$$

und der Vektor $b_1v_1 + \cdots + b_{n-r}v_{n-r}$ gehört zum Kern von f . Da die Vektoren k_1, \dots, k_r eine Basis von $\text{Ker}(f)$ sind, läßt sich der Vektor $b_1v_1 + \cdots + b_{n-r}v_{n-r}$ als eine Linearkombination der k_i 's darstellen:

$$b_1v_1 + \cdots + b_{n-r}v_{n-r} = c_1k_1 + \cdots + c_rk_r.$$

Umstellen liefert

$$b_1v_1 + \cdots + b_{n-r}v_{n-r} - c_1k_1 - \cdots - c_rk_r = 0.$$

Da die v_i 's und die k_i 's zusammen eine Basis von V bilden und daher linear unabhängig sind, folgt $b_1 = \cdots = b_{n-r} = 0$ und damit die lineare Unabhängigkeit von $f(v_1), \dots, f(v_{n-r})$. Es gilt also $\dim(\text{Bild}(f)) = n - r$.

□

2.6 Euklidische Vektorräume

In Kapitel 1 haben wir für den \mathbb{R}^n bereits geometrische Begriffe wie die Länge eines Vektors und den Abstand zweier Vektoren definiert. Die Definition war dabei vor allem von der geometrischen Anschauung im \mathbb{R}^2 bzw. \mathbb{R}^3 motiviert. Häufig werden aber abstraktere Definitionen benötigt. In diesem Abschnitt definieren wir für allgemeine Vektorräume Länge und Abstand von Vektoren wobei wir uns auf Vektorräume V bezüglich des Körpers $K = \mathbb{R}$ beschränken. In diesem Fall nennt man V einen *reellen Vektorraum*.

Definition 2.14 (Normierter Vektorraum)

Sei V ein reeller Vektorraum. Eine Abbildung

$$\begin{aligned}
\| \cdot \| : V &\mapsto \mathbb{R} \\
x &\mapsto \|x\|
\end{aligned}$$

heißt Norm auf V , falls für alle $x, y \in V$ und $a \in K$ gilt:

1. $\|x\| \geq 0$ und $\|x\| = 0 \iff x = 0$
2. $\|ax\| = a\|x\|$
3. $\|x + y\| \leq \|x\| + \|y\|$ (Dreiecksungleichung)

Die reelle Zahl $\|x\|$ heißt Norm (auch Betrag, Länge) des Vektors x . Das Paar $(V, \|\cdot\|)$ heißt normierter Vektorraum. Ist klar, welche Norm gemeint ist, so schreibt man kurz V statt $(V, \|\cdot\|)$.

Beispiel 2.22 (Normen)

Sei $V = \mathbb{R}^n$ und $1 \leq p \leq \infty$. Dann wird durch

$$\|x\|_p = \begin{cases} \left(\sum_{i=1}^n |x_i|^p \right)^{\frac{1}{p}} & 1 \leq p < \infty \\ \max\{|x_1|, \dots, |x_n|\} & p = \infty \end{cases}$$

eine Norm auf V definiert. Für $p = 2$ heißt die Norm euklidische Norm und für $p = \infty$ Tschebyscheff Norm oder Unendlichnorm. In Abbildung 2.1 sind für $V = \mathbb{R}^2$ und $p = 1, 2, \infty$ die Normen veranschaulicht. Der Länge des Vektors entspricht dabei die Länge der dickgezeichneten Linie. Im Fall $p = 2$ entspricht die Norm bzw. die Länge eines Vektors der Länge des jeweiligen Ortsvektors.

△

Definition 2.15 (Metrik)

Sei V eine Menge. Unter einer Metrik auf V versteht man eine Abbildung

$$\begin{aligned} d: V \times V &\mapsto \mathbb{R} \\ (x, y) &\mapsto d(x, y) \end{aligned}$$

mit folgenden Eigenschaften:

1. $d(x, y) \geq 0$, $d(x, y) = 0 \iff x = y$
2. $d(x, y) = d(y, x)$ für alle $x, y \in V$ (Symmetrie)
3. $d(x, z) \leq d(x, y) + d(y, z)$, für alle $x, y, z \in V$ (Dreiecksungleichung)

Das Paar (V, d) heißt metrischer Raum. Man nennt d auch den Abstand der Punkte x und y .

Ist $\|\cdot\|$ eine Norm auf V , so ist durch $d(x, y) := \|y - x\|$ für $x, y \in V$ eine Metrik auf V erklärt. In Abbildung 2.2 sind die durch die Norm $\|\cdot\|_p$ induzierten Metriken für $p = 1, 2, \infty$ veranschaulicht. Die jeweiligen Abstände ergeben sich als die Länge der dick gezeichneten Linien. Im Fall $p = 2$ heißt der Abstand zwischen den Punkten auch euklidischer Abstand.

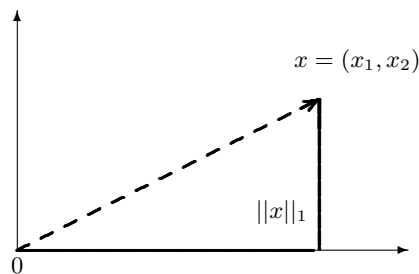
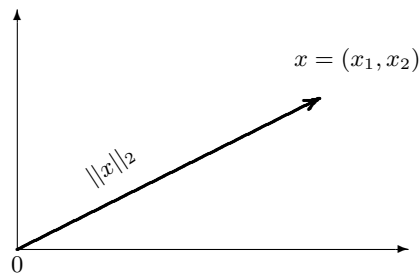
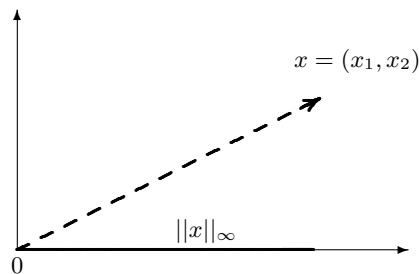
a) $\|x\|_1$ b) $\|x\|_2$ c) $\|x\|_\infty$ 

Abbildung 2.1. Veranschaulichung der Normen $\|x\|_1$, $\|x\|_2$ und $\|x\|_\infty$ im \mathbb{R}^2 .

Definition 2.16 (inneres Produkt)

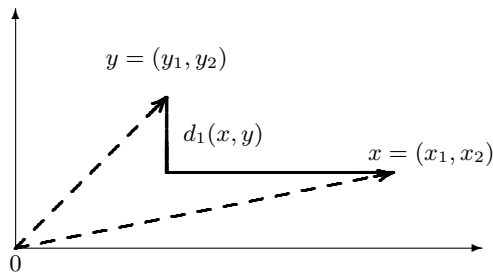
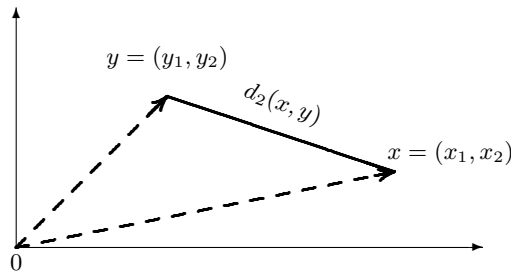
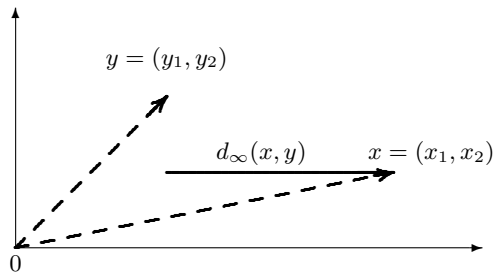
Sei V ein reeller Vektorraum. Eine Abbildung $\langle \cdot, \cdot \rangle: V \times V \mapsto \mathbf{R}$ heißt inneres Produkt oder Skalarprodukt, wenn sie die folgenden Axiome erfüllt:

1. $\langle ax_1 + bx_2, y \rangle = a \langle x_1, y \rangle + b \langle x_2, y \rangle$
2. $\langle x, y \rangle = \langle y, x \rangle$
3. $\langle x, x \rangle \geq 0$, $\langle x, x \rangle = 0 \iff x = 0$

Ein reeller Vektorraum versehen mit einem inneren Produkt heißt euklidischer Vektorraum.

Bemerkung:

Jeder euklidische Vektorraum wird durch $\|v\| := \langle v, v \rangle^{\frac{1}{2}}$ zu einem normierten Vektorraum.

a) $d_1(x, y)$ b) $d_2(x, y)$ c) $d_\infty(x, y)$ Abbildung 2.2. Veranschaulichung der Metriken $d_1(x, y)$, $d_2(x, y)$ und $d_\infty(x, y)$ im \mathbb{R}^2 .**Beispiel 2.23 (Standardskalarprodukt)**Sei $V = \mathbb{R}^n$. Dann ist das Skalarprodukt

$$\langle x, y \rangle := \sum_{i=1}^n x_i y_i = x' y$$

aus Kapitel 1 ein inneres Produkt und wird als Standardskalarprodukt bezeichnet. Bei der durch das Standardskalarprodukt induzierten Norm bzw. Metrik handelt es sich um die euklidische Norm bzw. den euklidischen Abstand.

△

Beispiel 2.24

Wir betrachten den Vektorraum $V = C[a, b]$ der auf $[a, b]$ stetigen Funktionen. Für diesen Vektorraum stellt zum Beispiel

$$\langle f, g \rangle := \int f(x)g(x)dx$$

ein Skalarprodukt dar.

Satz 2.13 (Cauchy–Schwarzsche Ungleichung)

Sei V ein euklidischer Vektorraum, dann gilt für alle $x, y \in V$:

$$| \langle x, y \rangle | \leq \|x\| \|y\|$$

Beweis:

Für $y = 0$ ist die Aussage klar. Sei also $y \neq 0$. Wir definieren

$$\lambda := \frac{\langle x, y \rangle}{\|y\|^2}.$$

Damit gilt:

$$\begin{aligned} 0 &\leq \langle x - \lambda y, x - \lambda y \rangle \\ &= \langle x, x \rangle - 2\lambda \langle x, y \rangle + \lambda^2 \langle y, y \rangle \\ &= \|x\|^2 - 2 \frac{\langle x, y \rangle^2}{\|y\|^2} + \frac{\langle x, y \rangle^2}{\|y\|^2} \\ &= \|x\|^2 - \frac{\langle x, y \rangle^2}{\|y\|^2} \end{aligned}$$

Umstellen liefert $\langle x, y \rangle^2 \leq \|x\|^2 \|y\|^2$ und damit $| \langle x, y \rangle | \leq \|x\| \|y\|$.

□

Definition 2.17 (Winkel zwischen zwei Vektoren)

Sei V ein reeller Vektorraum. Dann ist der Winkel ϕ zwischen zwei von Null verschiedenen Vektoren v_1 und v_2 definiert durch

$$\cos \phi = \frac{\langle v_1, v_2 \rangle}{\|v_1\| \|v_2\|}.$$

Definition 2.18 (Orthogonalität, Orthonormalität)

Sei V ein euklidischer Vektorraum. Zwei Vektoren $x_1, x_2 \in V$ werden orthogonal genannt, geschrieben $x_1 \perp x_2$, wenn $\langle x_1, x_2 \rangle = 0$ gilt. Sie heißen orthonormal, wenn sie zusätzlich die Länge Eins haben, d.h. $\|x_i\| = 1$. Eine Menge x_1, \dots, x_n heißt orthogonal, wenn ihre verschiedenen Elemente paarweise orthogonal sind, d.h. $\langle x_i, x_j \rangle = 0$ für $i \neq j$. Die Menge heißt orthonormal, wenn jedes x_i die Länge 1 hat. Es ist klar, dass jede orthogonale Menge von Vektoren durch Normierung zu einer orthonormalen Menge gemacht werden kann.

Bemerkung:

Der Nullvektor 0 ist zu jedem Vektor $v \in V$ orthogonal, denn

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

▽

Bemerkung:

Ist das Standardskalarprodukt zugrundegelegt, so bedeutet die Orthogonalität zweier Vektoren im \mathbb{R}^2 bzw. im \mathbb{R}^3 , dass die beiden Ortsvektoren senkrecht aufeinander stehen (vgl. Abbildung 1.4).

▽

Für orthonormale Mengen gilt der folgende

Satz 2.14

Sei V ein euklidischer Vektorraum. Dann gilt:

1. Eine orthonormale Menge x_1, \dots, x_r ist linear unabhängig. Für einen beliebigen Vektor $x \in V$ ist der Vektor

$$w = x - \langle x, x_1 \rangle x_1 - \dots - \langle x, x_r \rangle x_r$$

zu jedem der x_i orthogonal.

2. Sei v_1, \dots, v_r eine beliebige lineare unabhängige Menge von V . Dann lässt sich diese stets in eine orthonormale Menge x_1, \dots, x_n überführen, die denselben Raum aufspannt.

Beweis:

zu 1): Es gelte $0 = \lambda_1 x_1 + \dots + \lambda_r x_r$. Es ist zu zeigen, dass $\lambda_i = 0$ folgt. Für $i = 1, \dots, r$ bilden wir auf beiden Seiten das Skalarprodukt bezüglich x_i und erhalten

$$\begin{aligned} \langle 0, x_i \rangle = 0 &= \langle \lambda_1 x_1 + \dots + \lambda_r x_r, x_i \rangle \\ &= \lambda_1 \langle x_1, x_i \rangle + \lambda_i \langle x_i, x_i \rangle + \lambda_r \langle x_r, x_i \rangle \\ &= \lambda_i \langle x_i, x_i \rangle = \lambda_i, \end{aligned}$$

woraus der erste Teil von Aussage 1) folgt. Weiter ist zu zeigen, dass w und x_i orthogonal sind. Die folgende Rechnung liefert den Nachweis:

$$\begin{aligned}
\langle w, x_i \rangle &= \langle x - \langle x, x_1 \rangle x_1 - \cdots - \langle x, x_r \rangle x_r, x_i \rangle \\
&= \langle x, x_i \rangle - \langle x, x_1 \rangle \langle x_1, x_i \rangle - \cdots - \langle x, x_i \rangle \langle x_i, x_i \rangle - \\
&\quad \cdots - \langle x, x_r \rangle \langle x_r, x_i \rangle \\
&= \langle x, x_i \rangle - \langle x, x_i \rangle \langle x_i, x_i \rangle \\
&= \langle x, x_i \rangle - \langle x, x_i \rangle \\
&= 0
\end{aligned}$$

zu 2): Setze

$$\tilde{v}_1 = \frac{v_1}{\|v_1\|}.$$

Offensichtlich ist \tilde{v}_1 orthonormal. Weiter setzen wir

$$w_2 = v_2 - \langle v_2, \tilde{v}_1 \rangle \tilde{v}_1$$

und

$$\tilde{v}_2 = \frac{w_2}{\|w_2\|}.$$

Wegen Aussage 1) des Satzes ist w_2 und damit \tilde{v}_2 orthogonal zu \tilde{v}_1 , und \tilde{v}_1, \tilde{v}_2 sind orthonormal. Die Fortführung dieses Konstruktionsprinzips liefert nach Erhalt des Orthonormalsystems $\tilde{v}_1, \dots, \tilde{v}_i$

$$w_{i+1} = v_{i+1} - \langle v_{i+1}, \tilde{v}_1 \rangle \tilde{v}_1 - \cdots - \langle v_{i+1}, \tilde{v}_i \rangle \tilde{v}_i$$

und

$$\tilde{v}_{i+1} = \frac{w_{i+1}}{\|w_{i+1}\|}.$$

(Etwas formaler wird der Beweis durch Induktion geführt.)

□

Der Beweis des Satz liefert also auch ein Orthonormalisierungsverfahren mit dessen Hilfe jede beliebige Basis v_1, \dots, v_n in eine orthonormale Basis $\tilde{v}_1, \dots, \tilde{v}_n$ überführt werden kann. Dieser Algorithmus ist als Gram-Schmidtsches Orthonormalisierungsverfahren bekannt.

Algorithmus 2.1 (Gram-Schmidtsches Orthonormalisierungsverfahren)

Sei V ein endlichdimensionaler Vektorraum und v_1, \dots, v_n eine Basis. Folgender Algorithmus liefert eine orthonormale Basis $\tilde{v}_1, \dots, \tilde{v}_n$.

1. Setze $\tilde{v}_1 = \frac{v_1}{\|v_1\|}$.
2. Setze $i = 2$.

3. Setze $w_i = v_i - \langle v_i, \tilde{v}_1 \rangle \tilde{v}_1 - \dots - \langle v_i, \tilde{v}_{i-1} \rangle \tilde{v}_{i-1}$ und damit $\tilde{v}_i = \frac{w_i}{\|w_i\|}$.
4. Setze $i = i + 1$. Falls $i > n$ beende den Algorithmus, $\tilde{v}_1, \dots, \tilde{v}_n$ ist dann eine orthonormale Basis von V . Ansonsten fahre fort mit 3).

Beispiel 2.25

Betrachte die Vektoren

$$v_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \quad v_2 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \quad v_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

Diese bilden eine Basis des \mathbb{R}^3 . Wir bestimmen eine Orthonormalbasis $\tilde{v}_1, \tilde{v}_2, \tilde{v}_3$.

Setze

$$\tilde{v}_1 = \frac{v_1}{\|v_1\|} = \frac{v_1}{\sqrt{3}} = \begin{pmatrix} \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{pmatrix}.$$

Weiter berechnen wir

$$w_2 = v_2 - \langle v_2, \tilde{v}_1 \rangle \tilde{v}_1 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} - \frac{2}{\sqrt{3}} \begin{pmatrix} \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{pmatrix} = \begin{pmatrix} -\frac{2}{3} \\ \frac{1}{3} \\ \frac{1}{3} \end{pmatrix}$$

$$\tilde{v}_2 = \frac{w_2}{\|w_2\|} = \begin{pmatrix} -\frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \end{pmatrix}$$

und

$$\begin{aligned} w_3 &= v_3 - \langle v_3, \tilde{v}_1 \rangle \tilde{v}_1 - \langle v_3, \tilde{v}_2 \rangle \tilde{v}_2 \\ &= \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} - \frac{1}{\sqrt{3}} \begin{pmatrix} \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{pmatrix} - \frac{1}{\sqrt{6}} \begin{pmatrix} -\frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \end{pmatrix} = \begin{pmatrix} 0 \\ -\frac{1}{2} \\ \frac{1}{2} \end{pmatrix} \end{aligned}$$

$$\tilde{v}_3 = \frac{w_3}{\|w_3\|} = \begin{pmatrix} 0 \\ -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}.$$

△

Definition 2.19 (orthogonale Unterräume, orthogonales Komplement)

Sei V ein euklidischer Vektorraum. Ein Unterraum $U \subseteq V$ heißt orthogonal zu einem Unterraum $W \subseteq V$, geschrieben $U \perp W$, wenn gilt: $u \perp w$ für alle $u \in U, w \in W$.

Das orthogonale Komplement eines Unterraumes $U \subseteq V$ ist definiert als:

$$U_{\perp} := \{x \in V : x \perp u \text{ für alle } u \in U\}.$$

Satz 2.15

Sei V ein euklidischer Vektorraum und sei U ein Unterraum von V . Dann ist U_{\perp} ein Unterraum von V .

Beweis:

$0 \in U_{\perp}$, da 0 zu allen Vektoren orthogonal ist. Seien $u_1, u_2 \in U_{\perp}$ und u beliebig in U .

Dann gilt

$$\begin{aligned} \langle u_1 + u_2, u \rangle &= \langle u_1, u \rangle + \langle u_2, u \rangle \\ &= 0 + 0 = 0 \end{aligned}$$

und

$$\langle \lambda u_1, u \rangle = \lambda \langle u_1, u \rangle = \lambda \cdot 0 = 0$$

d.h. $u_1 + u_2$ und λu_1 sind in U_{\perp} enthalten. U_{\perp} ist also ein Unterraum.

□

3

Der Rang einer Matrix

In diesem Kapitel befassen wir uns mit einer wichtigen Kennzahl von Matrizen, dem sogenannten Rang einer Matrix. Der Rang einer Matrix ist eng verknüpft mit den Begriffen der Basis und Dimension von Vektorräumen, die in Kapitel 2.4 behandelt wurden. Wir beginnen mit der Definition und wichtigen Eigenschaften des Rangs einer Matrix (Abschnitt 3.1). Im darauf folgenden Abschnitt 3.2 beschäftigen wir uns mit der sogenannten Inverse einer Matrix. Es existiert jedoch nicht zu jeder Matrix eine Inverse. Die Existenz einer Inverse ist eng mit dem Rang einer Matrix verknüpft. Abschnitt 3.3 beschreibt wie der Rang einer Matrix algorithmisch bestimmt bzw. berechnet werden kann. Schließlich behandelt Abschnitt 3.4 eine wichtige Zerlegung von Matrizen, die sogenannte Vollrangzerlegung.

3.1 Definition und Eigenschaften des Rangs

Definition 3.1 (Zeilenrang, Spaltenrang, Zeilenraum, Spaltenraum)

Sei \mathbf{A} eine $m \times n$ Matrix. Die Maximalzahl linear unabhängiger Spaltenvektoren des \mathbf{R}^m heißt Spaltenrang von \mathbf{A} , geschrieben $\text{rgs}(\mathbf{A})$. Der von den (linear unabhängigen) Spaltenvektoren aufgespannte Unterraum heißt Spaltenraum, geschrieben $S(\mathbf{A})$. Es gilt:

$$S(\mathbf{A}) = \left\{ z \in \mathbf{R}^m : z = \mathbf{A}x = \sum_{i=1}^n a_i x_i, x \in \mathbf{R}^n \right\}$$

Entsprechend kann man den Zeilenrang $\text{rgz}(\mathbf{A})$ von \mathbf{A} als die Maximalzahl linear unabhängiger Zeilen von \mathbf{A} definieren. Der von den (linear unabhängigen) Zeilen aufgespannte Unterraum $Z(\mathbf{A})$ heißt Zeilenraum. Es gilt:

$$Z(\mathbf{A}) = \left\{ z \in \mathbf{R}^n : z = \mathbf{A}'x = \sum_{i=1}^m (a^i)' x_i, x \in \mathbf{R}^m \right\}$$

Für den Spalten- und Zeilenrang gilt:

Satz 3.1

Spaltenrang und Zeilenrang einer $m \times n$ Matrix \mathbf{A} sind gleich, d.h.

$$rgs(\mathbf{A}) = rgz(\mathbf{A}).$$

Beweis:

Sei $rgz(\mathbf{A}) = z$. Dann bilden z linear unabhängige Zeilenvektoren eine Basis des Zeilenraumes. O.B.d.A. seien dies die ersten z . Nun lässt sich jeder Zeilenvektor a^i , $i = 1, \dots, m$, als Linearkombination der Basisvektoren darstellen:

$$a^i = (a_{i1}, \dots, a_{in}) = \sum_{j=1}^z b_{ij}(a_{j1}, \dots, a_{jn}) = \sum_{j=1}^z b_{ij}a^j$$

Für jedes Element a_{ik} von A gilt also

$$a_{ik} = \sum_{j=1}^z b_{ij}a_{jk}. \quad (3.1)$$

Definiert man die $m \times 1$ Vektoren $b_j := (b_{1j} \dots b_{mj})'$, $j = 1, \dots, z$, so lässt sich nun andererseits jeder Spaltenvektor a_k , $k = 1, \dots, n$, wegen (3.1) als Linearkombination der Vektoren b_1, \dots, b_z darstellen:

$$a_k = (a_{1k}, \dots, a_{mk})' = \sum_{j=1}^z a_{jk}(b_{1j}, \dots, b_{mj})' = \sum_{j=1}^z a_{jk}b_j$$

Daraus folgt, dass $rgs(\mathbf{A}) \leq rgz(\mathbf{A})$. Eine entsprechende Überlegung für \mathbf{A}' liefert $rgz(\mathbf{A}') \leq rgs(\mathbf{A}')$ und damit:

$$rgz(\mathbf{A}) = rgs(\mathbf{A}') \leq rgz(\mathbf{A}') = rgs(\mathbf{A}).$$

□

Damit ist folgende Definition gerechtfertigt:

Definition 3.2 (Rang einer Matrix)

Der Rang $rg(\mathbf{A})$ einer $m \times n$ Matrix \mathbf{A} ist definiert als die Dimension des Spalten- bzw. Zeilenraumes von \mathbf{A} :

$$rg(\mathbf{A}) := rgs(\mathbf{A}) = rgz(\mathbf{A}) \leq \min\{m, n\}$$

Für $rg(\mathbf{A}) = m$ ($rg(\mathbf{A}) = n$) heißt \mathbf{A} zeilenregulär (spaltenregulär).

Zur praktischen Bestimmung des Rangs einer Matrix vergleiche Kapitel 3.3. Im Folgenden beweisen wir zunächst einige allgemeine Rangbeziehungen:

Satz 3.2 (allgemeine Rangbeziehungen)

Für Matrizen $\mathbf{A}, \mathbf{B}, \mathbf{C}$ passender Ordnung gilt:

1. $rg(\mathbf{A}) = rg(-\mathbf{A})$
2. $rg(\mathbf{A}') = rg(\mathbf{A})$
3. $rg(\mathbf{A}) - rg(\mathbf{B}) \leq rg(\mathbf{A} + \mathbf{B}) \leq rg(\mathbf{A}) + rg(\mathbf{B})$
4. $rg(\mathbf{AB}) \leq \min \{rg(\mathbf{A}), rg(\mathbf{B})\}$
5. $rg(\mathbf{I}_n) = n$

Beweis:

Die Behauptungen 1) und 2) sind offensichtlich.

zu 3) : Ein Element z des Unterraumes $U := S(\mathbf{A}) + S(\mathbf{B})$ lässt sich darstellen als

$$z = k_1 \mathbf{A}x_1 + k_2 \mathbf{B}x_2$$

mit $k_1, k_2 \in \mathbb{R}$ und $x_1, x_2 \in \mathbb{R}^n$. Ein Element w des Spaltenraumes von $\mathbf{A} + \mathbf{B}$ lässt sich darstellen als

$$w = (\mathbf{A} + \mathbf{B})x = 1 \cdot \mathbf{A}x + 1 \cdot \mathbf{B}x,$$

so dass also alle Elemente w von $S(\mathbf{A} + \mathbf{B})$ in U enthalten sind, d.h es gilt $S(\mathbf{A} + \mathbf{B}) \subset S(\mathbf{A}) + S(\mathbf{B})$. Damit folgt unter Zuhilfenahme von Satz ??

$$\begin{aligned} rg(\mathbf{A} + \mathbf{B}) &= \dim(S(\mathbf{A} + \mathbf{B})) \leq \dim(S(\mathbf{A}) + S(\mathbf{B})) \\ &= \dim(S(\mathbf{A})) + \dim(S(\mathbf{B})) - \dim(S(\mathbf{A}) \cap S(\mathbf{B})) \\ &\leq \dim(S(\mathbf{A})) + \dim(S(\mathbf{B})) = rg(\mathbf{A}) + rg(\mathbf{B}), \end{aligned}$$

womit die zweite Ungleichung bewiesen ist. Weiter gilt unter Zuhilfenahme von 1) und des soeben bewiesenen

$$rg(\mathbf{A}) = rg(\mathbf{A} + \mathbf{B} - \mathbf{B}) \leq rg(\mathbf{A} + \mathbf{B}) + rg(-\mathbf{B}) = rg(\mathbf{A} + \mathbf{B}) + rg(\mathbf{B}).$$

Umstellen liefert die erste Ungleichung.

zu 4) : Schreibt man \mathbf{A} als Matrix der Spalten von \mathbf{A} , dann folgt

$$\mathbf{AB} = (b_{11}a_1 + \cdots + b_{n1}a_n, \dots, b_{1p}a_1 + \cdots + b_{np}a_n).$$

Die Spalten von \mathbf{AB} sind also Linearkombinationen der Spalten von \mathbf{A} und damit im Spaltenraum von \mathbf{A} enthalten. Damit besitzt \mathbf{AB} höchstens so viele linear unabhängige Spalten wie \mathbf{A} , es gilt also $rg(\mathbf{AB}) \leq rg(\mathbf{A})$. Analog zeigt man $rg(\mathbf{AB}) \leq rg(\mathbf{B})$, indem man \mathbf{B} als Matrix der Zeilen von \mathbf{B} schreibt.

zu 5): Die Spalten der Einheitsmatrix sind offensichtlich linear unabhängig, so dass unmittelbar die Behauptung folgt.

□

Definition 3.3 (Nullraum)

Der Nullraum $N(\mathbf{A})$ einer $m \times n$ Matrix \mathbf{A} ist definiert als die Menge

$$N(\mathbf{A}) := \{x \in \mathbf{R}^n : \mathbf{A}x = 0\}.$$

Satz 3.3 (Eigenschaften des Nullraums)

Sei \mathbf{A} eine $m \times n$ Matrix. Dann gilt:

1. Der Nullraum ist ein Unterraum des \mathbf{R}^n .
2. $rg(\mathbf{A}) + \dim(N(\mathbf{A})) = \dim(\mathbf{R}^n) = n$ bzw. $\dim(N(\mathbf{A})) = n - rg(\mathbf{A})$. Die Dimension des Nullraums $N(\mathbf{A})$ wird als Defekt von \mathbf{A} bezeichnet.
3. Der Nullraum $N(\mathbf{A})$ ist das orthogonale Komplement des Zeilenraums $Z(\mathbf{A})$ von \mathbf{A} .
4. $N(\mathbf{A}'\mathbf{A}) = N(\mathbf{A})$.

Beweis:

zu 1) : Offensichtlich gilt $0 \in N(\mathbf{A})$. Seien nun $x_1 \in N(\mathbf{A})$ und $x_2 \in N(\mathbf{A})$ und λ_1, λ_2 Skalare. Dann gilt

$$\mathbf{A}(\lambda_1 x_1 + \lambda_2 x_2) = \lambda_1 \underbrace{\mathbf{A}x_1}_{=0} + \lambda_2 \underbrace{\mathbf{A}x_2}_{=0} = 0,$$

d.h. $\lambda_1 x_1 + \lambda_2 x_2$ ist wieder ein Element von $N(\mathbf{A})$ und somit $N(\mathbf{A})$ ein Unterraum.

zu 2) : Wir definieren die lineare Abbildung $F(x) = \mathbf{A}x$. Offensichtlich ist $N(\mathbf{A})$ der Kern und der Spaltenraum $S(\mathbf{A})$ von \mathbf{A} das Bild von F . Damit folgt unter Zuhilfenahme von Satz 2.12 2) die Behauptung:

$$\dim(\ker(F)) = \dim(N(\mathbf{A})) = \dim(\mathbf{R}^n) - \dim(\text{bild}(F)) = n - rg(\mathbf{A})$$

zu 3): Seien $z \in Z(\mathbf{A})$ und $x \in N(\mathbf{A})$. Dann existiert ein Vektor $y \in \mathbf{R}^m$ mit $z = \mathbf{A}'y$ und es folgt unter Zuhilfenahme von $\mathbf{A}x = 0$

$$x'z = x'\mathbf{A}'y = (\mathbf{A}x)'y = 0,$$

d.h. x und z sind orthogonal, woraus die Behauptung folgt.

zu 4) Sei zunächst $x \in N(\mathbf{A})$. Dann folgt wegen $\mathbf{A}x = 0$ auch $\mathbf{A}'\mathbf{A}x = 0$ und damit $N(\mathbf{A}) \subset N(\mathbf{A}'\mathbf{A})$. Sei nun umgekehrt $x \in N(\mathbf{A}'\mathbf{A})$. Mit $y := \mathbf{A}x \in \mathbb{R}^n$ folgt unter Zuhilfenahme von $\mathbf{A}'\mathbf{A}x = 0$

$$0 = x'\mathbf{A}'\mathbf{A}x = y'y = \sum_{i=1}^n y_i^2,$$

woraus für $i = 1, \dots, n$, $y_i = 0$ und damit $\mathbf{A}x = 0$ folgt. Es gilt also $N(\mathbf{A}'\mathbf{A}) \subset N(\mathbf{A})$. Zusammen mit $N(\mathbf{A}) \subset N(\mathbf{A}'\mathbf{A})$ folgt daraus die Behauptung. □

3.2 Inverse einer Matrix

In diesem Abschnitt befassen wir uns mit der sogenannten Inverse einer quadratischen $n \times n$ Matrix. Es existiert jedoch nicht zu jeder Matrix eine Inverse. Entscheidend für die Existenz der Inverse ist der Rang einer Matrix. Hat eine quadratische Matrix maximalen Rang, so existiert auch die Inverse, andernfalls nicht. Im Falle ihrer Existenz ist die Inverse einer Matrix aber eindeutig bestimmt.

Definition 3.4 (inverse Matrix)

Sei \mathbf{A} eine quadratische Matrix. Die Matrix \mathbf{A}^{-1} heißt Inverse zur Matrix \mathbf{A} , falls gilt:

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

Für die Inverse gilt folgende Existenz und Eindeutigkeitsaussage:

Satz 3.4

Die Inverse einer quadratischen $n \times n$ Matrix \mathbf{A} existiert genau dann, wenn $\text{rg}(\mathbf{A}) = n$ gilt. Sie ist dann eindeutig bestimmt. Eine Matrix, deren Inverse existiert heißt auch regulär.

Beweis:

Existiert die Inverse \mathbf{A}^{-1} , dann gilt

$$n = \text{rg}(\mathbf{I}) = \text{rg}(\mathbf{A}\mathbf{A}^{-1}) \leq \min \{ \text{rg}(\mathbf{A}), \text{rg}(\mathbf{A}^{-1}) \} \leq n,$$

woraus $\text{rg}(\mathbf{A}) = n$ folgt. Als Nebenprodukt erhält man $\text{rg}(\mathbf{A}^{-1}) = n$.

Sei nun $\text{rg}(\mathbf{A}) = n$. Dann bilden die Spalten von \mathbf{A} eine Basis des \mathbb{R}^n und jeder Vektor $z \in \mathbb{R}^n$ lässt sich eindeutig als Linearkombination der Spaltenvektoren darstellen, d.h.

$z = \mathbf{A}x$, $x \in \mathbb{R}^n$. Wählt man speziell für z die Einheitsvektoren e_i , so gilt $e_i = \mathbf{A}x_i$, $i = 1, \dots, n$. In Matrixnotation lässt sich dies mit $\mathbf{X} := (x_1 \dots x_n)$ schreiben als $\mathbf{A}\mathbf{X} = (e_1 \dots e_n) = \mathbf{I}$. Da die Zeilen von \mathbf{A} die Spalten von \mathbf{A}' sind, können die e_i ebenso eindeutig als Linearkombination der Spalten von \mathbf{A}' dargestellt werden, d.h. $\mathbf{A}'y_i = e_i$, $y \in \mathbb{R}^n$. Mit $\mathbf{Y} := (y_1 \dots y_n)$ ergibt sich in Matrixnotation $\mathbf{A}'\mathbf{Y} = \mathbf{Y}'\mathbf{A} = \mathbf{I}$. Nun erhält man

$$\mathbf{Y}' = \mathbf{Y}'\mathbf{I} = \mathbf{Y}'\mathbf{A}\mathbf{X} = \mathbf{I}\mathbf{X} = \mathbf{X}$$

und folglich $\mathbf{A}\mathbf{X} = \mathbf{X}\mathbf{A} = \mathbf{I}$, so dass \mathbf{X} eine Inverse von \mathbf{A} darstellt.

Eindeutigkeit: Seien \mathbf{B} und \mathbf{C} Inversen von \mathbf{A} . Dann gilt $\mathbf{A}\mathbf{C} = \mathbf{I}$. Multiplikation mit \mathbf{B} von links liefert $\mathbf{B}\mathbf{A}\mathbf{C} = \mathbf{B}\mathbf{I} = \mathbf{B}$, woraus wegen $\mathbf{B}\mathbf{A} = \mathbf{I}$ $\mathbf{C} = \mathbf{B}$ folgt.

□

Beispiel 3.1

Betrachte die Matrix

$$\mathbf{A} = \begin{pmatrix} 2 & 3 & 1 \\ 1 & 0 & 1 \\ 3 & 5 & 1 \end{pmatrix}.$$

Durch Multiplikation verifiziert man leicht, dass

$$\mathbf{A}^{-1} = \begin{pmatrix} -5 & 2 & 3 \\ 2 & -1 & -1 \\ 5 & -1 & -3 \end{pmatrix}.$$

die Inverse zur Matrix \mathbf{A} ist.

△

Eine allgemeine praktische Berechnungsmöglichkeit für die Inverse einer Matrix werden wir in Kapitel 5.2 behandeln.

Satz 3.5 (Rechenregeln für Inverse)

Von den folgenden Matrizen wird angenommen, dass deren Inverse jeweils existiere. Dann gilt:

1. $(\mathbf{A}^{-1})^{-1} = \mathbf{A}$
2. $(k\mathbf{A})^{-1} = k^{-1}\mathbf{A}^{-1} = \frac{1}{k}\mathbf{A}^{-1}$
3. $(\mathbf{A}')^{-1} = (\mathbf{A}^{-1})'$
4. $(\mathbf{A}\mathbf{B})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$

$$5. (\mathbf{ABC})^{-1} = \mathbf{C}^{-1}\mathbf{B}^{-1}\mathbf{A}^{-1}$$

6. \mathbf{A} symmetrisch $\implies \mathbf{A}^{-1}$ ist auch symmetrisch.

7. Sei $\mathbf{A} = \text{diag}(a_1, \dots, a_n)$ eine Diagonalmatrix. Dann gilt $\mathbf{A}^{-1} = \text{diag}(a_1^{-1}, \dots, a_n^{-1})$.

8. Falls \mathbf{A} orthogonal, gilt $\mathbf{A} = \mathbf{A}^{-1} = \mathbf{I}$.

Beweis:

Die Aussagen ergeben sich unmittelbar durch Anwendung der Definition der Inversen. □

Mit Hilfe der Matrixmultiplikation und der Inversen einer Matrix kann der empirische Erwartungswert und die empirische Streuung in Matrixnotation geschrieben werden. Sei $x = (x_1, \dots, x_n)'$. Dann gilt:

$$1. \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{n} \mathbf{1}'x = (\mathbf{1}'\mathbf{1})^{-1} \mathbf{1}'x$$

$$2. \sum_{i=1}^n (x_i - \bar{x})^2 = x'x - (\mathbf{1}'\mathbf{1})^{-1} (\mathbf{1}'x)^2.$$

3.3 Praktische Bestimmung des Rangs einer Matrix

Bei der praktischen Bestimmung des Rangs einer Matrix spielt die Reduktion einer Matrix auf Dreiecksform (vgl. Definition 1.13) die entscheidende Rolle. Es zeigt sich nämlich, dass die Reduktion auf Dreiecksform durch elementare Zeilen- und Spaltenoperationen den Rang einer Matrix unverändert läßt (Satz 3.6 1). Andererseits kann man den Rang einer Matrix in Dreiecksform leicht ablesen, nämlich als die Anzahl der von Null verschiedenen Zeilen (Satz 3.6 2). Zur Bestimmung des Rangs einer Matrix reduzieren wir diese also auf Dreiecksform und lesen anschließend den Rang ab.

Zur Vorbereitung des entscheidenden Satzes 3.6 befassen wir uns zunächst mit der genaueren Charakterisierung von elementaren Zeilen- und Spaltenoperationen mit Hilfe sogenannter Elementarmatrizen. Es zeigt sich nämlich, dass elementare Matrixoperationen formal durch die Rechts- bzw. Linksmultiplikation mit eben diesen Elementarmatrizen durchgeführt werden können. Mit Hilfe der Elementarmatrix \mathbf{E}_{ij} können Zeilenvertauschungen vorgenommen werden. Sie entsteht aus der Einheitsmatrix \mathbf{I} durch Vertauschen der i -ten und j -ten Zeile. Die Matrix \mathbf{E}_{ij} besitzt also folgende Gestalt:

1. $\mathbf{E}'_{ij} = \mathbf{E}_{ij}$ und $\mathbf{E}^{-1}_{ij} = \mathbf{E}_{ij}$.
2. $\mathbf{R}_{ii}(\lambda)' = \mathbf{R}_{ii}(\lambda)$. und $\mathbf{R}_{ii}(\lambda)^{-1} = \mathbf{R}_{ii}(\frac{1}{\lambda})$
3. $\mathbf{P}_{ij}(\lambda)' = \mathbf{P}_{ji}(\lambda)$ und $\mathbf{P}_{ij}(\lambda)^{-1} = \mathbf{P}_{ij}(-\lambda)$.

Damit bewirkt Rechtsmultiplikation der $n \times n$ Matrizen \mathbf{E}_{ij} bzw. $\mathbf{R}_{ii}(\lambda)$ die entsprechende Matrixoperation für die Spalten von \mathbf{A} . Eine Addition des λ -fachen der i -ten Spalte zur j -ten Spalte erhält man durch Rechtsmultiplikation der $n \times n$ Matrix $\mathbf{P}_{ji}(\lambda)$.

Neben der Charakterisierung von elementaren Zeilen- und Spaltenoperationen durch Elementarmatrizen liefert der folgende Satz die Grundlage zur Berechnung des Rangs einer Matrix.

Satz 3.6

1. Die Multiplikation einer Matrix \mathbf{A} mit einer regulären Matrix \mathbf{T} ändert nicht den Rang, d.h. $rg(\mathbf{A}) = rg(\mathbf{TA})$.
2. Der Rang einer Matrix in Dreieckform ist gleich der Anzahl der von Null verschiedenen Zeilen.

Beweis:

zu 1) : Nach Satz 3.2.4 gilt $rg(\mathbf{TA}) \leq rg(\mathbf{A})$. Andererseits gilt wegen $\mathbf{A} = \mathbf{T}^{-1}\mathbf{TA}$ auch $rg(\mathbf{A}) = rg(\mathbf{T}^{-1}(\mathbf{TA})) \leq rg(\mathbf{TA})$, so dass $rg(\mathbf{A}) = rg(\mathbf{TA})$ folgt.

zu 2) : Es ist zu zeigen, dass die von Null verschiedenen Zeilenvektoren a^r, \dots, a^1 linear unabhängig sind. Angenommen die Vektoren seien linear abhängig. Dann ist nach Satz 2.4 einer der Vektoren, z. B. a^i , $i \leq r$, eine Linearkombination der vorherigen, d.h.

$$a^i = b_{i+1}a^{i+1} + \dots + b_r a^r.$$

Da \mathbf{A} eine Matrix in Dreiecksform ist, sind jeweils die i -ten Komponenten von a^{i+1}, \dots, a^r Null. Dies bedeutet aber auch, dass dann die i -te Komponente von a^i Null ist, was aber im Widerspruch zur Dreiecksgestalt von \mathbf{A} steht. Folglich sind a^r, \dots, a^1 linear unabhängig und damit $rg(\mathbf{A}) = r$.

□

Die Aussage 1) des Satzes bedeutet insbesondere, dass elementare Matrixoperationen den Rang einer Matrix unverändert lassen, da elementare Matrixoperationen (formal) durch Multiplikation mit regulären (Elementar-) Matrizen durchgeführt werden. Damit haben wir gezeigt, dass der Rang einer Matrix bestimmt werden kann, indem diese zunächst auf

Dreiecksform reduziert wird. Die Anzahl der von Null verschiedenen Zeilen ergeben dann gemäß der zweiten Aussage des Satzes den Rang der Matrix.

Beispiel 3.2

Wir betrachten die Matrix

$$\mathbf{A} = \begin{pmatrix} 2 & 3 & 1 \\ 1 & 1 & 1 \\ 3 & 5 & 1 \end{pmatrix}.$$

In Beispiel 1.18 wurde \mathbf{A} auf Dreiecksform reduziert, wobei wir folgende Matrix erhalten haben:

$$\tilde{\mathbf{A}} = \begin{pmatrix} 2 & 3 & 1 \\ 0 & -\frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 0 \end{pmatrix}.$$

Damit besitzt \mathbf{A} den Rang 2, da 2 Zeilen der Matrix in Dreiecksform von Null verschieden sind. \mathbf{A} ist also keine reguläre Matrix.

△

3.4 Vollrang Zerlegung einer Matrix

Bei der Herleitung des zentralen Satzes 3.8 (Vollrangzerlegung einer Matrix) spielt die in Kapitel 1.4 behandelte Reduzierung auf Diagonalform eine entscheidende Rolle.

Sind für eine Reduktion einer Matrix \mathbf{A} auf Diagonalform \mathbf{D} insgesamt l Zeilenoperationen und k Spaltenoperationen notwendig und bezeichnet die Matrix \mathbf{B}_i eine elementare Zeilenoperation und \mathbf{C}_i eine elementare Spaltenoperation, so entsteht die Matrix \mathbf{D} aus \mathbf{A} durch Links- bzw. Rechtsmultiplikation mit den Matrizen B_i bzw. C_i . Das heißt, \mathbf{D} lässt sich darstellen als

$$\mathbf{D} = \begin{pmatrix} \mathbf{D}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} = \mathbf{B}_l \dots \mathbf{B}_1 \mathbf{A} \mathbf{C}_1 \dots \mathbf{C}_k.$$

Mit den regulären Matrizen $\mathbf{P} := \mathbf{B}_l \dots \mathbf{B}_1$ und $\mathbf{Q} := \mathbf{C}_1 \dots \mathbf{C}_k$ erhält man

$$\mathbf{D} = \begin{pmatrix} \mathbf{D}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} = \mathbf{P} \mathbf{A} \mathbf{Q} \tag{3.2}$$

bzw.

$$\mathbf{A} = \mathbf{P}^{-1} \mathbf{D} \mathbf{Q}^{-1} \tag{3.3}$$

wobei $rg(\mathbf{D}) = rg(\mathbf{A}) = r$ ist. Dies liefert folgenden Satz:

Satz 3.7

Zu jeder von Null verschiedenen $m \times n$ Matrix \mathbf{A} mit $\text{rg}(\mathbf{A}) = r$ existieren reguläre Matrizen \mathbf{P} und \mathbf{Q} , so dass gilt:

$$\mathbf{PAQ} = \begin{pmatrix} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}$$

Beweis:

Ausgehend von (3.2), müssen nur noch die ersten r Zeilen mit dem Reziproken des Diagonalelements multiplizieren werden um zu der gewünschten Darstellung zu gelangen. Bei der Multiplikation handelt es sich um eine elementare Matrixoperation, die durch Multiplikation der entsprechenden Matrizen \mathbf{B}_i zur Matrix \mathbf{P} in (3.2) hinzugefügt werden.

□

Beispiel 3.3

Für die Matrix \mathbf{A} aus Beispiel 1.18 und 3.2 ergeben sich \mathbf{P} und \mathbf{Q} zu:

$$\begin{aligned} \mathbf{P} &= \mathbf{P}_{23}(1)\mathbf{P}_{13}\left(-\frac{3}{2}\right)\mathbf{P}_{12}\left(-\frac{1}{2}\right) \\ &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -\frac{3}{2} & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ -\frac{1}{2} & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0 & 0 \\ -\frac{1}{2} & 1 & 0 \\ -2 & 1 & 1 \end{pmatrix} \end{aligned}$$

$$\begin{aligned} \mathbf{Q} &= \mathbf{P}_{21}\left(-\frac{3}{2}\right)\mathbf{P}_{31}\left(-\frac{1}{2}\right)\mathbf{P}_{32}(1) \\ &= \begin{pmatrix} 1 & -\frac{3}{2} & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & -\frac{1}{2} \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 1 & -\frac{3}{2} & -2 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix} \end{aligned}$$

Damit gilt

$$\mathbf{PAQ} = \begin{pmatrix} 2 & 0 & 0 \\ 0 & -0.5 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Zusätzliche Multiplikation der beiden Matrizen

$$\mathbf{R}_{11}\left(\frac{1}{2}\right) = \begin{pmatrix} \frac{1}{2} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

und

$$\mathbf{R}_{22}(-2) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

liefert

$$\mathbf{P} = \mathbf{R}_{22}(-2) \cdot \mathbf{R}_{11}\left(\frac{1}{2}\right) \cdot \mathbf{P}_{23}(1) \mathbf{P}_{13}\left(-\frac{3}{2}\right) \mathbf{P}_{12}\left(-\frac{1}{2}\right) = \begin{pmatrix} \frac{1}{2} & 0 & 0 \\ 1 & -2 & 0 \\ -2 & 1 & 1 \end{pmatrix}.$$

Damit erhalten wir schließlich die Zerlegung

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} = \mathbf{P}\mathbf{A}\mathbf{Q} = \begin{pmatrix} \frac{1}{2} & 0 & 0 \\ 1 & -2 & 0 \\ -2 & 1 & 1 \end{pmatrix} \begin{pmatrix} 2 & 3 & 1 \\ 1 & 1 & 1 \\ 3 & 5 & 1 \end{pmatrix} \begin{pmatrix} 1 & -\frac{3}{2} & -2 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$$

bzw.

$$\mathbf{A} = \mathbf{P}^{-1} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \mathbf{Q}^{-1} = \begin{pmatrix} 2 & 0 & 0 \\ 1 & -\frac{1}{2} & 0 \\ 3 & \frac{1}{2} & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & \frac{3}{2} & \frac{1}{2} \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{pmatrix}.$$

△

Bemerkung:

Handelt es sich bei \mathbf{A} um eine reguläre $n \times n$ Matrix, dann gilt sogar

$$\mathbf{P}\mathbf{A}\mathbf{Q} = \mathbf{I}$$

bzw.

$$\mathbf{A} = \mathbf{P}^{-1}\mathbf{I}\mathbf{Q}^{-1} = \mathbf{P}^{-1}\mathbf{Q}^{-1}.$$

Eine reguläre Matrix \mathbf{A} lässt sich also immer als Produkt von Elementarmatrizen schreiben. Diese Tatsache erweist sich bei Beweisen häufig als nützlich (vergleiche zum Beispiel den Beweis zu Satz 4.5).

▽

Schließlich erhält man folgenden Satz:

Satz 3.8 (Vollrang Zerlegung)

Jede $m \times n$ Matrix \mathbf{A} mit $\mathbf{A} \neq \mathbf{0}$ und $\text{rg}(\mathbf{A}) = r$ lässt sich darstellen als Produkt einer spaltenregulären $m \times r$ Matrix \mathbf{K} und einer zeilenregulären $r \times n$ Matrix \mathbf{L} :

$$\mathbf{A} = \mathbf{KL}$$

Beweis:

Aufgrund von Satz 5.2 existieren reguläre Matrizen \mathbf{P} ($m \times m$) und \mathbf{Q} ($n \times n$), so dass

$$\mathbf{A} = \mathbf{P}^{-1} \begin{pmatrix} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \mathbf{Q}^{-1}.$$

Mit Hilfe der Partitionierung von \mathbf{P}^{-1} in eine $m \times r$ Matrix \mathbf{K} und eine $m \times (m-r)$ Matrix \mathbf{W}

$$\mathbf{P}^{-1} = \begin{pmatrix} \mathbf{K} & \mathbf{W} \end{pmatrix}$$

und der Partitionierung von \mathbf{Q}^{-1} in eine $r \times n$ Matrix \mathbf{L} und eine $(n-r) \times n$ Matrix \mathbf{Z}

$$\mathbf{Q}^{-1} = \begin{pmatrix} \mathbf{L} \\ \mathbf{Z} \end{pmatrix}$$

erhält man

$$\mathbf{A} = \begin{pmatrix} \mathbf{K} & \mathbf{W} \end{pmatrix} \begin{pmatrix} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{L} \\ \mathbf{Z} \end{pmatrix} = \begin{pmatrix} \mathbf{K} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{L} \\ \mathbf{Z} \end{pmatrix} = \mathbf{KL}$$

Da \mathbf{P}^{-1} regulär ist, sind die Spalten von \mathbf{P}^{-1} linear unabhängig und damit erst recht die Spalten von \mathbf{K} , so dass \mathbf{K} spaltenregulär ist. Genauso ist wegen der Regularität von \mathbf{Q}^{-1} \mathbf{L} zeilenregulär.

□

4

Determinante und Spur

In diesem Kapitel behandeln wir zusätzlich zum Rang einer Matrix (vgl. das vorangegangene Kapitel) zwei weitere Kennzahlen von Matrizen, nämlich die Determinante und die Spur von quadratischen Matrizen. Die Abschnitte 4.1 und 4.2 beschäftigen sich mit der Determinante und der Abschnitt 4.3 mit der Spur von Matrizen.

4.1 Permutationen

Dieser Abschnitt über Permutationen dient als Vorbereitung zur Definition der Determinante einer Matrix im nächsten Abschnitt.

Definition 4.1 (Permutation)

Eine Permutation ist eine bijektive Abbildung σ der Menge $\{1, \dots, n\}$ auf sich selbst. Man bezeichnet die Permutation σ mit $\sigma = j_1 j_2 \dots j_n$, wobei $j_i = \sigma(i)$. Wegen der Eineindeutigkeit ist die Folge j_1, \dots, j_n also einfach eine Umordnung der Zahlen $1, 2, \dots, n$. Die Menge aller Permutationen wird mit S_n bezeichnet. Mit Hilfe von Regeln der Kombinatorik ergibt sich die Anzahl der möglichen Permutationen zu $n!$.

Definition 4.2 (Signum einer Permutation σ)

Das Signum $\text{sign}(\sigma)$ einer Permutation ist $+1$ (-1), wenn j_1, \dots, j_n durch eine gerade (ungerade) Anzahl von Vertauschungen benachbarter Ziffern in die natürliche Reihenfolge $1, 2, \dots, n$ gebracht werden kann. Man sagt dann auch σ sei gerade (ungerade).

Beispiel 4.1

1. Sei $\epsilon = 1, 2, \dots, n$ die identische Abbildung. Es sind keine Vertauschungen benachbarter Ziffern nötig, da sich die Ziffern bereits in ihrer natürlichen Reihenfolge befinden. ϵ ist also gerade, d.h. $\text{sign}(\epsilon) = 1$.
2. Sei τ die Permutation, die lediglich zwei Zahlen vertauscht und alle anderen Zahlen in ihrer natürlichen Reihenfolge behält:

$$\tau(i) = j \quad \tau(j) = i \quad \tau(k) = k \quad k \neq i, j$$

Man nennt τ eine Transposition. Durch die Vertauschung der Ziffern i und j kann die natürliche Reihenfolge wiederhergestellt werden. τ ist also ungerade, d.h. $\text{sign}(\tau) = -1$.

△

Wir benötigen folgenden Satz:

Satz 4.1

Sei $\sigma = j_1 \dots j_n$ eine Permutation und $\sigma^{-1} = k_1 \dots k_n$ die dazugehörige Umkehrabbildung. Dann gilt $\text{sign}(\sigma^{-1}) = \text{sign}(\sigma)$ und für beliebige Skalare a_{ij} , $i, j = 1, \dots, n$,

$$a_{j_1 1} a_{j_2 2} \dots a_{j_n n} = a_{1 k_1} a_{2 k_2} \dots a_{n k_n}.$$

Beweis:

Sei $\epsilon = 1, 2, \dots, n$ die identische Abbildung. Man beachte, dass $\epsilon = \sigma \circ \sigma^{-1}$. Da ϵ gerade ist, sind σ und σ^{-1} entweder gerade oder ungerade, so dass $\text{sgn}(\sigma^{-1}) = \text{sgn}(\sigma)$. Da σ eine Permutation ist, gilt

$$a_{j_1 1} a_{j_2 2} \dots a_{j_n n} = a_{1 k_1} a_{2 k_2} \dots a_{n k_n}$$

und folglich für die Zahlen $k_1 \dots k_n$:

$$\sigma(k_1) = 1, \quad \sigma(k_2) = 2, \dots, \sigma(k_n) = n$$

Sei nun $\delta = k_1 \dots k_n$. Dann gilt für $i = 1, \dots, n$

$$(\sigma \circ \delta)(i) = \sigma(\delta(i)) = \sigma(k_i) = i$$

und somit $\sigma \delta = \epsilon$, so dass schließlich folgt: $\delta = \sigma^{-1}$.

□

4.2 Determinante einer Matrix

Nach den Vorbemerkungen über Permutationen können wir jetzt die Determinante einer Matrix definieren.

Definition 4.3 (Determinante)

Jeder quadratischen Matrix \mathbf{A} ist eine reelle Zahl zugeordnet, die als Determinante von \mathbf{A} bezeichnet wird:

$$\det(\mathbf{A}) := \sum_{\sigma \in S_n} \text{sign}(\sigma) a_{1j_1} a_{2j_2} \cdots a_{nj_n}$$

Dabei ist $\sigma = j_1 \dots j_n$ eine Permutation der Zahlen $1, 2, \dots, n$.

Für Dimensionen $n \leq 3$ lässt sich die Determinante leicht ausrechnen wie folgendes Beispiel zeigt:

Beispiel 4.2

1. Für eine 2×2 Matrix gilt $\det(\mathbf{A}) = a_{11}a_{22} - a_{12}a_{21}$.
2. Für eine 3×3 Matrix gilt $\det(\mathbf{A}) = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31} - a_{23}a_{32}a_{11} - a_{33}a_{12}a_{21}$.

△

Ohne Beweis zeigen wir, dass sich die Determinante einer Matrix \mathbf{A} geometrisch interpretieren lässt. Wir veranschaulichen die geometrische Interpretation anhand der Determinante der 2×2 Matrix

$$\mathbf{A} = \begin{pmatrix} 4 & 2 \\ 1 & 3 \end{pmatrix}.$$

Die beiden Spaltenvektoren $a_1 = (4, 1)'$ und $a_2 = (2, 3)'$ der Matrix sind als Ortsvektoren in Abbildung 4.1 abgebildet. Die Determinante von \mathbf{A} ist gegeben durch

$$\det(\mathbf{A}) = 4 \cdot 3 - 2 \cdot 1.$$

Die Determinante von \mathbf{A} ist also gleich dem *Flächeninhalt* des von den beiden Spaltenvektoren gebildeten Parallelogramms. Diese Interpretation einer Determinante ist allgemeingültig. Bei 3×3 Matrizen handelt es sich bei der Determinante von \mathbf{A} um das Volumen des von den drei Spaltenvektoren aufgespannten Körpers. Für $n > 3$ ergeben sich analoge Interpretationen.

Im Folgenden wollen wir einige wichtige Eigenschaften von Determinanten zusammentragen. Wir beginnen mit der Determinante der transponierten Matrix \mathbf{A}' einer Matrix \mathbf{A} .

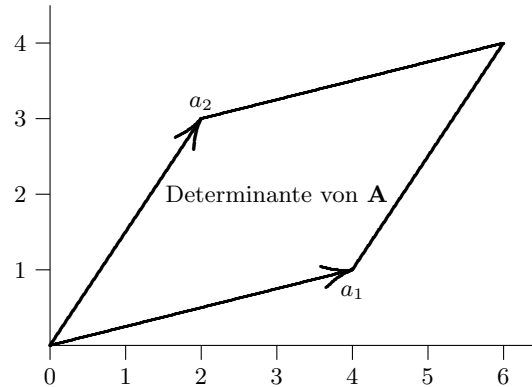


Abbildung 4.1. Geometrische Veranschaulichung der Determinante einer 2×2 Matrix.

Satz 4.2 (Determinante der Transponierten einer Matrix)

Für eine quadratische Matrix \mathbf{A} gilt $\det(\mathbf{A}') = \det(\mathbf{A})$

Beweis:

Sei $\mathbf{A} = (a_{ij})$ und $\mathbf{A}' = (b_{ij}) = (a_{ji})$. Dann gilt

$$\det(\mathbf{A}') = \sum_{\sigma \in S_n} \text{sign}(\sigma) b_{1j_1} b_{2j_2} \cdots b_{nj_n} = \sum_{\sigma \in S_n} \text{sign}(\sigma) a_{j_1 1} a_{j_2 2} \cdots a_{j_n n},$$

wobei $\sigma = j_1 \cdots j_n$ gilt. Sei nun $\sigma^{-1} = k_1 \cdots k_n$. Wegen Satz 4.1 gilt $\text{sign}(\sigma) = \text{sign}(\sigma^{-1})$ und $a_{j_1 1} \cdots a_{j_n n} = a_{1k_1} \cdots a_{nk_n}$ und damit

$$\det(\mathbf{A}') = \sum_{\sigma \in S_n} \text{sign}(\sigma^{-1}) a_{1k_1} a_{2k_2} \cdots a_{nk_n}.$$

Da σ durch alle Elemente von S_n geht, läuft auch σ^{-1} durch alle Elemente von S_n , woraus die Behauptung folgt. □

Aufgrund des Satzes müssen zukünftig Sätze über die Determinante, die sowohl Spalten als auch Zeilen einer Matrix betreffen, nur entweder für die Spalte oder die Zeile bewiesen werden.

Für einige spezielle Matrizen lässt sich die Determinante sofort angeben:

Satz 4.3 (Determinante einiger bestimmter Matrizen)

Sei \mathbf{A} eine quadratische Matrix. Dann gilt:

1. Wenn eine Zeile (Spalte) von \mathbf{A} aus Nullen besteht, dann gilt $\det(\mathbf{A}) = 0$.
2. Wenn \mathbf{A} zwei identische Zeilen (Spalten) besitzt, dann gilt $\det(\mathbf{A}) = 0$.
3. Die Determinante einer Matrix in Dreiecksform ist das Produkt der Diagonalelemente.

4. $\det(\mathbf{I}) = 1$

Der folgende Satz zeigt die Auswirkung elementarer Matrixoperationen auf die Determinante:

Satz 4.4

Sei \mathbf{B} die Matrix, die man aus der $n \times n$ Matrix \mathbf{A} erhält, wenn man

1. eine Zeile (Spalte) von \mathbf{A} mit λ multipliziert. Dann gilt $\det(\mathbf{B}) = \lambda \det(\mathbf{A})$.
2. zwei Zeilen (Spalten) von \mathbf{A} vertauscht. Dann gilt $\det(\mathbf{B}) = -\det(\mathbf{A})$.
3. das λ -fache der i -ten Zeile (Spalte) zur j -ten Zeile (Spalte) addiert. Dann gilt $\det(\mathbf{B}) = \det(\mathbf{A})$.

Beweis der Sätze 4.3 und 4.4:

zu 1) Satz 4.3: Jeder Summand in $\det(\mathbf{A})$ enthält einen Faktor aus jeder Zeile von \mathbf{A} . Somit ist jeder Summand in $\det(\mathbf{A})$ Null und folglich $\det(\mathbf{A}) = 0$.

zu 2) Satz 4.4: Beweis der Aussage für den Fall, dass zwei Spalten vertauscht werden. Sei τ die Transposition, welche die zwei Zahlen vertauscht, die zu den zwei Spalten von \mathbf{A} gehört, die vertauscht werden. Für jedes Element b_{ij} von B gilt $b_{ij} = a_{i\tau(j)}$ und es folgt für $\sigma = j_1 \cdots j_n$

$$b_{1j_1} \cdots b_{nj_n} = a_{1\tau(j_1)} \cdots a_{n\tau(j_n)}.$$

Somit gilt

$$\begin{aligned} \det(B) &= \sum_{\sigma \in S_n} \text{sign}(\sigma) b_{1j_1} \cdots b_{nj_n} \\ &= \sum_{\sigma \in S_n} \text{sign}(\sigma) a_{1\tau(j_1)} \cdots a_{n\tau(j_n)} \end{aligned}$$

Da τ ungerade ist gilt $\text{sign}(\tau\sigma) = \text{sign}(\tau)\text{sign}(\sigma) = -\text{sign}(\sigma)$ und damit

$$\det(B) = - \sum_{\sigma \in S_n} \text{sign}(\tau\sigma) a_{1\tau(j_1)} \cdots a_{n\tau(j_n)}$$

Da σ alle Permutationen in S_n durchläuft, durchläuft auch $\tau\sigma$ alle Permutationen in S_n , so dass schließlich die Behauptung folgt.

zu 2) Satz 4.3: Vertauscht man die zwei identischen Zeilen, so erhält man wieder die Matrix \mathbf{A} . Folglich gilt nach Satz 4.4.2 $\det(\mathbf{A}) = -\det(\mathbf{A})$, woraus $\det(\mathbf{A}) = 0$ folgt.

zu 3) Satz 4.3: Wegen der Dreiecksform von \mathbf{A} wird in $t = \text{sign}(\sigma) a_{1j_1} \cdots a_{nj_n}$ der Faktor a_{ij_i} immer Null, wenn $j_i < i$. Für j_n muss also $j_n = n$ gelten, damit $t \neq 0$, für j_{n-1} muss

dann $j_{n-1} = n - 1$ gelten, damit $t \neq 0$, usw.... Das heißt nur für $\sigma = 1, 2, \dots, n$ ist $t \neq 0$. Da außerdem $\text{sign}(1, 2, \dots, n) = 1$ gilt, folgt $\det(\mathbf{A}) = a_{11} \dots a_{nn}$.

zu 4) **Satz 4.3:** Die Behauptung folgt aus Satz 4.3 3).

zu 1) **Satz 4.4:** Multipliziert man Zeile i mit λ , so folgt:

$$\begin{aligned} \det(\mathbf{B}) &= \sum_{\sigma \in S_n} \text{sign}(\sigma) a_{1j_1} \cdots \lambda a_{ij_i} \cdots a_{nj_n} \\ &= \lambda \sum_{\sigma \in S_n} \text{sign}(\sigma) a_{1j_1} \cdots a_{ij_i} \cdots a_{nj_n} = \lambda \det(\mathbf{A}). \end{aligned}$$

zu 3) **Satz 4.4:** Unter Verwendung des Symbols $\hat{\cdot}$, um die j -te Position im Ausdruck der Determinante anzuzeigen, gilt:

$$\begin{aligned} \det(\mathbf{B}) &= \sum_{\sigma \in S_n} \text{sign}(\sigma) a_{1j_1} \cdots (a_{jj_j} + \lambda a_{ij_j}) \cdots a_{nj_n} \\ &= \sum_{\sigma \in S_n} \text{sign}(\sigma) a_{1j_1} \cdots \hat{a}_{jj_j} \cdots a_{nj_n} + \\ &\quad \lambda \sum_{\sigma \in S_n} \text{sign}(\sigma) a_{1j_1} \cdots \hat{a}_{ij_j} \cdots a_{nj_n} \end{aligned}$$

Die zweite Summe ist die Determinante einer Matrix, deren i -te und j -te Zeile gleich ist. Nach Satz 4.4.2 ist diese aber Null, so dass die Behauptung folgt.

□

Aufgrund von Satz 4.4 können nun auch die Determinanten der Elementarmatrizen \mathbf{E}_{ij} , $\mathbf{R}_{ii}(\lambda)$ und $\mathbf{P}_{ij}(\lambda)$ (siehe Kapitel 3.3) angegeben werden:

1. $\det(\mathbf{E}_{ij}) = -1 \det(\mathbf{I}) = -1$
2. $\det(\mathbf{R}_{ii}(\lambda)) = \lambda$
3. $\det(\mathbf{P}_{ij}(\lambda)) = 1$, da $\mathbf{P}_{ij}(\lambda)$ oder $\mathbf{P}_{ij}(\lambda)'$ eine Dreiecksmatrix ist, deren Diagonalelemente sämtlich gleich eins sind.

Bezeichne nun \mathbf{B} eine der drei Matrizen \mathbf{E}_{ij} , $\mathbf{R}_{ii}(\lambda)$, $\mathbf{P}_{ij}(\lambda)$, dann kann man leicht durch nachrechnen zeigen, dass

$$\det(\mathbf{BA}) = \det(\mathbf{B})\det(\mathbf{A}) \quad \text{bzw.} \quad \det(\mathbf{AB}) = \det(\mathbf{A})\det(\mathbf{B}) \quad (4.1)$$

gilt. Diese Tatsache wird im Beweis von Satz 4.5, der weitere Eigenschaften von Determinanten beinhaltet, benötigt.

Berechnung der Determinante:

Satz 4.4 liefert auch eine Berechnungsmöglichkeit der Determinante einer Matrix \mathbf{A} .

Durch Zeilen- bzw. Spaltenvertauschungen und Addition des λ -fachen der i -ten Zeile zur j -ten Zeile kann die Matrix \mathbf{A} zunächst auf Dreiecksform gebracht werden. Bezeichnet man die Matrix in Dreiecksform mit \mathbf{B} , so stimmen aufgrund von Satz 4.4.2 und 4.4.3 die Determinanten beider Matrizen bis auf das Vorzeichen überein. Bezeichne s die Anzahl der Zeilen und Spaltenvertauschungen, die nötig sind um \mathbf{A} auf Dreiecksgestalt zu bringen und seien b_{11}, \dots, b_{nn} die Diagonalelemente von \mathbf{B} , dann ist die Determinante von \mathbf{A} gegeben durch:

$$\det(\mathbf{A}) = (-1)^s b_{11} \cdots b_{nn}$$

Beispiel 4.3

Betrachte die Matrix

$$\mathbf{A} = \begin{pmatrix} 6 & -4 & -10 & 4 \\ -5 & 2 & 8 & -5 \\ -2 & 4 & 7 & -3 \\ 2 & -3 & -5 & 8 \end{pmatrix}.$$

Wir bringen die Matrix durch folgende elementare Zeilen- und Spaltenoperationen auf Dreiecksform:

- Addiere das $\frac{5}{6}$ fache der 1. Zeile zur 2. Zeile
- Addiere das $\frac{2}{6} = \frac{1}{3}$ fache der 1. Zeile zur 3. Zeile
- Addiere das $-\frac{2}{6} = -\frac{1}{3}$ fache der 1. Zeile zur 4. Zeile
- Addiere das $\frac{8}{3} \cdot \frac{3}{4} = 2$ fache der 2. Zeile zur 3. Zeile
- Addiere das $\frac{5}{3} \cdot \frac{3}{8} = \frac{5}{8}$ der 2. Zeile zur 4. Zeile
- Addiere das $-\frac{15}{24 \cdot 3} = -\frac{5}{24}$ fache der 3. Zeile zur 4. Zeile

Wir erhalten die Matrix:

$$\tilde{\mathbf{A}} = \begin{pmatrix} 6 & -4 & -10 & 4 \\ 0 & -\frac{4}{3} & -\frac{1}{3} & -\frac{5}{3} \\ 0 & 0 & 3 & -5 \\ 0 & 0 & 0 & \frac{160}{24} \end{pmatrix}$$

Da keine Zeilen und Spaltenvertauschungen notwendig waren, um \mathbf{A} auf Dreiecksform zu bringen, folgt

$$\det(\mathbf{A}) = \det(\tilde{\mathbf{A}}) = 6 \cdot \left(-\frac{4}{3}\right) \cdot 3 \cdot \frac{160}{24} = -160.$$

△

Es folgen noch einige wichtige Eigenschaften von Determinanten:

Satz 4.5 (Eigenschaften von Determinanten)

Für die Determinante einer $n \times n$ Matrix \mathbf{A} gilt:

1. $\det(k\mathbf{A}) = k^n \det(\mathbf{A})$
2. $\det(\mathbf{A}) \neq 0 \iff \text{rg}(\mathbf{A}) = n$
3. $\det(\mathbf{AB}) = \det(\mathbf{A})\det(\mathbf{B})$
4. $\det(\mathbf{A}^{-1}) = \frac{1}{\det(\mathbf{A})}$
5. \mathbf{A} orthogonal $\implies \det(\mathbf{A}) = \pm 1$

Beweis:

zu 1) : Die Behauptung folgt unmittelbar aus Satz 4.4.1.

zu 2) : Sei \mathbf{B} die zu \mathbf{A} gehörige Matrix in Dreiecksform. Die Determinanten beider Matrizen stimmen bis auf das Vorzeichen überein. Ist \mathbf{A} regulär d.h. gilt $\text{rg}(\mathbf{A}) = n$, so sind sämtliche Zeilen von \mathbf{B} nach Satz 3.6.2 von Null verschieden und folglich wegen der Dreiecksgestalt von \mathbf{B} alle Diagonalelemente. Da die Determinante von \mathbf{B} nach Satz 4.3.3 das Produkt der Diagonalelemente ist, gilt $\det(\mathbf{B}) \neq 0$ und folglich auch $\det(\mathbf{A}) \neq 0$.

Sei nun umgekehrt $\det(\mathbf{A}) \neq 0$. Dann ist auch $\det(\mathbf{B}) \neq 0$. Folglich sind alle Diagonalelemente von \mathbf{B} ungleich Null und demzufolge auch alle Zeilen von \mathbf{B} von Null verschieden, \mathbf{A} ist also regulär.

zu 3): Ist \mathbf{A} singulär, dann ist wegen $\text{rg}(\mathbf{AB}) \leq \text{rg}(\mathbf{A})$ auch \mathbf{AB} singulär und es gilt wegen 2)

$$\det(\mathbf{AB}) = 0 = \det(\mathbf{A})\det(\mathbf{B}).$$

Ist \mathbf{A} regulär, dann ist \mathbf{A} darstellbar als Produkt von Elementarmatrizen $\mathbf{C}_1, \dots, \mathbf{C}_r$ (vgl. hierzu die Bemerkung in Kapitel 3.4) und es gilt:

$$\det(\mathbf{A}) = \det(\mathbf{C}_1 \cdots \mathbf{C}_r) = \det(\mathbf{C}_1) \cdots \det(\mathbf{C}_r)$$

Nun folgt unter Zuhilfenahme von (4.1):

$$\det(\mathbf{AB}) = \det(\mathbf{C}_1 \cdots \mathbf{C}_r \mathbf{B}) = \det(\mathbf{C}_1) \cdots \det(\mathbf{C}_r) \det(\mathbf{B}) = \det(\mathbf{A}) \det(\mathbf{B})$$

zu 4) : Es gilt $\det(\mathbf{A})\det(\mathbf{A}^{-1}) = \det(\mathbf{AA}^{-1}) = \det(\mathbf{I}) = 1$. Umstellen liefert die Behauptung.

zu 5) : Es gilt $1 = \det(\mathbf{I}) = \det(\mathbf{AA}') = \det(\mathbf{A})\det(\mathbf{A}') = \det(\mathbf{A})^2$, woraus die Behauptung folgt.

□

Definition 4.4 (Minoren und Kofaktoren)

Sei \mathbf{A} eine quadratische $n \times n$ Matrix und sei \mathbf{M}_{ij} die Teilmatrix von \mathbf{A} , die man durch Streichen der i -ten Zeile und der j -ten Spalte erhält. Die Determinante von \mathbf{M}_{ij} heißt Minor des Elements a_{ij} von \mathbf{A} . Der Kofaktor \mathbf{A}_{ij} von a_{ij} ist definiert als

$$\mathbf{A}_{ij} := (-1)^{i+j} \det(\mathbf{M}_{ij})$$

Der folgende Satz liefert eine weitere Berechnungsmöglichkeit für die Determinante:

Satz 4.6

Sei \mathbf{A} eine $n \times n$ Matrix. Dann lässt sich die Determinante von \mathbf{A} berechnen als die Summe der Produkte, die man erhält, wenn man die Elemente einer beliebigen Zeile i (Spalte j) mit ihren Kofaktoren multipliziert:

$$\det(\mathbf{A}) = a_{i1}\mathbf{A}_{i1} + a_{i2}\mathbf{A}_{i2} + \dots + a_{in}\mathbf{A}_{in} = a_{1j}\mathbf{A}_{1j} + a_{2j}\mathbf{A}_{2j} + \dots + a_{nj}\mathbf{A}_{nj}$$

Beweis:

Jeder Summand $\text{sign}(\sigma)a_{1j_1} \cdot a_{2j_2} \cdots a_{nj_n}$ in $\det(\mathbf{A})$ enthält genau ein Element der i -ten Zeile (a_{i1}, \dots, a_{in}) von \mathbf{A} . $\det(\mathbf{A})$ lässt sich daher in der Form

$$\det(\mathbf{A}) = a_{i1}\mathbf{A}_{i1}^* + \dots + a_{in}\mathbf{A}_{in}^*$$

schreiben. Die Terme \mathbf{A}_{ij}^* sind dabei jeweils Summen von Ausdrücken, die kein Element der i -ten Zeile von \mathbf{A} enthalten. Wir zeigen, dass $\mathbf{A}_{ij}^* = (-1)^{i+j} \det(\mathbf{M}_{ij})$ gilt.

Sei zunächst $i = j = n$. Dann gilt

$$a_{nn} \cdot \mathbf{A}_{nn}^* = a_{nn} \sum_{\sigma} \text{sign}(\sigma) a_{1\sigma(1)} \cdots a_{n-1\sigma(n-1)},$$

wobei über alle $\sigma \in S_n$ summiert wird, für die $\sigma(n) = n$ gilt. Da dies gleichwertig mit der Summation über alle Permutationen von $1, \dots, n-1$ ist, folgt:

$$\mathbf{A}_{nn}^* = \det(\mathbf{M}_{nn}) = (-1)^{n+n} \det(\mathbf{M}_{nn}).$$

Der Fall, dass i und j beliebig sind, führen wir durch Zeilen- und Spaltenvertauschungen auf obigen Fall zurück. Wir vertauschen die i -te Zeile mit jeder folgenden bis zur letzten, genauso die j -te Spalte mit jeder folgenden bis zur letzten. Dadurch bleibt $\det(\mathbf{M}_{ij})$ unbeeinflusst. Lediglich das Vorzeichen von $\det(\mathbf{A})$ verändert sich durch die Zeilen- und Spaltenvertauschungen $n-i$ und $n-j$ mal. Also folgt

$$\mathbf{A}_{ij}^* = (-1)^{n-i+n-j} \det(\mathbf{M}_{ij}) = (-1)^{i+j} \det(\mathbf{M}_{ij}).$$

Beispiel 4.4

Betrachte die Matrix

$$\mathbf{A} = \begin{pmatrix} 6 & -4 & -10 & 4 \\ -5 & 2 & 8 & -5 \\ 0 & 1 & 0 & 0 \\ 2 & -3 & -5 & 8 \end{pmatrix}.$$

Da in der 3. Zeile 3 mal die Null steht und einmal die Eins, entwickeln wir zweckmäßigerweise nach der 3. Zeile. Wir erhalten

$$\begin{aligned} \det(A) &= a_{31} \cdot \mathbf{A}_{31} + a_{32} \cdot \mathbf{A}_{32} + a_{33} \cdot \mathbf{A}_{33} + a_{34} \cdot \mathbf{A}_{34} \\ &= a_{32} \cdot \mathbf{A}_{32} \\ &= (-1)^{3+2} \det(\mathbf{M}_{32}) \\ &= -1 \cdot \det \begin{pmatrix} 6 & -10 & 4 \\ -5 & 8 & -5 \\ 2 & -5 & 8 \end{pmatrix} \\ &= (-1)(6 \cdot 8 \cdot 8 - 10(-5) \cdot 2 + 4(-5)(-5) - 2 \cdot 8 \cdot 4 + 5(-5)6 - 8(-5)(-10)) \\ &= (-1)(384 + 100 + 100 - 64 - 150 - 400) \\ &= (-1)(-30) = 30. \end{aligned}$$

Dabei wurde die explizite Form der Determinante einer 3×3 Matrix aus Beispiel 4.2 benutzt.

△

Beispiel 4.5

Wir betrachten wieder die Matrix

$$\mathbf{A} = \begin{pmatrix} 6 & -4 & -10 & 4 \\ -5 & 2 & 8 & -5 \\ -2 & 4 & 7 & -3 \\ 2 & -3 & -5 & 8 \end{pmatrix}$$

aus Beispiel 4.3. Aus Beispiel 4.4 wissen wir, dass die Berechnung der Determinante gemäss Zeilen- oder Spaltenentwicklung genau dann besonders einfach ist, wenn eine Zeile oder Spalte aus genau einem Element ungleich Null besteht. Wir führen zunächst die folgenden elementaren Zeilen- und Spaltenoperationen durch:

- Addiere das $\frac{5}{6}$ fache der 1. Zeile zur 2. Zeile
- Addiere das $\frac{2}{6} = \frac{1}{3}$ fache der 1. Zeile zur 3. Zeile

– Addiere das $-\frac{2}{6} = -\frac{1}{3}$ fache der 1. Zeile zur 4. Zeile

Damit erhalten wir die Matrix

$$\tilde{\mathbf{A}} = \begin{pmatrix} 6 & -4 & -10 & 4 \\ 0 & -\frac{4}{3} & -\frac{1}{3} & -\frac{5}{3} \\ 0 & \frac{8}{3} & \frac{11}{3} & -\frac{5}{3} \\ 0 & -\frac{5}{3} & -\frac{5}{3} & \frac{20}{3} \end{pmatrix}.$$

Wir entwickeln nach der 1.Spalte und erhalten

$$\begin{aligned} \det(\mathbf{A}) &= \det(\tilde{\mathbf{A}}) \\ &= 6 \cdot (-1)^{1+1} \det(\mathbf{M}_{11}) \\ &= 6 \cdot \det \begin{pmatrix} -\frac{4}{3} & -\frac{1}{3} & -\frac{5}{3} \\ \frac{8}{3} & \frac{11}{3} & -\frac{5}{3} \\ -\frac{5}{3} & -\frac{5}{3} & \frac{20}{3} \end{pmatrix} \\ &= 6 \cdot (-26\frac{2}{3}) = -160. \end{aligned}$$

△

4.3 Die Spur einer Matrix

Definition 4.5 (Spur einer Matrix)

Sei $\mathbf{A} = (a_{ij})$ eine quadratische $n \times n$ Matrix. Dann heißt die Summe der Diagonalelemente Spur von \mathbf{A} , d.h.

$$sp(\mathbf{A}) = \sum_{i=1}^n a_{ii}.$$

Satz 4.7 (Eigenschaften der Spur)

Für die Spur der $n \times n$ Matrizen \mathbf{A}, \mathbf{B} gilt:

1. $sp(\mathbf{A} + \mathbf{B}) = sp(\mathbf{A}) + sp(\mathbf{B})$
2. $sp(\mathbf{A}) = sp(\mathbf{A}')$
3. $sp(k\mathbf{A}) = k \cdot sp(\mathbf{A})$
4. $sp(\mathbf{AB}) = sp(\mathbf{BA})$. Dies bleibt auch für den Fall gültig, dass \mathbf{A} eine $m \times n$ und \mathbf{B} eine $n \times m$ Matrix ist.
5. Seien $x, y \in \mathbb{R}^n$. Dann gilt $sp(xy') = sp(yx') = sp(x'y) = x'y$

Beweis:

Die Aussagen 1) - 3) folgen unmittelbar aus der Definition der Spur einer Matrix.

zu 4) : Es gilt

$$\mathbf{AB} = (c_{ik}) = \left(\sum_{j=1}^n a_{ij} b_{jk} \right).$$

und

$$\mathbf{BA} = (d_{ik}) = \left(\sum_{r=1}^m b_{ir} a_{rk} \right).$$

Nun folgt:

$$\begin{aligned} sp(\mathbf{AB}) &= \sum_{r=1}^m c_{rr} = \sum_{r=1}^m \sum_{j=1}^n a_{rj} b_{jr} = \sum_{j=1}^n \sum_{r=1}^m b_{jr} a_{rj} \\ &= \sum_{j=1}^n d_{jj} = sp(\mathbf{BA}) \end{aligned}$$

zu 5) : Die Behauptung folgt aus 2) und 4).

□

Beispiel 4.6

Wir betrachten wieder die Matrix

$$\mathbf{A} = \begin{pmatrix} 6 & -4 & -10 & 4 \\ -5 & 2 & 8 & -5 \\ -2 & 4 & 7 & -3 \\ 2 & -3 & -5 & 8 \end{pmatrix}$$

aus den Beispielen 4.3 und 4.5. Als Spur von \mathbf{A} erhalten wir

$$sp(\mathbf{A}) = 6 + 2 + 7 + 8 = 23.$$

△

5

Lineare Gleichungssysteme

Mit linearen Gleichungssystemen und deren Lösung haben wir uns bereits in Abschnitt 1.4 befasst. In diesem Kapitel gehen wir darauf genauer ein. Wir wiederholen zunächst die Definition eines linearen Gleichungssystems und befassen uns mit der allgemeinen Struktur der Lösungen (Abschnitt 5.1). Im darauf folgenden Abschnitt 5.2 beschäftigen wir uns mit der Lösung linearer Gleichungssysteme durch Dreieckszerlegung der entsprechenden Koeffizientenmatrix. Eine alternative Lösungstechnik für spezielle Gleichungssysteme (mit positiv definiter Koeffizientenmatrix) werden wir in Kapitel 7.2 kennenlernen. Der letzte Abschnitt dieses Kapitels behandelt verallgemeinerte Inversen von Matrizen und stellt den Zusammenhang mit der Lösung linearer Gleichungssysteme her.

5.1 Definition und allgemeine Eigenschaften

Definition 5.1 (Lineares Gleichungssystem)

Unter einem linearen Gleichungssystem mit Unbekannten $x_1, \dots, x_n \in \mathbf{R}$ versteht man ein System von Gleichungen der Form

$$a_{i1}x_1 + a_{i2}x_2 + \dots + a_{in}x_n = c_i \quad i = 1, \dots, m,$$

wobei die Skalare $a_{ij}, c_i \in \mathbf{R}$ bekannte Koeffizienten sind. Fasst man die Skalare a_{ij} zur $m \times n$ Matrix \mathbf{A} und x_i und c_i zu den $n \times 1$ bzw. $m \times 1$ Spaltenvektoren x und c zusammen so lässt sich ein lineares Gleichungssystem durch

$$\mathbf{A}x = c$$

in Matrixnotation schreiben. Für $c = 0$, heißt das Gleichungssystem homogen, andernfalls inhomogen. Für ein inhomogenes Gleichungssystem heißt $\mathbf{A}x = 0$ das zu $\mathbf{A}x = c$ gehörende homogene System. Ein lösbares lineares Gleichungssystem heißt konsistent, andernfalls inkonsistent. Offensichtlich ist ein homogenes Gleichungssystem stets konsistent.

Bemerkung:

Bei der Lösungsmenge L_0 eines homogenen linearen Gleichungssystems handelt es sich um

den Nullraum von \mathbf{A} (vergleiche Definition 3.3). Aufgrund von Satz 3.3 wissen wir bereits, dass L_0 ein Unterraum des \mathbb{R}^n ist mit

$$\dim(L_0) = \dim(N(\mathbf{A})) = n - \text{rg}(\mathbf{A}).$$

▽

Der folgende Satz liefert ein Kriterium für die Lösbarkeit linearer Gleichungssysteme:

Satz 5.1 (Kriterium für die Lösbarkeit)

Das Gleichungssystem $\mathbf{A}x = c$ ist genau dann lösbar, wenn $\text{rg}(\begin{pmatrix} \mathbf{A} \\ c \end{pmatrix}) = \text{rg}(\mathbf{A})$. Die Matrix $\begin{pmatrix} \mathbf{A} \\ c \end{pmatrix}$ heißt erweiterte Koeffizientenmatrix.

Beweis:

Der Vektor $\mathbf{A}x = a_1x_1 + \dots + a_nx_n$ ist eine Linearkombination der Spalten von \mathbf{A} , d.h. $\mathbf{A}x = c$ ist genau dann lösbar, wenn c im Spaltenraum von \mathbf{A} enthalten ist. Daraus folgt aber, dass $\text{rg}(\begin{pmatrix} \mathbf{A} \\ c \end{pmatrix}) = \text{rg}(\mathbf{A})$ gelten muss.

□

Bei der Bestimmung der Lösungsmenge L des Gleichungssystems $\mathbf{A}x = c$ kann man sich im wesentlichen auf die Bestimmung der Lösungsmenge des zugehörigen homogenen Systems $\mathbf{A}x = 0$ beschränken. Die Rechtfertigung dafür liefert folgender Satz:

Satz 5.2 (Lösungsstruktur linearer Gleichungssysteme)

Sei $\mathbf{A}x = c$ ein lineares Gleichungssystem. Dann ist die Lösungsmenge L des Gleichungssystems gegeben durch

$$L = x_0 + L_0 = \{x_0 + x : x \in L_0\},$$

wobei x_0 eine partikuläre Lösung des Gleichungssystems ist und L_0 die Lösungsmenge des dazugehörigen homogenen Gleichungssystems.

Beweis:

Sei $x \in L_0$ eine Lösung des homogenen Systems. Dann gilt für $x_0 + x$

$$\mathbf{A}(x_0 + x) = \mathbf{A}x_0 + \mathbf{A}x = c + 0 = c,$$

d.h. $x_0 + x$ ist eine Lösung des inhomogenen Systems. Es gilt also $x_0 + L_0 \subset L$.

Sei nun y eine beliebige Lösung des inhomogenen Systems. Es gilt $y = x_0 + (y - x_0)$ und

$$\mathbf{A}(y - x_0) = \mathbf{A}y - \mathbf{A}x_0 = c - c = 0,$$

d.h. $y - x_0$ ist ein Element von L_0 und damit $x_0 + y - x_0 = y$ ein Element aus $x_0 + L_0$,

d.h. $L \subset x_0 + L_0$. Mit $x_0 + L_0 \subset L$ folgt $x_0 + L_0 = L$.

□

5.2 Lösen von linearen Gleichungssystemen

Um die Lösungsmenge eines allgemeinen linearen Gleichungssystems bestimmen zu können, muss nach Satz 5.2 zunächst die Lösungsmenge des homogenen Systems bestimmt werden. Folgender Satz liefert die Grundlage für einen Algorithmus:

Satz 5.3

Sei $\mathbf{A}x = 0$ ein homogenes lineares Gleichungssystem und \mathbf{P} eine reguläre $m \times m$ Matrix. Dann haben $\mathbf{A}x = 0$ und $\mathbf{P}\mathbf{A}x = 0$ die gleiche Lösungsmenge.

Beweis:

Sei $\mathbf{A}x = 0$. Dann gilt auch $\mathbf{P}\mathbf{A}x = \mathbf{P}0 = 0$. Sei umgekehrt $\mathbf{P}\mathbf{A}x = 0$. Dann gilt $\mathbf{A}x = \mathbf{P}^{-1}\mathbf{P}\mathbf{A}x = \mathbf{P}^{-1}0 = 0$.

□

Die Aussage des Satzes gewährleistet insbesondere, dass elementare Zeilenoperationen die Lösungsmenge eines homogenen Gleichungssystems nicht ändern. Da auch Spaltenvertauschungen nur eine Umnummerierung der Unbekannten bewirken, gelangt man schließlich zu folgendem Algorithmus:

Algorithmus 5.1 (Bestimmung einer partikulären Lösung von $\mathbf{A}x = 0$)

Bezeichne $x = (x_1, \dots, x_n)'$ eine partikuläre Lösung des Gleichungssystems $\mathbf{A}x = 0$. Dann lässt sich x gemäß dem folgenden Algorithmus bestimmen:

1. Reduziere \mathbf{A} gemäß dem Algorithmus 1.1 aus Kapitel 1.4 auf Dreiecksgestalt und merke die dabei nötigen Spaltenvertauschungen. Die dabei entstehende Matrix \mathbf{B} hat folgende Gestalt:

$$\mathbf{B} = \begin{pmatrix} b_{11} & & & & & \\ 0 & b_{22} & & & & \\ \vdots & & \ddots & & & \\ 0 & 0 & \dots & b_{rr} & \dots & b_{rn} \\ \vdots & & & & \ddots & \vdots \\ 0 & \dots & & & \dots & 0 \end{pmatrix}$$

2. Die Unbekannten x_{r+1}, \dots, x_n sind frei wählbar, setze also dafür beliebige Werte ein.
3. Setze $i = r$
4. $x_i = -\frac{(b_{i,i+1}x_{i+1} + \dots + b_{in}x_n)}{b_{ii}}$
5. Setze $i = i - 1$. Für $i = 0$ fahre fort mit 6, ansonsten Rücksprung auf 4.

6. Mache bei der erhaltenen Lösung x die Spaltenvertauschungen wieder rückgängig.

Beispiel 5.1

Betrachte das lineare Gleichungssystem

$$\begin{pmatrix} 2 & 4 & 16 \\ 1 & -3 & -7 \\ -2 & 2 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

Wir reduzieren zunächst die Koeffizientenmatrix auf Dreiecksform. Durch Addition des $-\frac{1}{2}$ -fachen der 1. Zeile zur 2. Zeile und des $-\frac{-2}{2} = 1$ -fachen der 1. Zeile zur 3. Zeile erhalten wir das äquivalente System

$$\begin{pmatrix} 2 & 4 & 16 \\ 0 & -5 & -15 \\ 0 & 6 & 18 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

Schließlich erhalten wir durch Addition des $-\frac{6}{-5} = \frac{6}{5}$ -fachen der 2. Zeile zur 3. Zeile

$$\begin{pmatrix} 2 & 4 & 16 \\ 0 & -5 & -15 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

Damit können wir x_3 frei wählen. Wir setzen $x_3 = 1$. Für x_2 und x_3 erhalten wir

$$\begin{aligned} x_2 &= -\frac{15x_3}{5} = -\frac{15 \cdot 1}{5} = -3 \\ x_1 &= -\frac{4x_2 + 16x_3}{2} = -\frac{4 \cdot (-3) + 16 \cdot 1}{2} = -2 \end{aligned}$$

△

Folgender Algorithmus liefert eine Basis des Lösungsraumes eines homogenen linearen Gleichungssystems:

Algorithmus 5.2 (Basis des Lösungsraumes von $\mathbf{A}x = 0$)

Bezeichne x_1, \dots, x_{n-r} eine Basis des Lösungsraumes L_0 von $\mathbf{A}x = 0$. Diese lässt sich mit Hilfe des folgenden Algorithmus bestimmen:

1. Setze $i = r + 1$ und $j = 1$
2. Bestimme eine partikuläre Lösung x_j durch Anwendung des vorangegangenen Algorithmus. Setze für die frei wählbaren Unbekannten $x_{ji} = 1$ und $x_{j,i+1} = \dots = x_{j,n} = 0$.

3. Setze $i = i + 1$ und $j = j + 1$. Der Algorithmus ist beendet, wenn $i = n + 1$. Die Vektoren x_1, \dots, x_{n-r} sind dann eine Basis für L_0 . Ansonsten Rücksprung auf 2.

Bemerkung:

Es muss noch bewiesen werden, dass die Vektoren x_1, \dots, x_{n-r} tatsächlich eine Basis des Lösungsraumes darstellen. Betrachte dazu die Matrix

$$\mathbf{X} = (x'_1 \dots x'_{n-r}) = \begin{pmatrix} x_{11} & \cdots & x_{1r} & 1 & \cdots & 0 \\ \vdots & & \vdots & \vdots & \ddots & \vdots \\ x_{n-r,1} & \cdots & x_{n-r,r} & 0 & \cdots & 1 \end{pmatrix}$$

Es ist zu zeigen, dass die Zeilen von \mathbf{X} linear unabhängig sind, d.h. $rg(\mathbf{X}) = n - r$. Durch Spaltenvertauschungen kann man die ersten r Spalten ans Ende der Matrix bringen, so dass eine Dreiecksmatrix $\tilde{\mathbf{X}}$ entsteht, die den gleichen Rang wie \mathbf{X} hat (da Spaltenvertauschungen den Rang nicht ändern, vgl. Abschnitt 3.3):

$$\tilde{\mathbf{X}} = \begin{pmatrix} 1 & \cdots & 0 & x_{11} & \cdots & x_{1r} \\ & \ddots & & \vdots & & \vdots \\ 0 & \cdots & 1 & x_{n-r,1} & \cdots & x_{n-r,r} \end{pmatrix}$$

Da der Rang einer Dreiecksmatrix gleich der Anzahl der von Null verschiedenen Zeilen ist, folgt $rg(\tilde{\mathbf{X}}) = n - r = rg(\mathbf{X})$ und damit die Behauptung.

△

Beispiel 5.2

Wir betrachten wieder das Gleichungssystem aus Beispiel 5.1. Offensichtlich gilt $\dim(L_0) = 1$, so dass die gefundene Lösung $x = (-2, -3, 1)'$ zugleich eine Basis des Lösungsraums L_0 ist.

△

Beispiel 5.3

Betrachte das Gleichungssystem

$$\begin{pmatrix} 4 & 2 & -4 & 6 \\ 3 & 2 & -1 & 2 \\ 1 & 1 & 1 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

Wir reduzieren die Koeffizientenmatrix durch folgende Operationen auf Dreiecksform:

– Addition des $-\frac{3}{4}$ fachen der 1. Zeile zur 2. Zeile

– Addition des $-\frac{1}{4}$ fachen der 1. Zeile zur 3. Zeile

– Addition des -1 fachen der 2. Zeile zur 3. Zeile

Wir erhalten das folgende äquivalente System:

$$\begin{pmatrix} 4 & 2 & -4 & 6 \\ 0 & \frac{1}{2} & 2 & -\frac{5}{2} \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

Offenbar gilt $\dim(L_0) = 2$. Zur Bestimmung einer Basis bestimmen wir zunächst eine Lösung x_1 , wobei wir (gemäß Algorithmus) für die beiden freien Komponenten $x_{13} = 1$ und $x_{14} = 0$ wählen. Für x_{12} und x_{11} erhalten wir

$$x_{12} = -2 \cdot (2x_{13} - \frac{5}{2}x_{14}) = -4,$$

$$x_{11} = -\frac{2x_{12} - 4x_{13} + 6x_{14}}{4} = 3.$$

Analog setzen wir für die freien Komponenten der zweiten Lösung x_2 $x_{23} = 0$ und $x_{24} = 1$ und erhalten weiter

$$x_{22} = -2 \cdot (2x_{23} - \frac{5}{2}x_{24}) = 5,$$

$$x_{21} = -\frac{2x_{22} - 4x_{23} + 6x_{24}}{4} = -4.$$

Damit bilden die Vektoren $x_1 = (3, -4, 1, 0)'$ und $x_2 = (-4, 5, 0, 1)'$ eine Basis des Lösungsraumes L_0 .

△

Nach der Lösung homogener linearer Gleichungssysteme wenden wir uns im Folgenden der Lösung inhomogener Systeme $\mathbf{A}x = c$ zu. Sei $(\mathbf{B} \ d)$ die Matrix, die man aus der erweiterten Matrix $(\mathbf{A} \ c)$ erhält, indem man \mathbf{A} durch Zeilenoperationen auf Dreiecksform reduziert und die dafür nötigen Operationen auch auf c anwendet. Dann lässt sich $(\mathbf{B} \ d)$ schreiben als $(\mathbf{B} \ d) = \mathbf{P}(\mathbf{A} \ c) = (\mathbf{PA} \ \mathbf{P}c)$, wobei \mathbf{P} eine Matrix elementarer Zeilenoperationen ist. Gelte nun $\mathbf{B}v = d$. Dann folgt:

$$\mathbf{A}v = \mathbf{P}^{-1}\mathbf{P}\mathbf{A}v = \mathbf{P}^{-1}\mathbf{B}v = \mathbf{P}^{-1}d = \mathbf{P}^{-1}\mathbf{P}c = c$$

Eine Lösung v des Systems $\mathbf{B}v = d$ ist also gleichzeitig auch eine Lösung von $\mathbf{A}x = c$, so dass man analog zur Bestimmung einer partikulären Lösung bei homogenen Gleichungssystemen eine Lösung eines inhomogenen Gleichungssystems durch Dreieckszerlegung von

($\mathbf{A} \ c$) erhält. Mit Hilfe des folgenden Algorithmus kann die Lösungsmenge eines inhomogenen Gleichungssystems bestimmt werden:

Algorithmus 5.3 (Bestimmung der Lösungsmenge von $\mathbf{A}x = c$)

Bezeichne x_0 eine partikuläre Lösung des Gleichungssystems $\mathbf{A}x = c$. Dann kann die Lösungsmenge L des Systems wie folgt bestimmt werden:

1. Reduziere \mathbf{A} auf Dreiecksgestalt und wende die dabei nötigen Operationen auch auf c an. Man erhält:

$$(\mathbf{B}, d) = \begin{pmatrix} b_{11} & & & & & & d_1 \\ 0 & b_{22} & & & & & d_2 \\ \vdots & & \ddots & & & & \vdots \\ 0 & 0 & \dots & b_{rr} & & b_{rn} & d_r \\ 0 & 0 & \dots & 0 & 0 & \dots & 0 & d_{r+1} \\ \vdots & & & & & \ddots & \vdots & \vdots \\ 0 & & \dots & & \dots & 0 & d_m \end{pmatrix}$$

Eventuell nötige Spaltenvertauschungen muss man sich merken.

2. Man erkennt, dass das Gleichungssystem unlösbar ist, d.h. $L = \emptyset$, wenn mindestens ein d_i , $i = r + 1, \dots, m$ ungleich Null ist. In diesem Fall ist der Algorithmus beendet. Ansonsten sind bei der Bestimmung einer partikulären Lösung die Unbekannten $x_{0,r+1}, \dots, x_{0n}$ frei wählbar, man setze etwa $x_{0,r+1} = \dots = x_{0n} = 0$.
3. Setze $i = r$.
4. $x_{0i} = \frac{d_i - (b_{i,i+1}x_{0,i+1} + \dots + b_{in}x_{0n})}{b_{ii}}$
5. Setze $i = i - 1$. Gilt $i = 0$, so ist eine partikuläre Lösung x_0 von $\mathbf{A}x = c$ bereits gefunden. Fahre in diesem Fall fort mit 6. Ansonsten Rücksprung auf 4.
6. Bestimme mit Hilfe von Algorithmus 5.2 die Lösungsmenge L_0 des homogenen Systems und setze $L = x_0 + L_0$.

Bemerkung:

Aus dem Algorithmus ist ersichtlich, dass das Gleichungssystem genau dann eindeutig lösbar ist, wenn $m = n$ und $r = n$ gilt, d.h. wenn \mathbf{A} eine reguläre $n \times n$ Matrix ist. In diesem Fall entfällt Schritt 6 und die Lösungsmenge ist $L = x_0$.

△

Beispiel 5.4

Betrachte das Gleichungssystem

$$\begin{pmatrix} 2 & 3 & -2 \\ 1 & -2 & 3 \\ 4 & -1 & 4 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 5 \\ 2 \\ 1 \end{pmatrix}.$$

Wir reduzieren zunächst die Koeffizientenmatrix auf Dreiecksform und wenden die dazu nötigen Operationen auch auf $b = (5, 2, 1)'$ an. Die Koeffizientenmatrix wird durch folgende Operationen auf Dreiecksform reduziert:

- Addition des $-\frac{1}{2}$ fachen der 1. Zeile zur 2. Zeile
- Addition des $-\frac{4}{2} = -2$ fachen der 1. Zeile zur 3. Zeile
- Addition des -2 fachen der 2. Zeile zur 3. Zeile

Wir erhalten das äquivalente Gleichungssystem

$$\begin{pmatrix} 2 & 3 & -2 \\ 0 & -\frac{7}{2} & 4 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 5 \\ -\frac{1}{2} \\ -8 \end{pmatrix}.$$

Da $d_3 = -8 \neq 0$, ist das Gleichungssystem inkonsistent.

△

Beispiel 5.5

Betrachte das lineare Gleichungssystem

$$\begin{pmatrix} 2 & 1 & -2 \\ 6 & 4 & 4 \\ 5 & 4 & 3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 10 \\ 2 \\ 4 \end{pmatrix}.$$

Zur Lösung des Gleichungssystems reduzieren wir wieder die Koeffizientenmatrix auf Dreiecksform und wenden die dazu nötigen Operationen auch auf $b = (10, 2, 4)'$ an. Wir erhalten das äquivalente System

$$\begin{pmatrix} 2 & 1 & -2 \\ 0 & 1 & 10 \\ 0 & 0 & -7 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 10 \\ -28 \\ \frac{42}{2} \end{pmatrix}.$$

Wir erkennen, dass das Gleichungssystem eindeutig lösbar ist und erhalten

$$\begin{aligned}x_3 &= -\frac{42}{2 \cdot 7} = -3, \\x_2 &= -28 + 10 \cdot 3 = 2, \\x_1 &= \frac{10 - 2 \cdot 3 - 2}{2} = 1.\end{aligned}$$

△

Beispiel 5.6

Betrachte das Gleichungssystem

$$\begin{pmatrix} 2 & 4 & -6 \\ 2 & -1 & 4 \\ 4 & 3 & -2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 12 \\ 2 \\ 14 \end{pmatrix}.$$

Reduktion auf Dreiecksform liefert das äquivalente System

$$\begin{pmatrix} 2 & 4 & -6 \\ 0 & -5 & 10 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 12 \\ -10 \\ 0 \end{pmatrix}.$$

Das Gleichungssystem besitzt also eine frei wählbare Variable x_3 und hat somit unendlich viele Lösungen. Zur Bestimmung einer partikulären Lösung setzen wir für die frei wählbare Variable $x_3 = 0$ und erhalten weiter

$$\begin{aligned}x_2 &= \frac{-10}{-5} = 2, \\x_1 &= \frac{12 - 4 \cdot 2}{2} = 2.\end{aligned}$$

Die Dimension des Lösungsraumes des zugehörigen homogenen Systems ist $\dim(L_0) = 1$. Gemäß Algorithmus 5.2 erhalten wir als Basis den Vektor $(-1, 2, 1)'$. Damit erhalten wir als Lösungsmenge des Gleichungssystems

$$L = \left\{ \begin{pmatrix} 2 \\ 2 \\ 0 \end{pmatrix} + \lambda \begin{pmatrix} -1 \\ 2 \\ 1 \end{pmatrix} : \lambda \in \mathbf{R} \right\}.$$

△

5.3 Verallgemeinerte Inverse

Lineare Gleichungssysteme lassen sich auch auf andere Weise lösen, indem man auf die Theorie der verallgemeinerten Inverse zurückgreift. Insbesondere ist damit die Lösungsmenge eines Gleichungssystems auf einfache Weise darstellbar.

Definition 5.2 (verallgemeinerte Inverse)

Sei \mathbf{A} eine beliebige $m \times n$ Matrix mit $m \leq n$. Dann heißt die $n \times m$ Matrix \mathbf{A}^- verallgemeinerte Inverse oder g -Inverse (generalized Inverse) von \mathbf{A} falls

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

gilt.

Satz 5.4 (Existenz der g -Inversen)

Zu jeder Matrix \mathbf{A} existiert eine verallgemeinerte Inverse, die aber im allgemeinen nicht eindeutig ist.

Beweis:

\mathbf{A} lässt sich gemäß Kapitel 3.4 schreiben als

$$\mathbf{A} = \mathbf{P}^{-1} \begin{pmatrix} \mathbf{D}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \mathbf{Q}^{-1} = \mathbf{P}^{-1} \mathbf{D} \mathbf{Q}^{-1}.$$

Dann kann man leicht nachrechnen, dass

$$\mathbf{A}^- = \mathbf{Q} \begin{pmatrix} \mathbf{D}_r^{-1} & \mathbf{X} \\ \mathbf{Y} & \mathbf{Z} \end{pmatrix} \mathbf{P},$$

wobei \mathbf{X} , \mathbf{Y} , \mathbf{Z} beliebige Matrizen passender Ordnung sind, eine g -Inverse zu \mathbf{A} ist. □

Damit liefert obiger Beweis auch eine Berechnungsmöglichkeit für die g -Inverse einer Matrix \mathbf{A} . Man bestimme durch elementare Matrixoperationen gemäß Kapitel 5 die Matrizen \mathbf{P} , \mathbf{Q} und \mathbf{D}_r und berechne damit die g -Inverse von \mathbf{A} . Durch spezielle Wahl von $\mathbf{X} = \mathbf{Y} = \mathbf{Z} = \mathbf{0}$ erhält man eine besonders einfach zu bestimmende g -Inverse:

$$\mathbf{A}^- = \mathbf{Q} \begin{pmatrix} \mathbf{D}_r^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \mathbf{P}$$

Beispiel 5.7

Wir betrachten die Matrix

$$\mathbf{A} = \begin{pmatrix} 2 & 3 & 1 \\ 1 & 1 & 1 \\ 3 & 5 & 1 \end{pmatrix}.$$

In Beispiel 3.3 haben wir die Zerlegung

$$\mathbf{P}\mathbf{A}\mathbf{Q} = \begin{pmatrix} 2 & 0 & 0 \\ 0 & -0.5 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

erhalten mit

$$\mathbf{P} \begin{pmatrix} 1 & 0 & 0 \\ -\frac{1}{2} & 1 & 0 \\ -2 & 1 & 1 \end{pmatrix} \quad \text{und} \quad \mathbf{Q} = \begin{pmatrix} 1 & -\frac{3}{2} & -2 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}.$$

Damit erhalten wir als g-Inverse die Matrix

$$\mathbf{A}^- = \begin{pmatrix} 1 & -\frac{3}{2} & -2 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{1}{2} & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ -\frac{1}{2} & 1 & 0 \\ -2 & 1 & 1 \end{pmatrix} = \begin{pmatrix} -1 & 3 & 0 \\ 1 & -2 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

△

Satz 5.5 (Eigenschaften der g-Inverse)

Sei \mathbf{A}^- eine g-Inverse der Matrix \mathbf{A} . Dann gilt:

1. $rg(\mathbf{A}) = rg(\mathbf{A}\mathbf{A}^-) = rg(\mathbf{A}^-\mathbf{A})$
2. $rg(\mathbf{A}) \leq rg(\mathbf{A}^-)$
3. \mathbf{A} regulär $\implies \mathbf{A}^- = \mathbf{A}^{-1}$
4. $\mathbf{A}^-\mathbf{A}$ und $\mathbf{A}\mathbf{A}^-$ sind idempotent.

Beweis:

zu 1) : Es gilt

$$rg(\mathbf{A}) = rg(\mathbf{A}\mathbf{A}^-\mathbf{A}) \leq rg(\mathbf{A}\mathbf{A}^-) \leq rg(\mathbf{A}),$$

so dass $rg(\mathbf{A}\mathbf{A}^-) = rg(\mathbf{A})$ folgt. Analog zeigt man $rg(\mathbf{A}^-\mathbf{A}) = rg(\mathbf{A})$.

zu 2) : Gleiche Argumentation wie bei 1).

zu 3) : Es gilt: $\mathbf{A}\mathbf{A}^-\mathbf{A} = \mathbf{A}$. Rechts- und Linksmultiplikation mit \mathbf{A}^{-1} ergibt

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

woraus die Behauptung folgt.

zu 4) : Die Aussage folgt unmittelbar aus der Definition der g-Inversen.

□

Der nachfolgende Satz liefert nun eine Verbindung zu linearen Gleichungssystemen und deren Lösungsmenge:

Satz 5.6 (Gleichungssysteme und verallgemeinerte Inversen)

Gegeben sei das Gleichungssystem $\mathbf{A}x = c$, wobei \mathbf{A} eine $m \times n$ Matrix und x ein $n \times 1$ Spaltenvektor sei. Dann gilt:

1. Das Gleichungssystem ist genau dann lösbar, wenn gilt: $\mathbf{A}\mathbf{A}^-c = c$
2. Im Falle der Lösbarkeit erhält man als Lösungsmenge des linearen Gleichungssystems

$$x = \mathbf{A}^-c + (\mathbf{I} - \mathbf{A}^-\mathbf{A})w,$$

mit $w \in \mathbb{R}^n$ beliebig.

Beweis:

zu 1) : Sei $\mathbf{A}x = c$ lösbar. Dann folgt

$$\mathbf{A}\mathbf{A}^-c = \mathbf{A}\mathbf{A}^-\mathbf{A}x = \mathbf{A}x = c.$$

Gilt umgekehrt $\mathbf{A}\mathbf{A}^-c = c$ so ist das Gleichungssystem lösbar, indem man zum Beispiel $x = \mathbf{A}^-c$ setzt, denn $\mathbf{A}x = \mathbf{A}\mathbf{A}^-c = c$.

zu 2) : Sei $x = \mathbf{A}^-c + (\mathbf{I} - \mathbf{A}^-\mathbf{A})w$. Dann gilt:

$$\mathbf{A}x = \mathbf{A}(\mathbf{A}^-c + (\mathbf{I} - \mathbf{A}^-\mathbf{A})w) = \mathbf{A}\mathbf{A}^-c + \mathbf{A}(\mathbf{I} - \mathbf{A}^-\mathbf{A})w = c + \underbrace{(\mathbf{A} - \mathbf{A}\mathbf{A}^-\mathbf{A})}_{=0}w = c$$

Es ist noch zu zeigen, dass jede Lösung x obige Form besitzt. Sei also x_0 eine Lösung des Gleichungssystems. Dann gilt:

$$x_0 = \mathbf{A}^-c + x_0 - \mathbf{A}^-c = \mathbf{A}^-c + x_0 - \mathbf{A}^-\mathbf{A}x_0 = \mathbf{A}^-c + (\mathbf{I} - \mathbf{A}^-\mathbf{A})x_0$$

□

Bemerkung:

Ist \mathbf{A} regulär, dann ist das Gleichungssystem eindeutig lösbar und es gilt unter Verwendung von Satz 5.5.3

$$x = \mathbf{A}^-c + (\mathbf{I} - \mathbf{A}^-\mathbf{A})w = \mathbf{A}^{-1}c + (\mathbf{I} - \mathbf{I})w = \mathbf{A}^{-1}c.$$

△

Eine spezielle g-Inverse, die sogenannte Moore–Penrose–Inverse, ist eindeutig bestimmt:

Definition 5.3 (Moore–Penrose–Inverse)

Eine Matrix \mathbf{A}^+ heißt Moore–Penrose–Inverse, wenn folgende Bedingungen erfüllt sind:

1. $\mathbf{AA}^+\mathbf{A} = \mathbf{A}$, d.h. \mathbf{A}^+ ist eine g -Inverse zu \mathbf{A} .
2. $\mathbf{A}^+\mathbf{AA}^+ = \mathbf{A}^+$, d.h. \mathbf{A} ist g -Inverse von \mathbf{A}^+ .
3. $(\mathbf{AA}^+)' = \mathbf{AA}^+$, d.h. \mathbf{AA}^+ ist symmetrisch.
4. $(\mathbf{A}^+\mathbf{A})' = \mathbf{A}^+\mathbf{A}$, d.h. $\mathbf{A}^+\mathbf{A}$ ist symmetrisch.

Satz 5.7 (Existenz und Eindeutigkeit)

Die Moore–Penrose–Inverse \mathbf{A}^+ einer Matrix \mathbf{A} existiert immer und ist eindeutig bestimmt. Sie ist zum Beispiel gegeben durch $\mathbf{A}^+ = \mathbf{L}'(\mathbf{K}'\mathbf{A}\mathbf{L}')^{-1}\mathbf{K}'$, wobei \mathbf{K} und \mathbf{L} aus der Vollrang Zerlegung von \mathbf{A} stammen.

Beweis:

Es ist zu zeigen, dass $\mathbf{A}^+ = \mathbf{L}'(\mathbf{K}'\mathbf{A}\mathbf{L}')^{-1}\mathbf{K}'$ die vier Eigenschaften der Moore–Penrose–Inversen erfüllt.

zu 1) : Unter Beachtung der Regularität von $\mathbf{K}'\mathbf{K}$ und $\mathbf{L}\mathbf{L}'$ gilt

$$\begin{aligned}\mathbf{AA}^+\mathbf{A} &= \mathbf{A}\mathbf{L}'(\mathbf{K}'\mathbf{A}\mathbf{L}')^{-1}\mathbf{K}'\mathbf{A} = \mathbf{K}\mathbf{L}\mathbf{L}'(\mathbf{K}'\mathbf{K}\mathbf{L}\mathbf{L}')^{-1}\mathbf{K}'\mathbf{K}\mathbf{L} \\ &= \mathbf{K}\mathbf{L}\mathbf{L}'(\mathbf{L}\mathbf{L}')^{-1}(\mathbf{K}'\mathbf{K})^{-1}\mathbf{K}'\mathbf{K}\mathbf{L} \\ &= \mathbf{K}\mathbf{L} = \mathbf{A}.\end{aligned}$$

zu 2) : Es gilt

$$\mathbf{A}^+\mathbf{AA}^+ = \mathbf{L}'(\mathbf{K}'\mathbf{A}\mathbf{L}')^{-1}\mathbf{K}'\mathbf{A}\mathbf{L}'(\mathbf{K}'\mathbf{A}\mathbf{L}')^{-1}\mathbf{K}' = \mathbf{L}'(\mathbf{K}'\mathbf{A}\mathbf{L}')^{-1}\mathbf{K}' = \mathbf{A}^+.$$

zu 3) : Wir erhalten

$$\mathbf{AA}^+ = \mathbf{K}\mathbf{L}\mathbf{L}'(\mathbf{K}'\mathbf{K}\mathbf{L}\mathbf{L}')^{-1}\mathbf{K}' = \mathbf{K}\mathbf{L}\mathbf{L}'(\mathbf{L}\mathbf{L}')^{-1}(\mathbf{K}'\mathbf{K})^{-1}\mathbf{K}' = \mathbf{K}(\mathbf{K}'\mathbf{K})^{-1}\mathbf{K}'.$$

Man erkennt sofort, dass dies eine symmetrische Matrix ist.

zu 4) : Beweis analog zu 3).

zur Eindeutigkeit: Seien \mathbf{B} und \mathbf{C} zwei Moore–Penrose–Inverse. Dann gilt

$$\mathbf{A}\mathbf{B} \stackrel{4.)}{=} (\mathbf{A}\mathbf{B})' = \mathbf{B}' \overbrace{\mathbf{A}'}^{\mathbf{A}=\mathbf{A}\mathbf{C}\mathbf{A}} = \mathbf{B}'(\mathbf{A}\mathbf{C}\mathbf{A})' = \mathbf{B}'\mathbf{A}'(\mathbf{A}\mathbf{C})' \stackrel{4.)}{=} (\mathbf{A}\mathbf{B})'\mathbf{A}\mathbf{C} \stackrel{4.)}{=} \overbrace{\mathbf{A}\mathbf{B}\mathbf{A}}^{\mathbf{A}}\mathbf{C} = \mathbf{A}\mathbf{C}$$

und

$$\mathbf{B}\mathbf{A} \stackrel{3.)}{=} (\mathbf{B}\mathbf{A})' = \overbrace{\mathbf{A}'}^{\mathbf{A}=\mathbf{A}\mathbf{C}\mathbf{A}}\mathbf{B}' = (\mathbf{A}\mathbf{C}\mathbf{A})'\mathbf{B}' = (\mathbf{C}\mathbf{A})'\mathbf{A}'\mathbf{B}' \stackrel{3.)}{=} \mathbf{C}\mathbf{A}\mathbf{B}\mathbf{A} = \mathbf{C}\mathbf{A}.$$

Damit folgt

$$\mathbf{B} \stackrel{2.)}{\underset{\downarrow}{=}} \mathbf{BAB} = \mathbf{BAC} = \mathbf{CAC} = \mathbf{C}.$$

□

Satz 5.8 (Eigenschaften der Moore–Penrose–Inverse)

Die Moore–Penrose–Inverse \mathbf{A}^+ einer Matrix \mathbf{A} besitzt folgende Eigenschaften:

1. $(\mathbf{A}^+)^+ = \mathbf{A}$
2. $(\mathbf{A}^+)' = (\mathbf{A}')^+$
3. $\text{rg}(\mathbf{A}) = m \implies \mathbf{A}^+ = \mathbf{A}'(\mathbf{A}\mathbf{A}')^{-1}$ und $\mathbf{A}\mathbf{A}^+ = \mathbf{I}_m$
4. $\text{rg}(\mathbf{A}) = n \implies \mathbf{A}^+ = (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'$ und $\mathbf{A}^+\mathbf{A} = \mathbf{I}_n$
5. \mathbf{A} symmetrisch und idempotent $\implies \mathbf{A}^+ = \mathbf{A}$

Beweis:

Die Aussagen können unmittelbar durch Einsetzen in die Definition bewiesen werden.

□

6

Eigenwerte und Eigenvektoren

Dieses Kapitel befasst sich mit Eigenwerten und Eigenvektoren von Matrizen. Definitionen und allgemeine Eigenschaften sind Gegenstand von Abschnitt 6.1. Der folgende Abschnitt 6.2 befasst sich mit sogenannten ähnlichen Matrizen. Diese spielen bei der praktischen Berechnung von Eigenwerten eine wichtige Rolle. Der letzte Abschnitt 6.3 behandelt Eigenwerte von symmetrischen Matrizen. Besondere Bedeutung hat die Spektralzerlegung einer Matrix, die ebenfalls in diesem Abschnitt hergeleitet wird.

6.1 Definition und allgemeine Eigenschaften

Definition 6.1 (Eigenwert und Eigenvektor)

Sei \mathbf{A} eine quadratische $n \times n$ Matrix. $\lambda \in \mathbb{C}$ heißt Eigenwert von \mathbf{A} , wenn ein Vektor $x \in \mathbb{C}^n$ mit $x \neq 0$ existiert, so dass gilt:

$$\mathbf{A}x = \lambda x \quad \text{bzw.} \quad (\mathbf{A} - \lambda \mathbf{I})x = 0$$

Der Vektor x heißt dann Eigenvektor zum Eigenwert λ . Die Aufgabe zu einer vorgegebenen Matrix \mathbf{A} die Eigenwerte und Eigenvektoren zu bestimmen, heißt Eigenwertproblem.

Bei der Berechnung der Eigenwerte einer Matrix \mathbf{A} spielt folgende Determinante eine herausragende Rolle:

Definition 6.2 (Charakteristisches Polynom)

Sei \mathbf{A} eine quadratische $n \times n$ Matrix. Dann heißt

$$q(\lambda) := \det(\mathbf{A} - \lambda \mathbf{I})$$

charakteristisches Polynom von \mathbf{A} .

Bemerkung:

- Vergegenwärtigt man sich die Definition der Determinante (siehe Definition 4.3), dann macht man sich leicht klar, dass $q(\lambda)$ tatsächlich ein Polynom vom Grad n ist. Wir können also $q(\lambda)$ äquivalent darstellen als

$$q(\lambda) = (-\lambda)^n + \alpha_{m-1}(-\lambda)^{m-1} + \cdots + \alpha_1(-\lambda) + \alpha_0, \quad (6.1)$$

wobei die Skalare $\alpha_0, \dots, \alpha_{m-1}$ zunächst unspezifiziert sind.

– Das Polynom $q(\lambda) := \det(\mathbf{A} - \lambda\mathbf{I})$ lässt sich stets auch in die Gestalt

$$q(\lambda) = \det(\mathbf{A} - \lambda\mathbf{I}) = \prod_{i=1}^n (\lambda_i - \lambda) \quad (6.2)$$

bringen, wobei $\lambda_1, \dots, \lambda_n$ die Nullstellen des Polynoms sind. Nach dem Fundamentalsatz der Algebra hat dieses Polynom genau n nicht notwendig verschiedene und auch nicht notwendig reellwertige Nullstellen. Vergleiche hierzu zum Beispiel Bronstein, Semendjajew (1991) Seite 134.

▽

Der folgende Satz liefert nun eine Berechnungsmöglichkeit für die Eigenwerte einer Matrix:

Satz 6.1 (Berechnung über das charakteristische Polynom)

Die Eigenwerte einer quadratischen Matrix \mathbf{A} sind die Nullstellen des sogenannten charakteristischen Polynoms

$$\det(\mathbf{A} - \lambda\mathbf{I}) = 0.$$

Beweis:

Ist $\mathbf{A} - \lambda\mathbf{I}$ regulär, so ist das Gleichungssystem $(\mathbf{A} - \lambda\mathbf{I})x = 0$ nur für $x = 0$ lösbar. Es muss also λ so bestimmt werden, dass $(\mathbf{A} - \lambda\mathbf{I})x = 0$ auch Lösungen $x \neq 0$ besitzt. Dies ist äquivalent dazu, dass $(\mathbf{A} - \lambda\mathbf{I})$ singulär ist, d.h. $\det(\mathbf{A} - \lambda\mathbf{I}) = 0$.

□

Beispiel 6.1

Betrachte die Matrix

$$\mathbf{A} = \begin{pmatrix} 2 & 1 \\ 2 & -2 \end{pmatrix}.$$

Wir bestimmen die Eigenwerte von \mathbf{A} . Dazu berechnen wir zunächst das charakteristische Polynom

$$\det(\mathbf{A} - \lambda\mathbf{I}) = \det \begin{pmatrix} 2 - \lambda & 1 \\ 2 & -2 - \lambda \end{pmatrix} = (2 - \lambda)(-2 - \lambda) - 2 \cdot 1 = \lambda^2 - 6.$$

Nullsetzen und Auflösen nach λ liefert die Eigenwerte

$$\begin{aligned} \lambda_1 &= \sqrt{6}, \\ \lambda_2 &= -\sqrt{6}. \end{aligned}$$

Beispiel 6.2

Betrachte die Matrix

$$\mathbf{A} = \begin{pmatrix} 2 & -1 \\ 8 & -2 \end{pmatrix}.$$

Wir berechnen wieder das charakteristische Polynom

$$\det(\mathbf{A} - \lambda \mathbf{I}) = \det \begin{pmatrix} 2 - \lambda & -1 \\ 8 & -2 - \lambda \end{pmatrix} = (2 - \lambda)(-2 - \lambda) + 8 = \lambda^2 + 4.$$

Nullsetzen liefert die komplexen Eigenwerte

$$\begin{aligned} \lambda_1 &= 2i, \\ \lambda_2 &= -2i. \end{aligned}$$

△

Satz 6.2 (Allgemeine Eigenschaften von Eigenwerten)

Für die Eigenwerte λ_i einer $n \times n$ Matrix gelten folgende Eigenschaften:

1. $\det(\mathbf{A}) = \prod_{i=1}^n \lambda_i$
2. $sp(\mathbf{A}) = \sum_{i=1}^n \lambda_i$
3. \mathbf{A} ist genau dann regulär, wenn alle Eigenwerte ungleich Null sind.
4. Die Matrizen \mathbf{A} und \mathbf{A}' besitzen dasselbe charakteristische Polynom und damit dieselben Eigenwerte.
5. Ist λ ein Eigenwert einer regulären Matrix \mathbf{A} , dann ist $\frac{1}{\lambda}$ ein Eigenwert von \mathbf{A}^{-1} .
6. Die Eigenwerte einer Diagonalmatrix \mathbf{D} sind gerade die Hauptdiagonalelemente.
7. Für die Eigenwerte λ_i einer orthogonalen Matrix \mathbf{A} gilt $\lambda_i = \pm 1$
8. Die Eigenwerte einer idempotenten Matrix \mathbf{A} sind 1 oder 0.

Beweis:

zu 1) und 2) : Wir schreiben $\det(\mathbf{A} - \lambda \mathbf{I}) = 0$ in polynomialer Form

$$\det(\mathbf{A} - \lambda \mathbf{I}) = (-\lambda) + \alpha_{n-1}(-\lambda)^{n-1} + \dots + \alpha_1(-\lambda) + \alpha_0 = 0 \quad (6.3)$$

und bestimmen die Koeffizienten α_0 und α_{n-1} . Einsetzen von $\lambda = 0$ in 6.3 liefert

$$\det(\mathbf{A} - 0 \cdot \mathbf{I}) = \det(\mathbf{A}) = \alpha_0.$$

Zur Bestimmung von α_{n-1} vergegenwärtige man sich die Definition 4.3 der Determinante. Die Determinante ist definiert als eine Summe von Termen über alle Permutationen von $1, \dots, n$. α_{n-1} ist der Koeffizient von $(-\lambda)^{n-1}$, d.h. wir haben nur Summanden zu berücksichtigen, in denen $n-1$ der Diagonalelemente von $\mathbf{A} - \lambda\mathbf{I}$ vorkommen. Da die jeweiligen Summanden Produkte von jeweils genau einem Element in jeder Zeile von $\mathbf{A} - \lambda\mathbf{I}$ sind, kommt nur der Summand mit sämtlichen Diagonalelementen von $\mathbf{A} - \lambda\mathbf{I}$ in Frage. α_{n-1} ist also der Koeffizient von $(-\lambda)^{n-1}$ in

$$(a_{11} - \lambda)(a_{22} - \lambda) \cdots (a_{nn} - \lambda),$$

also $\alpha_{n-1} = a_{11} + a_{22} + \dots + a_{nn} = sp(\mathbf{A})$.

Gemäß (6.2) gilt

$$\det(\mathbf{A} - \lambda\mathbf{I}) = (\lambda_1 - \lambda)(\lambda_2 - \lambda) \cdots (\lambda_n - \lambda) = 0,$$

wobei $\lambda_1, \dots, \lambda_n$ die Nullstellen der charakteristischen Polynoms also die Eigenwerte von \mathbf{A} sind. Ausmultiplizieren und Koeffizientenvergleich mit 6.3 liefert schließlich

$$\det(\mathbf{A}) = \alpha_0 = \prod_{i=1}^n \lambda_i$$

und

$$sp(\mathbf{A}) = \alpha_{n-1} = \sum_{i=1}^n \lambda_i.$$

zu 3) : Die Behauptung folgt sofort aus 1).

zu 4) : Wegen Satz 4.2 gilt

$$\det(\mathbf{A} - \lambda\mathbf{I}) = \det((\mathbf{A} - \lambda\mathbf{I})') = \det(\mathbf{A}' - \lambda\mathbf{I}),$$

woraus die Behauptung folgt.

zu 5) : Es gilt $\mathbf{A}x = \lambda x$. Multiplikation von links mit \mathbf{A}^{-1} ergibt $x = \mathbf{A}^{-1}\lambda x$, woraus die Behauptung folgt.

zu 6) : Ist \mathbf{D} eine Diagonalmatrix, so ist auch $\mathbf{D} - \lambda\mathbf{I}$ eine Diagonalmatrix. Nach Satz 4.3.3 ist die Determinante einer Diagonalmatrix gleich dem Produkt der Diagonalelemente, d.h.

$$\det(\mathbf{D} - \lambda\mathbf{I}) = (d_1 - \lambda)(d_2 - \lambda) \cdots (d_n - \lambda),$$

wobei die d_i die Diagonalelemente von \mathbf{D} sind. Die Nullstellen dieses Polynoms sind aber gerade die Diagonalelemente d_i .

zu 7) : Sei λ ein beliebiger Eigenwert von \mathbf{A} . Dann gilt für einen Vektor $x \neq 0$

$$\mathbf{A}x = \lambda x.$$

Wegen der Orthogonalität von \mathbf{A} gilt $\mathbf{A}^{-1} = \mathbf{A}'$ und wegen 5) gilt

$$\mathbf{A}'x = \frac{1}{\lambda}x.$$

Da aber wegen 4) \mathbf{A} und \mathbf{A}' dasselbe charakteristische Polynom besitzen folgt

$$\lambda = \frac{1}{\lambda},$$

und daraus die Behauptung.

zu 8) : Es gilt

$$\mathbf{A}x = \lambda x$$

und

$$\mathbf{A}x = \mathbf{A}\mathbf{A}x = \mathbf{A}\lambda x = \lambda^2 x.$$

Gleichsetzen beider Gleichungen liefert $\lambda x = \lambda^2 x$, woraus $\lambda(\lambda - 1) = 0$ folgt, d.h. $\lambda = 0$ oder $\lambda = 1$.

□

Definition 6.3 (Eigenraum)

Sei \mathbf{A} eine quadratische Matrix und λ ein Eigenwert von \mathbf{A} . Die Menge $\mathbf{A}_\lambda := \{x \in \mathbb{C}^n \mid x \text{ Eigenvektor zu } \lambda\} \cup \{0\}$ heißt Eigenraum zum Eigenwert λ .

Satz 6.3

Jeder Eigenraum \mathbf{A}_λ ist ein Unterraum des \mathbb{R}^n .

Beweis:

Seien x und y zwei Eigenvektoren zum Eigenwert λ . Dann gilt

$$\mathbf{A}(x + y) = \mathbf{A}x + \mathbf{A}y = \lambda x + \lambda y = \lambda(x + y),$$

woraus die Abgeschlossenheit bezüglich der Vektoraddition folgt. Analog ergibt sich die Abgeschlossenheit bezüglich der skalaren Multiplikation mit $k \in \mathbb{C}$:

$$\mathbf{A}(kx) = k\mathbf{A}x = k\lambda x = \lambda(kx).$$

□

Beispiel 6.3 (Fortsetzung von Beispiel 6.1)

Wir betrachten wieder die Matrix \mathbf{A} aus Beispiel 6.1. Wir bestimmen jeweils eine Basis des Eigenraumes zu den Eigenwerten $\lambda_1 = \sqrt{6}$ und $\lambda_2 = -\sqrt{6}$. Der Eigenraum zu $\lambda_1 = \sqrt{6}$ ist die Lösungsmenge des homogenen Gleichungssystems

$$(\mathbf{A} - \lambda_1 \mathbf{I}) \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 2 - \sqrt{6} & 1 \\ 2 & -2 - \sqrt{6} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Subtraktion des $\frac{2}{2-\sqrt{6}}$ -fachen der 1. Gleichung von der 2. Gleichung liefert das äquivalente Gleichungssystem

$$\begin{pmatrix} 2 - \sqrt{6} & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Damit hat der Lösungsraum die Dimension eins, und jede von Null verschiedene Lösung des Gleichungssystems ist eine Basis. Wir setzen $x_2 = 1$ und erhalten $x_1 = -\frac{1}{2-\sqrt{6}}$. Analog bestimmen wir eine Basis des Eigenraumes von $\lambda_2 = -\sqrt{6}$ und erhalten

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} -\frac{1}{2+\sqrt{6}} \\ 1 \end{pmatrix}$$

als Basisvektor.

△

Beispiel 6.4 (Fortsetzung von Beispiel 6.2)

Wir bestimmen zu den Eigenwerten $\lambda_1 = 2i$ und $\lambda_2 = -2i$ der Matrix \mathbf{A} aus Beispiel 6.2 die Eigenräume. Der Eigenraum zum Eigenwert $2i$ ist die Lösungsmenge des Gleichungssystems

$$(\mathbf{A} - \lambda_1 \mathbf{I}) \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 2 - 2i & -1 \\ 8 & -2 - 2i \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Durch Subtraktion des $\frac{8}{2-2i}$ fachen der 1. Zeile von der 2. Zeile wird die 2. Zeile Null. Die Dimension des Eigenraumes ist also eins. Wir setzen $x_2 = 1$ und erhalten

$$x_1 = \frac{1}{2-2i} = \frac{2+2i}{(2-2i)(2+2i)} = \frac{2+2i}{8} = \frac{1}{4} + \frac{1}{4}i.$$

Damit ist der Basisvektor gegeben durch

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \frac{1}{4} + \frac{1}{4}i \\ 1 \end{pmatrix}.$$

Analog erhalten wir den Basisvektor des Eigenraumes zum Eigenvektor $\lambda_2 = -2i$. Wir erhalten

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \frac{1}{4} - \frac{1}{4}i \\ 1 \end{pmatrix}.$$

Definition 6.4 (algebraische und geometrische Vielfachheit)

Sei λ ein Eigenwert der Matrix \mathbf{A} . Die arithmetische (algebraische) Vielfachheit $V_a(\lambda)$ ist definiert als die Vielfachheit der Nullstelle λ des charakteristischen Polynoms. Die geometrische Vielfachheit $V_g(\lambda)$ ist definiert als die Dimension des dazugehörigen Eigenraumes \mathbf{A}_λ .

6.2 Ähnliche Matrizen

Definition 6.5 (Ähnliche Matrizen)

Zwei Matrizen \mathbf{A} und \mathbf{B} heißen ähnlich, wenn eine reguläre Matrix \mathbf{C} existiert, so dass $\mathbf{B} = \mathbf{C}\mathbf{A}\mathbf{C}^{-1}$. Schreibweise: $\mathbf{A} \sim \mathbf{B}$.

Bemerkung:

Die Ähnlichkeit von Matrizen ist eine Äquivalenzrelation, d.h. es gilt:

1. $\mathbf{A} \sim \mathbf{A}$
2. $\mathbf{A} \sim \mathbf{B} \implies \mathbf{B} \sim \mathbf{A}$
3. $\mathbf{A} \sim \mathbf{B}$ und $\mathbf{B} \sim \mathbf{C} \implies \mathbf{A} \sim \mathbf{C}$

▽

Satz 6.4 (Eigenwerte ähnlicher Matrizen)

Für ähnliche Matrizen \mathbf{A} und \mathbf{B} gilt:

1. \mathbf{A} und \mathbf{B} haben dasselbe charakteristische Polynom, d.h. sie haben dieselben Eigenwerte.
2. Ist x ein Eigenvektor zum Eigenwert λ , so ist $\mathbf{C}x$ ein Eigenvektor der Matrix $\mathbf{B} = \mathbf{C}\mathbf{A}\mathbf{C}^{-1}$.

Beweis:

zu 1) : Es gilt

$$\begin{aligned} \det(\mathbf{C}\mathbf{A}\mathbf{C}^{-1} - \lambda\mathbf{I}) &= \det(\mathbf{C}(\mathbf{A} - \lambda\mathbf{I})\mathbf{C}^{-1}) = \det(\mathbf{C})\det(\mathbf{A} - \lambda\mathbf{I})\det(\mathbf{C}^{-1}) \\ &= \det(\mathbf{C})\det(\mathbf{A} - \lambda\mathbf{I})\frac{1}{\det(\mathbf{C})} = \det(\mathbf{A} - \lambda\mathbf{I}). \end{aligned}$$

zu 2) : Es gilt

$$\mathbf{B}\mathbf{C}x = \mathbf{C}\mathbf{A}\mathbf{C}^{-1}\mathbf{C}x = \mathbf{C}\mathbf{A}x = \mathbf{C}\lambda x = \lambda\mathbf{C}x.$$

□

Ähnliche Matrizen spielen bei der effizienten numerischen Berechnung von Eigenwerten eine bedeutende Rolle, da Ähnlichkeitstransformationen die Eigenwerte einer Matrix nicht ändern. Entsprechende Algorithmen finden sich z.B. in Hämmerlin, Hoffman (1990).

6.3 Eigenwerte symmetrischer Matrizen

Satz 6.5 (Eigenwerte und Eigenvektoren symmetrischer Matrizen)

Sei \mathbf{A} eine symmetrische $n \times n$ Matrix. Dann gilt:

1. Alle Eigenwerte sind reell.
2. Die zu verschiedenen Eigenwerten gehörenden Eigenvektoren sind paarweise orthogonal.

Beweis:

zu 1) : Angenommen $\lambda := a + ib$ sei ein komplexer Eigenwert zu \mathbf{A} und $\gamma := x + iy$ sei ein dazugehöriger komplexer Eigenvektor. Dann gilt

$$\mathbf{A}(x + iy) = (a + ib)(x + iy)$$

und folglich für Real- und Imaginärteil

$$\mathbf{A}x = ax - by \quad \text{und} \quad \mathbf{A}y = bx + ay.$$

Linksmultiplikation mit y' bzw. x' liefert

$$y'\mathbf{A}x = ay'x - by'y \quad \text{und} \quad x'\mathbf{A}y = bx'x + ax'y$$

Aufgrund der Symmetrie von \mathbf{A} gilt $y'\mathbf{A}x = x'\mathbf{A}y$ und folglich

$$0 = y'\mathbf{A}x - x'\mathbf{A}y = ay'x - by'y - bx'x - ax'y = -b(y'y + x'x),$$

woraus $b = 0$ und damit die Behauptung folgt.

zu 2) : Seien $\lambda_1 \neq \lambda_2$ zwei verschiedene Eigenwerte zur Matrix \mathbf{A} und x_1 und x_2 dazugehörige Eigenvektoren. Es gilt

$$\mathbf{A}x_1 = \lambda_1 x_1 \quad \text{und} \quad \mathbf{A}x_2 = \lambda_2 x_2.$$

Linksmultiplikation von x_2' bzw. x_1' liefert

$$x_2'\mathbf{A}x_1 = \lambda_1 x_2'x_1 \quad \text{und} \quad x_1'\mathbf{A}x_2 = \lambda_2 x_1'x_2.$$

Aufgrund der Symmetrie von \mathbf{A} gilt $x_2' \mathbf{A} x_1 = x_1' \mathbf{A} x_2$ und damit

$$\lambda_1 x_2' x_1 = \lambda_2 x_1' x_2.$$

Wegen $\lambda_1 \neq \lambda_2$ folgt daraus $x_1' x_2 = 0$ und damit die Orthogonalität der Eigenvektoren x_1 und x_2 . □

Der folgende Satz, die Spektralzerlegung, spielt in vielen Bereichen der Statistik (und anderen Wissenschaften) eine bedeutende Rolle. In diesem Skript werden wir die Spektralzerlegung in den folgenden Anwendungen benutzen:

- beim Beweis der Singulärwertzerlegung einer Matrix, vergleiche Satz 7.6 in Kapitel 7;
- bei der Herleitung von Verteilungseigenschaften quadratischer Formen von multivariat normalverteilten Zufallsvektoren vergleiche die Sätze 9.13 und 9.16 in Kapitel 9;

Satz 6.6 (Spektralzerlegung)

Sei \mathbf{A} eine symmetrische $n \times n$ Matrix mit $\text{rg}(\mathbf{A}) = r$. Dann existiert eine $n \times r$ Matrix \mathbf{P} , so dass gilt:

$$\mathbf{P}' \mathbf{A} \mathbf{P} = \text{diag}(\lambda_1, \dots, \lambda_r) \quad \text{bzw.} \quad \mathbf{A} = \mathbf{P} \text{diag}(\lambda_1, \dots, \lambda_r) \mathbf{P}'$$

Dabei sind die λ_i die von Null verschiedenen Eigenwerte von \mathbf{A} . Die Spaltenvektoren von \mathbf{P} bestehen aus paarweise orthonormalen Eigenvektoren von \mathbf{A} .

Beweis:

Wegen Satz 6.5.2 sind die zu den von Null verschiedenen Eigenwerten gehörenden Eigenvektoren x_1, \dots, x_r paarweise orthogonal. Durch geeignete Normierung bilden diese ein System orthonormaler Eigenvektoren. Fasst man diese zur Matrix $\mathbf{P} := (x_1, \dots, x_r)$ zusammen, so folgt für das Produkt $\mathbf{A} \mathbf{P}$

$$\begin{aligned} \mathbf{A} \mathbf{P} &= (\mathbf{A} x_1, \mathbf{A} x_2, \dots, \mathbf{A} x_r) = (\lambda_1 x_1, \dots, \lambda_r x_r) \\ &= (x_1, \dots, x_r) \text{diag}(\lambda_1, \dots, \lambda_r) = \mathbf{P} \text{diag}(\lambda_1, \dots, \lambda_r) \end{aligned} \tag{6.4}$$

Da die Spalten von \mathbf{P} paarweise orthonormal sind, gilt $\mathbf{P} \mathbf{P}' = \mathbf{I}_n$ und es folgt durch Rechtsmultiplikation auf beiden Seiten von (6.4) mit \mathbf{P}' die Behauptung. □

Korollar 6.1

Sei \mathbf{A} symmetrisch und regulär. Dann kann man Potenzen von \mathbf{A} definieren:

$$\mathbf{A}^z = \mathbf{P} \text{diag}(\lambda_1^z, \dots, \lambda_r^z) \mathbf{P}'$$

Dabei gilt $z \in \mathbf{Z}$.

Sind sogar alle Eigenwerte positiv, so kann man auch rationale Potenzen definieren

$$\mathbf{A}^{\frac{k}{s}} = \mathbf{P} \operatorname{diag}(\lambda_1^{\frac{k}{s}}, \dots, \lambda_r^{\frac{k}{s}}) \mathbf{P}',$$

mit ganzen Zahlen $s > 0$ und k .

Beweis:

Die Behauptung kann leicht mit Hilfe der Spektralzerlegung durch vollständige Induktion bewiesen werden. □

Bemerkung:

Wichtige Spezialfälle sind

$$\mathbf{A}^{-1} = \mathbf{P} \operatorname{diag}(\lambda_1^{-1}, \dots, \lambda_r^{-1}) \mathbf{P}'$$

und für $\lambda_i > 0$ die symmetrische Wurzelzerlegung

$$\mathbf{A}^{\frac{1}{2}} = \mathbf{P} \operatorname{diag}(\lambda_1^{\frac{1}{2}}, \dots, \lambda_r^{\frac{1}{2}}) \mathbf{P}'$$

bzw. (für $\lambda_i > 0$)

$$\mathbf{A}^{-\frac{1}{2}} = \mathbf{P} \operatorname{diag}(\lambda_1^{-\frac{1}{2}}, \dots, \lambda_r^{-\frac{1}{2}}) \mathbf{P}'.$$

Für die Determinanten von $\mathbf{A}^{\frac{1}{2}}$ erhalten wir

$$\det(\mathbf{A}) = \det(\mathbf{A}^{\frac{1}{2}} \mathbf{A}^{\frac{1}{2}}) = \det(\mathbf{A}^{\frac{1}{2}}) \det(\mathbf{A}^{\frac{1}{2}})$$

also

$$\det(\mathbf{A}^{\frac{1}{2}}) = \sqrt{\det(\mathbf{A})}. \tag{6.5}$$

Analog erhalten wir

$$\det(\mathbf{A}^{-\frac{1}{2}}) = \sqrt{\frac{1}{\det(\mathbf{A})}}. \tag{6.6}$$

▽

Korollar 6.2

Sei \mathbf{A} symmetrisch. Dann ist der Rang von \mathbf{A} gleich der Anzahl der von Null verschiedenen Eigenwerte.

Beweis:

Die $n \times r$ Matrix \mathbf{P} ist spaltenregulär, da eine Menge von paarweise orthonormalen Vektoren linear unabhängig ist (vgl. Satz A11). Folglich ist \mathbf{P}' zeilenregulär. Da nach Satz

3.6.1 die Multiplikation von links mit einer spaltenregulären Matrix und Multiplikation von rechts mit einer zeilenregulären Matrix den Rang nicht ändert, gilt

$$rg(\text{diag}(\lambda_1, \dots, \lambda_r)) = rg(\mathbf{P} \text{diag}(\lambda_1, \dots, \lambda_r) \mathbf{P}') = rg(\mathbf{A})$$

woraus die Behauptung folgt. □

Korollar 6.3 (Spektralzerlegung einer idempotenten Matrix)

Sei \mathbf{A} eine symmetrische und idempotente $n \times n$ Matrix mit $rg(\mathbf{A}) = r$. Dann existiert eine orthogonale Matrix \mathbf{P} so dass gilt

$$\mathbf{P}' \mathbf{A} \mathbf{P} = \begin{pmatrix} \mathbf{I}_r & 0 \\ 0 & 0 \end{pmatrix}$$

Weiter folgern wir:

$$rg(\mathbf{A}) = sp(\mathbf{A})$$

Beweis:

Wegen $rg(\mathbf{A}) = r$ besitzt \mathbf{A} genau r von Null verschiedene Eigenwerte. Gemäß Satz 6.2 8) besitzt \mathbf{A} also genau r Eigenwerte mit dem Wert Eins und $n - r$ Eigenwerte mit dem Wert Null. Die erste Behauptung folgt damit durch Anwendung der Spektralzerlegung. Die zweite Behauptung folgt aus Satz 6.2 2). □

Beispiel 6.5

Wir betrachten wieder die Matrix

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

aus Beispiel 1.12. Da \mathbf{C} idempotent ist erhalten wir für den Rang von \mathbf{C} :

$$rg(\mathbf{C}) = sp(\mathbf{C}) = \sum_{i=1}^n \left(1 - \frac{1}{n}\right) = n - 1$$

△

Quadratische Formen und definite Matrizen

Dieses Kapitel behandelt sogenannte quadratische Formen von Matrizen. Eine wichtige Rolle spielt in diesem Zusammenhang die Definitheit von Matrizen (Abschnitt 7.1). Der zweite Abschnitt dieses Kapitels befasst sich mit der Chleskyzerlegung positiv definitiver Matrizen. Diese Matrixzerlegung spielt insbesondere bei der Lösung linearer Gleichungssysteme eine wichtige Rolle.

7.1 Definition und allgemeine Eigenschaften

Definition 7.1 (quadratische Form)

Sei \mathbf{A} eine symmetrische $n \times n$ Matrix. Eine quadratische Form in einem Vektor $x \in \mathbf{R}^n$ ist definiert durch:

$$\begin{aligned} Q(x) &= x' \mathbf{A} x = \sum_{i=1}^n \sum_{j=1}^n a_{ij} x_i x_j = \\ &= \sum_{i=1}^n a_{ii} x_i^2 + 2 \cdot \sum_{i=1}^n \sum_{j>i}^n a_{ij} x_i x_j \end{aligned}$$

Definition 7.2 (definite Matrizen)

Die quadratische Form $x' \mathbf{A} x$ und die Matrix \mathbf{A} heißen

1. *positiv definit*, falls $x' \mathbf{A} x > 0$ für alle $x \neq 0$. Schreibweise: $\mathbf{A} > 0$.
2. *positiv semidefinit*, falls $x' \mathbf{A} x \geq 0$ und $x' \mathbf{A} x = 0$ für mindestens ein $x \neq 0$.
3. *nichtnegativ definit*, falls $x' \mathbf{A} x$ bzw. \mathbf{A} entweder positiv oder positiv semidefinit ist.
Schreibweise: $\mathbf{A} \geq 0$.
4. *negativ definit*, wenn $-\mathbf{A}$ positiv definit ist.
5. *negativ semidefinit*, wenn $-\mathbf{A}$ positiv semidefinit ist.
6. *indefinit* in allen anderen Fällen.

Satz 7.1 (Kriterium für die Definitheit einer Matrix)

Sei \mathbf{A} eine symmetrische Matrix mit den (reellen) Eigenwerten $\lambda_1, \dots, \lambda_n$. Dann ist \mathbf{A} genau dann

1. positiv definit, wenn $\lambda_i > 0$ für $i = 1, \dots, n$,
2. positiv semidefinit, wenn $\lambda_i \geq 0$ für $i = 1, \dots, n$ und mindestens ein $\lambda_i = 0$,
3. negativ definit, wenn $\lambda_i < 0$ für alle $i = 1, \dots, n$,
4. negativ semidefinit, wenn $\lambda_i \leq 0$ für $i = 1, \dots, n$ und mindestens ein $\lambda_i = 0$,
5. indefinit, wenn \mathbf{A} mindestens einen positiven und einen negativen Eigenwert besitzt.

Beweis:

zu 1) : Die Behauptung folgt aus der Spektralzerlegung $\mathbf{A} = \mathbf{P} \text{diag}(\lambda_1, \dots, \lambda_n) \mathbf{P}'$. Sei $y := \mathbf{P}'x$ für einen beliebigen Vektor $x \in \mathbb{R}^n$. Dann gilt

$$x' \mathbf{A} x = x' \mathbf{P} \text{diag}(\lambda_1, \dots, \lambda_n) \mathbf{P}' x = y' \text{diag}(\lambda_1, \dots, \lambda_n) y = \sum_{i=1}^n \lambda_i y_i^2$$

Sind nun alle Eigenwerte größer als Null, so folgt sofort $x' \mathbf{A} x > 0$. Sei nun umgekehrt $x' \mathbf{A} x > 0$. Nimmt man zunächst an, dass einer der Eigenwerte $\lambda_i \leq 0$ ist, dann folgt für den dazugehörigen Eigenvektor x_i

$$x_i' \mathbf{A} x_i = \lambda_i x_i' x_i \leq 0,$$

was im Widerspruch zu $x' \mathbf{A} x > 0$ steht und es folgt daher $\lambda_i > 0$.

zu 2) - 5) : Die Behauptung folgt durch zu 1) analoger Argumentation. □

Satz 7.2 (Eigenschaften positiv definiter Matrizen)

Sei \mathbf{A} positiv definit. Dann gilt:

1. \mathbf{A} ist regulär.
2. Für die Diagonalelemente a_{ii} , $i = 1, \dots, n$ gilt: $a_{ii} > 0$
3. $\text{sp}(\mathbf{A}) > 0$
4. Sei \mathbf{B} positiv semidefinit. Dann ist $\mathbf{A} + \mathbf{B}$ positiv definit.

Beweis:

zu 1) : Da die Eigenwerte λ_i , $i = 1, \dots, n$ einer positiv definiten Matrix alle größer als Null sind, folgt die Behauptung unmittelbar aus Korollar 6.2.

zu 2) : Da \mathbf{A} positiv definit ist, gilt für alle $x \neq 0$ $x' \mathbf{A} x > 0$. Wählt man speziell den Vektor $e_i := (0, \dots, 0, 1, 0, \dots, 0)'$, der lediglich an der i -ten Position von Null verschieden ist, so folgt

$$e_i' \mathbf{A} e_i = a_{ii} > 0.$$

zu 3) : Die Behauptung folgt unmittelbar aus 2).

zu 4): Wegen $x' \mathbf{A} x > 0$ und $x' \mathbf{B} x \geq 0$ folgt

$$x' (\mathbf{A} + \mathbf{B}) x = x' \mathbf{A} x + x' \mathbf{B} x > 0.$$

□

Bemerkung

Dem Beweis des Satzes entnimmt man, dass die Eigenschaften 2) und 3) analog auf eine positiv semidefinite Matrix übertragen werden können. Es gilt dann:

1. $a_{ii} \geq 0, i = 1, \dots, n$.
2. $sp(\mathbf{A}) \geq 0$.

▽

Satz 7.3

Sei \mathbf{A} eine $n \times n$ Matrix und sei weiterhin \mathbf{Q} eine $n \times m$ Matrix. Dann gilt:

1. Ist \mathbf{A} nichtnegativ definit, so ist auch $\mathbf{Q}' \mathbf{A} \mathbf{Q}$ nichtnegativ definit.
2. Ist \mathbf{A} positiv definit und \mathbf{Q} spaltenregulär, so ist auch $\mathbf{Q}' \mathbf{A} \mathbf{Q}$ positiv definit.
3. Ist \mathbf{A} positiv definit, dann ist auch \mathbf{A}^{-1} positiv definit.

Beweis:

zu 1) : Sei $x \in \mathbb{R}^m$ ($x \neq 0$) und $y := \mathbf{Q}x \in \mathbb{R}^n$. Dann gilt

$$x' \mathbf{Q}' \mathbf{A} \mathbf{Q} x = y' \mathbf{A} y \geq 0.$$

zu 2) : Seien x und y wie im Beweis von 1) definiert. Da \mathbf{Q} spaltenregulär ist, folgt $y \neq 0$ und damit

$$x' \mathbf{Q}' \mathbf{A} \mathbf{Q} x = y' \mathbf{A} y > 0.$$

zu 3) : Setzt man in 2) $\mathbf{Q} = \mathbf{A}^{-1}$, so folgt (unter Beachtung der Symmetrie von \mathbf{A})

$$x'(\mathbf{A}^{-1})'\mathbf{A}\mathbf{A}^{-1}x = x'\mathbf{A}^{-1}x > 0.$$

□

Wir zeigen im Folgenden einige Eigenschaften der Matrizen $\mathbf{B}'\mathbf{B}$ und $\mathbf{B}\mathbf{B}'$ die aus der $m \times n$ Matrix \mathbf{B} gebildet sind :

Satz 7.4

Sei \mathbf{B} eine $m \times n$ Matrix. Dann ist die Matrix $\mathbf{B}'\mathbf{B}$ symmetrisch und nicht negativ definit. Sie ist positiv definit, wenn \mathbf{B} spaltenregulär ist. Neben $\mathbf{B}'\mathbf{B}$ ist dann auch $\mathbf{B}\mathbf{B}'$ nichtnegativ definit.

Beweis:

Die Symmetrie von $\mathbf{B}'\mathbf{B}$ ist trivial. Setzt man in Satz 7.3.1 $\mathbf{A} = \mathbf{I}$, so folgt

$$\mathbf{B}'\mathbf{I}\mathbf{B} = \mathbf{B}'\mathbf{B} \geq 0,$$

d.h. $\mathbf{B}'\mathbf{B}$ ist nichtnegativ definit. Ist \mathbf{B} zusätzlich spaltenregulär, dann lässt sich Satz 7.3.2 ebenfalls mit $\mathbf{A} = \mathbf{I}$ anwenden.

□

Satz 7.5 (Eigenwerte von $\mathbf{B}'\mathbf{B}$ und $\mathbf{B}\mathbf{B}'$)

Sei \mathbf{B} eine $m \times n$ Matrix mit $\text{rg}(\mathbf{B}) = r$.

1. Die Matrizen $\mathbf{B}\mathbf{B}'$ und $\mathbf{B}'\mathbf{B}$ besitzen identische Eigenwerte. Die r von Null verschiedenen Eigenwerte λ_j , $j = 1, \dots, r$ sind positiv.
2. Falls v ein Eigenvektor von $\mathbf{B}'\mathbf{B}$ zum Eigenwert λ ist, dann ist

$$u := \frac{1}{\sqrt{\lambda}}\mathbf{B}v$$

ein Eigenvektor von $\mathbf{B}\mathbf{B}'$ zum Eigenwert λ .

Beweis:

zu 1): Wegen Satz 3.2 gilt $\text{rg}(\mathbf{B}) = \text{rg}(\mathbf{B}'\mathbf{B}) = \text{rg}(\mathbf{B}\mathbf{B}') = r$. Aufgrund von Korollar 6.2 besitzen $\mathbf{B}'\mathbf{B}$ und $\mathbf{B}\mathbf{B}'$ genau r von Null verschiedene Eigenwerte. Da $\mathbf{B}'\mathbf{B}$ und $\mathbf{B}\mathbf{B}'$ nach Satz 7.4 nicht negativ definit sind, sind gemäß Satz 7.1 sämtliche von Null verschiedene Eigenwerte positiv.

Sei $\lambda > 0$ ein Eigenwert von $\mathbf{B}'\mathbf{B}$. Dann gilt aufgrund der Definition von Eigenwerten

$$\mathbf{B}'\mathbf{B}v = \lambda v$$

für ein $v \neq 0$. Daraus folgt $\mathbf{B}v \neq 0$. Also gilt

$$\mathbf{B}\mathbf{B}'\mathbf{B}v = \lambda\mathbf{B}v,$$

d.h. $\mathbf{B}v$ ist ein Eigenvektor von $\mathbf{B}\mathbf{B}'$ zum Eigenwert λ . $\mathbf{B}'\mathbf{B}$ und $\mathbf{B}\mathbf{B}'$ haben also dieselben Eigenwerte.

zu 2): In 1) wurde bereits gezeigt, dass $\mathbf{B}v$ ein Eigenvektor von $\mathbf{B}\mathbf{B}'$ ist. Damit ist auch $u = \frac{1}{\sqrt{\lambda}}\mathbf{B}v$ ein Eigenvektor (da die Eigenvektoren von $\mathbf{B}\mathbf{B}'$ einen Vektorraum bilden). \square

Die Aussagen aus Satz 7.5 werden vor allem zum Beweis des folgenden Satzes benötigt:

Satz 7.6 (Singularwertzerlegung)

Zu jeder $m \times n$ Matrix \mathbf{B} mit $\text{rg}(\mathbf{B}) = r$ existieren $m \times r$ und $n \times r$ Matrizen \mathbf{U} und \mathbf{V} mit $\mathbf{U}'\mathbf{U} = \mathbf{V}'\mathbf{V} = \mathbf{I}_r$, so dass gilt:

$$\mathbf{B} = \mathbf{U}\mathbf{L}\mathbf{V}'.$$

Die dabei auftretende Matrix \mathbf{L} ist eine Diagonalmatrix, deren Diagonalelemente aus den Wurzeln der positiven Eigenwerte von $\mathbf{B}'\mathbf{B}$ bzw. $\mathbf{B}\mathbf{B}'$ bestehen. Die Spalten von \mathbf{U} bestehen aus Eigenvektoren von $\mathbf{B}\mathbf{B}'$, während die Spalten von \mathbf{V} aus Eigenvektoren von $\mathbf{B}'\mathbf{B}$ bestehen. Die Diagonalelemente von \mathbf{L} heißen Singularwerte von \mathbf{B} .

Beweis:

Wegen Satz 7.5 sind die Eigenwerte von $\mathbf{B}'\mathbf{B}$ und $\mathbf{B}\mathbf{B}'$ identisch und gemäß Satz 6.6 (Spektralzerlegung) existieren Matrizen \mathbf{U} und \mathbf{V} mit

$$\mathbf{U}'\mathbf{B}\mathbf{B}'\mathbf{U} = \text{diag}(\lambda_1, \dots, \lambda_r)$$

und

$$\mathbf{V}'\mathbf{B}'\mathbf{B}\mathbf{V} = \text{diag}(\lambda_1, \dots, \lambda_r).$$

Die Spalten der $m \times r$ Matrix \mathbf{U} bestehen dabei (gemäß Spektralzerlegung) aus paarweise orthogonalen Eigenvektoren von $\mathbf{B}\mathbf{B}'$.

Die Spalten der $n \times r$ Matrix \mathbf{V} bestehen aus paarweise orthogonalen Eigenvektoren von $\mathbf{B}'\mathbf{B}$. Somit gilt $\mathbf{U}'\mathbf{U} = \mathbf{V}'\mathbf{V} = \mathbf{I}_r$ (bzw. $\mathbf{U}\mathbf{U}' = \mathbf{V}\mathbf{V}' = \mathbf{I}$). Wir zeigen dass die Spalten $u_j, j = 1, \dots, r$, von \mathbf{U} als

$$u_j := \frac{1}{\sqrt{\lambda_j}}\mathbf{B}v_j \tag{7.1}$$

definiert werden können, wobei v_j die Spalten von \mathbf{V} sind.

Wegen Satz 7.5 sind die u_j Eigenvektoren von $\mathbf{B}\mathbf{B}'$.

Es bleibt zu zeigen, dass die u_j orthogonal sind. Es gilt

$$u_j' u_j = \left(\frac{1}{\sqrt{\lambda_j}} \right)^2 v_j' \underbrace{\mathbf{B}'\mathbf{B} v_j}_{=\lambda_j v_j} = \frac{1}{\lambda_j} \lambda_j \underbrace{v_j' v_j}_{=1} = 1$$

und

$$u_j' u_i = \frac{1}{\sqrt{\lambda_j} \sqrt{\lambda_i}} v_j' \mathbf{B}' \mathbf{B} v_i = \frac{1}{\sqrt{\lambda_j} \sqrt{\lambda_i}} \lambda_i \underbrace{v_j' v_i}_{=0} = 0,$$

d.h. die Vektoren u_1, \dots, u_r sind paarweise orthogonal (bzw. \mathbf{U} eine orthogonale Matrix).

Aus (7.1) folgt

$$u_j \sqrt{\lambda_j} = \mathbf{B} v_j,$$

also

$$\mathbf{U} \underbrace{\text{diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_r})}_{=L} = \mathbf{B}\mathbf{V}.$$

Daraus folgt

$$\mathbf{B} = \mathbf{U}\mathbf{L}\mathbf{V}'.$$

□

7.2 Choleskyzerlegung

Satz 7.7 (Choleskyzerlegung)

Jede symmetrische und positiv definite $n \times n$ Matrix \mathbf{A} lässt sich eindeutig darstellen als

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

wobei \mathbf{L} die Gestalt einer unteren Dreiecksmatrix besitzt und positive Diagonalelemente hat.

Beweis:

Durch Induktion über n . Für $n = 1$ ist die Behauptung trivial.

Sei nun die Behauptung für $(n-1) \times (n-1)$ Matrizen als wahr vorausgesetzt (Induktionsvoraussetzung). Wir partitionieren \mathbf{A} in

$$\mathbf{A} = \begin{pmatrix} d & \nu' \\ \nu & \tilde{\mathbf{H}} \end{pmatrix},$$

wobei aufgrund von Satz 7.2.2 $d > 0$ gilt. Offenbar gilt

$$\mathbf{A} = \begin{pmatrix} d & \nu' \\ \nu & \tilde{\mathbf{H}} \end{pmatrix} = \begin{pmatrix} \sqrt{d} & 0 \\ \frac{\nu}{\sqrt{d}} & \mathbf{I}_{n-1} \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 \\ 0 & \mathbf{H} \end{pmatrix} \cdot \begin{pmatrix} \sqrt{d} & \frac{\nu'}{\sqrt{d}} \\ 0 & \mathbf{I}_{n-1} \end{pmatrix},$$

wobei $\mathbf{H} = \tilde{\mathbf{H}} - \frac{\nu\nu'}{d}$. Die Matrix \mathbf{H} ist offenbar symmetrisch und auch positiv definit, denn für beliebige $x \in \mathbb{R}^{n-1}$ gilt

$$\begin{pmatrix} -\frac{x'\nu}{d} & x' \end{pmatrix} \begin{pmatrix} d & \nu' \\ \nu & \tilde{\mathbf{H}} \end{pmatrix} \begin{pmatrix} -\frac{x'\nu}{d} \\ x \end{pmatrix} = x' \left(\tilde{\mathbf{H}} - \frac{\nu\nu'}{d} \right) x = x' \mathbf{H} x,$$

woraus $x' \mathbf{H} x > 0$ folgt. Aufgrund der Induktionsvoraussetzung lässt sich \mathbf{H} zerlegen in $\mathbf{H} = \mathbf{L}_H \mathbf{L}'_H$, wobei \mathbf{L}_H eine untere Dreiecksmatrix mit positiven Diagonalelementen ist. Damit erhalten wir

$$\begin{aligned} \mathbf{A} &= \begin{pmatrix} \sqrt{d} & 0 \\ \frac{\nu}{\sqrt{d}} & \mathbf{I}_{n-1} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & \mathbf{L}_H \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & \mathbf{L}'_H \end{pmatrix} \begin{pmatrix} \sqrt{d} & \frac{\nu'}{\sqrt{d}} \\ 0 & \mathbf{I}_{n-1} \end{pmatrix} = \\ &= \begin{pmatrix} \sqrt{d} & 0 \\ \frac{\nu}{\sqrt{d}} & \mathbf{L}_H \end{pmatrix} \begin{pmatrix} \sqrt{d} & \frac{\nu'}{\sqrt{d}} \\ 0 & \mathbf{L}'_H \end{pmatrix} = \mathbf{L} \mathbf{L}'. \end{aligned}$$

□

Bemerkung

Mit Hilfe der Choleskyzerlegung kann die Determinante einer Matrix \mathbf{A} leicht berechnet werden. Aufgrund der Dreiecksform von \mathbf{L} erhalten wir

$$\det(\mathbf{A}) = \det(\mathbf{L} \mathbf{L}') = \det(\mathbf{L}) \det(\mathbf{L}') = (l_{11} \cdot l_{22} \cdots l_{nn})^2.$$

▽

Beispiel 7.1

Gegeben sei die symmetrische Matrix

$$\mathbf{A} = \begin{pmatrix} 4 & 6 & 6 \\ 6 & 13 & 11 \\ 6 & 11 & 14 \end{pmatrix}.$$

Dann lässt sich \mathbf{A} zerlegen in

$$\mathbf{A} = \mathbf{L} \mathbf{L}' = \begin{pmatrix} 2 & 0 & 0 \\ 3 & 2 & 0 \\ 3 & 1 & 2 \end{pmatrix} \begin{pmatrix} 2 & 3 & 3 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{pmatrix}.$$

Für die Determinante von \mathbf{A} erhalten wir

$$\det(\mathbf{A}) = (2 \cdot 2 \cdot 2)^2 = 64.$$

△

Praktische Berechnung der Choleskyzerlegung

Aus der Darstellung

$$\begin{pmatrix} a_{11} & \cdots & \cdots & a_{1n} \\ a_{21} & \cdots & \cdots & a_{2n} \\ \vdots & & & \\ a_{n1} & \cdots & \cdots & a_{nn} \end{pmatrix} = \begin{pmatrix} l_{11} & \cdots & \cdots & 0 \\ l_{21} & l_{22} & & \vdots \\ \vdots & & \ddots & \vdots \\ l_{n1} & l_{n2} & \cdots & l_{nn} \end{pmatrix} \begin{pmatrix} l_{11} & l_{21} & \cdots & l_{n1} \\ 0 & l_{22} & \cdots & l_{n2} \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & \cdots & l_{nn} \end{pmatrix}$$

erhält man für $i = 1, \dots, n$, $j = i + 1, \dots, n$:

$$l_{ii} = \left(a_{ii} - \sum_{k=1}^{i-1} l_{ik}^2 \right)^{\frac{1}{2}} \quad (7.2)$$

$$l_{ji} = \frac{1}{l_{ii}} \left(a_{ji} - \sum_{k=1}^{i-1} l_{jk} l_{ik} \right) \quad (7.3)$$

Man berechnet also zunächst die Elemente der 1. Spalte

$$\begin{aligned} l_{11} &= \sqrt{a_{11}}, \\ l_{21} &= \frac{1}{l_{11}} a_{21}, \\ l_{31} &= \frac{1}{l_{11}} a_{31}, \\ &\vdots \\ l_{n1} &= \frac{1}{l_{11}} a_{n1}. \end{aligned}$$

Anschließend berechnet man die Elemente in der 2. Spalte

$$\begin{aligned} l_{22} &= (a_{22} - l_{21}^2)^{\frac{1}{2}}, \\ l_{32} &= \frac{1}{l_{22}} (a_{32} - l_{31} \cdot l_{21}), \\ &\vdots \\ l_{n2} &= \frac{1}{l_{22}} (a_{n2} - l_{n1} \cdot l_{21}) \end{aligned}$$

usw.

Alternativ könnte man auch zuerst das Element in der 1. Zeile berechnen, also

$$l_{11} = \sqrt{a_{11}}.$$

Anschließend werden die Elemente in der 2. Zeile berechnet, also

$$\begin{aligned} l_{21} &= \frac{1}{l_{11}} a_{21} \\ l_{22} &= (a_{22} - l_{21}^2)^{\frac{1}{2}} \end{aligned}$$

usw.

Ob man spalten- oder zeilenweise vorgeht, hängt davon ab, wie die Elemente von \mathbf{L} (im Computer) gespeichert werden.

Beispiel 7.2

Betrachte die symmetrische Matrix

$$\mathbf{A} = \begin{pmatrix} 4 & 2 & 4 & 4 \\ 2 & 10 & 17 & 11 \\ 4 & 17 & 33 & 29 \\ 4 & 11 & 29 & 39 \end{pmatrix}.$$

Wir berechnen

$$\begin{aligned} l_{11} &= \sqrt{a_{11}} = \sqrt{4} = 2 \\ l_{21} &= \frac{1}{l_{11}}a_{21} = \frac{1}{2} \cdot 2 = 1 \\ l_{22} &= (a_{22} - l_{21}^2)^{\frac{1}{2}} = (10 - 1^2)^{\frac{1}{2}} = 3 \\ l_{31} &= \frac{1}{l_{11}}a_{31} = \frac{1}{2} \cdot 4 = 2 \\ l_{32} &= \frac{1}{l_{22}}(a_{32} - l_{31}l_{21}) = \frac{1}{3}(17 - 2 \cdot 1) = 5 \end{aligned}$$

usw.

Schließlich erhalten wir

$$\mathbf{L} = \begin{pmatrix} 2 & 0 & 0 & 0 \\ 1 & 3 & 0 & 0 \\ 2 & 5 & 2 & 0 \\ 2 & 3 & 5 & 1 \end{pmatrix}.$$

△

Lineare Gleichungssysteme:

Mit Hilfe der Choleskyzerlegung einer positiv definiten Matrix \mathbf{A} können auch Gleichungssysteme der Form $\mathbf{A}\mathbf{x} = \mathbf{b}$ gelöst werden. Da \mathbf{A} regulär ist, besitzt das Gleichungssystem stets eine eindeutig bestimmte Lösung. Zur Lösung des Gleichungssystems bestimme man zunächst die Choleskyzerlegung $\mathbf{A} = \mathbf{L}\mathbf{L}'$. Anschließend löse man das System

$$\mathbf{L}\mathbf{y} = \mathbf{b}$$

und schließlich

$$\mathbf{L}'\mathbf{x} = \mathbf{y}.$$

Einsetzen von $y = \mathbf{L}'x$ in $\mathbf{L}y = b$ ergibt $\mathbf{L}\mathbf{L}'x = b$ also $\mathbf{A}x = b$, so dass x tatsächlich eine Lösung des Gleichungssystems darstellt. Da \mathbf{L} und \mathbf{L}' Dreiecksmatrizen sind, können wir die Lösungen y und x explizit angeben. Aus

$$\begin{pmatrix} l_{11} & \cdots & \cdots & 0 \\ \vdots & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ l_{n1} & \cdots & \cdots & l_{nn} \end{pmatrix} \cdot \begin{pmatrix} y_1 \\ \vdots \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} b_1 \\ \vdots \\ \vdots \\ b_n \end{pmatrix}$$

erhalten wir

$$\begin{aligned} y_1 &= \frac{b_1}{l_{11}}, \\ y_2 &= \frac{1}{l_{22}}(b_2 - l_{21}y_1) \end{aligned}$$

usw.

Allgemein gilt für $i = 1, \dots, n$

$$y_i = \frac{1}{l_{ii}}(b_i - \sum_{j=1}^{i-1} l_{ij}y_j).$$

Aus

$$\begin{pmatrix} l_{11} & \cdots & \cdots & l_{n1} \\ \vdots & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & \cdots & l_{nn} \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ \vdots \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} y_1 \\ \vdots \\ \vdots \\ y_n \end{pmatrix}$$

erhalten wir

$$\begin{aligned} x_n &= \frac{y_n}{l_{nn}}, \\ x_{n-1} &= \frac{1}{l_{n-1,n-1}}(y_{n-1} - l_{n,n-1}x_n) \end{aligned}$$

usw.

Allgemein erhält man für $i = n, n-1, \dots, 1$

$$x_i = \frac{1}{l_{ii}}(y_i - \sum_{j=i+1}^n l_{ji}x_j).$$

Wir fassen den soeben entwickelten Algorithmus noch einmal übersichtlich zusammen:

Algorithmus 7.1 (Lösung von $\mathbf{Ax} = b$, falls $\mathbf{A} > 0$)

1. Berechne die Choleskyzerlegung von $\mathbf{A} = \mathbf{LL}'$. Für $i = 1, \dots, n$ berechne

$$l_{ii} = (a_{ii} - \sum_{k=1}^{i-1} l_{ik}^2)^{\frac{1}{2}}$$

und für $j = i + 1, \dots, n$:

$$l_{ji} = \frac{1}{l_{ii}}(a_{ji} - \sum_{k=1}^{i-1} l_{jk}l_{ik})$$

2. Löse das Gleichungssystem $\mathbf{Ly} = b$ durch Vorwärtsselektion. Für $i = 1, \dots, n$ gilt

$$y_i = \frac{1}{l_{ii}}(b_i - \sum_{j=1}^{i-1} l_{ij}y_j).$$

3. Löse das Gleichungssystem $\mathbf{L}'x = y$ durch Rückwärtselimination.

Für $i = n, n - 1, \dots, 1$ gilt

$$x_i = \frac{1}{l_{ii}}(y_i - \sum_{j=i+1}^n l_{ji}x_j).$$

Beispiel 7.3

Betrachte das lineare Gleichungssystem $\mathbf{Ax} = b$ gegeben durch

$$\begin{pmatrix} 4 & 2 & 4 & 4 \\ 2 & 10 & 17 & 11 \\ 4 & 17 & 33 & 29 \\ 4 & 11 & 29 & 39 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 44 \\ 133 \\ 269 \\ 257 \end{pmatrix}.$$

Die Choleskyzerlegung von \mathbf{A} wurde bereits in Beispiel 7.2 berechnet. Es gilt

$$\mathbf{L} = \begin{pmatrix} 2 & 0 & 0 & 0 \\ 1 & 3 & 0 & 0 \\ 2 & 5 & 2 & 0 \\ 2 & 3 & 5 & 1 \end{pmatrix}$$

und

$$\mathbf{L}' = \begin{pmatrix} 2 & 1 & 2 & 2 \\ 0 & 3 & 5 & 3 \\ 0 & 0 & 2 & 5 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Durch Vorwärtsselektion lösen wir zunächst $\mathbf{L}y = b$. Wir erhalten:

$$\begin{aligned}y_1 &= \frac{44}{2} = 22 \\y_2 &= \frac{1}{3}(133 - 1 \cdot 22) = 37 \\y_3 &= \frac{1}{2}(269 - 2 \cdot 22 - 5 \cdot 37) = 20 \\y_4 &= 1 \cdot (257 - 2 \cdot 22 - 3 \cdot 37 - 5 \cdot 20) = 2\end{aligned}$$

Schließlich berechnet sich die Lösung von $\mathbf{L}'x = y$ zu

$$\begin{aligned}x_4 &= 2 \\x_3 &= \frac{1}{2}(20 - 5 \cdot 2) = 5 \\x_2 &= \frac{1}{3}(37 - 5 \cdot 5 - 3 \cdot 2) = 2 \\x_1 &= \frac{1}{2}(22 - 1 \cdot 2 - 2 \cdot 5 - 2 \cdot 2) = 3.\end{aligned}$$

△

Bandmatrizen

Häufig treten in der Statistik Situationen auf, in denen Gleichungssysteme mit (positiv definiten) Bandmatrizen zu lösen sind. In diesem Fall vereinfacht sich die Choleskyzerlegung und das anschließende Lösen von Gleichungssystemen erheblich.

Wir betrachten zunächst den einfachsten Fall, wenn \mathbf{A} eine symmetrische Tridiagonalmatrix ist, d.h.

$$\mathbf{A} = \begin{pmatrix} a_{11} & & & & & & \\ a_{21} & a_{22} & & & & & \\ 0 & a_{32} & a_{33} & & & & \\ \vdots & & \ddots & \ddots & & & \\ \vdots & & & \ddots & \ddots & & \\ 0 & & & & a_{n,n-1} & a_{nn} & \end{pmatrix},$$

wobei in obiger Darstellung wegen der Symmetrie von \mathbf{A} nur das untere Dreieck von \mathbf{A} dargestellt ist. Wir zeigen zunächst, dass dann der Choleskyfaktor \mathbf{L} ebenfalls eine Bandmatrix ist:

Satz 7.8

Sei \mathbf{A} eine positiv definite Tridiagonalmatrix der Dimension $n \times n$. Dann hat in der Choleskyzerlegung $\mathbf{A} = \mathbf{L}\mathbf{L}'$ die Matrix \mathbf{L} dieselbe Bandstruktur, d.h. \mathbf{L} hat die Gestalt

$$\mathbf{L} = \begin{pmatrix} l_{11} & & & & & \\ l_{21} & l_{22} & & & & \\ 0 & l_{32} & l_{33} & & & \\ \vdots & & \ddots & \ddots & & \\ \vdots & & & \ddots & \ddots & \\ 0 & & & & l_{n,n-1} & l_{nn} \end{pmatrix} \quad (7.4)$$

Beweis:

Durch Induktion über n . Für $n = 1$ ist die Aussage offenbar richtig. Sei nun die Behauptung für $(n-1) \times (n-1)$ Matrizen als wahr vorausgesetzt (Induktionsvoraussetzung). Wie im Beweis zu Satz 7.7 (Choleskyzerlegung) partitionieren wir \mathbf{A} in

$$\mathbf{A} = \begin{pmatrix} d & \nu' \\ \nu & \tilde{\mathbf{H}} \end{pmatrix},$$

wobei wegen der Tridiagonalgestalt von \mathbf{A} die Matrix $\tilde{\mathbf{H}}$ ebenfalls eine Tridiagonalmatrix ist und $\nu = (a_{21}, 0, \dots, 0)$ gilt. Analog zum Beweis von Satz 7.7 erhalten wir

$$\mathbf{A} = \begin{pmatrix} \sqrt{d} & 0 \\ \frac{\nu}{\sqrt{d}} & \mathbf{L}_H \end{pmatrix} \cdot \begin{pmatrix} \sqrt{d} & \frac{\nu'}{\sqrt{d}} \\ 0 & \mathbf{L}'_H \end{pmatrix} = \mathbf{L}\mathbf{L}',$$

wobei $\mathbf{H} = \tilde{\mathbf{H}} - \frac{\nu\nu'}{d}$ und $\mathbf{H} = \mathbf{L}_H\mathbf{L}'_H$. Nach Induktionsvoraussetzung ist \mathbf{L}_H eine Matrix der Gestalt (7.4). Wegen $\nu = (a_{21}, 0, \dots, 0)$ folgt unmittelbar, dass \mathbf{L} die Gestalt (7.4) besitzt. □

Die praktische Berechnung der Choleskyzerlegung vereinfacht sich im Vergleich zum allgemeinen Fall erheblich. Da $l_{ik} = 0$ für $k < i-1$, erhält man aus den allgemeinen Formeln (7.2) und (7.3) folgenden Algorithmus zur Choleskyzerlegung bei Tridiagonalmatrizen:

$$l_{11} = \sqrt{a_{11}}$$

Für $i = 2, \dots, n$

$$l_{i,i-1} = \frac{1}{l_{i-1,i-1}} a_{i,i-1}$$

und

$$l_{ii} = (a_{ii} - l_{i,i-1}^2)^{\frac{1}{2}}.$$

Beispiel 7.4

Betrachte die Tridiagonalmatrix

$$\mathbf{A} = \begin{pmatrix} 4 & 2 & 0 & 0 & 0 \\ 2 & 17 & 8 & 0 & 0 \\ 0 & 8 & 53 & 28 & 0 \\ 0 & 0 & 28 & 25 & 3 \\ 0 & 0 & 0 & 3 & 2 \end{pmatrix}$$

Wir erhalten

$$\begin{aligned} l_{11} &= \sqrt{a_{11}} = 2 \\ l_{21} &= \frac{1}{l_{11}} \cdot a_{21} = \frac{1}{2} \cdot 2 = 1 \\ l_{22} &= (a_{22} - l_{21}^2)^{\frac{1}{2}} = (17 - 1^2)^{\frac{1}{2}} = 4 \\ l_{32} &= \frac{1}{l_{22}} \cdot a_{32} = \frac{1}{4} \cdot 8 = 2 \\ l_{33} &= (a_{33} - l_{32}^2)^{\frac{1}{2}} = (53 - 2^2)^{\frac{1}{2}} = 7 \\ l_{43} &= \frac{1}{l_{33}} \cdot a_{43} = \frac{1}{7} \cdot 28 = 4 \\ l_{44} &= (a_{44} - l_{43}^2)^{\frac{1}{2}} = (25 - 4^2)^{\frac{1}{2}} = 3 \\ l_{54} &= \frac{1}{l_{44}} \cdot a_{54} = \frac{1}{3} \cdot 3 = 1 \\ l_{55} &= (a_{55} - l_{54}^2)^{\frac{1}{2}} = (2 - 1^2)^{\frac{1}{2}} = 1 \end{aligned}$$

und schließlich

$$\mathbf{L} = \begin{pmatrix} 2 & 0 & 0 & 0 & 0 \\ 1 & 4 & 0 & 0 & 0 \\ 0 & 2 & 7 & 0 & 0 \\ 0 & 0 & 4 & 3 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix}.$$

△

Neben der Choleskyzerlegung vereinfacht sich auch die Lösung linearer Gleichungssysteme $\mathbf{A}x = b$, wenn \mathbf{A} eine Tridiagonalmatrix ist.

Für die Lösung des Systems $\mathbf{L}y = b$ erhält man

$$y_1 = \frac{1}{l_{11}} \cdot b_1$$

und

$$y_i = \frac{1}{l_{ii}} (b_i - l_{i,i-1} y_{i-1})$$

für $i = 2, \dots, n$. Die Lösung des Systems $\mathbf{L}'x = y$ ergibt sich zu

$$x_n = \frac{1}{l_{nn}} \cdot y_n$$

und

$$x_i = \frac{1}{l_{ii}}(y_i - l_{i+1,i}x_{i+1})$$

für $i = n - 1, \dots, 1$.

Beispiel 7.5

Wir suchen eine Lösung des Systems $\mathbf{A}x = b$ mit \mathbf{A} aus Beispiel 7.4 und $b = (14, 63, 133, 90, 14)$. Wir lösen zunächst $\mathbf{L}y = b$ und erhalten:

$$\begin{aligned} y_1 &= \frac{1}{l_{11}}b_1 = \frac{1}{2} \cdot 14 = 7 \\ y_2 &= \frac{1}{l_{22}}(b_2 - l_{21}y_1) = \frac{1}{4}(63 - 1 \cdot 7) = 14 \\ y_3 &= \frac{1}{l_{33}}(b_3 - l_{32}y_2) = \frac{1}{7}(133 - 2 \cdot 14) = 15 \\ y_4 &= \frac{1}{l_{44}}(b_4 - l_{43}y_3) = \frac{1}{3}(90 - 4 \cdot 15) = 10 \\ y_5 &= \frac{1}{l_{55}}(b_5 - l_{54}y_4) = \frac{1}{1}(14 - 1 \cdot 10) = 4 \end{aligned}$$

Auflösen von $\mathbf{L}'x = y$ liefert schließlich:

$$\begin{aligned} x_5 &= \frac{1}{l_{55}}y_5 = \frac{1}{1} \cdot 4 = 4 \\ x_4 &= \frac{1}{l_{44}}(y_4 - l_{54}x_5) = \frac{1}{3}(10 - 1 \cdot 4) = 2 \\ x_3 &= \frac{1}{l_{33}}(y_3 - l_{43}x_4) = \frac{1}{7}(15 - 4 \cdot 2) = 1 \\ x_2 &= \frac{1}{l_{22}}(y_2 - l_{32}x_3) = \frac{1}{4}(14 - 2 \cdot 1) = 3 \\ x_1 &= \frac{1}{l_{11}}(y_1 - l_{21}x_2) = \frac{1}{2}(7 - 1 \cdot 3) = 2 \end{aligned}$$

△

Algorithmen zur Bestimmung der Choleskyzerlegung und zum Lösen von linearen Gleichungssystemen mit allgemeinen Bandmatrizen findet man zum Beispiel in George und Liu (1981).

Differenziation von Matrizen

Dieses kurze Kapitel stellt in Abschnitt 8.1 die wichtigsten Regeln für die Differenziation einer Matrix, deren Elemente reellwertige Funktionen sind, zusammen. Abschnitt 8.2 befasst sich dann mit der Differenziation einer Matrixfunktion nach den Elementen der Matrix. Eine Anwendung einiger der Regeln findet man in den Kapiteln 10.2 und 10.3.

8.1 Differenziation nach einem Skalar

Definition 8.1 (Differenziation nach einem Skalar)

Sei $\mathbf{A} = (a_{ij})$ eine $m \times n$ Matrix, deren Elemente differenzierbare Funktionen der reellen Variablen t seien. Dann heißt die Matrix

$$\frac{\delta \mathbf{A}}{\delta t} = \left(\frac{\delta a_{ij}}{\delta t} \right)$$

Ableitung von \mathbf{A} nach t .

Es ergeben sich folgende (leicht beweisbare) Rechenregeln:

Satz 8.1 (Rechenregeln)

Sei \mathbf{A} eine Matrix passenden Formats. Dann gilt:

1. $\frac{\delta \mathbf{A}}{\delta a_{ij}} = e_i e_j'$, wobei $e_i = (0, \dots, \underbrace{1}_i, \dots, 0)$.
2. $\frac{\delta \mathbf{A}'}{\delta a_{ij}} = e_j e_i'$
3. $\frac{\delta \mathbf{A}\mathbf{B}}{\delta t} = \frac{\delta \mathbf{A}}{\delta t} \mathbf{B} + \mathbf{A} \frac{\delta \mathbf{B}}{\delta t}$ (Produktregel)

8.2 Differenziation einer Matrixfunktion nach der Matrix

Definition 8.2 (Differenziation nach einer Matrix)

Sei $\mathbf{A} = (a_{ij})$ eine $m \times n$ Matrix und $f(\mathbf{A})$ eine differenzierbare reellwertige Funktion der mn Elemente a_{ij} . Dann heißt die $m \times n$ Matrix

$$\frac{\delta f}{\delta \mathbf{A}} = \left(\frac{\delta f}{\delta a_{ij}} \right)$$

Ableitung von f nach \mathbf{A} .

Es folgen umständlich aber leicht zu beweisende Rechenregeln:

Satz 8.2 (Rechenregeln)

Seien \mathbf{A}, \mathbf{B} Matrizen, f, g Funktionen von Matrizen und x, y Vektoren. Bei den folgenden Größen wird angenommen, dass sie existieren und von passender Ordnung sind. Dann gelten folgende Rechenregeln:

1. $\frac{\delta fg}{\delta \mathbf{A}} = \frac{\delta f}{\delta \mathbf{A}}g + f \frac{\delta g}{\delta \mathbf{A}}$
2. $\frac{\delta sp(\mathbf{A})}{\delta \mathbf{A}} = I$
3. $\frac{\delta sp(\mathbf{BA})}{\delta \mathbf{A}} = \mathbf{B}'$
4. $\frac{\delta sp(\mathbf{A}'\mathbf{BA})}{\delta \mathbf{A}} = (\mathbf{B} + \mathbf{B}')\mathbf{A}$
5. $\frac{\delta sp(\mathbf{ABA}')}{\delta \mathbf{A}} = \mathbf{A}'(\mathbf{B} + \mathbf{B}')$
6. $\frac{\delta sp(\mathbf{ABA})}{\delta \mathbf{A}} = \mathbf{A}'\mathbf{B}' + \mathbf{B}'\mathbf{A}'$
7. $\frac{\delta y'x}{\delta x} = y$
8. $\frac{\delta x'\mathbf{A}y}{\delta \mathbf{A}} = xy'$
9. $\frac{\delta x'\mathbf{A}x}{\delta x} = (\mathbf{A} + \mathbf{A}')x$
10. \mathbf{A} symmetrisch $\implies \frac{\delta x'\mathbf{A}x}{\delta x} = 2\mathbf{A}x = 2\mathbf{A}'x$

Die multivariate Normalverteilung

In diesem Kapitel soll die multivariate Normalverteilung als eine Anwendung der Matrixtheorie behandelt werden. Die multivariate Normalverteilung spielt eine zentrale Bedeutung in der Statistik, eine Vielzahl von statistischen Verfahren beruht auf der Annahme einer multivariaten Normalverteilung. Voraussetzung zum Verständnis dieses Kapitels sind Grundkenntnisse der Wahrscheinlichkeitsrechnung wie sie in Anfängervorlesungen behandelt werden, siehe etwa Fahrmeir et al. (2001). Da die multivariate Normalverteilung aus der univariaten Normalverteilung abgeleitet ist, stellen wir zunächst einige Eigenschaften der univariaten Normalverteilung zusammen (Abschnitt 9.1). Die multivariate Normalverteilung und ihre wichtigsten Eigenschaften sind dann Gegenstand von Abschnitt 9.2. Der letzte Abschnitt 9.3 stellt dann noch den Zusammenhang mit einigen Verteilungen her, die häufig die Verteilung der Teststatistik beim statistischen Testen von Hypothesen bilden, vergleiche hierzu auch Kapitel 11.

9.1 Die univariate Normalverteilung

Definition 9.1 (Standardnormalverteilung)

Eine stetige Zufallsvariable X mit der Dichte

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

heißt *standardnormalverteilt*. Schreibweise: $X \sim N(0, 1)$.

Damit dies wohldefiniert ist, muß gezeigt werden daß das Integral über f eins ergibt bzw. das Integral über $e^{-\frac{x^2}{2}} \sqrt{2\pi}$ ergibt.

Beweis:

Für den Nachweis wird auf die Gammafunktion zurückgegriffen, die wie folgt definiert ist:

$$\Gamma(x) = \int_0^{\infty} t^{x-1} e^{-t} dt$$

Man kann zeigen (vgl. z.B. Forster, 1999), dass $\Gamma(\frac{1}{2}) = \sqrt{\pi}$. Damit gilt:

$$\begin{aligned} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx &= 2 \int_0^{\infty} e^{-\frac{x^2}{2}} dx = 2 \underbrace{\int_0^{\infty} \frac{1}{\sqrt{2t}} e^{-t} dt}_{\text{subst. : } t = \frac{x^2}{2}} = \frac{2}{\sqrt{2}} \int_0^{\infty} t^{1-\frac{1}{2}} e^{-t} dt \\ &= \frac{2}{\sqrt{2}} \Gamma(\frac{1}{2}) = \sqrt{2\pi}. \end{aligned}$$

Eigenschaften von Zufallsvariablen bzw. Zufallsvektoren lassen sich häufig leicht über die sogenannte momenterzeugende Funktion bestimmen:

Definition 9.2 (momenterzeugende Funktion)

Sei X ein n -dimensionaler Zufallsvektor. Dann ist die momenterzeugende Funktion $m_X : \mathbb{R}^n \mapsto \mathbb{R}$ von X definiert durch

$$m_X(t) = E(e^{t'X}),$$

sofern diese Größe in einer Umgebung um Null existiert.

Die momenterzeugende Funktion besitzt folgende Eigenschaften (zum Beweis vergleiche z.B. Billingsley (1985)):

Satz 9.1 (Eigenschaften der momenterzeugenden Funktion)

Seien $X \in \mathbb{R}^n$ und $Y \in \mathbb{R}^n$ Zufallsvektoren, mit existierenden momenterzeugenden Funktionen.

1. Die Verteilungen von X und Y stimmen genau dann überein, wenn die momenterzeugenden Funktionen gleich sind.
2. $E(X_1^{k_1} \dots X_n^{k_n}) = \left. \frac{\delta^{(k_1+\dots+k_n)} m_X(t)}{\delta^{(k_1)} t_1 \dots \delta^{(k_n)} t_n} \right|_{t=0}$
3. Sei $Y := aX + b$. Dann gilt: $m_Y(t) = e^{tb} m_X(at)$
4. X und Y sind genau dann unabhängig, wenn gilt: $m_{X,Y}(t) = m_X(t) m_Y(t)$
5. Seien X und Y unabhängig. Dann gilt für die momenterzeugende Funktion der Summe $X + Y$: $m_{X+Y}(t) = m_X(t) m_Y(t)$

Mit Hilfe der momenterzeugenden Funktionen können wir jetzt Eigenschaften der Standardnormalverteilung und anschließend auch der univariaten und multivariaten Normalverteilung herleiten:

Satz 9.2 (Eigenschaften der Standardnormalverteilung)

Sei $X \sim N(0, 1)$. Dann gilt:

1. $m_X(t) = e^{\frac{1}{2}t^2}$.
2. $E(X) = 0$.
3. $Var(X) = 1$.
4. Für die Verteilungsfunktion Φ gilt: $\Phi(-x) = 1 - \Phi(x)$.

Beweis:

zu 1) :

$$\begin{aligned} m_X(t) &= E(e^{tX}) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tx} e^{-\frac{x^2}{2}} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\frac{t^2 - (x-t)^2}{2}} dx \\ &= \frac{1}{\sqrt{2\pi}} e^{\frac{t^2}{2}} \int_{-\infty}^{\infty} e^{-\frac{(x-t)^2}{2}} dx = \frac{1}{\sqrt{2\pi}} e^{\frac{t^2}{2}} \underbrace{\int_{-\infty}^{\infty} e^{-\frac{u^2}{2}} du}_{\text{subst. : } u=x-t} = e^{\frac{t^2}{2}} \end{aligned}$$

zu 2) : Es gilt $m'_X(t) = te^{\frac{t^2}{2}}$ und damit nach Satz 9.2 2) für den Erwartungswert:

$$E(X) = m'_X(0) = 0.$$

zu 3) : Es gilt $m''_X(t) = t^2 e^{\frac{t^2}{2}} + e^{\frac{t^2}{2}} = e^{\frac{t^2}{2}}(t^2 + 1)$ und damit für die Varianz:

$$Var(X) = E(X^2) - E(X)^2 = m''_X(0) - 0 = 1.$$

zu 4) : Die Behauptung folgt aus der Tatsache, daß die Dichte von X eine gerade Funktion, also symmetrisch zur y -Achse, ist. □

Die Definition der univariaten Normalverteilung kann auf die Standardnormalverteilung zurückgeführt werden:

Definition 9.3 (univariate Normalverteilung)

Eine Zufallsvariable X heißt genau dann univariat normalverteilt, wenn sie als Linearkombination einer standardnormalverteilten Zufallsvariable Y darstellbar ist, d.h. $X = \sigma Y + \mu$ mit $\sigma > 0$.

Für den Beweis von Eigenschaften der univariaten und später der multivariaten Normalverteilung benötigen wir folgenden Satz aus der Wahrscheinlichkeitsrechnung über die lineare Transformation von Zufallsvektoren:

Satz 9.3

Sei X ein Zufallsvektor mit Dichte f . Für die lineare Transformation $Y = T(X) := \mathbf{A}X + b$, wobei \mathbf{A} eine invertierbare $n \times n$ Matrix und b ein $n \times 1$ Spaltenvektor seien, folgt für die Dichte von Y

$$h(y) = \frac{f(\mathbf{A}^{-1}(y - b))}{|\det(\mathbf{A})|}.$$

Satz 9.4 (Eigenschaften der univariaten Normalverteilung)

Sei X univariat normalverteilt. Dann gilt:

1. Die Dichte von X ist gegeben durch

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}.$$

2. $m_X(t) = e^{t\mu + \frac{1}{2}\sigma^2 t^2}$

3. $E(X) = \mu$

4. $\text{Var}(X) = \sigma^2$. Schreibweise: $X \sim N(\mu, \sigma)$

5. Für die Zufallsvariable $Z = aX + b$ gilt: $Z \sim N(a\mu + b, a^2\sigma^2)$

6. Ist $X' \sim N(\mu', \sigma'^2)$ eine weitere Zufallsvariable und sind X und X' unabhängig, dann gilt: $X + X' \sim N(\mu + \mu', \sigma^2 + \sigma'^2)$

Beweis:

zu 1) : Anwendung des Transformationssatzes für Dichten. Da X darstellbar ist als $X = \mu + \sigma Z$ mit $Z \sim N(0, 1)$ folgt:

$$f(x) = f_Z\left(\frac{x - \mu}{\sigma}\right) = \frac{\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}}{\sigma}$$

zu 2) : $m_X(t) = e^{t\mu} m_Y(\sigma t) = e^{t\mu} e^{\frac{1}{2}\sigma^2 t^2}$

zu 3) : $E(X) = E(\sigma Y + \mu) = \sigma E(Y) + \mu = \mu$

zu 4) : $\text{Var}(X) = \text{Var}(\sigma Y + \mu) = \sigma^2 \text{Var}(Y) = \sigma^2$

zu 5) : Für die momenterzeugende Funktion von Z gilt nach Satz 9.2 3):

$$m_Z(t) = e^{bt} e^{at\mu + \frac{1}{2}\sigma^2 a^2 t^2} = e^{t(a\mu + b) + \frac{1}{2}(a\sigma)^2 t^2}$$

Dies ist aber die momenterzeugende Funktion einer normalverteilten Zufallsvariable mit Erwartungswert $a\mu + b$ und Varianz $a^2\sigma^2$.

zu 6) : Für die momenterzeugende Funktion von $X + X'$ gilt

$$\begin{aligned} m_{X+X'}(t) &= m_X(t)m_{X'}(t) = e^{t\mu + \frac{1}{2}\sigma^2 t^2} e^{t\mu' + \frac{1}{2}\sigma'^2 t^2} \\ &= e^{t(\mu+\mu') + \frac{1}{2}(\sigma^2 + \sigma'^2)t^2}, \end{aligned}$$

woraus nach Satz 9.2 5) die Behauptung folgt. □

9.2 Die multivariate Normalverteilung

Nach den geleisteten Vorarbeiten definieren wir jetzt die multivariate Normalverteilung, wobei wir die Definition zurückführen auf die univariate Normalverteilung.

Definition 9.4 (multivariate Normalverteilung)

Ein Zufallsvektor $X = (X_1, X_2, \dots, X_n)'$ heißt genau dann multivariat normalverteilt, wenn für alle $a \in \mathbf{R}^n$ die Linearkombination $a_1X_1 + \dots + a_nX_n$ univariat normalverteilt oder einpunktverteilt ist. Dabei heißt eine Zufallsvariable Y einpunktverteilt, wenn sie nur für einen Wert c eine positive Wahrscheinlichkeit besitzt, d.h. $P(X = c) = 1$.

Aus der Definition ergeben sich unmittelbar zwei Folgerungen:

Korollar 9.1

Sei X ein Zufallsvektor. Sind die Komponenten von X unabhängig und univariat normalverteilt, dann ist X multivariat normalverteilt.

Beweis:

Aufgrund des Additionssatzes für univariat normalverteilte Zufallsvariablen (Satz 9.4 6) ist $a_1X_1 + \dots + a_nX_n$ univariat normalverteilt mit Erwartungswert $\sum a_i\mu_i$ und Varianz $\sum a_i^2\sigma_i^2$, so daß aus der Definition der multivariaten Normalverteilung unmittelbar die Behauptung folgt. □

Korollar 9.2

Sei X multivariat normalverteilt. Sei weiterhin \mathbf{D} eine $m \times n$ Matrix und d ein $m \times 1$ Spaltenvektor. Dann ist auch der m -dimensionale Zufallsvektor $Y = \mathbf{D}X + d$ multivariat normalverteilt.

Beweis:

Sei $a \in \mathbb{R}^m$. Dann ist $a'Y = a'DX + a'd$ eine Linearkombination von X und damit univariat normalverteilt, woraus die Behauptung folgt. □

Satz 9.5

Sei $\mathbf{A} > 0$ eine $n \times n$ Matrix und a ein $n \times 1$ Spaltenvektor. Dann gilt:

1. Es existiert ein normalverteilter Zufallsvektor X mit der momenterzeugenden Funktion

$$m_X(t) = e^{t'a + \frac{1}{2}t'\mathbf{A}t}$$

2. Sei X multivariat normalverteilt mit $E(X) = \mu$ und $Cov_X = \Sigma$. Dann hat die momenterzeugende Funktion von X die Form

$$m_X(t) = e^{t'\mu + \frac{1}{2}t'\Sigma t}$$

Schreibweise: $X \sim N_n(\mu, \Sigma)$.

Beweis:

zu 1) : Sei $Z = (Z_1, \dots, Z_n)'$ ein Zufallsvektor mit unabhängigen Z_i und $Z_i \sim N(0, 1)$. Dann ist Z nach Korollar 9.1 multivariat normalverteilt und damit nach Korollar 9.2 auch eine Linearkombination $X = \mathbf{A}^{\frac{1}{2}}Z + a$, wobei $\mathbf{A}^{\frac{1}{2}}$ wie auf Seite 118 definiert sei. Die momenterzeugende Funktion dieses Zufallsvektors ist mit $u := \mathbf{A}^{\frac{1}{2}}t$ gegeben durch

$$\begin{aligned} m_X(t) &= E\left(e^{t'X}\right) = E\left(e^{t'(\mathbf{A}^{\frac{1}{2}}Z + a)}\right) = E\left(e^{t'\mathbf{A}^{\frac{1}{2}}Z} e^{t'a}\right) \\ &= e^{t'a} E\left(e^{t'\mathbf{A}^{\frac{1}{2}}Z}\right) = e^{t'a} E\left(e^{u'Z}\right) = e^{t'a} E\left(e^{\sum u_i Z_i}\right) \\ &= e^{t'a} E\left(\prod e^{u_i Z_i}\right) = e^{t'a} \prod E\left(e^{u_i Z_i}\right) = e^{t'a} \prod m_{Z_i}(u_i) \\ &= e^{t'a} \prod e^{\frac{u_i^2}{2}} = e^{t'a} e^{\frac{1}{2} \sum u_i^2} = e^{t'a} e^{\frac{1}{2}u'u} \\ &= e^{t'a} e^{\frac{1}{2}t'\mathbf{A}^{\frac{1}{2}}\mathbf{A}^{\frac{1}{2}}t} = e^{t'a + \frac{1}{2}t'\mathbf{A}t}, \end{aligned}$$

wodurch 1) bewiesen ist.

zu 2) : Sei also X multivariat normalverteilt mit $E(X) = \mu$ und $Cov_X = \Sigma$ und sei zunächst $t'X$ univariat normalverteilt. Dann hat $t'X$ den Erwartungswert $t'\mu$, die Varianz $t'\Sigma t$ und die momenterzeugende Funktion:

$$m_{t'X}(k) = E(e^{kt'X}) = e^{kt'\mu + \frac{1}{2}t'\Sigma tk^2}$$

Nun gilt:

$$m_X(t) = E(e^{t'X}) = m_{t'X}(1) = e^{t'\mu + \frac{1}{2}t'\Sigma t}$$

Der Nachweis für den Fall, daß $t'X$ einpunktverteilt ist, verläuft analog. □

Satz 9.6 (Eigenschaften der multivariaten Normalverteilung)

Sei $X \sim N_n(\mu, \Sigma)$ mit $\Sigma > 0$ (positiv definit).

1. Sei $Y = \mathbf{D}X + d$ wobei \mathbf{D} eine $m \times n$ Matrix und d ein $m \times 1$ Spaltenvektor sei. Dann gilt:

$$Y \sim N_m(\mathbf{D}\mu + d, \mathbf{D}\Sigma\mathbf{D}')$$

2. X besitzt folgende Dichte:

$$f(x) = \frac{1}{\sqrt{2\pi}^n \sqrt{|\Sigma|}} e^{-\frac{1}{2}(x - \mu)'\Sigma^{-1}(x - \mu)}.$$

Beweis:

zu 1) : Die Behauptung folgt aus Folgerung 9.2 und durch Anwendung der Regeln über die Erwartungswertbildung und die Kovarianzbildung.

zu 2) : Sei $Z = (Z_1, \dots, Z_n)'$, wobei die Z_i unabhängig und standardnormalverteilt seien. Dann ist nach Folgerung 9.1 $Z \sim N_n(0, \mathbf{I}_n)$. Die Dichtefunktion von Z ist wegen der Unabhängigkeit der Z_i gegeben als das Produkt der einzelnen Dichten:

$$g(z) = \frac{1}{\sqrt{2\pi}^n} e^{-\frac{1}{2} \sum z_i^2} = \frac{1}{\sqrt{2\pi}^n} e^{-\frac{1}{2}z'z}$$

Sei nun $X = \Sigma^{\frac{1}{2}}Z + \mu$. X ist nach 1) $N_n\left(\Sigma^{\frac{1}{2}}0 + \mu, \Sigma^{\frac{1}{2}}\mathbf{I}_n\Sigma^{\frac{1}{2}}\right) = N_n(\mu, \Sigma)$ verteilt. Gemäß dem Transformationsatz für Dichten (Satz 9.3) erhalten wir:

$$f_X(x) = \frac{1}{|\Sigma^{\frac{1}{2}}|} g\left(\Sigma^{-\frac{1}{2}}(x - \mu)\right) = \frac{1}{\sqrt{|\Sigma|}\sqrt{2\pi}^n} e^{-\frac{1}{2}\left(\Sigma^{-\frac{1}{2}}(x - \mu)\right)'\left(\Sigma^{-\frac{1}{2}}(x - \mu)\right)},$$

woraus die Behauptung folgt. Dabei haben wir

$$|\Sigma^{\frac{1}{2}}| = \sqrt{|\Sigma|}$$

verwendet, siehe Gleichung (6.5) auf Seite 118. □

9.2.1 Marginalverteilungen, Unkorreliertheit, Unabhängigkeit

In diesem Abschnitt zeigen wir zunächst, dass die Marginalverteilungen der multivariaten Normalverteilung wieder normalverteilt sind:

Satz 9.7 (Marginalverteilungen)

Sei $X \sim N_n(\mu, \Sigma)$ mit $\Sigma > 0$. Dann ist jeder $r \leq n$ -dimensionale Subvektor $X^{(1)}$ wieder multivariat normalverteilt mit $X^{(1)} \sim N_r(\mu^{(1)}, \Sigma^{(1)})$. Dabei entstehen $\mu^{(1)}$ und $\Sigma^{(1)}$ durch Streichen der entsprechenden Zeilen und Spalten.

Beweis:

O.B.d.A. sei $X^{(1)} = (X_1, \dots, X_r)$. Dann läßt sich $X^{(1)}$ als Linearkombination von X darstellen $X^{(1)} = (\mathbf{I}_r \quad \mathbf{0}_{r, n-r})X$ wobei $\mathbf{0}_{r, n-r}$ eine $r \times n - r$ Matrix ist, deren Elemente sämtlich Null sind. Nach Satz 9.6 1) ist dies multivariat normalverteilt. Erwartungswert und Kovarianzmatrix ergeben sich entsprechend. □

Der folgende Satz besagt, dass Unabhängigkeit und Unkorreliertheit im Falle der Normalverteilung äquivalent sind, eine Eigenschaft die im Allgemeinen **nicht** gilt:

Satz 9.8 (Äquivalenz von Unabhängigkeit und Unkorreliertheit)

Sei $X \sim N_n(\mu, \Sigma)$ mit $\Sigma > 0$ und X partitioniert mit

$$X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \text{ und } \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{X_1} & \Sigma_{X_1 X_2} \\ \Sigma_{X_2 X_1} & \Sigma_{X_2} \end{pmatrix}.$$

Dann sind X_1 und X_2 genau dann unabhängig, wenn X_1 und X_2 unkorreliert sind, d.h. $\Sigma_{X_1 X_2} = \Sigma_{X_2 X_1} = \mathbf{0}$ gilt.

Beweis:

Die Behauptung, daß aus der Unabhängigkeit die Unkorreliertheit folgt, muß nicht bewiesen werden, da dies allgemein gilt, d.h. unabhängig davon welche Verteilung zugrundeliegt (vgl. z. B. Fahrmeir et al., 2001). Seien also nun X_1 und X_2 unkorreliert, d.h. $\Sigma_{X_1 X_2} = \Sigma_{X_2 X_1} = \mathbf{0}$, und $t' = (t_1, t_2)$ mit $t_1 \in \mathbf{R}^r$, $t_2 \in \mathbf{R}^{n-r}$. Dann gilt

$$t' \mu = t_1' \mu_1 + t_2' \mu_2$$

und

$$t' \Sigma t = t_1' \Sigma_{X_1} t_1 + t_2' \Sigma_{X_2} t_2.$$

Damit ergibt sich die momenterzeugende Funktion von X gemäß Satz 9.5 2) zu

$$m_X(t) = e^{t_1' \mu_1 + t_2' \mu_2 + \frac{1}{2} t_1' \Sigma_{X_1} t_1 + \frac{1}{2} t_2' \Sigma_{X_2} t_2} = m_{X_1}(t_1) m_{X_2}(t_2),$$

woraus nach Satz 9.2 4) die Behauptung folgt.

9.3 Testverteilungen: χ^2 - t - und F -Verteilung

Wir behandeln im Folgenden einige Verteilungen, die aus der Normalverteilung abgeleitet sind bzw. einen Zusammenhang mit der Normalverteilung aufweisen. Diese Verteilungen haben ihre Bedeutung insbesondere als Verteilungen von Teststatistiken bei bekannten Tests in der Statistik.

9.3.1 Die χ^2 -Verteilung

Bei der χ^2 -Verteilung handelt es sich um einen Spezialfall der Gammaverteilung. Wir definieren also zunächst die Gammaverteilung und beweisen einige Eigenschaften:

Definition 9.5 (Gammaverteilung)

Sei $p > 0$ und $\lambda > 0$. Eine stetige Zufallsvariable X mit der Dichte

$$f(x) = \begin{cases} \frac{\lambda^p}{\Gamma(p)} x^{p-1} e^{-\lambda x} & x \geq 0 \\ 0 & \text{sonst} \end{cases}$$

heißt *gammaverteilt*. Schreibweise: $X \sim G(p, \lambda)$

Damit dies wohldefiniert ist, muß gezeigt werden, daß das Integral über die Dichte eins ergibt.

Beweis:

Unter Verwendung der Gammafunktion erhalten wir:

$$\begin{aligned} \int_0^{\infty} \frac{\lambda^p}{\Gamma(p)} x^{p-1} e^{-\lambda x} dx &= \frac{\lambda^p}{\Gamma(p)} \int_0^{\infty} x^{p-1} e^{-\lambda x} dx = \frac{\lambda^p}{\lambda \Gamma(p)} \underbrace{\int_0^{\infty} \left(\frac{u}{\lambda}\right)^{p-1} e^{-u} du}_{\text{subst. : } u=\lambda x} \\ &= \frac{\lambda^p}{\lambda^p \Gamma(p)} \int_0^{\infty} u^{p-1} e^{-u} du = \frac{1}{\Gamma(p)} \Gamma(p) = 1. \end{aligned}$$

□

Satz 9.9 (Eigenschaften der Gammaverteilung)

Sei $X \sim G(p, \lambda)$. Dann gilt:

1. $m_X(t) = \left(\frac{\lambda}{\lambda - t}\right)^p$
2. $E(X) = \frac{p}{\lambda}$

$$3. \operatorname{Var}(X) = \frac{p}{\lambda^2}$$

4. Sei $Y \sim G(p', \lambda)$ und seien X und Y unabhängig. Dann gilt:

$$X + Y \sim G(p + p', \lambda)$$

Beweis:

zu 1) : Unter Verwendung der Gammafunktion gilt:

$$\begin{aligned} m_X(t) &= \frac{\lambda^p}{\Gamma(p)} \int_0^\infty x^{p-1} e^{-\lambda x} e^{tx} dx = \underbrace{\frac{\lambda^p}{\Gamma(p)(\lambda-t)} \int_0^\infty \left(\frac{u}{\lambda-t}\right)^{p-1} e^{-u} du}_{\text{subst. : } u=x(\lambda-t)} \\ &= \frac{\lambda^p}{\Gamma(p)(\lambda-t)^p} \int_0^\infty u^{p-1} e^{-u} du = \left(\frac{\lambda}{\lambda-t}\right)^p \end{aligned}$$

zu 2) : Es gilt $m'_X(t) = \lambda^p p (\lambda-t)^{-(p+1)}$ und damit nach Satz 9.2 2) für den Erwartungswert:

$$E(X) = m'_X(0) = \frac{p}{\lambda}$$

zu 3) : Es gilt $m''_X(t) = \lambda^p p(p+1)(\lambda-t)^{-(p+2)}$ und damit nach Satz 9.2 2) für die Varianz:

$$\begin{aligned} \operatorname{Var}(X) &= E(X^2) - E(X)^2 = m''_X(0) - \frac{p^2}{\lambda^2} = \frac{p(p+1)}{\lambda^2} - \frac{p^2}{\lambda^2} \\ &= ds \frac{p^2 + p - p^2}{\lambda^2} = \frac{p}{\lambda^2} \end{aligned}$$

zu 4) : Anwendung von Satz 9.2 5). □

Mit Hilfe der Gammaverteilung definieren wir jetzt die χ^2 -Verteilung:

Definition 9.6 (χ^2 -Verteilung)

Eine gammaverteilte Zufallsvariable X mit $p = \frac{n}{2}$, $n \in \mathbf{N}$ und $\lambda = \frac{1}{2}$ heißt χ^2 -verteilt mit n Freiheitsgraden. Eine χ^2 -verteilte Zufallsvariable hat also die Dichte

$$f(x) = \begin{cases} \frac{1}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-\frac{1}{2}x} & x \geq 0 \\ 0 & \text{sonst} \end{cases}$$

Schreibweise: $X \sim \chi_n^2$

Aus den Eigenschaftender Gammaverteilung ergeben sich folgende Eigenschaften der χ^2 -Verteilung:

Satz 9.10 (Eigenschaften der χ^2 -Verteilung)

Sei X eine χ^2 -verteilte Zufallsvariable. Dann gilt:

1. $m_X(t) = \left(\frac{1}{1-2t}\right)^{\frac{n}{2}}$
2. $E(X) = n$
3. $\text{Var}(X) = 2n$
4. Sei $Y \sim \chi_m^2$ und seien X und Y unabhängig, dann gilt: $X + Y \sim \chi_{n+m}^2$

In den folgenden Sätzen stellen wir den Zusammenhang zur Normalverteilung her:

Satz 9.11 (Zusammenhang zur Normalverteilung)

Sei $X = (X_1, \dots, X_n)' \sim N(0, I)$. Dann ist

$$Y := \sum_{i=1}^n X_i^2 = X'X$$

χ_n^2 -verteilt.

Beweis:

Es genügt zu zeigen, daß $Z := X_1^2 \sim \chi_1^2$ verteilt ist, da dann aus dem Additionssatz 9.10 4) die Behauptung folgt. Dies kann mit Hilfe des Transformationssatzes für Dichten (Satz ??) bewiesen werden:

Mit $T(x) = x^2$ und der Zerlegung von \mathbf{R} in $G_1 =]-\infty; 0[$ und $G_2 =]0; \infty[$ sowie der Nullmenge $N = \{0\}$ gilt $T'_i \neq 0$ für alle $x \in G_i$. Damit folgt für die Dichte von Z :

$$h(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{\sqrt{z}^2}{2}} \frac{1}{2\sqrt{z}} I_{]0; \infty[} + \frac{1}{\sqrt{2\pi}} e^{-\frac{\sqrt{z}^2}{2}} \frac{1}{2\sqrt{z}} I_{]0; \infty[} = \frac{1}{\sqrt{2\pi}} z^{-\frac{1}{2}} e^{-\frac{z}{2}} I_{]0; \infty[}$$

Dies ist wegen $\Gamma(\frac{1}{2}) = \sqrt{\pi}$ die Dichte der χ_1^2 Verteilung.

□

Satz 9.12 (Zusammenhang mit der Normalverteilung)

Sei $X \sim N_n(\mu, \Sigma)$ mit $\Sigma > 0$. Dann gilt:

$$Y := (X - \mu)' \Sigma^{-1} (X - \mu) \sim \chi_n^2$$

Beweis:

Sei $Z := \Sigma^{-\frac{1}{2}}(X - \mu) = \Sigma^{-\frac{1}{2}}X - \Sigma^{-\frac{1}{2}}\mu$. Z ist nach Satz 9.6 1) multivariat normalverteilt mit

$$E(Z) = \Sigma^{-\frac{1}{2}}E(X) - \Sigma^{-\frac{1}{2}}\mu = 0$$

und

$$\text{Var}(Z) = \Sigma^{-\frac{1}{2}}\Sigma\Sigma^{-\frac{1}{2}} = \Sigma^{-\frac{1}{2}}\Sigma^{\frac{1}{2}}\Sigma^{\frac{1}{2}}\Sigma^{-\frac{1}{2}} = I,$$

d.h. $Z \sim N_n(0, \mathbf{I})$ mit unabhängigen Z_i . Die Behauptung folgt nun unter Zuhilfenahme von Satz 9.12 aus:

$$Z'Z = \left(\Sigma^{-\frac{1}{2}}(X - \mu)\right)' \Sigma^{-\frac{1}{2}}(X - \mu) = (X - \mu)' \Sigma^{-1}(X - \mu).$$

□

Satz 9.13

Sei $X \sim N_n(0, \mathbf{I})$, \mathbf{B} eine $m \times n$ ($m \leq n$) Matrix und \mathbf{R} eine symmetrische idempotente $n \times n$ Matrix mit $\text{Rg}(\mathbf{R}) = r$. Dann gilt:

1. $X'\mathbf{R}X \sim \chi_r^2$
2. Aus $\mathbf{B}\mathbf{R} = 0$ folgt die Unabhängigkeit der quadratischen Form $X'\mathbf{R}X$ von der Linearform $\mathbf{B}X$.

Beweis:

zu 1) : Wegen Korollar 6.3 existiert eine orthogonale Matrix \mathbf{P} mit $\mathbf{P}'\mathbf{R}\mathbf{P} = \Lambda$, wobei

$$\Lambda = \begin{pmatrix} \mathbf{I}_r & 0 \\ 0 & 0 \end{pmatrix}.$$

Sei $Y := \mathbf{P}'X$, d.h. $X = \mathbf{P}Y$. Offensichtlich ist $Y \sim N_n(0, \mathbf{I})$ verteilt. Mit der Partitionierung $Y' = (Y_1' \ Y_2')$, wobei Y_1 ein $r \times 1$ und Y_2 ein $(n - r) \times 1$ Spaltenvektor ist, gilt:

$$\begin{aligned} X'\mathbf{R}X &= Y'\mathbf{P}'\mathbf{R}\mathbf{P}Y = (Y_1' \ Y_2') \begin{pmatrix} \mathbf{I}_r & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \\ &= (Y_1'\mathbf{I}_r \ 0) \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = Y_1'Y_1 \sim \chi_r^2 \end{aligned}$$

Dabei wurde im letzten Schritt auf Satz 9.12 zurückgegriffen.

zu 2) : Ziel ist es $\mathbf{B}X$ als Linearkombination des Vektors Y_2 darzustellen. Wegen $X'\mathbf{R}X = Y_1'Y_1$ und der Unabhängigkeit von Y_1 und Y_2 sind dann auch $\mathbf{B}X$ und $X'\mathbf{R}X$ unabhängig. Es gilt:

$$\begin{aligned}
 \mathbf{BR} &= 0 && \iff \\
 \mathbf{BI}_n\mathbf{R} &= 0 && \iff \\
 \mathbf{BPP}'\mathbf{R} &= 0 && \iff \\
 \mathbf{BPP}'\mathbf{RP} &= 0 && \iff \\
 \mathbf{BP} \begin{pmatrix} \mathbf{I}_r & 0 \\ 0 & 0 \end{pmatrix} &= 0 &&
 \end{aligned}$$

Nun folgt mit Hilfe der Partitionierung von $\mathbf{C} := \mathbf{BP}$ in eine $m \times r$ Matrix \mathbf{C}_1 und eine $m \times (n-r)$ Matrix \mathbf{C}_2 :

$$0 = \mathbf{C} \begin{pmatrix} \mathbf{I}_r & 0 \\ 0 & 0 \end{pmatrix} = (\mathbf{C}_1 \ \mathbf{C}_2) \begin{pmatrix} \mathbf{I}_r & 0 \\ 0 & 0 \end{pmatrix} = \mathbf{C}_1$$

Damit ist wegen $\mathbf{C} = (\mathbf{C}_1 \ \mathbf{C}_2)$ und $\mathbf{C}_1 = 0$ gezeigt, daß \mathbf{C} die Gestalt

$$\mathbf{C} = (0 \ \mathbf{C}_2)$$

besitzt. Die Behauptung folgt nun aus

$$\mathbf{BX} = \mathbf{BPY} = \mathbf{CY} = (0 \ \mathbf{C}_2) \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \mathbf{C}_2 Y_2.$$

□

Satz 9.14

Seien X_1, \dots, X_n unabhängige Zufallsvariablen mit $X_i \sim N(\mu, \sigma^2)$ und sei S^2 gegeben durch

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2.$$

Dann gilt:

1. $\frac{n-1}{\sigma^2} S^2 \sim \chi_{n-1}^2$
2. S^2 und \bar{X} sind unabhängig.

Beweis:

Der Beweis geschieht durch eine Übertragung auf den allgemeineren Fall in Satz 9.13. Dazu verwenden wir im Folgenden die idempotente Matrix

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

aus den Beispielen 1.12 und 6.5. Für $\mathbf{B} := \frac{1}{n} \mathbf{1}'$ gilt wegen 1.12 4) $\mathbf{BC} = 0$.

Sei nun $Z_i := \frac{X_i - \mu}{\sigma}$. Z_i ist standardnormalverteilt und $Z := (Z_1, \dots, Z_n) \sim N_n(0, \mathbf{I})$. Wegen 1.12 6) erhalten wir

$$Z' \mathbf{C} Z = \sum_{i=1}^n (Z_i - \bar{Z})^2.$$

Weiter gilt

$$\mathbf{B} Z = \frac{1}{n} \mathbf{1}' Z = \bar{Z}.$$

Damit sind die Voraussetzungen von Satz 9.13 erfüllt und wir erhalten: $\sum_{i=1}^n (Z_i - \bar{Z})^2$ ist χ_{n-1}^2 verteilt und ist unabhängig von \bar{Z} . Daraus folgt die Behauptung für X unter Beachtung von $\bar{X} = \sigma \bar{Z} + \mu$ und $S^2 = \frac{\sigma^2}{n-1} \sum_{i=1}^n (Z_i - \bar{Z})^2$. □

9.3.2 Die t-Verteilung

Definition 9.7 (t-Verteilung)

Sei X eine standardnormalverteilte Zufallsvariable und $Y \chi_n^2$ verteilt und seien X und Y unabhängig, dann heißt

$$T = \frac{X}{\sqrt{\frac{Y}{n}}}$$

t-verteilt mit n Freiheitsgraden. Schreibweise: $T \sim t_n$

Satz 9.15 (von Student)

Seien X_1, \dots, X_n unabhängige $N(\mu, \sigma^2)$ verteilte Zufallsvariablen. Dann gilt:

$$\frac{\bar{X} - \mu}{S} \sqrt{n} \sim t_{n-1}$$

Beweis:

Sei $U := \frac{\bar{X} - \mu}{\sigma} \sqrt{n}$ und $V := \frac{n-1}{\sigma^2} S^2$. Unter Verwendung von Satz 9.3 5) ist $U \sim N(0, 1)$ verteilt, während V nach Satz 9.14 1) χ_{n-1}^2 -verteilt ist. Nach Definition der t-Verteilung gilt:

$$\frac{U}{\sqrt{\frac{V}{n-1}}} \sim t_{n-1}$$

Die Behauptung folgt nun aus der folgenden Umformung:

$$\frac{U}{\sqrt{\frac{V}{n-1}}} = \frac{\frac{\bar{X} - \mu}{\sigma} \sqrt{n}}{\sqrt{\frac{\frac{n-1}{\sigma^2} S^2}{n-1}}} = \frac{\bar{X} - \mu}{S} \sqrt{n}.$$

□

9.3.3 Die F -Verteilung

Definition 9.8 (F -Verteilung)

Seien $X_1 \sim \chi_n^2$ und $X_2 \sim \chi_m^2$ unabhängige Zufallsvariablen. Dann heißt

$$F := \frac{X_1}{X_2} \frac{m}{n}$$

F -verteilt mit n und m Freiheitsgraden. Schreibweise: $F \sim F_{n,m}$

Satz 9.16

Sei $X \sim N_n(0, I)$ und seien die $n \times n$ Matrizen \mathbf{R} und \mathbf{S} symmetrisch und idempotent mit $\text{rg}(\mathbf{R}) = r$ und $\text{rg}(\mathbf{S}) = s$, und es gelte $\mathbf{RS} = 0$. Dann gilt:

1. $X'\mathbf{R}X$ und $X'\mathbf{S}X$ sind unabhängig.

2. $\frac{s}{r} \frac{X'\mathbf{R}X}{X'\mathbf{S}X} \sim F_{r,s}$

Beweis :

zu 1) : Wegen Korollar 6.3 existiert eine orthogonale Matrix \mathbf{P} mit

$$\mathbf{P}'\mathbf{R}\mathbf{P} = \Lambda,$$

wobei

$$\Lambda = \begin{pmatrix} \mathbf{I}_r & 0 \\ 0 & 0 \end{pmatrix}.$$

Sei analog zum Beweis von Satz 9.13 $Y := \mathbf{P}'X$ und Y partitioniert mit $Y = (Y_1 \ Y_2)$. Wie im Beweis von Satz 9.13 1) gezeigt, gilt:

$$X'\mathbf{R}X = Y_1'Y_1$$

Ziel ist es nun wiederum $X'\mathbf{S}X$ in einen Ausdruck umzuformen, der nur von Y_2 (und nicht von Y_1) abhängt, da dann wegen der Unabhängigkeit von Y_1 und Y_2 die Behauptung folgt.

Sei also $\mathbf{G} := \mathbf{P}'\mathbf{S}\mathbf{P}$. Offensichtlich ist \mathbf{G} symmetrisch und es gilt:

$$\mathbf{G}\Lambda = \mathbf{P}'\mathbf{S}\mathbf{P}\mathbf{P}'\mathbf{R}\mathbf{P} = \mathbf{P}'\mathbf{S}\mathbf{R}\mathbf{P} = \mathbf{P}'(\mathbf{RS})'\mathbf{P} = 0$$

In partitionierter Form läßt sich dies darstellen als

$$\begin{pmatrix} \mathbf{G}_{11} & \mathbf{G}_{12} \\ \mathbf{G}_{21} & \mathbf{G}_{22} \end{pmatrix} \begin{pmatrix} \mathbf{I}_r & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix},$$

woraus sofort $\mathbf{G}_{11} = 0$ und $\mathbf{G}_{12} = \mathbf{G}_{21} = 0$ folgt. \mathbf{G} besitzt also folgende Gestalt:

$$\mathbf{G} = \begin{pmatrix} 0 & 0 \\ 0 & \mathbf{G}_{22} \end{pmatrix}$$

Nun folgt die Behauptung aus folgender Umformung:

$$X'\mathbf{S}X = Y'\mathbf{P}'\mathbf{S}\mathbf{P}Y = Y'\mathbf{G}Y = Y_2'\mathbf{G}_{22}Y_2$$

zu 2) : Wegen Satz 9.13 1) ist $X'\mathbf{R}X$ χ_r^2 verteilt und $X'\mathbf{S}X$ χ_s^2 verteilt. Aufgrund der in 1) gezeigten Unabhängigkeit folgt die Behauptung direkt aus der Definition der F-Verteilung.

□

Das klassische lineare Regressionsmodell

In diesem Kapitel befassen wir uns mit dem linearen Regressionsmodell. Die lineare Regression spielt zweifellos eine Hauptrolle in der Statistik. Allgemein geht es bei Regressionsmodellen um die Beschreibung des Zusammenhangs zwischen einer primär interessierenden Variable Y und einer Reihe von beeinflussenden Variablen X_1, \dots, X_K . Wie üblich in der Statistik ist der Zusammenhang nicht exakt, sondern vielmehr überlagert von unbeobachteten zufälligen Störgrößen. Dieses Kapitel soll keine umfassende Abhandlung linearer Regressionsmodelle bieten, sondern als (durchaus komplexe) Anwendung der Matrixalgebra dienen.

10.1 Modelldefinition

Gegeben sei eine primär interessierende Variable Y und eine Menge $X = (X_1, \dots, X_K)'$ von sogenannten *Kovariablen* (auch *unabhängige Variablen*). Y heißt *Responsevariable* (kurz: *Response*) oder auch *abhängige Variable*. Man nimmt an, daß ein funktionaler Zusammenhang zwischen Y und den Kovariablen besteht, d.h.

$$Y = f(X) = f(X_1, \dots, X_K).$$

Im Rahmen der linearen Modelle wird speziell von einem linearen Zusammenhang zwischen Y und X ausgegangen, d.h.

$$Y = \beta_1 X_1 + \dots + \beta_K X_K. \quad (10.1)$$

In der Regel gilt der Zusammenhang nicht exakt, sondern wird durch eine zufällige Störgröße ε kontaminiert/überlagert/gestört wird. Wir gehen im Folgenden von einer *additiven Überlagerung* des Zusammenhangs zwischen Y und X aus, d.h. das Modell (10.1) wird zu

$$Y = \beta_1 X_1 + \dots + \beta_K X_K + \varepsilon.$$

Aufgabe der Statistik ist es die Art und Weise des Zusammenhangs zu bestimmen. Dies ist gleichbedeutend mit der geeigneten Schätzung des Parametervektors $\beta = (\beta_1, \dots, \beta_K)'$.

Zu diesem Zweck werden Daten y_t und $x_t = (x_{t1}, \dots, x_{tK})'$, $t = 1, \dots, T$, erhoben, so daß man für jeden Beobachtungspunkt die Gleichung

$$y_t = \beta_1 x_{t1} + \dots + \beta_K x_{tK} + \varepsilon_t \quad (10.2)$$

erhält. Definiert man die $T \times 1$ Vektoren

$$y = \begin{pmatrix} y_1 \\ \vdots \\ y_T \end{pmatrix} \quad \text{und} \quad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_T \end{pmatrix},$$

sowie die *Designmatrix* \mathbf{X} der Dimension $T \times K$

$$\mathbf{X} = \begin{pmatrix} x_{11} & \cdots & x_{1K} \\ \vdots & & \vdots \\ x_{T1} & \cdots & x_{TK} \end{pmatrix},$$

so lassen sich die T Gleichungen aus (10.2) kompakt in Matrixnotation schreiben:

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

Im Rahmen des klassischen linearen Modells werden über den Vektor ε der Störgrößen folgende Annahmen getroffen:

- $E(\varepsilon) = 0$, d.h. die Störungen sind im Mittel Null;
- $E(\varepsilon\varepsilon') = \text{Cov}(\varepsilon) = \sigma^2\mathbf{I}$, d.h. die Varianz der Störgrößen bleibt konstant und die Störungen sind von Beobachtung zu Beobachtung unkorreliert;

Für die Designmatrix \mathbf{X} nehmen wir zusätzlich an, dass

- \mathbf{X} nichtstochastisch ist und
- $\text{rg}(\mathbf{X}) = K$, d.h. \mathbf{X} hat vollen Spaltenrang bzw. ist spaltenregulär.

Insgesamt erhalten wir das *klassische lineare Regressionsmodell*:

1. $y = \mathbf{X}\beta + \varepsilon$
2. $E(\varepsilon) = 0$
3. $E(\varepsilon\varepsilon') = \sigma^2\mathbf{I}$
4. \mathbf{X} ist nichtstochastisch und besitzt vollen Spaltenrang.

Als einfache Folgerungen erhält man

$$E(y) = E(\mathbf{X}\beta + \varepsilon) = \mathbf{X}\beta + E(\varepsilon) = \mathbf{X}\beta$$

und

$$\text{Cov}(y) = \text{Cov}(\mathbf{X}\beta + \varepsilon) = \text{Cov}(\varepsilon) = \sigma^2 \mathbf{I}.$$

Beispiel 10.1 (univariates Regressionsmodell)

Einen wichtigen Spezialfall des linearen Modells stellt das univariate Regressionsmodell dar, das eine Konstante und nur eine unabhängige Variable X enthält:

$$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t \quad (t = 1, \dots, T)$$

Die Designmatrix hat in diesem Fall die Gestalt

$$X = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_T \end{pmatrix}.$$

△

Beispiel 10.2 (multiples Regressionsmodell mit Intercept)

Das multiple Regressionsmodell mit konstantem Glied (sogenannter Intercept) ist gegeben durch

$$y_t = \beta_0 + \beta_1 x_{t1} + \dots + \beta_K x_{tK} \quad (t = 1, \dots, T)$$

Für die Designmatrix X gilt in diesem Fall

$$X = \begin{pmatrix} 1 & x_{11} & \dots & x_{1K} \\ \vdots & \vdots & & \vdots \\ 1 & x_{T1} & \dots & x_{TK} \end{pmatrix}.$$

△

Beispiel 10.3 (nichtlineare Beziehungen)

Im Rahmen der linearen Modelle können durchaus auch nichtlineare Beziehungen zwischen der abhängigen Variable und den Kovariablen behandelt werden. Betrachte zum Beispiel das folgende Modell

$$y_t = f(z_t) + \varepsilon_t = \beta_0 + \beta_1 z_t + \beta_2 z_t^2 + \beta_3 z_t^3 + \varepsilon_t,$$

indem die Funktion f ein Polynom dritten Grades ist. Wir können dieses Modell auf ein einfaches lineares Modell zurückführen, indem wir die Variablen $x_{1t} := z_t$, $x_{2t} := z_t^2$ und

$x_{3t} := z_t^3$ definieren. Damit erhalten wir wieder ein lineares Modell. In Abhängigkeit der Beobachtungen z_t ergibt sich die Designmatrix zu

$$X = \begin{pmatrix} 1 & z_1 & z_1^2 & z_1^3 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & z_T & z_T^2 & z_T^3 \end{pmatrix}.$$

Im Allgemeinen lassen sich alle nichtlinearen Beziehungen auf ein einfaches lineares Modell zurückführen, solange sie linear in den Parametern sind. Ein Beispiel für ein echtes nichtlineares Modell ist gegeben durch

$$y_t = f(z_t) + \varepsilon_t = \beta_0 + \beta_1 \sin(\beta_2 z_t) + \varepsilon_t.$$

△

10.2 Die Methode der kleinsten Quadrate

In diesem Abschnitt stellen wir die Methode der kleinsten Quadrate dar, mit deren Hilfe aus den Daten $y_t, x_{1t}, \dots, x_{Kt}$, $t = 1, \dots, T$, die unbekannt Parameter $\beta = (\beta_1, \dots, \beta_K)'$ geschätzt werden können.

Wir bezeichnen im Folgenden den geschätzten Parametervektor mit $\hat{\beta} = (\hat{\beta}_1, \dots, \hat{\beta}_K)'$. Diese Unterscheidung ist notwendig, da es wohl kaum je gelingen wird den "wahren" Parametervektor β ohne Fehler zu schätzen, so dass stets $\beta \neq \hat{\beta}$ gilt. Mit Hilfe des geschätzten Parametervektors können wir dann auch für jede der T Beobachtungen einen Schätzwert \hat{y}_t für y_t bestimmen. Es liegt nahe

$$\hat{y}_t = \hat{\beta}_1 x_{1t} + \dots + \hat{\beta}_K x_{Kt}$$

zu verwenden. Der Schätzfehler, d.h. die Abweichung des wahren Wertes y_t vom Schätzwert \hat{y}_t heißt Residuum und wird mit $\hat{\varepsilon}_t$ bezeichnet. Es gilt

$$\hat{\varepsilon}_t := y_t - \hat{y}_t = y_t - x_t' \hat{\beta}_t.$$

Es ist wichtig zu verstehen, dass die Residuen $\hat{\varepsilon}_t$ nicht gleich den Störgrößen ε_t sind, die wie der Parametervektor β unbekannt sind. Vielmehr können die Residuen $\hat{\varepsilon}_t$ als Schätzungen für ε_t angesehen werden.

Unser Ziel ist es, den Vektor $\hat{\beta}$ der geschätzten Regressionsparameter so zu bestimmen, daß die Abweichungen

$$y_t - \hat{y}_t = y_t - x_t' \hat{\beta} = \hat{\varepsilon}$$

der Schätzwerte \hat{y}_t von den wahren Werten y_t in ihrer Gesamtheit möglichst klein werden. Im folgenden bezeichne $\tilde{\beta}$ ein beliebige Wahl für den Vektor der Regressionskoeffizienten. Unter anderem sind folgende Minimierungsansätze denkbar:

1. $S_1(\tilde{\beta}) := \sum_{t=1}^T |y_t - x_t' \tilde{\beta}| = \sum_{t=1}^T |\tilde{\varepsilon}_t| = \|\tilde{\varepsilon}\|_1 \longrightarrow \min_{\tilde{\beta}}$
2. $S_2(\tilde{\beta}) := \sum_{t=1}^T (y_t - x_t' \tilde{\beta})^2 = \sum_{t=1}^T \tilde{\varepsilon}_t^2 = \tilde{\varepsilon}' \tilde{\varepsilon} = \|\tilde{\varepsilon}\|_2^2 \longrightarrow \min_{\tilde{\beta}}$
3. $S_\infty(\tilde{\beta}) := \max_{1 \leq t \leq T} |y_t - x_t' \tilde{\beta}| = \max_{1 \leq t \leq T} |\tilde{\varepsilon}_t| = \|\tilde{\varepsilon}\|_\infty \longrightarrow \min_{\tilde{\beta}}$

Der zweite Ansatz ist als Methode der kleinsten Quadrate (KQ-Methode) bekannt und wird am häufigsten verwendet. Einer der Hauptgründe dabei ist sicherlich, dass die Methode der kleinsten Quadrate mathematisch am besten handhabbar ist. Darüberhinaus besitzt der nach der KQ-Methode gewonnene Schätzer eine Reihe wünschenswerter statistischer Eigenschaften, vergleiche Abschnitt 10.5.

Bei der Bestimmung des Minimums von $S_2(\tilde{\beta})$ wollen wir zunächst noch nicht voraussetzen, dass \mathbf{X} vollen Spaltenrang besitzt. Wir formen zunächst $S_2(\tilde{\beta})$ um:

$$\begin{aligned} S_2(\tilde{\beta}) &= \tilde{\varepsilon}' \tilde{\varepsilon} & (10.3) \\ &= (y - \mathbf{X}\tilde{\beta})'(y - \mathbf{X}\tilde{\beta}) \\ &= y'y - \tilde{\beta}' \mathbf{X}' y - y' \mathbf{X} \tilde{\beta} + \tilde{\beta}' \mathbf{X}' \mathbf{X} \tilde{\beta} \\ &= y'y - 2y' \mathbf{X} \tilde{\beta} + \tilde{\beta}' \mathbf{X}' \mathbf{X} \tilde{\beta}. \end{aligned}$$

Man überzeugt sich leicht, daß alle vorkommenden Größen Skalare sind. Unter Verwendung von Satz 8.2 7) und 8.2 10) erhalten wir

$$\frac{\partial S(\tilde{\beta})}{\partial \tilde{\beta}} = -2 \mathbf{X}' y + 2 \mathbf{X}' \mathbf{X} \tilde{\beta}. \quad (10.4)$$

sowie

$$\frac{\partial^2 S(\tilde{\beta})}{\partial \tilde{\beta} \partial \tilde{\beta}'} = 2 \mathbf{X}' \mathbf{X}.$$

Da die Hesse Matrix nach Satz 7.4 nichtnegativ definit ist, erhalten wir die Lösungen $\hat{\beta}$ des Minimierungsproblems $S(\tilde{\beta}) \longrightarrow \min_{\tilde{\beta}}$ durch Nullsetzen von (10.4). Wir erhalten also $\hat{\beta}$ als Lösungen der sogenannten Normalgleichungen

$$\mathbf{X}' \mathbf{X} \hat{\beta} = \mathbf{X}' y.$$

Im folgenden Satz zeigen wir, dass die Normalgleichungen immer lösbar sind und bestimmen die Lösungsmenge:

Satz 10.1 (Lösungen der Normalgleichungen)

Die Normalgleichungen $\mathbf{X}'\mathbf{X}\tilde{\beta} = \mathbf{X}'y$ sind stets lösbar mit der Lösungsmenge

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'y + (\mathbf{I} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X})w, \quad (10.5)$$

wobei $w \in \mathbf{R}^K$ beliebig ist.

Beweis

Allgemein ist nach Satz 5.6 ein Gleichungssystem $\mathbf{A}x = c$ genau dann lösbar, wenn $\mathbf{A}\mathbf{A}^{-1}c = c$ gilt. Das heißt, daß $\mathbf{X}'\mathbf{X}\tilde{\beta} = \mathbf{X}'y$ genau dann lösbar ist, wenn $\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'y = \mathbf{X}'y$. Zunächst gilt (aufgrund der Definition der verallgemeinerten Inversen) $\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X} = \mathbf{X}'\mathbf{X}$. Daraus folgt unter Zuhilfenahme der Kürzungsregel (vgl. Satz 1.4) die Beziehung $\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X} = \mathbf{X}$. Durch transponieren auf beiden Seiten erhalten wir $\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' = \mathbf{X}'$ und schließlich durch Rechtsmultiplikation von y die gewünschte Beziehung $\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'y = \mathbf{X}'y$. Die Normalgleichungen sind also stets lösbar. Die Lösungsmenge erhält man wieder durch Anwendung von Satz 5.6 auf den vorliegenden Fall. □

Bemerkung

Besitzt die Matrix \mathbf{X} vollen Spaltenrang, d.h. $\text{rg}(\mathbf{X}) = K$, so ist $\mathbf{X}'\mathbf{X}$ regulär und damit invertierbar. In diesem Fall sind die Normalgleichungen eindeutig lösbar und es gilt

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'y. \quad (10.6)$$

Um den Lösungsvektor $\hat{\beta}$ zu berechnen, ist es natürlich nicht notwendig die Inverse von $\mathbf{X}'\mathbf{X}$ zu berechnen, es genügt die Normalgleichungen zu lösen. Man beachte, dass zur Berechnung der Inversen mehrere Gleichungssysteme gelöst werden müssen, die Berechnung von $\hat{\beta}$ über (10.6) ist also wenig effizient. Bei Gleichung (10.6) handelt es sich lediglich um eine kompakte Darstellung des KQ-Schätzers. Zur Berechnung des KQ-Schätzers (10.6) kann man beispielsweise das in Abschnitt 7.2 auf Seite 129 dargestellte Verfahren zur Lösung linearer Gleichungssysteme verwenden, das auf der Choleskyzerlegung beruht.

Satz 10.2

Für zwei Lösungen $\hat{\beta}_1$ und $\hat{\beta}_2$ der Normalgleichungen gilt die Beziehung

$$\mathbf{X}\hat{\beta}_1 = \mathbf{X}\hat{\beta}_2.$$

Beweis

Jede Lösung der Normalgleichungen $\hat{\beta}$ hat die Gestalt (10.5) und wir erhalten

$$\begin{aligned} X\hat{\beta} &= \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'y + \mathbf{X}(\mathbf{I} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X})w \\ &= \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'y + (\mathbf{X} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X})w. \end{aligned}$$

Im Beweis von Satz 10.1 haben wir mit Hilfe der Kürzungsregel 1.4 gesehen, dass $\mathbf{X} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}$. Damit erhalten wir

$$X\hat{\beta} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'y,$$

d.h. $X\hat{\beta}$ ist unabhängig von w woraus die Behauptung folgt. □

Zur Interpretation des Satzes:

Selbst wenn die Normalgleichungen nicht eindeutig lösbar sind, so sind zumindest die geschätzten Werte

$$\hat{y} = \mathbf{X}\hat{\beta}_1 = \mathbf{X}\hat{\beta}_2$$

für alle Lösungen der Normalgleichungen gleich.

Beispiel 10.4 (univariate Regression)

Für das Modell

$$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t \quad t = 1, \dots, T$$

haben wir in Beispiel 10.1 für die Designmatrix

$$\mathbf{X} = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_T \end{pmatrix}$$

erhalten. Eine einfache Rechnung ergibt

$$\mathbf{X}'\mathbf{X} = \begin{pmatrix} T & \sum x_t \\ \sum x_t & \sum x_t^2 \end{pmatrix},$$

sowie

$$\mathbf{X}'y = \begin{pmatrix} \sum y_t \\ \sum x_t y_t \end{pmatrix}.$$

Damit erhalten wir für die die Normalgleichungen

- I. $T\tilde{\beta}_0 + \sum x_t \tilde{\beta}_1 = \sum y_t$
- II. $\sum x_t \tilde{\beta}_0 + \sum x_t^2 \tilde{\beta}_1 = \sum x_t y_t.$

bzw.

$$\begin{aligned} \text{I.} \quad & \tilde{\beta}_0 + \tilde{\beta}_1 \bar{x} = \bar{y} \\ \text{II.} \quad & b_0 \bar{x} + \tilde{\beta}_1 \bar{x}^2 = \bar{x}\bar{y}. \end{aligned}$$

Durch Auflösen erhalten wir als Lösungen

$$\begin{aligned} \hat{\beta}_0 &= \bar{y} - \frac{\sum (x_t - \bar{x})(y_t - \bar{y})}{\sum (x_t - \bar{x})^2} \bar{x} \\ \hat{\beta}_1 &= \frac{\sum (x_t - \bar{x})(y_t - \bar{y})}{\sum (x_t - \bar{x})^2}, \end{aligned}$$

bzw. mit den Abkürzungen

$$\begin{aligned} S_{XX} &= \sum_{t=1}^T (x_t - \bar{x})^2 = \sum_{t=1}^T x_t^2 - T \bar{x}^2 \\ S_{YY} &= \sum_{t=1}^T (y_t - \bar{y})^2 = \sum_{t=1}^T y_t^2 - T \bar{y}^2 \\ S_{XY} &= \sum_{t=1}^T (x_t - \bar{x})(y_t - \bar{y}) = \sum_{t=1}^T y_t x_t - T \bar{y} \bar{x} \end{aligned}$$

die beiden einprägsameren Formeln

$$\begin{aligned} \tilde{\beta}_0 &= \bar{y} - \frac{S_{XY}}{S_{XX}} \bar{x} \\ \tilde{\beta}_1 &= \frac{S_{XY}}{S_{XX}}. \end{aligned}$$

△

10.3 Die KQ-Methode unter linearen Nebenbedingungen

Gelegentlich kann der Vektor β der unbekannt Parameter nicht jeden beliebigen Wert des \mathbb{R}^K annehmen, sondern nur einen eingeschränkten Wertebereich. Ziel dieses Abschnittes ist die Bestimmung des KQ-Schätzers im Modell

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

wenn für β die lineare Restriktion

$$\underset{(J \times K)}{\mathbf{R}} \quad \underset{(K \times 1)}{\beta} = \underset{(J \times 1)}{r}$$

mit $\text{rg}(R) = J$ gilt (d.h. es müssen J linear unabhängige Nebenbedingungen erfüllt sein). Zusätzlich sei in diesem Abschnitt vorausgesetzt, daß die Designmatrix \mathbf{X} vollen Spaltenrang besitzt, d.h. $\text{rg}(\mathbf{X}) = K$. Wir werden die KQ-Methode unter linearen Nebenbedingungen insbesondere auch in Abschnitt 11.3 beim Testen von linearen Hypothesen über die Regressionsparameter brauchen. Zunächst folgen einige Beispiele für lineare Restriktionen.

Beispiel 10.5 (lineare Restriktionen)

1. Kenntnis einer Komponente, z. B. $\beta_i = \beta_i^*$. In diesem Fall erhält man

$$\underset{(1 \times 1)}{r} = \beta_i^*$$

sowie für die Matrix \mathbf{R}

$$\underset{(1 \times K)}{\mathbf{R}} = (0, \dots, 0, \underset{\uparrow}{1}, 0, \dots, 0).$$

2. Kenntnis eines Subvektors, O.B.d.A. seien die ersten J Komponenten von β bekannt, d.h. $\beta_i = \beta_i^*$ $i = 1, \dots, J$. Wir setzen $\beta_1^* := (\beta_1^*, \dots, \beta_J^*)'$. Dann erhalten wir

$$\underset{(J \times 1)}{r} = \beta_1^*$$

und

$$\underset{(J \times K)}{\mathbf{R}} = \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & & \ddots & & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 1 & 0 & \cdots & 0 \end{pmatrix} \begin{matrix} \text{Zeile 1} \\ \\ \\ \text{Zeile } J \end{matrix}.$$

3. Gleichheit zweier Komponenten, z.B. sei $\beta_1 = \beta_2$ bzw. $\beta_1 - \beta_2 = 0$. Dann erhalten wir

$$\underset{(1 \times 1)}{r} = 0$$

sowie

$$\underset{(1 \times K)}{\mathbf{R}} = (1, -1, 0, \dots, 0).$$

4. Gleichheit dreier Komponenten, z.B. sei $\beta_1 = \beta_2 = \beta_3$. Diese Bedingung ist äquivalent zum Gleichungssystem

$$\begin{aligned} \beta_1 - \beta_2 &= 0 \\ \beta_2 - \beta_3 &= 0 \end{aligned}$$

und wir erhalten

$$\underset{(2 \times 1)}{r} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

und

$$\underset{(2 \times K)}{\mathbf{R}} = \begin{pmatrix} 1 & -1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & -1 & 0 & \cdots & 0 \end{pmatrix}$$

Man beachte, dass die Darstellung linearer Restriktionen i.a. *nicht* eindeutig ist. Zur Verdeutlichung betrachten wir folgendes Beispiel.

Beispiel 10.6

Die Nebenbedingung

$$\beta_1 = 1 \quad r = 1 \quad \mathbf{R} = (1, 0, \dots, 0)$$

ist offensichtlich äquivalent zu

$$2\beta_1 = 2 \quad r = 2 \quad \mathbf{R} = (2, 0, \dots, 0).$$

△

Im Folgenden soll nun der KQ-Schätzer unter linearen Nebenbedingungen berechnet werden. Ziel ist die Minimierung der Zielfunktion

$$S_2(\tilde{\beta}) = (y - \mathbf{X}\tilde{\beta})'(y - \mathbf{X}\tilde{\beta}) = y'y - 2y'\mathbf{X}\tilde{\beta} + \tilde{\beta}'\mathbf{X}'\mathbf{X}\tilde{\beta}$$

unter der Nebenbedingung

$$\mathbf{R}\tilde{\beta} = r.$$

Wir beweisen folgenden Satz.

Satz 10.3 (KQ-Schätzer unter linearen Restriktionen)

Der KQ-Schätzer $\hat{\beta}^R$ unter linearen Restriktionen ist gegeben durch

$$\hat{\beta}^R = \hat{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'(\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}')^{-1}(r - \mathbf{R}\hat{\beta}),$$

wobei $\hat{\beta}$ der unrestringierte KQ-Schätzer ist.

Beweis:

Die Lösung erfolgt mit dem sogenannten *Lagrange-Ansatz* (vergleiche z.B. Forster 1999):

$$\begin{aligned} S_2(\tilde{\beta}; \lambda) &= S_2(\tilde{\beta}) - 2\lambda'(\mathbf{R}\tilde{\beta} - r) \\ &= y'y - 2y'\mathbf{X}\tilde{\beta} + \tilde{\beta}'\mathbf{X}'\mathbf{X}\tilde{\beta} - 2\lambda'\mathbf{R}\tilde{\beta} + 2\lambda'r, \end{aligned}$$

mit einem Spaltenvektor λ von Lagrange-Multiplikatoren der Dimension $(J \times 1)$. Unter Verwendung von Satz 8.2 7) und 8.2 10) erhalten wir:

$$\begin{aligned} \frac{\partial S_2(\tilde{\beta}; \lambda)}{\partial \tilde{\beta}} &= -2\mathbf{X}'y + 2\mathbf{X}'\mathbf{X}\tilde{\beta} - 2\mathbf{R}'\lambda \\ \frac{\partial S_2(\tilde{\beta}; \lambda)}{\partial \lambda} &= -2\mathbf{R}\tilde{\beta} + 2r \end{aligned}$$

Nullsetzen liefert die folgenden beiden Gleichungen:

$$\text{I.} \quad \mathbf{X}'\mathbf{X}\tilde{\beta} - \mathbf{X}'y = \mathbf{R}'\lambda$$

$$\text{II.} \quad \mathbf{R}\tilde{\beta} = r$$

Wir lösen beide Gleichungen zuerst nach λ und anschließend nach $\tilde{\beta}$ auf. Multiplikation von I. mit $(\mathbf{X}'\mathbf{X})^{-1}$ von links liefert

$$\tilde{\beta} - \hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'\lambda.$$

Multiplikation dieser Gleichung mit der Matrix \mathbf{R} von links ergibt

$$\mathbf{R}\tilde{\beta} - \mathbf{R}\hat{\beta} = \mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'\lambda.$$

Durch Einsetzen von II erhalten wir

$$r - \mathbf{R}\hat{\beta} = \mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'\lambda$$

und damit

$$\lambda = (\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}')^{-1}(r - \mathbf{R}\hat{\beta}).$$

Dabei haben wir ausgenutzt, dass die Matrix $\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'$ gemäß Satz 7.3 2) positive definit ist und damit invertierbar.

Einsetzen von λ in I liefert

$$\mathbf{X}'\mathbf{X}\tilde{\beta} - \mathbf{X}'y = \mathbf{R}'(\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}')^{-1}(r - \mathbf{R}\hat{\beta})$$

und schließlich

$$\tilde{\beta} = \hat{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'(\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}')^{-1}(r - \mathbf{R}\hat{\beta}).$$

□

Im folgenden Beispiel betrachten wir noch einen wichtigen Spezialfall:

Beispiel 10.7

Sei β unterteilt in zwei Subvektoren β_1 und β_2 , d.h. $\beta = (\beta_1', \beta_2')'$. Betrachte die Restriktion $\beta_1 = 0$. Dann läßt sich zeigen, daß die restriktive KQ-Schätzung von β_2 mit der gewöhnlichen KQ-Schätzung im reduzierten Modell $y = \mathbf{X}_2\beta_2 + \varepsilon$ übereinstimmt, d.h.

$$\hat{\beta}_2^R = (\mathbf{X}_2'\mathbf{X}_2)^{-1}\mathbf{X}_2'y,$$

wobei \mathbf{X}_2 aus den Spalten von \mathbf{X} besteht, die β_2 betreffen.

△

10.4 Geschätzte Werte, Residuen, Streuungszerlegung

10.4.1 Geschätzte Werte

Eine naheliegende (*ex post*-)Vorhersage \hat{y} von y ist

$$\hat{y} = X\hat{\beta},$$

wobei wie bisher $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'y$ die gewöhnliche KQ-Schätzung für β bezeichnet. Für \hat{y} gilt

$$\hat{y} = \mathbf{X}\hat{\beta} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'y = \mathbf{P}y,$$

wobei die $T \times T$ Matrix

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

die sogenannte *Prediction-Matrix* oder auch *Hat-Matrix* ist. Es gilt:

Satz 10.4 (Eigenschaften der Hat-Matrix)

Die Hat-Matrix \mathbf{P} besitzt folgende Eigenschaften:

1. \mathbf{P} ist symmetrisch.
2. \mathbf{P} ist idempotent.
3. $rg(\mathbf{P}) = sp(\mathbf{P}) = K$.

Beweis:

Die Eigenschaften 1) und 2) sieht man durch einfaches nachrechnen. Die Eigenschaft 3) folgt unter Zuhilfenahme von Korrolar 6.3 und Satz 4.7 4) aus

$$rg(\mathbf{P}) = sp(\mathbf{P}) = sp(\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}') = sp(\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}) = sp(\mathbf{I}_K) = K.$$

□

10.4.2 Residuen

Eine naheliegende Schätzung des Fehlerterms ε sind die Residuen $\hat{\varepsilon} = y - \hat{y}$. Es gilt

$$\hat{\varepsilon} = y - \hat{y} = y - \mathbf{P}y = (\mathbf{I} - \mathbf{P})y = \mathbf{Q}y,$$

wobei die $T \times T$ Matrix \mathbf{Q} gegeben ist durch

$$\mathbf{Q} = \mathbf{I} - \mathbf{P} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'.$$

Für \mathbf{Q} gilt:

Satz 10.5 (Eigenschaften von \mathbf{Q})

Die Matrix \mathbf{Q} besitzt folgende Eigenschaften:

1. \mathbf{Q} ist symmetrisch.
2. \mathbf{Q} ist idempotent.
3. $\text{rg}(\mathbf{Q}) = \text{sp}(\mathbf{Q}) = T - K$.

Beweis:

Analog zu Satz 10.4.

□

10.4.3 Geometrische Eigenschaften der KQ-Schätzung

Im folgenden Satz zeigen wir wichtige geometrische Eigenschaften der KQ-Schätzung:

Satz 10.6 (Geometrische Eigenschaften der KQ-Schätzung)

Für den Zusammenhang zwischen Designmatrix \mathbf{X} und den Residuen $\hat{\varepsilon}$ bzw. zwischen Residuen und \hat{y} gilt:

1. $\mathbf{X}'\hat{\varepsilon} = 0$, d.h. die Spalten von \mathbf{X} sind orthogonal zu den Residuen.
2. $\hat{y}'\hat{\varepsilon} = 0$, d.h. die geschätzten Werte sind orthogonal zu den Residuen.

Beweis:

zu 1): Unter Zuhilfenahme von Eigenschaften der Hat-Matrix \mathbf{P} (siehe Satz 10.4) gilt

$$\begin{aligned}
 \mathbf{X}'\hat{\varepsilon} &= \mathbf{X}'(\mathbf{I} - \mathbf{P})y \\
 &= \mathbf{X}'y - \mathbf{X}'\mathbf{P}y \\
 &= \mathbf{X}'y - \mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'y \\
 &= \mathbf{X}'y - \mathbf{X}'y \\
 &= 0.
 \end{aligned}$$

zu 2): Wiederum unter Zuhilfenahme von Eigenschaften der Hat-Matrix erhalten wir:

$$\begin{aligned}
 \hat{y}'\hat{\varepsilon} &= \overbrace{y'\mathbf{P}}^{\hat{y}'} \overbrace{(\mathbf{I} - \mathbf{P})y}^{\hat{\varepsilon}} \\
 &= y'\mathbf{P}y - y'\mathbf{P}\mathbf{P}y
 \end{aligned}$$

$$\begin{aligned}
&= \mathbf{y}'\mathbf{P}\mathbf{y} - \mathbf{y}'\mathbf{P}\mathbf{y} \\
&= 0.
\end{aligned}$$

□

Im folgenden Satz zeigen wir noch Implikationen von $X'\hat{\varepsilon} = 0$ für ein Modell mit Intercept:

Satz 10.7

Im Modell

$$y_t = \beta_0 + \beta_1 x_{1t} + \cdots + \beta_K x_{Kt} + \varepsilon_t \quad (t = 1, \dots, T)$$

mit Intercept gilt:

1. $\sum_{t=1}^T \hat{\varepsilon}_t = 0$ bzw. $\bar{\hat{\varepsilon}} = \frac{1}{T} \sum_{t=1}^T \hat{\varepsilon}_t = 0$, d.h. die Residuen sind im Mittel Null.
2. $\bar{\hat{y}} = \frac{1}{T} \sum_{t=1}^T \hat{y}_t = \bar{y}$, d.h. der Mittelwert der Geschätzten Werte ist gleich dem Mittelwert der beobachteten Werte.
3. Die Regressionshyperebene geht durch den Schwerpunkt der Daten d.h.

$$\bar{y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x}_1 + \cdots + \hat{\beta}_K \bar{x}_K.$$

Beweis:

zu 1) Da die erste Spalte x_1 der Designmatrix der **1**-Vektor ist, folgt die Behauptung unter Zuhilfenahme von Satz 10.6 1) aus

$$0 = x_1' \hat{\varepsilon} = \mathbf{1}' \hat{\varepsilon} = \sum_{t=1}^T \hat{\varepsilon}_t.$$

zu 2) Mit 1) gilt

$$\sum_{t=1}^T \hat{y}_t = \sum_{t=1}^T y_t - \sum_{t=1}^T \hat{\varepsilon}_t = \sum_{t=1}^T y_t.$$

zu 3) Es gilt

$$\begin{aligned}
\bar{y} &= \frac{1}{T} \sum_{t=1}^T y_t \\
&= \frac{1}{T} \sum_{t=1}^T (\hat{y}_t + y_t - \hat{y}_t) \\
&= \frac{1}{T} \sum_{t=1}^T (\hat{\beta}_0 + x_{1t} \hat{\beta}_1 + \cdots + x_{Kt} \hat{\beta}_K + \hat{\varepsilon}_t)
\end{aligned}$$

$$= \frac{1}{T} \sum_{t=1}^T (\hat{\beta}_0 + x_{1t}\hat{\beta}_1 + \cdots + x_{Kt}\hat{\beta}_K).$$

□

10.4.4 Streuungszerlegung

Im allgemeinen linearen Regressionsmodell gilt die Beziehung

$$y = \hat{y} + \hat{\varepsilon}, \quad (10.7)$$

und damit natürlich auch

$$y' = \hat{y}' + \hat{\varepsilon}'.$$

Aus diesen beiden Aussagen folgt unmittelbar

$$\begin{aligned} y'y &= (\hat{y}' + \hat{\varepsilon}')(\hat{y} + \hat{\varepsilon}) \\ &= \hat{y}'\hat{y} + \underbrace{\hat{y}'\hat{\varepsilon}}_{=0} + \underbrace{\hat{\varepsilon}'\hat{y}}_{=0} + \hat{\varepsilon}'\hat{\varepsilon} \\ &= \hat{y}'\hat{y} + \hat{\varepsilon}'\hat{\varepsilon}. \end{aligned}$$

Somit erhält man die sogenannte *Streuungs-Zerlegung*

$$y'y = \hat{y}'\hat{y} + \hat{\varepsilon}'\hat{\varepsilon},$$

bzw. in Komponentendarstellung

$$\sum_{t=1}^T y_t^2 = \sum_{t=1}^T \hat{y}_t^2 + \sum_{t=1}^T \hat{\varepsilon}_t^2.$$

Im Folgenden wollen wir eine Formel für die Streuungszerlegung herleiten, wenn das Modell einen Intercept enthält. Hier gilt:

Satz 10.8 (Streuungszerlegung im Modell mit Intercept)

Im linearen Modell mit Intercept

$$y_t = \beta_0 + \beta_1 x_{1t} + \cdots + \beta_K x_{Kt} + \varepsilon_t \quad (t = 1, \dots, T)$$

gilt die Streuungszerlegungsformel

$$\sum_{t=1}^T (y_t - \bar{y})^2 = \sum_{t=1}^T (\hat{y}_t - \bar{y})^2 + \sum_{t=1}^T \hat{\varepsilon}_t^2. \quad (10.8)$$

Mit den Abkürzungen

$$\begin{aligned}
SST &= \sum_{t=1}^T (y_t - \bar{y})^2 && \text{Total Sum of Squares} \\
SSR &= \sum_{t=1}^T (\hat{y}_t - \bar{y})^2 && \text{Regression Sum of Squares} \\
SSE &= \sum_{t=1}^T \hat{\varepsilon}_t^2 && \text{Error Sum of Squares oder} \\
&&& \text{Residual Sum of Squares}
\end{aligned}$$

geht (10.8) über in

$$\begin{aligned}
SST &= SSR + SSE. \\
\left\{ \begin{array}{l} \text{Gesamt-} \\ \text{variation} \end{array} \right\} &= \left\{ \begin{array}{l} \text{erklärte} \\ \text{Streuung} \end{array} \right\} + \left\{ \begin{array}{l} \text{Rest-} \\ \text{Streuung} \end{array} \right\}
\end{aligned}$$

Beweis:

Bei der Herleitung der Streuungserlegung benutzen wir wieder die spezielle idempotente Matrix \mathbf{C} aus Beispiel 1.12. Multiplikation von (10.7) mit \mathbf{C} ergibt

$$\mathbf{C}y = \mathbf{C}\hat{y} + \mathbf{C}\hat{\varepsilon}.$$

Aufgrund von Eigenschaft 1) in Beispiel 1.12 und Satz 10.7 1) gilt $\mathbf{C}\hat{\varepsilon} = \hat{\varepsilon}$ und es folgt

$$\mathbf{C}y = \mathbf{C}\hat{y} + \hat{\varepsilon},$$

bzw.

$$y'\mathbf{C} = \hat{y}'\mathbf{C} + \hat{\varepsilon}'.$$

Damit folgt

$$\begin{aligned}
y'\mathbf{C}y &= (\hat{y}'\mathbf{C} + \hat{\varepsilon}')(\mathbf{C}\hat{y} + \hat{\varepsilon}) \\
&= \hat{y}'\mathbf{C}\mathbf{C}\hat{y} + \hat{y}'\mathbf{C}\hat{\varepsilon} + \hat{\varepsilon}'\mathbf{C}\hat{y} + \hat{\varepsilon}'\hat{\varepsilon} \\
&= \hat{y}'\mathbf{C}\hat{y} + \hat{y}'\hat{\varepsilon} + \hat{\varepsilon}'\hat{y} + \hat{\varepsilon}'\hat{\varepsilon}.
\end{aligned}$$

Gemäß Eigenschaft 6) aus Beispiel 1.12 gilt $y'\mathbf{C}y = y'\mathbf{C}y = \sum (y_i - \bar{y})^2$ und unter zusätzlicher Beachtung von $\bar{\hat{y}} = \bar{y}$ folgt $\hat{y}'\mathbf{C}\hat{y} = \sum (\hat{y}_i - \bar{y})^2$. Nach Satz 10.6 2) folgt $\hat{y}'\hat{\varepsilon} = \hat{\varepsilon}'\hat{y} = 0$ und wir erhalten

$$\sum_{t=1}^T (y_t - \bar{y})^2 = \sum_{t=1}^T (\hat{y}_t - \bar{y})^2 + \sum_{t=1}^T \hat{\varepsilon}_t^2.$$

Man beachte, daß bei der Herleitung dieser Streuungszerlegungsformel entscheidend mit eingeht, daß das Modell ein konstantes Glied enthält!

□

Mit Hilfe der Streuungszerlegungsformel läßt sich ein Maß definieren, welches die Güte der Anpassung der Regression an die Daten mißt, das sogenannte *Bestimmtheitsmaß* B :

Definition 10.1 (Bestimmtheitsmaß)

Im linearen Modell mit Interecept ist das Bestimmtheitsmaß definiert als

$$B = R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}.$$

Wegen $SST = SSR + SSE$ gilt die Ungleichungskette

$$0 \leq R^2 \leq 1.$$

Je näher R^2 bei 1 liegt (oder $1 - R^2$ bei 0), desto kleiner ist die Residuenquadratsumme SSE , d.h. desto besser ist die Anpassung an die Daten. Ist umgekehrt R^2 nahe bei 0 (d.h. $1 - R^2$ nahe bei 1), so ist die Residuenquadratsumme groß, und damit die Anpassung des Modells an die Daten gering.

Bemerkung:

Im Spezialfall des univariaten Regressionsmodells mit konstantem Glied

$$y = \beta_0 + \beta_1 x + \varepsilon$$

kann man zeigen, dass

$$B = R^2 = \hat{\rho}^2 = \frac{S_{XY}^2}{S_{XX}S_{YY}},$$

wobei $\hat{\rho}^2$ den quadrierten empirischen Korrelationskoeffizienten zwischen X und Y bezeichnet.

Allgemein gilt für das Bestimmtheitsmaß:

Satz 10.9

Sei $y = X_1\beta_1 + X_2\beta_2 + \varepsilon_X = X\beta + \varepsilon_X$ ein volles Modell, und $y = X_1\beta_1 + \varepsilon_{X_1}$ ein darin enthaltenes Submodell. Dann gilt:

$$R_X^2 - R_{X_1}^2 \geq 0,$$

d.h. das multiple Bestimmtheitsmaß B steigt mit zunehmender Anzahl der Regressoren (Kovariablen) automatisch an, ohne dass sich dabei die Güte der Anpassung signifikant verbessern muß.

Beweis:

Wegen

$$R_X^2 = 1 - \frac{\hat{\varepsilon}'_X \hat{\varepsilon}_X}{SST}$$

und

$$R_{X_1}^2 = 1 - \frac{\hat{\varepsilon}'_{X_1} \hat{\varepsilon}_{X_1}}{SST}$$

ist die Aussage

$$R_X^2 - R_{X_1}^2 \geq 0$$

äquivalent zu

$$\hat{\varepsilon}'_{X_1} \hat{\varepsilon}_{X_1} \geq \hat{\varepsilon}'_X \hat{\varepsilon}_X.$$

Nun kann aber die KQ-Schätzung im Submodell

$$y = X_1 \beta_1 + \varepsilon_{X_1}$$

auch durch eine restringierte KQ-Schätzung im vollen Modell

$$y = X\beta + \varepsilon_X$$

unter der Nebenbedingung

$$\beta_2 = 0$$

gewonnen werden. Die Behauptung folgt dann aus den Ausführungen im Abschnitt 11.3.1, wo gezeigt wird, dass die Differenz zwischen der Residuenquadratsumme im restringierten Modell und der Residuenquadratsumme im vollen Modell stets größer oder gleich Null ist.

□

Den oben gezeigten Nachteil des Bestimmtheitsmaßes $B = R^2$ gleicht das sogenannte *adjustierte Bestimmtheitsmaß*

$$\bar{R}^2 = 1 - \frac{T-1}{T-K} (1 - R^2)$$

aus. Dieser Term wird nicht automatisch größer, wenn eine oder mehrere zusätzliche Variablen in das Modell aufgenommen werden. Mehr Details zum adjustierten Bestimmtheitsmaß findet man in ??.

10.5 Eigenschaften des KQ-Schätzers

In diesem Abschnitt sollen die wichtigsten statistischen Eigenschaften des KQ-Schätzers behandelt werden. Ziel ist es insbesondere den gewöhnliche KQ-Schätzer

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

im klassischen linearen Modell mit der Klasse der *linearen Schätzer* hinsichtlich seiner "Güte" zu vergleichen. Ein linearer Schätzer hat dabei die Gestalt

$$b = \underset{(K \times 1)}{\mathbf{C}} \underset{(T \times 1)}{y} + \underset{(K \times 1)}{d}.$$

Offensichtlich handelt es sich beim KQ-Schätzer $\hat{\beta}$ um einen linearen Schätzer mit $\mathbf{C} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ und $d = 0$.

Wir führen zunächst einige gebräuchliche Gütekriterien ein:

10.5.1 Gütekriterien

Definition 10.2 (Bias einer Schätzung)

Der Bias (die Verzerrung) einer Schätzung b für β ist definiert als

$$\text{Bias}(b, \beta) = E(b) - \beta$$

Eine Schätzung b heißt *erwartungstreu* für β , falls

$$\text{Bias}(b, \beta) = 0,$$

d.h. $E(b) = \beta$ gilt.

Definition 10.3 (MSE = Mean Squared Error)

Der Mean-Squared-Error (MSE) einer Schätzung b für β ist definiert als

$$\text{MSE}(b, \beta) = E(b - \beta)(b - \beta)'$$

Im folgenden Satz zeigen wir, dass der MSE einer Schätzung zerlegt werden kann in eine Varianzkomponente und eine Biaskomponente.

Satz 10.10 (Zerlegungsformel für den MSE)

Der MSE eines Schätzers setzt sich wie folgt aus Varianz und Bias des Schätzers zusammen:

$$\text{MSE}(b, \beta) = \text{Cov}(b) + \text{Bias}(b, \beta) \text{Bias}(b, \beta)'$$

Beweis:

Sei $E(b) = \mu$. Dann gilt:

$$\begin{aligned}
 E(b - \beta)(b - \beta)' &= E\{(b - \mu) + (\mu - \beta)\}\{(b - \mu) + (\mu - \beta)\}' \\
 &= E(b - \mu)(b - \mu)' + E(\mu - \beta)(b - \mu)' \\
 &\quad + E(b - \mu)(\mu - \beta)' + E(\mu - \beta)(\mu - \beta)' \\
 &= \text{Cov}(b) + (\mu - \beta)E(b - \mu)' \\
 &\quad + E(b - \mu)(\mu - \beta)' + (\mu - \beta)(\mu - \beta)' \\
 &= \text{Cov}(b) + 0 + 0 + \text{Bias}(b, \beta)\text{Bias}(b, \beta)'.
 \end{aligned}$$

□

10.5.2 Statistische Eigenschaften linearer Schätzer

Wir betrachten im Folgenden zunächst allgemein lineare Schätzer

$$b = \mathbf{C}y + d$$

als Schätzfunktionen für die unbekanntenen Regressionskoeffizienten im linearen Modell. Es gilt:

Satz 10.11 (Eigenschaften linearer Schätzer)

Im linearen Regressionsmodell gelten für lineare Schätzer $b = \mathbf{C}y + d$ für die Regressionskoeffizienten β die folgenden Eigenschaften:

1. $E(b) = \mathbf{C}\mathbf{X}\beta + d$
2. $\text{Bias}(b, \beta) = (\mathbf{C}\mathbf{X} - \mathbf{I})\beta + d$
3. $\text{Cov}(b) = \sigma^2 \mathbf{C}\mathbf{C}'$
4. $\text{MSE}(b, \beta) = \sigma^2 \mathbf{C}\mathbf{C}' + \{(\mathbf{C}\mathbf{X} - \mathbf{I})\beta + d\}\{(\mathbf{C}\mathbf{X} - \mathbf{I})\beta + d\}'$
5. Die folgenden Bedingungen sind notwendig für die Erwartungstreue eines linearen Schätzers:
 - $d = 0$
 - $\mathbf{C}\mathbf{X} = \mathbf{I}_K$
 - $\text{rg}(\mathbf{X}) = K$

Beweis:

zu 1) $E(b) = E(\mathbf{C}y + d) = \mathbf{C} E(y) + d = \mathbf{C}\mathbf{X}\beta + d.$

zu 2) $\text{Bias}(b, \beta) = \mathbf{C}\mathbf{X}\beta + d - \beta = (\mathbf{C}\mathbf{X} - \mathbf{I})\beta + d.$

zu 3) $\text{Cov}(b) = \text{Cov}(\mathbf{C}y + d) = \mathbf{C} \text{Cov}(y) \mathbf{C}' = \sigma^2 \mathbf{C}\mathbf{C}' = \sigma^2 \mathbf{C}\mathbf{C}'.$

zu 4) Folgt in Verbindung mit Satz 10.10 unmittelbar aus 2) und 3).

zu 5) Soll der Schätzer b erwartungstreu sein, so muß $E(b) = \mathbf{C}\mathbf{X}\beta + d = \beta$ für alle $\beta \in \mathbb{R}^K$ gelten. Setzt man speziell $\beta = 0$, so folgt $d = 0$ als notwendige Bedingung für die Erwartungstreue von b . Umformen der Bedingung $\mathbf{C}\mathbf{X}\beta = \beta$ liefert $(\mathbf{C}\mathbf{X} - \mathbf{I}_K)\beta = 0$, so daß als weitere Bedingung für die Erwartungstreue von b zwingend $\mathbf{C}\mathbf{X} = \mathbf{I}_K$ gelten muß. Wegen $\text{rg}(\mathbf{C}\mathbf{X}) = \min(\text{rg}(\mathbf{X}), \text{rg}(\mathbf{C})) = \text{rg}(\mathbf{I}_K) = K$ muß auch $\text{rg}(\mathbf{X}) = K$ gelten, falls der Schätzer b erwartungstreu sein soll.

□

Da der KQ-Schätzer ein Spezialfall linearer Schätzer ist, erhalten wir unmittelbar

Korollar 10.1 (Eigenschaften des KQ-Schätzers)

Im linearen Modell besitzt der KQ-Schätzer $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'y$ folgende Eigenschaften:

1. $E(\hat{\beta}) = \beta$, d.h. der KQ-Schätzer ist erwartungstreu.
2. $\text{Bias}(\hat{\beta}, \beta) = 0.$
3. $\text{Cov}(\hat{\beta}) = \sigma^2 (\mathbf{X}'\mathbf{X})^{-1}.$
4. $\text{MSE}(\hat{\beta}, \beta) = \sigma^2 (\mathbf{X}'\mathbf{X})^{-1}.$

Im folgenden Satz zeigen wir, dass der KQ-Schätzer unter allen linearen, erwartungstreuen Schätzern eine herausragende Rolle spielt:

Satz 10.12 (Gauß-Markov-Theorem)

Im klassischen linearen Regressionsmodell ist die KQ-Schätzung $\hat{\beta}$ unter allen linearen, erwartungstreuen und homogenen Schätzern $b = \mathbf{C}y$ (also $d = 0$) die beste Schätzung, d.h. es gilt

$$\text{MSE}(b, \beta) - \text{MSE}(\hat{\beta}, \beta) \geq 0$$

bzw.

$$\text{Cov}(b) - \text{Cov}(\hat{\beta}) \geq 0$$

Beweis:

Sei o.B.d.A. die Matrix \mathbf{C} von der Form $\mathbf{C} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' + \mathbf{D}$. Dann folgt zunächst aus der Unverzerrtheitsbedingung $\mathbf{C}\mathbf{X} = \mathbf{I}_K$ (vergleiche Satz 10.11 5)):

$$\begin{aligned}\mathbf{C}\mathbf{X} = \mathbf{I}_K & \iff \\ (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X} + \mathbf{D}\mathbf{X} = \mathbf{I}_K & \iff \\ \mathbf{I}_K + \mathbf{D}\mathbf{X} = \mathbf{I}_K & \iff \\ \mathbf{D}\mathbf{X} = 0 \quad \text{bzw.} \quad \mathbf{X}'\mathbf{D}' = 0 & \end{aligned}$$

Damit folgt für die Kovarianzmatrix von b :

$$\begin{aligned}\text{Cov}(b) &= \sigma^2 \mathbf{C}\mathbf{C}' \\ &= \sigma^2 \{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' + \mathbf{D}\} \{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' + \mathbf{D}\}' \\ &= \sigma^2 \left\{ (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} + (\mathbf{X}'\mathbf{X})^{-1}\underbrace{\mathbf{X}'\mathbf{D}'}_{=0} + \underbrace{\mathbf{D}\mathbf{X}}_{=0}(\mathbf{X}'\mathbf{X})^{-1} + \mathbf{D}\mathbf{D}' \right\} \\ &= \sigma^2 (\mathbf{X}'\mathbf{X})^{-1} + \sigma^2 \mathbf{D}\mathbf{D}' \\ &= \text{Cov}(\hat{\beta}) + \sigma^2 \mathbf{D}\mathbf{D}'\end{aligned}$$

Nach Satz 7.4 gilt $\mathbf{D}\mathbf{D}' \geq 0$, so dass wir schließlich durch Umstellen

$$\text{Cov}(b) - \text{Cov}(\hat{\beta}) = \sigma^2 \mathbf{D}\mathbf{D}' \geq 0,$$

erhalten. □

Interpretation des Satzes

Da $\text{Cov}(b) - \text{Cov}(\hat{\beta}) \geq 0$ gilt wegen Satz 7.2 2) (vergleiche dort auch die nachfolgende Bemerkung) insbesondere

$$\text{Var}(b_i) \geq \text{Var}(\hat{\beta}_i), \quad i = 1, \dots, K,$$

d.h. der KQ-Schätzer besitzt unter allen linearen erwartungstreuen homogenen Schätzern die kleinsten Varianzen.

Der folgende Satz beschäftigt sich damit, eine möglichst gute Schätzung für eine Linearkombination

$$l = a'\beta$$

der Komponenten des Parametervektors β zu finden. Dabei sei a ein $K \times 1$ Vektor. Es wird sich zeigen, dass der Schätzer $\hat{l} := a'\hat{\beta}$, der auf dem KQ-Schätzer für β beruht, optimal ist. Offensichtlich ist \hat{l} erwartungstreu.

Satz 10.13

Im klassischen linearen Modell besitzt der Schätzer $\hat{l} = a' \hat{\beta}$ mit der Varianz

$$\text{Var}(\hat{l}) = \sigma^2 a' (X'X)^{-1} a$$

unter allen linearen und (für l) erwartungstreuen Schätzern der Gestalt $\tilde{l} = c' y$ die kleinste Varianz.

Beweis:

Für den Erwartungswert von \tilde{l} gilt

$$E(\tilde{l}) = c' E(y) = c' \mathbf{X} \beta.$$

Da wir annehmen, dass \tilde{l} unverzerrt ist, muss $c' \mathbf{X} \beta = a' \beta$ gelten und es folgt als Bedingung

$$c' \mathbf{X} = a'. \quad (10.9)$$

Setze nun wieder o.B.d.A.

$$c' = a' (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' + \tilde{c}'.$$

Einsetzen in (10.9) liefert

$$a' (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{X} + \tilde{c}' \mathbf{X} = a' + \tilde{c}' \mathbf{X} = a'$$

und folglich $\tilde{c}' \mathbf{X} = 0$ bzw. $\mathbf{X}' \tilde{c} = 0$. Damit erhält man für die Varianz von \tilde{l} :

$$\begin{aligned} \text{Var}(\tilde{l}) &= \text{Var} \left\{ \left(a' (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' + \tilde{c}' \right) y \right\} \\ &= \left\{ a' (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' + \tilde{c}' \right\} \text{Var}(y) \left\{ a' (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' + \tilde{c}' \right\}' \\ &= \sigma^2 \left\{ a' (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' + \tilde{c}' \right\} \left\{ a' (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' + \tilde{c}' \right\}' \\ &= \sigma^2 \left\{ a' (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1} a + a' (\mathbf{X}' \mathbf{X})^{-1} \underbrace{\mathbf{X}' \tilde{c}}_{=0} + \underbrace{\tilde{c}' \mathbf{X}}_{=0} (\mathbf{X}' \mathbf{X})^{-1} a + \tilde{c}' \tilde{c} \right\} \\ &= \sigma^2 a' (\mathbf{X}' \mathbf{X})^{-1} a + \sigma^2 \tilde{c}' \tilde{c} \\ &= \text{Var}(\hat{l}) + \sigma^2 \tilde{c}' \tilde{c}. \end{aligned}$$

Daraus folgt durch Umstellen die Behauptung. □

Satz 10.13 findet seine Anwendung u.a. bei der Bestimmung von optimalen Prognosen einer neuen Responsevariable y^* mit (nichtstochastischem, bekanntem) Kovariablenvektor x^* . Es gilt

$$E(y^*) = x^{*'} \beta.$$

Damit ist die optimale Schätzung des Erwartungswertes von y^* (im Sinne von Satz 10.13) gegeben durch

$$\hat{y}^* = x^{*'} \hat{\beta}.$$

10.5.3 Eine Schätzung für die Varianz σ^2

Bis jetzt haben wir uns ausschließlich darum bemüht einen möglichst optimalen Schätzer für die unbekanntes Regressionskoeffizienten β zu finden. In diesem Abschnitt beschäftigen wir uns mit der Schätzung der Varianz σ^2 , die im Allgemeinen auch unbekannt ist.

Zunächst jedoch folgender Hilfssatz über den Erwartungswert quadratischer Formen:

Satz 10.14 (Erwartungswert quadratischer Formen)

Sei Z ein K -dimensionaler Zufallsvektor mit $E(Z) = \mu$ und $Cov(Z) = \Sigma_Z$. Sei weiterhin \mathbf{A} eine symmetrische Matrix der Dimension $K \times K$. Dann gilt für den Erwartungswert der quadratischen Form $Z' \mathbf{A} Z$

$$E(Z' \mathbf{A} Z) = sp(\mathbf{A} \Sigma_Z) + \mu' \mathbf{A} \mu.$$

Beweis:

Da $Z' \mathbf{A} Z$ und $\mu' \mathbf{A} \mu$ Skalare sind gilt $sp(Z' \mathbf{A} Z) = Z' \mathbf{A} Z$ bzw. $sp(\mu' \mathbf{A} \mu) = \mu' \mathbf{A} \mu$. Unter Verwendung von Eigenschaften der Spur (Satz 4.7) erhalten wir

$$\begin{aligned} E(Z' \mathbf{A} Z) &= E(sp(Z' \mathbf{A} Z)) \\ &= E(sp(\mathbf{A} Z Z')) \\ &= sp \{ \mathbf{A} (\Sigma_Z + \mu \mu') \} \\ &= sp(\mathbf{A} \Sigma_Z) + sp(\mathbf{A} \mu \mu') \\ &= sp(\mathbf{A} \Sigma_Z) + sp(\mu' \mathbf{A} \mu) \\ &= sp(\mathbf{A} \Sigma_Z) + \mu' \mathbf{A} \mu. \end{aligned}$$

□

Wir widmen uns jetzt wieder der Bestimmung einer Schätzung für σ^2 . Es ist naheliegend, eine Schätzung $\hat{\sigma}^2$ für die Varianz σ^2 auf der Residuenquadratsumme

$$\hat{\varepsilon}' \hat{\varepsilon} = (y - \hat{y})'(y - \hat{y}) = y' \mathbf{Q} y$$

aufzubauen. Für den Erwartungswert der Residuenquadratsumme $\hat{\varepsilon}' \hat{\varepsilon}$ erhalten wir unter Verwendung von Satz 10.14 und Eigenschaften der Matrix \mathbf{Q} (Satz 10.5)

$$\begin{aligned} E \hat{\varepsilon}' \hat{\varepsilon} &= E y' \mathbf{Q} y \\ &= \sigma^2 sp \mathbf{Q} + \beta' \mathbf{X}' \mathbf{Q} \mathbf{X} \beta \end{aligned}$$

$$\begin{aligned}
&= \sigma^2(T - K) + \beta' \mathbf{X}' (\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}') \mathbf{X} \beta \\
&= \sigma^2(T - K) + \beta' \mathbf{X}' \mathbf{X} \beta - \beta' \mathbf{X}' \mathbf{X} (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \mathbf{X} \beta \\
&= \sigma^2(T - K) + \beta' \mathbf{X}' \mathbf{X} \beta - \beta' \mathbf{X}' \mathbf{X} \beta \\
&= \sigma^2(T - K).
\end{aligned} \tag{10.10}$$

Damit erhalten wir unmittelbar den folgenden Satz:

Satz 10.15 (Erwartungstreue Schätzung für σ^2)

Im klassischen linearen Modell ist der Schätzer

$$\hat{\sigma}^2 = \frac{1}{T - K} \hat{\varepsilon}' \hat{\varepsilon} \tag{10.11}$$

für die Varianz σ^2 erwartungstreu, d.h. $E(\hat{\sigma}^2) = \sigma^2$.

Als Folgerung erhalten wir:

Korollar 10.2 (Schätzer für die Kovarianzmatrix)

Im klassischen linearen Modell ist

$$\widehat{Cov}(\hat{\beta}) = \hat{\sigma}^2 (\mathbf{X}'\mathbf{X})^{-1}.$$

ein unverzerrter Schätzer für die Kovarianzmatrix $Cov(\hat{\beta})$ von $\hat{\beta}$.

Beispiel 10.8 (Schätzung der Kovarianzmatrix im univariaten Modell)

Gegeben sei das Modell

$$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t, \quad t = 1, \dots, T.$$

Für den Fall, dass die Varianz der Störungen σ^2 bekannt ist, gilt mit $\bar{x}^2 = \frac{1}{T} \sum x_t^2$ und unter Verwendung der in Beispiel 10.4 eingeführten Größe S_{XX}

$$\begin{aligned}
\text{Cov} \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix} &= \sigma^2 (X'X)^{-1} \\
&= \sigma^2 \begin{pmatrix} T & \sum x_t \\ \sum x_t & \sum x_t^2 \end{pmatrix}^{-1} \\
&= \frac{\sigma^2}{T \sum x_t^2 - T^2 \bar{x}^2} \begin{pmatrix} \sum x_t^2 & -\sum x_t \\ -\sum x_t & T \end{pmatrix} \\
&= \frac{\sigma^2}{S_{XX}} \begin{pmatrix} \bar{x}^2 & -\bar{x} \\ -\bar{x} & 1 \end{pmatrix}.
\end{aligned}$$

Ersetzt man darin die im allgemeinen unbekannte Varianz σ^2 durch die Schätzung

$$\hat{\sigma}^2 = \frac{1}{T - K} \hat{\varepsilon}' \hat{\varepsilon},$$

so erhält man eine (erwartungstreue) Schätzung für $\text{Cov}\begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix}$. Die Residuenquadratsumme kann berechnet werden als

$$\begin{aligned} \hat{\varepsilon}' \hat{\varepsilon} &= \sum_{t=1}^T \{y_t - \hat{y}_t\}^2 \\ &= \sum_{t=1}^T \{y_t - (\hat{\beta}_0 + \hat{\beta}_1 x_t)\}^2 \\ &= \sum_{t=1}^T \{(y_t - \bar{y}) - \hat{\beta}_1 (x_t - \bar{x})\}^2 \\ &= \sum_{t=1}^T \{y_t - \bar{y}\}^2 - 2\hat{\beta}_1 \sum_{t=1}^T \{x_t - \bar{x}\} \{y_t - \bar{y}\} + \hat{\beta}_1^2 \sum_{t=1}^T \{x_t - \bar{x}\}^2 \\ &= S_{YY} - 2\hat{\beta}_1 S_{XY} + \hat{\beta}_1^2 S_{XX} \\ &= S_{YY} - \hat{\beta}_1^2 S_{XX} \\ &= S_{XY} - \frac{S_{XY}^2}{S_{XX}}. \end{aligned}$$

Dabei haben wir in der 3. Zeile $\bar{y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x}$ aus Satz 10.7 3) und in der vorletzten Zeile $\hat{\beta}_1 = \frac{S_{XY}}{S_{XX}}$ bzw. $S_{XY} = \hat{\beta}_1 S_{XX}$ gemäß Beispiel 10.4 verwendet. Damit erhalten wir als Schätzer für die Kovarianzmatrix des KQ-Schätzers

$$\begin{aligned} \widehat{\text{Cov}}(\hat{\beta}) &= \hat{\sigma}^2 (\mathbf{X}'\mathbf{X})^{-1} \\ &= \frac{1}{T - 2} \left(S_{YY} - \frac{S_{XY}^2}{S_{XX}} \right) \cdot \frac{1}{S_{XX}} \begin{pmatrix} \bar{x}^2 & -\bar{x} \\ -\bar{x} & 1 \end{pmatrix} \\ &= \frac{1}{T - 2} \left(\frac{S_{YY}}{S_{XX}} - \frac{S_{XY}^2}{S_{XX}^2} \right) \begin{pmatrix} \bar{x}^2 & -\bar{x} \\ -\bar{x} & 1 \end{pmatrix}. \end{aligned}$$

10.5.4 Vergleich des unrestringierten mit dem KQ-Schätzer unter linearen Nebenbedingungen

Im Folgenden soll der gewöhnliche KQ-Schätzer

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

mit dem restringierten KQ-Schätzer

$$\hat{\beta}^R = \hat{\beta} + (\mathbf{X}'\mathbf{X})^{-1} \mathbf{R}' \left\{ \mathbf{R}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{R}' \right\}^{-1} (r - \mathbf{R}\hat{\beta})$$

aus Satz 10.3 verglichen werden, falls die lineare Restriktion

$$\mathbf{R}\beta = r$$

besteht.

Satz 10.16 (Eigenschaften des restringierten KQ-Schätzers)

Im klassischen linearen Regressionsmodell gelten für den restringierten KQ-Schätzer folgende Eigenschaften:

1. Falls die Restriktion erfüllt ist, gilt $E(\hat{\beta}^R) = \beta$.
2. Unabhängig davon, ob die Restriktion erfüllt ist, gilt

$$\text{Cov}(\hat{\beta}^R) = \sigma^2 \mathbf{S}^{-1} - \sigma^2 \mathbf{S}^{-1} \mathbf{R}' (\mathbf{R} \mathbf{S}^{-1} \mathbf{R}')^{-1} \mathbf{R} \mathbf{S}^{-1}$$

wobei $\mathbf{S} = \mathbf{X}'\mathbf{X}$.

Beweis:

zu 1)

$$\begin{aligned} E(\hat{\beta}^R) &= E \left\{ \hat{\beta} + (\mathbf{X}'\mathbf{X})^{-1} \mathbf{R}' \left\{ \mathbf{R} (\mathbf{X}'\mathbf{X})^{-1} \mathbf{R}' \right\}^{-1} (r - \mathbf{R}\hat{\beta}) \right\} \\ &= E(\hat{\beta}) + (\mathbf{X}'\mathbf{X})^{-1} \mathbf{R}' \left\{ \mathbf{R} (\mathbf{X}'\mathbf{X})^{-1} \mathbf{R}' \right\}^{-1} (r - \mathbf{R}E(\hat{\beta})) \\ &= \beta + (\mathbf{X}'\mathbf{X})^{-1} \mathbf{R}' \left\{ \mathbf{R} (\mathbf{X}'\mathbf{X})^{-1} \mathbf{R}' \right\}^{-1} (r - \mathbf{R}\beta) \\ &= \beta \end{aligned}$$

zu 2)

$$\begin{aligned} \text{Cov}(\hat{\beta}^R) &= \text{Cov} \left\{ \hat{\beta} + \mathbf{S}^{-1} \mathbf{R}' (\mathbf{R} \mathbf{S}^{-1} \mathbf{R}')^{-1} (r - \mathbf{R}\hat{\beta}) \right\} \\ &= \text{Cov} \left\{ \hat{\beta} - \mathbf{S}^{-1} \mathbf{R}' (\mathbf{R} \mathbf{S}^{-1} \mathbf{R}')^{-1} \mathbf{R} \hat{\beta} \right\} \\ &= \text{Cov} \left\{ \left(\mathbf{I} - \mathbf{S}^{-1} \mathbf{R}' (\mathbf{R} \mathbf{S}^{-1} \mathbf{R}')^{-1} \mathbf{R} \right) \hat{\beta} \right\} \\ &= \left\{ \mathbf{I} - \mathbf{S}^{-1} \mathbf{R}' (\mathbf{R} \mathbf{S}^{-1} \mathbf{R}')^{-1} \mathbf{R} \right\} \text{Cov} \hat{\beta} \left\{ \mathbf{I} - \mathbf{S}^{-1} \mathbf{R}' (\mathbf{R} \mathbf{S}^{-1} \mathbf{R}')^{-1} \mathbf{R} \right\}' \\ &= \left\{ \mathbf{I} - \mathbf{S}^{-1} \mathbf{R}' (\mathbf{R} \mathbf{S}^{-1} \mathbf{R}')^{-1} \mathbf{R} \right\} \sigma^2 \mathbf{S}^{-1} \left\{ \mathbf{I} - \mathbf{S}^{-1} \mathbf{R}' (\mathbf{R} \mathbf{S}^{-1} \mathbf{R}')^{-1} \mathbf{R} \right\}' \\ &= \left\{ \sigma^2 \mathbf{S}^{-1} - \sigma^2 \mathbf{S}^{-1} \mathbf{R}' (\mathbf{R} \mathbf{S}^{-1} \mathbf{R}')^{-1} \mathbf{R} \mathbf{S}^{-1} \right\} \left\{ \mathbf{I} - \mathbf{R}' (\mathbf{R} \mathbf{S}^{-1} \mathbf{R}')^{-1} \mathbf{R} \mathbf{S}^{-1} \right\}' \\ &= \sigma^2 \mathbf{S}^{-1} - \\ &\quad \sigma^2 \mathbf{S}^{-1} \mathbf{R}' (\mathbf{R} \mathbf{S}^{-1} \mathbf{R}')^{-1} \mathbf{R} \mathbf{S}^{-1} - \\ &\quad \sigma^2 \mathbf{S}^{-1} \mathbf{R}' (\mathbf{R} \mathbf{S}^{-1} \mathbf{R}')^{-1} \mathbf{R} \mathbf{S}^{-1} + \\ &\quad \sigma^2 \mathbf{S}^{-1} \mathbf{R}' (\mathbf{R} \mathbf{S}^{-1} \mathbf{R}')^{-1} \mathbf{R} \mathbf{S}^{-1} \mathbf{R}' (\mathbf{R} \mathbf{S}^{-1} \mathbf{R}')^{-1} \mathbf{R} \mathbf{S}^{-1} \\ &= \sigma^2 \mathbf{S}^{-1} - \sigma^2 \mathbf{S}^{-1} \mathbf{R}' (\mathbf{R} \mathbf{S}^{-1} \mathbf{R}')^{-1} \mathbf{R} \mathbf{S}^{-1}. \end{aligned}$$

Damit folgt unmittelbar das folgende Korollar:

Korollar 10.3

Im klassischen linearen Regressionsmodell gilt für die Differenz der Kovarianzmatrizen des KQ-Schätzers und des restringierten KQ-Schätzers

$$\text{Cov}(\hat{\beta}) - \text{Cov}(\hat{\beta}^R) = \sigma^2 \mathbf{S}^{-1} \mathbf{R}' (\mathbf{R} \mathbf{S}^{-1} \mathbf{R}')^{-1} \mathbf{R} \mathbf{S}^{-1} \geq 0$$

Beweis:

Die Matrix $\mathbf{R} \mathbf{S}^{-1} \mathbf{R}'$ und damit auch $(\mathbf{R} \mathbf{S}^{-1} \mathbf{R}')^{-1}$ ist positiv definit (vergleiche hierzu auch den Beweis von Satz 10.3). Die Aussage folgt dann aus Satz 7.3 2). □

Damit ist gezeigt, daß durch die Beachtung einer linearen Restriktion, die Schätzung von β mittels einer restringierten KQ-Methode zu einem Effizienzgewinn führen kann, d.h. der Schätzer besitzt eine kleinere Varianz als die gewöhnliche KQ-Schätzung. Voraussetzung hierzu ist aber ein Vorwissen über den wahren Parameter (in Form der Restriktion $\mathbf{R}\beta = r$). Man beachte, dass der restringierte KQ-Schätzer nur besser ist, falls die Restriktion erfüllt ist, andernfalls ist $\hat{\beta}^R$ nicht erwartungstreu und eine Aussage über die MSE's beider Schätzer ist nicht so leicht möglich.

11

Klassische Normalregression

Die bisher dargestellten Eigenschaften des KQ-Schätzers wurden ohne spezielle Annahmen über die Verteilung der Störungen ε gewonnen. Im Folgenden wollen wir zusätzlich annehmen, dass

$$\varepsilon \sim N(0, \sigma^2 \mathbf{I})$$

gilt, d.h.

$$y \sim N(\mathbf{X}\beta, \sigma^2 \mathbf{I})$$

Unter der Normalverteilungsannahme lassen sich weitere Eigenschaften der KQ-Schätzung herleiten. Insbesondere werden damit *Tests von linearen Hypothesen* über die unbekannt Parameter und die *Bestimmung von Konfidenzbereichen* möglich.

Im folgenden Abschnitt bestimmen wir zunächst den Maximum-Likelihood (ML) Schätzer für β und σ^2 . Dabei stellt sich heraus, dass der ML-Schätzer für β gleich dem KQ-Schätzer $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'y$ ist.

11.1 Maximum Likelihood-Schätzung

Es gilt:

Satz 11.1 (ML-Schätzer im klassischen linearen Model)

Im klassischen linearen Regressionsmodell unter Normalverteilungsannahme sind die ML-Schätzer für β und σ^2 gegeben durch

$$\hat{\beta}_{ML} = \hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'y$$

und

$$\hat{\sigma}_{ML}^2 = \frac{\hat{\varepsilon}'\hat{\varepsilon}}{T}.$$

Beweis:

Wegen der Unabhängigkeit der Störgrößen ist die Likelihood das Produkt der individuellen Likelihoods und wir erhalten

$$\begin{aligned}
L(\beta; \sigma^2; y) &= \prod_{t=1}^T \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(y_t - x_t'\beta)^2\right) \\
&= \frac{1}{(2\pi\sigma^2)^{T/2}} \exp\left\{-\frac{1}{2\sigma^2}(y - \mathbf{X}\beta)'(y - \mathbf{X}\beta)\right\}. \quad (11.1)
\end{aligned}$$

Durch logarithmieren erhalten wir die log-Likelihood

$$\ell(\beta; \sigma^2; y) = -\frac{T}{2} \log(2\pi) - \frac{T}{2} \log(\sigma^2) - \frac{1}{2\sigma^2}(y - \mathbf{X}\beta)'(y - \mathbf{X}\beta). \quad (11.2)$$

Partielle Differentiation nach β liefert

$$\frac{\partial \ell(\beta; \sigma^2; y)}{\partial \beta} = -\frac{1}{2\sigma^2} \frac{\partial S_2(\beta)}{\partial \beta} = -\frac{1}{2\sigma^2} (2\mathbf{X}'\mathbf{X}\beta - 2\mathbf{X}'y),$$

wobei $S_2(\beta)$ der in (10.3) definierte Ausdruck ist, den wir bereits bei der Berechnung des KQ-Schätzers abgeleitet haben. Partielle Differentiation nach σ^2 liefert

$$\frac{\partial \ell(\beta; \sigma^2; y)}{\partial \sigma^2} = -\frac{T}{2\sigma^2} + \frac{1}{2\sigma^4}(y - \mathbf{X}\beta)'(y - \mathbf{X}\beta).$$

Nullsetzen ergibt die beiden Gleichungssysteme

$$\begin{aligned}
\text{I.} \quad & \mathbf{X}'\mathbf{X}\beta - \mathbf{X}'y = 0 \\
\text{II.} \quad & \frac{1}{\sigma^4}(y - \mathbf{X}\beta)'(y - \mathbf{X}\beta) = \frac{T}{\sigma^2}.
\end{aligned}$$

Aus I. folgt sofort, dass der ML-Schätzer $\hat{\beta}_{ML}$ für β mit dem KQ-Schätzer $\hat{\beta}$ übereinstimmt, d.h. es ist

$$\hat{\beta}_{ML} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'y.$$

Einsetzen von $\hat{\beta}_{ML}$ in das System II. liefert

$$\frac{1}{\sigma^4}(y - \hat{y})'(y - \hat{y}) = \frac{T}{\sigma^2}$$

und damit

$$\hat{\sigma}_{ML}^2 = \frac{\hat{\varepsilon}'\hat{\varepsilon}}{T}.$$

□

Bemerkung

Die Schätzung für σ^2 stimmt nicht mit der Schätzung für die Varianz in Abschnitt 10.5.3 überein. Insbesondere ist also $\hat{\sigma}_{ML}^2$ auch nicht erwartungstreu. Allerdings gilt wegen

$$\hat{\sigma}_{ML}^2 = \frac{T-K}{T}\hat{\sigma}^2$$

für den Grenzübergang $T \rightarrow \infty$:

$$\lim_{T \rightarrow \infty} E(\hat{\sigma}_{ML}^2) = \sigma^2.$$

Die ML-Schätzung für die Varianz ist also zumindest asymptotisch erwartungstreu.

11.2 Verteilungstheorie im klassischen linearen Modell

Im folgenden Satz bestimmen wir die Verteilungen einiger wichtiger Größen im linearen Modell unter Normalverteilungsannahme. Unter Anderem berechnen wir die Verteilung des KQ-Schätzers $\hat{\beta}$. Beim Beweis des Satzes werden einige in Kapitel 9 behandelte Verteilungseigenschaften der multivariaten Normalverteilung von Bedeutung sein.

Satz 11.2

Im klassischen linearen Modell gelten unter Normalverteilungsannahme folgende Aussagen:

1. $y \sim N(\mathbf{X}\beta, \sigma^2\mathbf{I})$
2. $\hat{\beta} \sim N(\beta, \sigma^2(\mathbf{X}'\mathbf{X})^{-1})$
3. $1/\sigma^2 \cdot (\hat{\beta} - \beta)'(\mathbf{X}'\mathbf{X})(\hat{\beta} - \beta) \sim \chi_K^2$
4. $1/\sigma^2 \cdot \hat{\varepsilon}'\hat{\varepsilon} \sim \chi_{T-K}^2$
5. $\hat{\varepsilon}'\hat{\varepsilon}$ und $\hat{\beta}$ sind unabhängig.

Beweis:

zu 1) Da $y = \mathbf{X}\beta + \varepsilon$ und $\varepsilon \sim N(0, \sigma^2\mathbf{I})$ können wir Satz 9.6 1) anwenden und erhalten

$$y \sim N(\mathbf{X}\beta + 0, \sigma^2\mathbf{I}) = N(\mathbf{X}\beta, \sigma^2\mathbf{I}).$$

zu 2) Da $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'y$ eine lineare Transformation von y ist, können wir wieder Satz 9.6 1) anwenden. Unter Zuhilfenahme von 1) erhalten wir

$$\hat{\beta} \sim N(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\beta, (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\sigma^2\mathbf{I}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}) = N(\beta, \sigma^2(\mathbf{X}'\mathbf{X})^{-1}).$$

zu 3) Die Behauptung ist ein Spezialfall von Satz 9.12.

zu 4) Zum Beweis verwenden wir die idempotente Matrix $\mathbf{Q} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ aus Abschnitt 10.4.2. Dafür gilt

$$\mathbf{Q}\mathbf{X} = \mathbf{X} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X} = 0.$$

Unter Verwendung von $\varepsilon/\sigma \sim N_T(0, \mathbf{I})$ und unter Zuhilfenahme von Eigenschaften von \mathbf{Q} folgt

$$\begin{aligned} \frac{1}{\sigma^2}\hat{\varepsilon}'\hat{\varepsilon} &= \frac{1}{\sigma^2}y'\mathbf{Q}y \\ &= \frac{1}{\sigma^2}(\mathbf{X}\beta + \varepsilon)'\mathbf{Q}(\mathbf{X}\beta + \varepsilon) \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{\sigma^2} (\beta' \mathbf{X}' + \varepsilon') \mathbf{Q} \mathbf{Q} (\mathbf{X} \beta + \varepsilon) \\
&= \frac{1}{\sigma^2} (\beta' \underbrace{\mathbf{X}' \mathbf{Q}}_{=0} + \varepsilon' \mathbf{Q}) (\underbrace{\mathbf{Q} \mathbf{X}}_{=0} \beta + \mathbf{Q} \varepsilon) \\
&= \frac{1}{\sigma^2} \varepsilon' \mathbf{Q} \varepsilon \\
&= \frac{\varepsilon'}{\sigma} \mathbf{Q} \frac{\varepsilon}{\sigma}.
\end{aligned}$$

Die Behauptung folgt nun durch Anwendung von Satz 9.13 1).

zu 5) Es ist zu zeigen, dass $\frac{1}{\sigma}(\hat{\beta} - \beta)$ und $\frac{1}{\sigma^2} \hat{\varepsilon}' \hat{\varepsilon}$ unabhängig sind. Ziel ist es dabei Satz 9.13 2) anzuwenden. Es gilt

$$\begin{aligned}
\frac{1}{\sigma}(\hat{\beta} - \beta) &= \frac{1}{\sigma} \{(\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' y - \beta\} \\
&= \frac{1}{\sigma} \{(\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' (\mathbf{X} \beta + \varepsilon) - \beta\} \\
&= \frac{1}{\sigma} (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \varepsilon \\
&= (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \frac{\varepsilon}{\sigma}.
\end{aligned}$$

Weiter gilt (vgl. den Beweis zu 1))

$$\frac{1}{\sigma^2} \hat{\varepsilon}' \hat{\varepsilon} = \frac{\varepsilon'}{\sigma} \mathbf{Q} \frac{\varepsilon}{\sigma}.$$

Darüberhinaus gilt

$$(\mathbf{X}' \mathbf{X})^{-1} \underbrace{\mathbf{X}' \mathbf{Q}}_{=0} = 0.$$

Damit können wir Satz 9.13 2) verwenden, indem wir dort $\mathbf{R} := \mathbf{Q}$ und $\mathbf{B} := (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}'$ setzen.

□

11.3 Tests für allgemeine lineare Hypothesen

Ziel dieses Abschnitts ist die Konstruktion eines statistischen Tests, mit dem allgemeine lineare Hypothesen der Form

$$H : \quad \mathbf{R} \quad \beta \quad = \quad r$$

$(J \times K) \quad (K \times 1) \quad (J \times 1)$

getestet werden können.

Zur Lösung dieses Problems werden wir in etwa wie folgt vorgehen:

1. Berechne mit $SSE = \hat{\varepsilon}'\hat{\varepsilon}$ die Residuenquadratsumme im vollen Modell.
2. Berechne mit $SSE_H = \hat{\varepsilon}'_H\hat{\varepsilon}_H$ die Residuenquadratsumme im Modell unter der Nullhypothese, d.h. wenn die Restriktion $\mathbf{R}\beta = r$ gilt. Dabei wird die restringierte KQ-Schätzung aus Abschnitt 10.3 eine zentrale Rolle spielen.
3. Verwende als Teststatistik einen Ausdruck der Form

$$\frac{\Delta SSE}{SSE} = \frac{SSE_H - SSE}{SSE},$$

d.h. die relative Differenz zwischen den Residuenquadratsummen im restringierten Modell und vollen im Modell. Die Differenz $SSE_H - SSE$ ist dabei stets größer oder gleich Null. Intuitiv ist das einleuchtend, denn die KQ-Schätzung unter Beachtung einer Restriktion für β kann niemals so "gut" sein wie eine Schätzung, bei der keinerlei Restriktionen zu beachten sind. Ist aber die KQ-Schätzung unter der Restriktion "schlechter", dann muss auch die Residuenquadratsumme SSE_H größer sein als die Residuenquadratsumme SSE für die unrestringierte Schätzung. Für diese intuitive Erkenntnis werden wir im Folgenden auch noch einen formalen Beweis erhalten.

Informell führt obiges Vorgehen dann zu folgender Entscheidungsregel:

Ist ΔSSE hinreichend klein, d.h. nahe bei 0, so ist die Vergrößerung der Residuenquadratsumme gegenüber dem vollen Modell vernachlässigbar und wir können die Nullhypothese H beibehalten bzw. nicht ablehnen.

Ist umgekehrt ΔSSE "groß", so ist die Verschlechterung gegenüber dem vollen Modell nicht mehr vernachlässigbar und die Nullhypothese H wird abgelehnt.

Das Ziel der beiden folgenden Abschnitte ist die Bestimmung der Teststatistik und insbesondere die Herleitung der Verteilung der Teststatistik unter der Nullhypothese. Schließlich ist die Verteilung der Teststatistik die Voraussetzung für die Bestimmung von Annahme- und Ablehnbereichen der Nullhypothese.

11.3.1 Bestimmung von SSE_H und ΔSSE

Unter der Hypothese H gilt gemäß Satz 10.3 für den restringierten KQ-Schätzer $\hat{\beta}_H$:

$$\begin{aligned}\hat{\beta}_H &= \hat{\beta} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'\left(\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'\right)^{-1}(\mathbf{R}\hat{\beta} - r) \\ &= \hat{\beta} - \Delta_H,\end{aligned}$$

wobei $\hat{\beta}$ der gewöhnliche KQ-Schätzer ist und Δ_H definiert ist als

$$\Delta_H := (\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}' \left(\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}' \right)^{-1} (\mathbf{R}\hat{\beta} - r).$$

Damit erhalten wir für die geschätzten Werte \hat{y}_H unter Beachtung der Restriktion (d.h. unter der Nullhypothese)

$$\hat{y}_H = \mathbf{X}\hat{\beta}_H = \mathbf{X}(\hat{\beta} - \Delta_H) = \mathbf{X}\hat{\beta} - \mathbf{X}\Delta_H = \hat{y} - \mathbf{X}\Delta_H$$

und für die Residuen $\hat{\varepsilon}_H$ unter H

$$\hat{\varepsilon}_H = y - \hat{y}_H = y - \hat{y} + \mathbf{X}\Delta_H = \hat{\varepsilon} + \mathbf{X}\Delta_H.$$

Für die Residuenquadratsumme SSE_H unter H erhalten wir

$$\begin{aligned} SSE_H &= \hat{\varepsilon}'_H \hat{\varepsilon}_H \\ &= (\hat{\varepsilon} + \mathbf{X}\Delta_H)' (\hat{\varepsilon} + \mathbf{X}\Delta_H) \\ &= \hat{\varepsilon}'\hat{\varepsilon} + \underbrace{\hat{\varepsilon}'\mathbf{X}}_{=0} \Delta_H + \Delta'_H \underbrace{\mathbf{X}'\hat{\varepsilon}}_{=0} + \Delta'_H \mathbf{X}'\mathbf{X}\Delta_H \\ &= \hat{\varepsilon}'\hat{\varepsilon} + \Delta'_H \mathbf{X}'\mathbf{X}\Delta_H. \end{aligned}$$

Da $\mathbf{X}'\mathbf{X}$ positiv definit ist und damit $\Delta'_H \mathbf{X}'\mathbf{X}\Delta_H > 0$ haben wir auch den formalen Beweis erbracht, dass die Residuenquadratsumme unter H stets größer ist als die unrestringierte KQ-Schätzung. Diese Aussage vervollständigt den Beweis von Satz 10.9.

Schließlich erhalten wir für die Differenz ΔSSE der Residuenquadratsummen:

$$\begin{aligned} \Delta SSE &= SSE_H - SSE \\ &= \hat{\varepsilon}'\hat{\varepsilon} + \Delta'_H \mathbf{X}'\mathbf{X}\Delta_H - \hat{\varepsilon}'\hat{\varepsilon} \\ &= \Delta'_H \mathbf{X}'\mathbf{X}\Delta_H \\ &= \left\{ (\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}' \left(\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}' \right)^{-1} (\mathbf{R}\hat{\beta} - r) \right\}' \mathbf{X}'\mathbf{X} \cdot \\ &\quad \cdot \left\{ (\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}' \left(\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}' \right)^{-1} (\mathbf{R}\hat{\beta} - r) \right\} \\ &= (\mathbf{R}\hat{\beta} - r)' \left(\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}' \right)^{-1} \mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}' \left(\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}' \right)^{-1} (\mathbf{R}\hat{\beta} - r) \\ &= (\mathbf{R}\hat{\beta} - r)' \left(\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}' \right)^{-1} (\mathbf{R}\hat{\beta} - r) \end{aligned}$$

Wir fassen das wesentliche Resultat dieses Abschnitts nochmal im folgenden Satz zusammen:

Satz 11.3

Im klassischen linearen Modell gilt für die Differenz der Residuenquadratsummen ΔSSE im restringierten und im unrestringierten Modell

$$\Delta SSE = (\mathbf{R}\hat{\beta} - r)' \left(\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}' \right)^{-1} (\mathbf{R}\hat{\beta} - r),$$

wobei die lineare Nebenbedingung gegeben ist durch $\mathbf{R}\beta = r$.

11.3.2 Stochastische Eigenschaften der Differenz ΔSSE

Satz 11.4 (Stochastische Eigenschaften von ΔSSE)

Im klassischen linearen Modell gelten für die Differenz der Residuenquadratsummen ΔSSE im restringierten und im unrestringierten Modell folgende stochastische Eigenschaften:

1. $E(\Delta SSE) = J\sigma^2 + (\mathbf{R}\beta - r)' \left(\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}' \right)^{-1} (\mathbf{R}\beta - r)$
2. Unter $H : \mathbf{R}\beta = r$ gilt: $1/\sigma^2 \cdot \Delta SSE \sim \chi_J^2$
3. ΔSSE und SSE sind stochastisch unabhängig.

Beweis:

zu 1) Zum Beweis dieser Aussage verwenden wir Satz 10.14 über den Erwartungswert quadratischer Formen. Es gilt

$$E(\mathbf{R}\hat{\beta} - r) = \mathbf{R}\beta - r$$

und

$$\text{Cov}(\mathbf{R}\hat{\beta} - r) = \sigma^2 \mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'.$$

Wir verwenden Satz 10.14, indem wir dort $Z := \mathbf{R}\hat{\beta} - r$ und $\mathbf{A} := \left(\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}' \right)^{-1}$ setzen und erhalten

$$\begin{aligned} E(\Delta SSE) &= E \left\{ (\mathbf{R}\hat{\beta} - r)' \left(\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}' \right)^{-1} (\mathbf{R}\hat{\beta} - r) \right\} \\ &= sp \left\{ \sigma^2 \left(\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}' \right)^{-1} \mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}' \right\} + \\ &\quad (\mathbf{R}\beta - r)' \left(\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}' \right)^{-1} (\mathbf{R}\beta - r) \\ &= sp(\sigma^2 I_J) + (\mathbf{R}\beta - r)' \left(\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}' \right)^{-1} (\mathbf{R}\beta - r) \\ &= J\sigma^2 + (\mathbf{R}\beta - r)' \left(\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}' \right)^{-1} (\mathbf{R}\beta - r). \end{aligned}$$

zu 2) Die Behauptung ist eine Folgerung aus Satz 9.12. Wir definieren den Zufallsvektor $Z := \mathbf{R}\hat{\beta}$. Unter der Nullhypothese H gilt

$$E(Z) = \mathbf{R}\beta = r$$

und

$$\text{Cov}(Z) = \sigma^2 \mathbf{R}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{R}'.$$

Da $\hat{\beta}$ normalverteilt ist, folgt weiterhin

$$Z \sim N(r, \sigma^2 \mathbf{R}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{R}').$$

Die Behauptung folgt nun unmittelbar durch Anwendung von Satz 9.12 auf den Zufallsvektor Z .

zu 3) ΔSSE ist eine Funktion von $\hat{\beta}$. Damit folgt die Behauptung unmittelbar aus Satz 11.2 5). □

Mit den Aussagen von Satz 11.4 können wir jetzt die Verteilung der Teststatistik unter der Nullhypothese bestimmen:

Satz 11.5 (Verteilung der Teststatistik unter H)

Im klassischen linearen Modell unter Normalverteilungsannahme ist die Teststatistik

$$F := \frac{\frac{1}{J} \Delta SSE}{\frac{1}{T-K} SSE} = \frac{T-K}{J} \frac{\Delta SSE}{SSE}$$

F verteilt mit J und $T-K$ Freiheitsgraden, d.h.

$$F \sim F_{J, T-K}.$$

Beweis:

Nach Satz 11.4 2) gilt

$$\frac{1}{\sigma^2} \Delta SSE \sim \chi_J^2$$

und nach Satz 11.2 3) gilt

$$\frac{1}{\sigma^2} SSE \sim \chi_{T-K}^2.$$

Darüberhinaus sind ΔSSE und SSE nach Satz 11.4 3) stochastisch unabhängig. Damit folgt die Behauptung aus der Definition der F-Verteilung (vgl Definition 9.8). □

Damit kommen wir zu folgendem Test: Die Nullhypothese wird abgelehnt, falls die Teststatistik größer als das $(1 - \alpha)$ -Quantil der entsprechenden F -Verteilung ist. Im vorliegenden Fall also, falls

$$F > F_{J,T-K}(1 - \alpha).$$

Dabei ist α das Signifikanzniveau des Tests.

Bemerkungen:

Der soeben hergeleitete F -Test kann auch als Likelihood-Quotienten-Test aufgefasst werden. Die Parameter des Modells seien zum Vektor $\theta = (\beta, \sigma^2)'$ zusammengefasst. Sei weiterhin

$$\Theta := \left\{ \theta \in \mathbf{R}^{K+1} \mid \beta \in \mathbf{R}^K, \sigma^2 > 0 \right\}$$

der Parameterraum des vollen Modells, und

$$\Theta_H := \left\{ \theta \in \mathbf{R}^{K+1} \mid R\beta = r, \sigma^2 > 0 \right\}$$

der Parameterraum des restringierten Modells.

Allgemein wird beim Likelihood-Quotienten-Test (LQ-Test) die maximale Log-likelihood des restringierten Modells mit der maximalen Log-Likelihood des vollen Modells verglichen, d.h. es wird die Teststatistik

$$\lambda = \frac{\max_{\theta \in \Theta_H} l(\theta)}{\max_{\theta \in \Theta} l(\theta)}.$$

verwendet. Im vorliegenden Fall ist die Log-likelihood $l(\theta)$ durch (11.2) gegeben. Für λ gilt

$$\lambda = \left(\frac{SSE_H}{SSE} \right)^{-T/2}$$

und wir erhalten den Zusammenhang

$$F = \left(\lambda^{-2/T} - 1 \right) \frac{T - K}{J}.$$

Die Teststatistik F ist also lediglich eine monotone Transformation der Teststatistik im Likelihood-Quotienten-Test, so daß der vorliegende F -Test auch als LQ -Test angesehen werden kann.

11.3.3 Einige spezielle Testprobleme

In diesem Abschnitt behandeln wir einige spezielle für die Praxis bedeutende Testprobleme etwas genauer.

Test einzelner Parameter auf Signifikanz (t-Test)

Im einfachsten Fall will man testen, ob eine bestimmte Einflußgröße, z.B. X_i , einen signifikanten Einfluß besitzt. Besitzt die Kovariable X_i keinen Einfluß, ist dies gleichbedeutend damit, dass der i -te Regressionskoeffizient β_i gleich Null ist. Die Nullhypothese lautet also:

$$H : \beta_i = 0$$

Wir behandeln hier gleich den allgemeineren Fall

$$H : \beta_i = \beta_i^*.$$

mit β_i^* beliebig. Für \mathbf{R} und r gilt in diesem Fall:

$$\begin{aligned} \mathbf{R} &= (0, \dots, 0, \underset{i}{1}, 0, \dots, 0) \\ (1 \times K) & \\ r &= \beta_i^* \\ (1 \times 1) & \end{aligned}$$

Damit folgt zunächst

$$\frac{(\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}')^{-1}}{\hat{\sigma}^2} = (\mathbf{R} \widehat{Cov}(\hat{\beta}) \mathbf{R})^{-1} = (\hat{\sigma}^2(\hat{\beta}_i))^{-1},$$

wobei $\hat{\sigma}^2(\hat{\beta}_i)$ die geschätzte Varianz von $\hat{\beta}_i$ ist (vergleiche auch Korollar 10.2). Damit erhalten wir

$$F = \frac{(\hat{\beta}_i - \beta_i^*)^2}{\hat{\sigma}^2(\hat{\beta}_i)} \sim F_{1, T-K}.$$

Äquivalent dazu kann man den Test auch auf der Wurzel von F aufbauen, die t verteilt ist:

$$t = \frac{b_i - \beta_i^*}{\hat{\sigma}(b_i)} \sim t_{T-K} \quad (11.3)$$

Den kritischen Wert für den Ablehnbereich der Nullhypothese erhält man bei dieser Vorgehensweise als $\alpha/2$ -Fraktil einer t -Verteilung mit $T - K$ Freiheitsgraden. Wenn dieses Fraktil abkürzend mit $\tau_{T-K}(\alpha/2)$ bezeichnet wird, so lautet die Entscheidungsregel beim t -Test:

$$H \text{ ablehnen falls } |t| > \tau_{T-K}(\alpha/2).$$

Test eines Subvektors

Der Parametervektor β sei partitioniert in

$$\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix},$$

wobei β_1 ein $(J \times 1)$ -Vektor, und β_2 ein $(K - J \times 1)$ -Vektor sei. Man beabsichtigt nun den Test von Hypothesen der Form

$$H : \beta_1 = \beta_1^*.$$

Die beiden Größen \mathbf{R} und r für die Restriktionsbedingung ergeben sich dann als:

$$\mathbf{R}_{(J \times K)} = \begin{pmatrix} 1 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & & \vdots \\ 0 & \cdots & 1 & 0 & \cdots & 0 \end{pmatrix}$$

$$r_{(J \times 1)} = \beta_1^*$$

Damit besteht die Matrix $\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'$ aus der $(J \times J)$ -Submatrix von $(\mathbf{X}'\mathbf{X})^{-1}$, welche den Vektor β_1 betrifft. Folglich handelt es sich bei der Matrix

$$\frac{(R(X'X)^{-1}R')^{-1}}{\hat{\sigma}^2}$$

um die Inverse der geschätzten Kovarianzmatrix $\widehat{\text{Cov}}(\hat{\beta}_1)$ des Schätzers $\hat{\beta}_1$ für β_1 . Damit gilt für die Teststatistik in dieser Testsituation:

$$F = \frac{1}{J}(\hat{\beta}_1 - \beta_1^*)' \widehat{\text{Cov}}(\hat{\beta}_1)^{-1} (\hat{\beta}_1 - \beta_1^*) \sim F_{J, T-K}. \tag{11.4}$$

Testen der Hypothese „kein funktionaler Zusammenhang“ im Regressionsmodell mit Intercept

Ausgangspunkt sei das multiple Regressionsmodell mit Intercept:

$$y_t = \beta_0 + \beta_1 x_{1t} + \cdots + \beta_K x_{Kt} + \varepsilon_t \quad (t = 1, \dots, T)$$

Getestet werden soll die Hypothese

$$H_0 : \beta_1 = \beta_2 = \cdots = \beta_K = 0,$$

d.h. keine der Kovariablen besitzt einen Einfluß. (Vorsicht: Die Ablehnung der Hypothese bedeutet nicht automatisch, dass *alle* Variablen einen Einfluß besitzen.) In diesem Fall besteht die restringierte KQ-Schätzung nur aus einer Schätzung $\hat{\beta}_0$ für β_0 und es gilt

$$\hat{\beta}_0 = \bar{y}.$$

Folglich erhalten wir für die Residuenquadratsumme SSE_H unter der Nullhypothese

$$SSE_H = \sum_{t=1}^T (y_t - \bar{y})^2 = SST.$$

Für die Differenz zwischen der Residuenquadratsumme unter H und derjenigen im vollen Modell gilt unter Verwendung von Satz 10.8

$$\Delta SSE = SSE_H - SSE = SSR.$$

Damit folgt für die Verteilung der Teststatistik F :

$$\begin{aligned} F &= \frac{T - K - 1}{K} \cdot \frac{SSR}{SSE} \\ &= \frac{T - K - 1}{K} \cdot \frac{SSR}{SST - SSR} \\ &= \frac{T - K - 1}{K} \cdot \frac{SSR/SST}{1 - SSR/SST} \\ &= \frac{T - K - 1}{K} \cdot \frac{B}{1 - B} \sim F_{K, T-K-1}. \end{aligned}$$

Interpretation:

Für ein kleines multiples Bestimmtheitsmaß wird die Hypothese „kein funktionaler Zusammenhang“ eher beibehalten (da F dann klein ist), als bei einem Bestimmtheitsmaß nahe bei 1 (in diesem Fall ist F vergleichsweise groß).

11.3.4 Konfidenzintervalle und Konfidenzellipsoide

Aufgrund der Dualität zwischen Tests und Konfidenzbereichen kann man sofort Konfidenzintervalle für einen einzelnen Parameter β_i bzw. Konfidenzellipsoide für einen Subvektor β_1 von β konstruieren.

Als Konfidenzintervall für β_i zum Vertrauensgrad $\gamma = 1 - \alpha$ erhält man unter Zuhilfenahme der Formel (11.3) folgendes Resultat:

$$\left[\hat{\beta}_i - \tau_{T-K}(\alpha/2) \hat{\sigma}(\hat{\beta}_i); \hat{\beta}_i + \tau_{T-K}(\alpha/2) \hat{\sigma}(\hat{\beta}_i) \right]$$

Dabei ist $\tau_{T-K}(\alpha/2)$ wieder das entsprechende Fraktile einer t -Verteilung mit $T - K$ Freiheitsgraden.

Sei nun β partitioniert in $\beta = (\beta_1', \beta_2')'$. Dann erhält man unter Benutzung der Aussage (11.4) das folgende Konfidenzellipsoid für den Subvektor β_1 :

$${}^1/J(\hat{\beta}_1 - \beta_1)' \widehat{\text{Cov}}(\hat{\beta}_1)^{-1} (\hat{\beta}_1 - \beta_1) \leq F_{J, T-K}(1 - \alpha).$$

A

Körper

Definition (Körper):

Ein Körper ist ein Tripel $(K, +, \cdot)$, bestehend aus einer Menge K und zwei Verknüpfungen $+$ und \cdot auf K (Addition und Multiplikation) d.h. einer Abbildung

$$\begin{aligned} + : K \times K &\mapsto K \\ (a, b) &\mapsto a + b \end{aligned}$$

und einer Abbildung

$$\begin{aligned} \cdot : K \times K &\mapsto K \\ (a, b) &\mapsto a \cdot b \end{aligned}$$

mit den folgenden Eigenschaften (Körperaxiomen):

I. Axiome der Addition

1. Assoziativgesetz: $a + (b + c) = (a + b) + c$ für alle $a, b, c \in K$.
2. Kommutativgesetz: $a + b = b + a$ für alle $a, b \in K$
3. Existenz der Null: Es existiert eine Zahl $0 \in K$ mit $a + 0 = a$ für alle $a \in K$
4. Existenz des Negativen: Zu jedem $a \in K$ existiert eine Zahl $-a \in K$ mit $a + (-a) = 0$.

II. Axiome der Multiplikation

1. Assoziativgesetz: $(ab)c = a(bc)$ für alle $a, b, c \in K$.
2. Kommutativgesetz: $ab = ba$ für alle $a, b \in K$
3. Existenz der Eins: Es gibt eine Zahl $1 \in K, 1 \neq 0$, so dass $a \cdot 1 = a$ für alle $a \in K$.
4. Existenz der Inversen: Zu jedem von Null verschiedenen $a \in K$ gibt es ein $a^{-1} \in K$ mit $a \cdot a^{-1} = 1$.

III. Distributivgesetz

$$a(b + c) = ab + ac \text{ für alle } a, b, c \in K.$$

Aus den Körperaxiomen lassen sich einige wichtige Rechenregeln ableiten, die im folgenden Satz zusammengefasst sind:

Satz A1:

Sei K ein Körper. Dann gilt für alle $a, b, c \in K$:

1. Die Zahlen Null und Eins sind eindeutig bestimmt.
2. Das Negative und das Inverse einer Zahl ist eindeutig bestimmt.
3. Die Gleichung $a + x = b$ hat eine eindeutige Lösung, nämlich $x = b - a$.
4. $-(-a) = a$
5. $-(a + b) = -a - b$
6. Die Gleichung $ax = b$ ist für $a \neq 0$ eindeutig durch $x = ba^{-1}$ lösbar.
7. $a \cdot 0 = 0$
8. $ab = 0 \leftrightarrow a = 0$ oder $b = 0$
9. $(-a)(-b) = ab$
10. $(a^{-1})^{-1} = a$
11. $(ab)^{-1} = a^{-1}b^{-1}$

Beispiele für Körper:

1. Die Menge \mathbb{R} der reellen Zahlen, versehen mit der üblichen Addition und Multiplikation.
2. Die Menge \mathbb{Q} der rationalen Zahlen, definiert durch

$$\mathbb{Q} := \{z/n \in \mathbb{R} : z, n \in \mathbb{N}\},$$

versehen mit der üblichen Addition und Multiplikation.

3. Die Menge \mathbb{C} aller geordneten Paare reeller Zahlen, versehen mit den beiden wie folgt definierten Verknüpfungen:

$$(a, b) + (c, d) = (a + c, b + c) \quad (\text{Addition})$$

$$(a, b) \cdot (c, d) = (ac - bd, ad + bc) \quad (\text{Multiplikation})$$

\mathbb{C} heißt Körper der komplexen Zahlen.

Literaturverzeichnis

- Billingsley, P., 1985 *Probability and Measure. (2nd ed.)* Wiley, New York.
- Bronstein, I.N. und Semendjajew, K.A., 1991: *Taschenbuch der Mathematik*. Teubner, Stuttgart.
- Fahrmeir, L., Künstler, R., Pigeot, I. und Tutz, G., 2001 (4. Auflage): *Statistik. Der Weg zur Datenanalyse*. Springer Verlag.
- Forster, O., 1999: *Analysis I*. Vieweg, Braunschweig.
- George, A. and Liu, J.W.H., 1981: *Computer Solution of Large Sparse Positive Definite Systems*. Prentice-Hall, London.
- Hämmerlin G. und Hoffman, K.H., 1990: *Numerische Mathematik*. Springer-Verlag, Berlin.
- Lipschutz S. (1990) *Lineare Algebra*. Mc-Graw-Hill, London.
- Schott, J.R. (1997) *Matrix Analysis for Statistics*. Wiley, New York.
- Searle, S.R. (1982) *Matrix Algebra for Statistics*. Wiley, New York.

Index

- χ^2 -Verteilung, 148
 - Eigenschaften, 149
- ähnliche Matrizen, 115

- Abstand zweier Vektors, 6
- adjustiertes Bestimmtheitsmas, 172
- algebraische Vielfachheit, 115
- Algorithmus zur Reduzierung auf Diagonalfom, 29
- Algorithmus zur Reduzierung auf Dreiecksform, 28
- Austauschsatz, 47

- Basis eines Vektorraums, 48
- Berechnung der Determinante, 88
- Bestimmtheitsmas, 171
 - adjustiertes, 172
- Betrag, 34
- Betrag einer komplexen Zahl, 35
- Bias einer Schätzung, 173
- Bild einer linearen Abbildung, 58

- Cauchy–Schwarzsche Ungleichung, 63
- charakteristisches Polynom, 110
- Choleskyzerlegung, 126

- Defekt einer Matrix, 72
- Determinante, 85
 - bei Cholesky Zerlegung, 127
 - Berechnung, 88
 - der Transponierten, 86
- Determinanten der Elementarmatrizen, 88
- Diagonalfom, 29
- Diagonalmatrix, 10
- Differenziation einer Matrix nach einem Skalar, 137
- Differenziation einer Matrixfunktion nach der Matrix, 138
- Dimension eines Vektorraums, 48
- direkte Summe, 42
- Dreiecksform einer Matrix, 21
- Durchschnitt von Unterräumen, 42
- dyadisches Produkt, 15

- Eigenraum, 113
- Eigenvektor, 109
- Eigenwert, 109
- Eigenwerte
 - Eigenschaften, 111
- Eigenwerte symmetrischer Matrizen, 116
- Eigenwertproblem, 109
- Einheitsmatrix, 10
- Einsvektor, 2
- Elementare Matrixoperationen, 27
- empirische Streuung, 75
- empirischer Erwartungswert, 75
- Erwartungstreue Schätzung, 173
- Erwartungswert quadratischer Formen, 178
- erweiterte Koeffizientenmatrix, 96
- Erzeugendensystem, 46
- euklidische Norm, 60
- euklidischer Abstand, 60
- euklidischer Raum, 6
- euklidischer Vektorraum, 61

- F-Verteilung, 153
- Fundamentalsatz der Algebra, 110

- g-Inverse, 104
- Gammafunktion, 139
- Gammaverteilung, 147
 - Eigenschaften, 147
- Gaus-Markov-Theorem, 175
- geometrische Vielfachheit, 115
- Gram–Schmidtsches Orthonormalisierungsverfahren, 65

- Hat-Matrix, 166
 - Eigenschaften, 166
- Hauptdiagonale, 10
- Homomorphismus, 54

- idempotente Matrix, 19
- Imaginärteil, 34
- indefinit, 121
- inneres Produkt, 61

- Intercept, 157
inverse Matrix, 73
- Körper, 197
Körperaxiome, 197
kanonische Basis, 49
Kern einer linearen Abbildung, 58
klassisches lineares Regressionsmodell, 156
– ML-Schatzer, 183
Kofaktoren, 91
komplexe Matrix, 36
komplexe Zahl, 34
Komplexe Zahlen, 33
konjugierter, 34
Koordinaten, 50
KQ Methode, 158
KQ-Schatzer
– Eigenschaften, 175
KQ-Schatzer unter linearen Restriktionen, 164
KQ-Schätzung
– Geometrische Eigenschaften, 167
Kroneckersymbol, 49
Kurzungsregel, 18
- Länge eines Vektors, 59
Lange eines Vektors, 6
lineare Abbildung, 54
– Matrixdarstellung, 55
lineare Abhängigkeit, 43
lineare Schätzer
– Eigenschaften, 174
lineare Unabhängigkeit, 43
linearer Operator, 54
linearer Schätzer, 173
lineares Gleichungssystem, 22, 95
– homogenes, 22, 95
– inhomogenes, 22, 95
– inkonsistent, 22, 95
– konsistent, 22, 95
Linearkombination von Vektoren, 43
Losungen der Normalgleichungen, 160
Lösungsraum, 98
- Matrix, 7
– ähnlich, 115
– definite, 121
– Diagonal-, 10
– idempotente, 19
– inverse, 73
– orthogonale, 19
– Potenzen, 117
– quadratische, 10
– reell, 7
– reguläre, 10
– symmetrische, 10
– transponiert, 9
– Wurzel, 117
Matrixdarstellung einer linearen Abbildung, 55
Matrixmultiplikation
– Rechenregeln, 17
Matrizenaddition, 12
Matrizenmultiplikation, 13
Mean Squared Error, 173
– Zerlegungsformel, 173
Methode der kleinsten Quadrate, 158
Metrik, 60
metrischer Raum, 60
Minoren, 91
ML-Schatzer im klassischen linearen Regressionsmodell, 183
momentenerzeugende Funktion, 140
– Eigenschaften, 140
Moore–Penrose–Inverse, 107
MSE, 173
– Zerlegungsformel, 173
multivariate Normalverteilung, 143
– Dichte, 145
– Eigenschaften, 145
– Marginalverteilungen, 146
- negativ definit, 121
Norm, 59
Normierter Vektorraum, 59
Nullraum einer Matrix, 72
Nullvektor, 2
- Ordnung einer Matrix, 7
orthnormales Komplement, 67
orthogonale Matrix, 19
orthogonale Menge, 63
orthogonale Unterräume, 67
Orthogonalität, 63
Orthonormalität, 63
- partitionierte Matrix, 11
Permutation, 83
positiv definit, 121
Potenzen einer Matrix, 117
Predicton-Matrix, 166
- quadratische Form, 121
quadratische Matrix, 10

- Rang
 - Berechnung, 77
- Rang einer Matrix, 70
- Realteil, 34
- Reduzierung auf Diagonalform, 29
- reelle Matrix, 7
- reeller Vektorraum, 59
- reguläre Inverse, 73
- reguläre Matrix, 10
- restringierter KQ-Schatzer
 - Eigenschaften, 181

- Satz von Student, 152
- Schätzung für σ^2 , 179
- semidefinit, 121
- Signum einer Permutation, 83
- Singulärwertzerlegung, 125
- Singularwerte, 125
- Skalare, 2
- skalare Multiplikation, 12
- Skalarmultiplikation, 2
- Skalarprodukt, 61
- Skalarprodukt im \mathbb{R}^n , 5
- Spaltenrang, 69
- Spaltenraum, 69
- spaltenregulär, 70
- Spaltenvektor, 7
- Spektralzerlegung, 117
 - idempotente Matrix, 119
- Spur, 93
- Standardbasis, 49
- Standardnormalverteilung, 139
 - Eigenschaften, 141
- Standardskalarprodukt, 5, 62
- Strengszerlegung, 169
 - Modell mit Intercept, 169
- Submatrix, 11
- symmetrische Matrix, 10

- t-Verteilung, 152
- Teilmatrix, 11
- transponierte Matrix, 9
- Transposition, 84
- Tschebyscheff Norm, 60

- Unendlichnorm, 60
- unitärer Vektorraum, 61
- univariate Normalverteilung, 141
 - Eigenschaften, 142
- univariates Regressionsmodell, 157

- Vektoraddition, 2
- Vektorraum, 37
 - normiert, 59
- Vektorraum der $n \times m$ Matrizen, 38
- Vektorraum der Polynome, 39
- Vektorraum-Homomorphismus, 54
- Vektorraumaxiome, 37
- verallgemeinerte Inverse, 104
- Vielfachheit, 115
 - algebraische, 115
 - geometrische, 115
- Vollrang Zerlegung, 81

- Wurzel einer Matrix, 117

- Zeilenrang, 69
- Zeilenraum, 69
- zeilenregulär, 70
- Zeilenvektor, 7