.

Einführung in die Vektor- und Matrixrechnung für die Multivariate Statistik

U. Mortensen

.

Letzte Korrektur 22. 01. 2020

Inhaltsverzeichnis

1 Vektoren und Vektorräume **5**

1.1 Zur Bedeutung der Vektor- und Matrixrechnung 5

1.2 Punkträume 7

1.3 Vektoren und ihre Verknüpfungen 10

1.4 Lineare Abhängigkeit und Unabhängigkeit von Vektoren 22

1.5 Vektorräume 28

1.6 Erzeugendensysteme und Basen von Vektorräumen 31

2 Matrizen **42**

2.1 Definitionen 42

2.2 Operationen mit Matrizen 43

2.2.1 Addition und Multiplikation mit einem Skalar 43

2.2.2 Multiplikation einer Matrix mit einem Vektor 43

2.2.3 Multiplikation einer Matrix mit einer Matrix 47

2.3 Zufällige Vektoren und Matrizen 50

2.4 Der Rang einer Matrix 53

2.5 Die Inverse einer Matrix 58

2.6 Eigenvektoren symmetrischer Matrizen 62

2.6.1 Rotationen 63

2.6.2 Rotationen und Koordinatentransformationen 64

2.6.3 Quadratische Formen und Eigenvektoren 65

2.6.4 Inverse und Wurzel einer symmetrischen Matrix 79

2.6.5 Der Rayleigh-Quotient 80

2.6.6 Die Singularwertzerlegung einer Matrix 83

2.7 Eigenvektoren und Eigenwerte nichtsymmetrischer Matrizen 87

2.7.1 Der allgemeine Fall 87

2.7.2 Mehrfache Eigenwerte 93

2.7.3 Das generalisierte Eigenvektorproblem 94

2.8 Weitere Befunde 97

2.8.1 Die Zentrierungsmatrix 97

2.8.2 Die Beziehungen zwischen verschiedenen Basen 100

2.8.3 Die Pseudoinverse einer Matrix . . . 102
2.8.4 Vektor- und Matrixnormen . . . 104
2.8.5 Die Approximation von Matrizen . . . 107
2.9 Lineare Gleichungssysteme . . . 111
2.10 Latente Variablen . . . 115
2.10.1 Hauptkomponenten und SVD . . . 115
2.10.2 Diskriminieren und klassifizieren . . . 122
2.11 Projektionen . . . 127
2.11.1 Orthogonale Projektion eines Vektors auf einen anderen . . 127
2.11.2 Projektionsmatrizen . . . 128
2.11.3 Projektionen auf Hauptachsen . . . 130

3 Funktionenräume und PCA **131**
3.1 Einführung . . . 131
3.2 Karhunen-Loève-Entwicklung und PCA . . . 133

4 Anhang **138**
4.1 Elementarmatrizen und elementare Operationen . . . 138
4.1.1 Lineare Gleichungen und Gauß-Algorithmus . . . 140
4.1.2 Die Cholesky-Zerlegung . . . 142
4.2 Steinerscher Austauschsatz und das Fundamental-Lemma . . . 144
4.3 Zur Berechnung von Ellipsen für eine Punktekonfiguration . . . 145
4.4 Die Differentiation von Vektoren . . . 147
4.4.1 Die allgemeine Differentiationsformel . . . 147
4.4.2 Die Differentiation quadratischer Formen . . . 148
4.4.3 Die Kleinste-Quadrate-Schätzung für das Lineare Modell . 148
4.4.4 Extrema unter Nebenbedingungen . . . 151
4.5 Alternativer Beweis von Satz 2.34 . . . 153
4.6 Vektortransformationen und Abbildungen . . . 155
4.7 Determinanten . . . 159
4.7.1 Transformationen und Volumen . . . 169
4.7.2 Transformationen und Orientierung . . . 172
4.7.3 Die Transformation zufälliger Veränderlicher . . . 175

Literatur **184**

Index **185**

Seit man begonnen hat, die einfachsten Behauptungen zu beweisen, erwiesen sich viele von ihnen als falsch.
Bertrand Russell

1 Vektoren und Vektorräume

1.1 Zur Bedeutung der Vektor- und Matrixrechnung

Der Beginn der Vektor- und Matrixrechnung wird oft mit der Publikation von H.G. Graßmanns[1] *Lineale Ausdehnungslehre* (1844), einem Lehrbuch der Geometrie, und W.R. Rowans[2] *Lectures on Quaternions* (1853), einem Buch über bestimmte Vierergruppen von Zahlen, angesetzt. Graßmanns und Hamiltons Arbeiten sind unabhängig voneinander entstanden, aber es war Hamilton, der den Ausdruck 'Vektor' einführte. Das Wort 'Vektor' kommt aus dem Lateinischen und bedeutet so viel wie Führer, Fahrer oder Träger, aber wie so oft in der Mathematik kann man von der Bedeutung des Wortes nicht auf die Bedeutung des damit bezeichneten mathematischen Konzepts schließen. Wie Barrantes (2012) feststellt liegen die Ursrpünge der Vektorrechnung deutlich weiter zurück in der Geschichte der Mathematik und wurzeln in Bemühungen, eine systematische Theorie über das Lösen von Gleichungen, insbesondere von Gleichungen mit mehreren Unbekannten zu schaffen. Erste Arbeiten zu diesem Thema wurden von den Babyloniern schon 1700 v. Chr. und den Chinesen (Shang Cang: *Mathematik in neun Büchern*, 180 v. Chr.) verfasst, wie Artmann und Törner (1991) ausführen. So hatte Leibniz die Aufgabe des Königs von Hannover, das Abpumpen von Grundwasser in den Bergwerken des Harzes zu organisieren, und in diesem Zusammenhang legte Leibniz erste Grundlagen für die Theorie der Determinanten (Hirsch (2000)); er fand auch bereits die später nach Cramer[3] benannte Regel (Cramersche Regel), die Cramer 1750 wohl ohne Kenntnis der Leibnizschen Arbeit publizierte. In der Tat beginnen viele Lehrbücher über die Vektor- und Matrixrechnung (Lineare Algebra) mit einem Kapitel über lineare Gleichungssysteme. Das Rechnen mit Vektoren fand aber nicht sofort zu einer weiteren Verbreitung, und dem englischen Physiker Lord Kelvin (1824 – 1907) wird die Aussage "Vektoren werden niemals für irgend jemanden den geringsten Nutzen gehabt haben". Diese Sichtweise hat sich aber nicht durchgesetzt, spätestens seit man physikalische Größen wie etwa die Kraft durch Vektoren repäsentierte. Eine Kraft hat eine bestimmte Größe und wirkt in eine bestimmte Richtung, und noch heute werden Vektoren oft als Objekte, die sowohl eine Größe wie auch eine Orientierung haben ("gerichtete Größen"), eingeführt; in diesem Sinne verwendete der schottische Physiker James Clerk Maxwell (1831 – 1879) den Vektorbegriff auch in seiner Entwick-

[1] Hermann Günther Graßmann (1609 – 1877), deutscher Mathematiker
[2] William Rowan Hamilton (1805 - 1865), irischer Mathematiker
[3] Gabriel Cramer (1704 – 1752), Genfer Mathematiker

lung der Elektrodynamik[4] . Vektoren werden dementsprechend graphisch durch ein Pfeil $\nearrow$, oder $\nwarrow$, oder $\uparrow$ etc. dargestellt, dessen Länge die Stärke der Kraft und dessen Orientierung die Richtung, in der sie wirkt anzeigen soll. Dementsprechend werden Vektoren, wenn sie in Formeln auftreten, oft ebenfalls durch Pfeile gekennzeichnet, so dass z.B. $\vec{K}$ oder $\vec{F}$, $\vec{v}$ oder $\vec{a}$ Vektoren bezeichnen. In diesem Text wird allerdings in erster Linie die im englischsprachigen Bereich übliche Bezeichnung durch fettgedruckte Buchstaben gewählt, also etwa $\mathbf{K}$, $\mathbf{k}$, $\mathbf{F}$, $\mathbf{v}$ etc., da sie oft zu lesbareren Formeln führt: so wird die Schätzung eines Parameters a in der Statistik mit $\hat{a}$ notiert. Will man die Schätzung eines ganzen Vektors anschreiben, käme man zur Notation $\vec{\hat{a}}$, die innerhalb einer Formel zu Unübersichtlicheiten führen kann.

Denn mit Vektoren kann man rechnen, so dass Formeln für Vektoren entstehen, in denen ebenfalls Vektoren vorkommen. Da Kräfte sich addieren, kann man eine Summenbildung für Vektoren erklären: sind etwa $\mathbf{k}_1$ und $\mathbf{k}_2$ zwei Kräfte, die auf einen Körper wirken:

$$\mathbf{k} = \mathbf{k}_1 + \mathbf{k}_2 = \begin{pmatrix} k_{11} \\ k_{21} \\ k_{31} \end{pmatrix} + \begin{pmatrix} k_{12} \\ k_{22} \\ k_{32} \end{pmatrix} = \begin{pmatrix} k_{11} + k_{12} \\ k_{21} + k_{22} \\ k_{31} + k_{32} \end{pmatrix} \tag{1.1}$$

Vektoren werden, wenn nichts anderes vereinbart wird, also als Spaltenvektoren angeschrieben, die Komponenten k_{1j}, k_{2j} und k_{3j} des j-ten Vektors ($j = 1, 2$) entsprechen den drei Raumdimensionen, und die Summe von zwei Vektoren definiert einen Vektor, dessen Komponenten die Summe $k_{i1} + k_{i2}$ der korrespondierenen Komponenten k_{i1} und k_{i2} ist, ($i = 1, 2, 3$). Man kann einen Vektor mit einer reellen Zahl $a \in \mathbb{R}$ multiplizieren ($\mathbb{R}$ bezeichnet die Menge der reellen Zahlen):

$$a\mathbf{k} = a \begin{pmatrix} k_1 \\ k_2 \\ k_3 \end{pmatrix} = \begin{pmatrix} ak_1 \\ ak_2 \\ ak_3 \end{pmatrix}. \tag{1.2}$$

Diese Operation verändert, wie noch gezeigt werden wird, die Länge des Vektors $\mathbf{k}$. Dann kann man Summen der Form $\mathbf{k} = a_1\mathbf{k}_1 + a_2\mathbf{k}_2$ bilden; $\mathbf{k}$ ist dann eine *Linearkombination* der Vektoren $\mathbf{k}_1$ und $\mathbf{k}_2$. Darüber hinaus läßt sich auch ein Produkt von Vektoren erklären: es gibt gleich drei mögliche Produkte, von denen eines das *Skalarprodukt* ist:

$$\mathbf{k}_1 \cdot \mathbf{k}_2 = k_{11}k_{12} + k_{21}k_{22} + k_{31}k_{32} \tag{1.3}$$

Ein *Skalar* ist, im Unterschied zu einem Vektor, eine einzelne reelle Zahl. Die Fälle einerseits und die Variablen andererseits lassen sich jeweils durch Vektoren repräsentieren, und es zeigt sich, dass das Skalarprodukt zweier Vektoren proportional zum Kosinus des Winkel ist, den die beiden Vektoren bilden, und darüber

[4] *Treatise on Electricity and Magnetism* (1873)

hinaus ist das Skalarprodukt proportional zur Kovarianz oder Korrelation zwischen den Variablen, falls die beiden Vektoren Variable abbilden, eine geeignete Skalierung der Komponenten vorausgesetzt. Dieser Sachverhalt illustriert die Rede vom Vektor als gerichteter Größe: je weniger sich die Vektoren hinsichtlich ihrer Orientierung voneinander unterscheiden, desto ähnlicher sind sich die von ihnen repräsentierten Größen. Dies gilt natürlich auch für die Fälle: je kleiner der Winkel zwischen zwei Vektoren ist, die jeweils einen Fall repräsentieren, desto ähnlicher sind sich die Fälle. Bei der Herleitung dieses Befundes wird davon Gebrauch gemacht, dass Vektoren durch Pfeile, also Linienelemente definiert sind. Würde man Variablen und Fälle nur durch Punkte in einem Raum abbilden, müßte man diese Linienelemente erst durch Distanzen zwischen jeweils zwei Punkten herstellen, würde man indirekt zu Vektoren übergehen, also kann man gleich auf den Vektorbegriff rekurrieren.

Mit der Bildung von Linearkombinationen, der Veränderung der Längen von Vektoren und dem Skalarprodukt hat man Verknüpfungs-, also Rechenregeln erklärt, die nicht nur Beziehungen zwischen physikalischen Größen abbilden, sondern sich auf alle Anwendungen des Vektorbegriffs erstrecken, bei denen die Verknüpfung der betrachteten Größen den Regeln der Vektorverknüpfung entsprechen. Durch die Definition von Vektoren, Matrizen und den Verknüpfungsregeln entsteht der *Matrizen- oder Matrixkalkül* (Vektoren sind Matrizen mit nur einer Spalte oder Zeile). Die Verfahren der multivariaten Statistik werden fast ausnahmslos mit den mitteln des Matrizenkaküls dargestellt. Die damit oft implizit Annahme der Additivität erweist sich dabei als weniger problematisch, als sie auf den ersten Augenblick erscheinen mag, denn zum einen liefert wie oft gute Approximationen, zum anderen hat man in den letzten Jahren Verfahren für möglicherweise nichtlineare Zusammenhänge erarbeitet, für deren Analyse der Matrixkalkül gleichwohl gut Dienste leistet. Diese Sachverhalte bilden den Hintergrund für die Bedeutung des Vektor- und Matrixkalküls in der multivariaten Statistik in vielen empirischen Wissenschaften.

1.2 Punkträume

Mit dem Symbol $\mathbb{R}$ wird die Menge der reellen Zahlen bezeichnet[5]. Die Zahl $x \in \mathbb{R}$ heißt auch *Skalar*, weil sie auf der "Skala" von $-\infty$ bis $+\infty$ liegt. Eine Ebene wird durch das Cartesische Produkt $\mathbb{R} \times \mathbb{R} = \mathbb{R}^2$ definiert: $\mathbb{R}^2 = \{(x, y)|x, y \in \mathbb{R}\}$, d.h. durch die Menge aller Paare von reellen Zahlen. Das Paar $(x, y) \in \mathbb{R}^2$ kann als

[5]Das ist die Vereinigung (i) der natürlichen Zahlen $\mathbb{N} = 0, 1, 2, \ldots$, (ii) die ganzen Zahlen $\mathbb{Z} = \{\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots\}$, (iii) die rationalen Zahlen $\mathbb{Q} = \{p/q|, p, q \in \mathbb{Z}, q \neq 0\}$, und (iv) die irrationalen Zahlen, die sich nicht als Quotient p/q mit $p, q \in \mathbb{Z}$ darstellen lassen (lat. ratio = Bruch, Quotient). Bekannte irrationale Zahlen sind π, die Eulersche Zahl e (Basis des natürlichen Logarithmus), $\sqrt{2}$, etc. Die Dezimaldarstellung einer irrationalen Zahl ist nicht periodisch und bricht nicht ab. Es läßt sich zeigen, dass zwischen irgendzwei rationalen Zahlen stets beliebig viele (genauer: überabzählbar viele) irrationale Zahlen liegen. $\mathbb{R}$ bildet ein Kontinuum, also eine lückenlos zusammenhängende Menge von Zahlen.

Paar von Koordinaten eines Punktes interpretiert werden. Analog dazu bezeichnet $\mathbb{R}\times\mathbb{R}\times\mathbb{R} = \mathbb{R}^3$ die Menge aller Tripel (x, y, z) mit $x, y, z \in \mathbb{R}$, die als Koordinaten eines Punktes im 3-dimensionalen Raum betrachtet werden können. Analog dazu wird mit

$$\mathbb{R}^n = \{(x_1, x_2, \ldots, x_n) | x_1, x_2, \ldots, x_n \in \mathbb{R}\}, \quad n \in \mathbb{N} \tag{1.4}$$

der n-dimensionale Punktraum bezeichnet. Wiederum in Analogie zu den anschaulichen Räumen mit $n \leq 3$ lassen sich die $x_1, x_2, \ldots, x_n$ als Koordinaten eines "Punktes" auffassen.

Koordinatenachsen lassen sich skalieren, d.h. mit einem Faktor multiplizieren (wenn man etwa von Zentimetern zu Millimetern übergeht) und die Koordinaten von Punkten lassen sich addieren: Sind $x = (x_1, \ldots, x_n)$ und $y = (y_1, \ldots, y_n)$ irgendzwei Punkte, so soll
1. $ax = (ax_1, ax_2, \ldots, ax_n)$, für $a \in \mathbb{R}$, und
2. $x + y = (x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n)$
gelten. $x + y$ definiert also ebenfalls einen Punkt. Es gilt der folgende

Satz 1.1 *Es seien x, y, z Punkte im $\mathbb{R}^n$. Dann gelten die Aussagen*
1. $x + y = y + x$, (Kommutativität)
2. $(x + y) + z = x + (y + z)$, (Assoziativität der Summation)
3. $x + 0 = x$; 0 ist der neutrale Punkt *(der Ursprung des Koordinatensystems)*
4. Für jeden Punkt x existiert ein Punkt $-x$ derart, dass $x + (-x) = 0$,
5. $1x = x$,
6. $(ab)x = a(bx)$, für $a, b \in \mathbb{R}$,
7. $(a + b)x = ax + bx$, für $a, b \in \mathbb{R}$,

Beweis: Es genügt ein einfaches Nachrechnen. □

Es $x = (x_1, \ldots, x_n)$ und $y = (y_1, \ldots, y_n)$ seien irgendzwei Punkte im $\mathbb{R}^n$. Die Punkte sind durch eine Distanz $d(x, y)$ voneinander getrennt. $d(x, y)$ ist ein von den Koordinaten x_i und y_i Maß für die Entfernung zwischen den Punkten x und y. Die Distanz zwischen den Punken hängt von der *Metrik* des Raumes ab:

Definition 1.1 *Die Distanz $d(x, y)$ ist eine Funktion der Differenzen $x_i - y_i$, $i = 1, \ldots, n$ der Koordinaten x_i, y_i der Punkte x und y; genügen die Distanzen den folgenden Axiomen*
1. $d(x, y) \geq 0$,
2. $d(x, y) = d(y, x)$ (Reflexivität)
3. $d(x, z) \leq d(x, y) + d(y, z)$ (Dreiecksungleichung),
so definiert d eine Metrik *des Raumes, in dem die Punkte x und y liegen.*

Die Distanzen zwischen den Punkten eines Raumes können auf verschiedene Weisen definiert sein, und dementpsrechend ist ein Raum durch eine bestimmte Metrik definiert, die die geometrische Struktur des Raumes kennzeichnet. So ist nach

Euklid ist die kürzeste Verbindung zwischen zwei Punkten eine Gerade. Für irgendzwei Punkte x und y gilt dann

$$d(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} = \left[\sum_{i=1}^{n}(x_i - y_i)^2\right]^{\frac{1}{2}}. \tag{1.5}$$

Hier wird implizit angenommen, dass die x_i und y_i Koordinaten in einem rechtwinkligen Koordinatensystem sind, und $x_i - y_i$ ist die Differenz der Koordinaten der Punkte x und y auf der i-ten Koordinate. Die Gleichung (1.5) ist dann der Satz des Pythagoras für den n-dimensionalen Raum. Ist die Metrik eines Raums durch (1.5) definiert, so heißt der Raum *euklidisch*. Die euklidische Metrik ist nicht die einzig mögliche Metrik, wie weiter unten illustriert wird.

Die Axiome in Definition 1.1 definieren die allgemeinen Eigenschaften einer Metrik. Die Forderung, dass eine Distanz $d(x,y)$ nicht negativ sein darf, entspricht dem umgangssprachlichen Begriff von "Distanz": die Länge eines Weges von Ort A zu Ort B kann nicht negativ sein. Die Forderung der Reflexivität $d(x,y) = d(y,x)$ stellt eine Einschränkung dar: will man in einer Stadt mit dem Auto von der Adresse A_1 zu Adresse A_2 fahren, so kann wegen eines Einbahnstrassensystems die Distanz $d(A_1, A_2)$ größer als die Distanz $d(A_2, A_1)$ sein. Die Dreiecksungleichung ist wiederum intuitiv einleuchtend. Sind A_1, A_2 und A_3 drei Adressen in einer Stadt, so ist es möglich, dass es zwischen A_1 und A_3 keinen direkten Weg gibt und man immer über die Adresse A_2 fahren muß; dann gilt eben $d(A_1, A_3) = d(A_1, A_2) + d(A_2, A_3)$. Es wird aber nie $d(A_1, A_3) > d(A_1, A_2) + d(A_2, A_3)$ gelten, d.h. es muß $d(A_1, A_3) \leq d(A_1, A_2) + d(A_2, A_3)$ gelten.

Euklidische Räume wurden lange Zeit als "natürliche" Räume betrachtet. Isaac Newton (1642 – 1727) nahm implizit eine euklidische Struktur des physikalischen Raumes an, und der Philosoph Immanuel Kant (1724 – 1804) erklärte, der euklidische Raum sei eine notwendige Vorstellung a priori. So notwendig wie von Kant angenommen ist diese Vorstellung allerdings nicht, schon in der ersten Hälfte des 19-ten Jahrhunderts schlugen der ungarische Mathematiker János Bolyai (1802 – 1860), der russische Mathematiker Nikolai Iwanowitsch Lobatschewski (1792 – 1856) sowie der Mathematiker Carl Friederich Gauß (1777 - 1855) nicht-euklidische Geometrien vor. Der Mathematiker Bernhard Riemann (1826 – 1866) hielt 1850 seinen Habilitationsvortrag über eine nicht-euklidische Geometrie, die später für die Weiterentwicklung der Relativitätstheorie wichtig wurde. Der Mathematiker Hermann Minkowski (1864 – 1909) modifizierte ebenfalls im Zusammenhang mit der Relativitätstheorie die Begriffe von Raum und Zeit (Raum-Zeit-Kontiuum) und definierte eine Metrik – die Minkowski-Metrik –, die durch

$$d(x,y) = \left[\sum_{i=1}^{n}|x_i - y_i|^p\right]^{\frac{1}{p}}, \quad 0 < p \in \mathbb{R} \tag{1.6}$$

definiert ist. Für $p = 2$ ergibt sich die euklidische Metrik (1.5), und die Minkowski-Metrik kann als Verallgemeinerung der euklidischen Metrik angesehen werden.

Für $p = 1$ ergibt sich die *City-Block-Metrik* oder *Manhattan*-Metrik, weil man von A_1 nach A_2 gelangt, indem man rechtwinklig zueinander liegende Strassenabschnitte durchfahren oder durchlaufen muß. Die Minkowski-Metrik erlaubt es, *psychologische Distanzen*, etwa zwischen Begriffen, Stereotypen etc, zu modellieren, für die sich die euklidische Metrik oft als inadäquat erweist. Die Minkowski-Metrik wird im Zusammenhang mit der multidimensionalen Skalierung behandelt.

In vielen Untersuchungen werden an jeweils einer Person (allgemein: einem "Fall", und ein Fall muß keine Person sein) mehrere Merkmale gemessen und man berechnet die Korrelationen zwischen den Merkmalen. Man kann die Messungen als Koordinaten eines Punktes interpretieren, der dann den Fall repräsentiert. Auf diese Weise entsteht eine Punktekonfiguration ("Punktwolke"). Die Distanzen zwischen den Punkten reflektieren die Relationen zwischen den Fällen. Mit dem Distanzbegriff ist aber der Begriff der Orientierung nicht verbunden, der wiederum für Fragen der Interpretation von Interesse ist. Deswegen wird der Begriff des Vektors eingeführt. Sind P und Q zwei Punkte der Konfiguration, so sei die euklidische Distanz durch $d(P, Q)$ gegeben. Zu dieser Distanz korrespondiert ein Vektor $\overrightarrow{PQ}$, der durch eine Länge und eine Orientierung definiert ist; die Länge ist durch die euklidische Distanz $d(P, Q)$ gegeben, und die Orientierung der durch $d(P, Q)$ definierten Geraden ist durch die Winkel zwischen diewer Geraden und den Koordinatenachsen gegeben. Für maximal drei Dimensionen (3 Variablen) kann der Vektor graphisch durch einen Pfeil repräsentiert werden (s. Abbildung 1). Die Orientierung hängt von gewählten Koordinatensystem ab. Insbesondere lassen sich Abhängigkeiten zwischen Variablen leicht mittels des Vektorbegriffs darstellen und latente, also nicht direkt gemessene Variablen, zur Erklärung von Korrelationen zwischen Variablen bestimmen. Analog zum Punktraum kann dann der Begriff des Vektorraums eingeführt werden, der isomorph zum jeweiligen Punktraum ist. Je nach Perspektive macht man in der multivariaten Analyse sowohl vom Begriff des Punkt- wie des Vektorraums Gebrauch. In den folgenden Abschnitten wird der Begriff des Vektors und der des Vektorraums eingeführt und einige Resultate aus der Vektor- und Matrixrechung vorgestellt, soweit sie für die üblichen multivariaten Verfahren notwendig sind. Neuere Verfahren wie zum Beispiel die Klassifikation von Mustern oder Objekten anhand von "Support Vector Machines" erfordern mathematische Grundlagen, die in einem gesonderten Skript vorgestellt werden.

1.3 Vektoren und ihre Verknüpfungen

Wie in Abschnitt 1.1 angedeutet wurde werden nur endliche Vektorräume über $\mathbb{R}$ (die Menge der reellen Zahlen) betrachtet.

Definition 1.2 *Ein n-dimensionaler Vektor ist ein n-Tupel reeller Zahlen:*

$$\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \tag{1.7}$$

wobei es auf die Reihenfolge der Komponenten ankommt. Die $x_1, \ldots, x_n$ *heißen* Komponenten *des Vektors. Es wird auch* $\boldsymbol{x} \in \mathbb{R}^n$ *geschrieben.*

Dementsprechend sind

$$\mathbf{x}_1 = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \quad \mathbf{x}_2 = \begin{pmatrix} 2 \\ 1 \\ 3 \end{pmatrix}$$

verschiedene 3-dimensionale Vektoren.

Anmerkungen:
1. Vektoren werden im Folgenden wie in (1.7) durch fette Buchstaben bezeichnet; diese Bezeichung ist mittlerweile allgemein üblich. Eine alternative Schreibweise ist $\vec{x}$; der Pfeil deutet an, dass mit x ein Vektor gemeint ist. Vor dieser Schreibweise wird ebenfalls Gebrauch gemacht. Die Fett-Schreibweise erleichtert Schreibweisen wie $\hat{\mathbf{x}}$, die Schätzungen eines Vektors $\mathbf{x}$ bezeichnen, oder $\tilde{\mathbf{x}}$, die Vektoren bezeichnen, die durch die die Zeilen einer Matrix definiert sind. Die Schreibweise $\vec{x}$ wird dann leicht unübersichtlich: $\hat{\vec{x}}$ oder $\vec{\hat{x}}$ und $\tilde{\vec{x}}$ bzw. $\vec{\tilde{x}}$ zeichnen sich durch sehr geringen graphischen Charme aus, insbesondere wenn transponierte Vektoren $\vec{\hat{x}}'$ (s. unten) betrachtet werden.
2. Ein Vektor $\mathbf{x}$ wird wie in Gleichung (1.7) stets als Spalte oder Kolumne angeschrieben, es sei denn, er wird ausdrücklich als *transponiert* oder *gestürzt* gekennzeichnet; in diesem Fall wird er als Zeile angeschrieben:

$$\mathbf{x}' = (x_1, x_2, \ldots, x_n). \tag{1.8}$$

Andere Schreibweisen für transponierte Vektoren sind $\mathbf{x}^t$, $\mathbf{x}^T$ oder $\mathbf{x}^\top$. Eine platzsparende Schreibweise ist

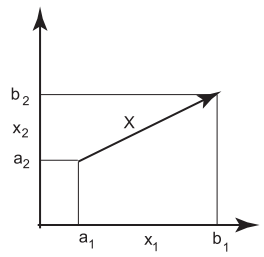
$$\mathbf{x} = (x_1, x_2, \ldots, x_n)'.$$

□

Vektoren werden graphisch oft durch Pfeile bestimmter Länge und bestimmter Orientierung repräsentiert; dem entspricht die Redeweise von Vektoren als 'gerichteten Größen'; in der Physik werden z.B. Kräfte durch derartige Vektoren (Pfeile) dargestellt: die Länge des Vektors entspricht der Größe der Kraft, die Orientierung der Richtung, in der die Kraft wirkt. Der Anfangspunkt des Pfeils habe die Koordinaten $a_1, \ldots, a_n$, der Endpunkt habe die Koordinaten $b_1, \ldots, b_n$. Dann sind die Komponenten durch

$$x_i = b_i - a_i, \quad i = 1, \ldots, n \tag{1.9}$$

Abbildung 1: Ein 2-dimensionaler -Vektor und seine Komponenten



gegeben, s. a. Abb. 1. Die x_i definieren damit nur die Länge (s. unten) und die Orientierung, nicht aber den genauen Ort des repräsentierenden Pfeils.

Da ein Vektor durch ein Zahlentupel $(x_1, \ldots, x_n)$ definiert ist, korrespondiert ein Vektor offenbar zu einem Punkt im n-dimensionalen Punktraum. Gleichwohl sind verschiedene Vorstellungen mit dem Punktebegriff einerseits und dem Vektorbegriff andererseits verbunden: da die Komponenten als Differenzen zwischen den Koordinaten des Anfangs- und des Endpunktes des Vektors definiert sind, fallen der Punkt mit den Koordinaten $(x_1, \ldots, x_n)$ und der Vektor genau dann zusammen, wenn der Anfangspunkt in den Nullpunkt des Koordinatensystems gelegt werden kann.

Spezielle Vektoren: Es seien

$$\vec{0} = (0, 0, \ldots, 0)' \tag{1.10}$$
$$\vec{1} = (1, 1, \ldots, 1)' \tag{1.11}$$
$$\mathbf{e}_j = (0, \ldots, 0, 1, 0, \ldots, 0)', \;\; j = 1, \ldots, n \tag{1.12}$$

$\vec{0}$ heißt der *Nullvektor*, $\vec{1}$ heißt *Einsvektor*; gelegentlich wird auch $\vec{1}_n$ bzw. $\vec{0}_n$ geschrieben, um anzudeuten, dass der Vektor n Komponenten hat. $\mathbf{e}_j$ ist der j-te *kanonische Einheitsvektor*, seine Komponenten sind alle gleich Null bis auf die j-te Komponente, die gleich 1 ist (s.auch die Definition des Einheitsvektors auf Seite 18).

Anmerkung: Der Nullvektor kann nicht durch einen Pfeil dargestellt werden, denn er hat weder eine Länge (dies wird in Gleichung (1.24), Seite 18, – Norm eines Vektors – noch explizit gemacht werden, noch eine Orientierung; dieser Sachverhalt wird in der Anmerkung 1.3, Seite 20 elaboriert. □

Multiplikation mit einem Skalar: Es sei $\mathbf{x}$ ein Vektor und $\lambda \in \mathbb{R}$ ein Skalar, also eine einzelne reelle Zahl. Dann bedeutet die Schreibweise $\lambda\mathbf{x}$, dass jede

Komponente von $\mathbf{x}$ mit λ multipliziert wird:

$$\lambda\mathbf{x} = \lambda \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} \lambda x_1 \\ \lambda x_2 \\ \vdots \\ \lambda x_n \end{pmatrix}, \quad \lambda \in \mathbb{R} \tag{1.13}$$

λ heißt *Skalar*, weil λ ein Wert der "Skala", d.h. der reellen Zahlen zwischen $-\infty$ und ∞ ist. Man kann λ als einkomponentigen Vektor auffassen.

Addition von Vektoren Es seien $\mathbf{x}$ und $\mathbf{y}$ zwei n-dimensionale Vektoren. Dann heißt

$$\mathbf{z} = \mathbf{x} + \mathbf{y} = \begin{pmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{pmatrix} = \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{pmatrix} \tag{1.14}$$

die Summe der Vektoren $\mathbf{x}$ und $\mathbf{y}$. Vektoren können also nur addiert werden, wenn sie dieselbe Anzahl von Komponenten haben. Abb. 2 zeigt ein sicherlich stark vereinfachtes Beispiel für die Summe zweier Vektoren, die bestimmte Fähigkeiten repräsentieren: Sprachliches Vermögen plus Phantasie ergeben poetisches Vermögen. Interpretiert man die Länge eines Vektors als Maß für die Ausprägung des durch den Vektor repräsentierten Merkmals, so bedeutet die vektorielle Addition offenbar, dass die Ausprägung des durch die Summe repräsentierten Merkmals (poetisches Vermögen) nicht gleich der Summe der Ausprägungen der Merkmale ist, die vektoriell addiert werden. Die Summe (2.5) kann verallgemeinert werden: es ist ja möglich, dass das sprachliche Vermögen und die Phantasie in gewichteter Form in das poetische Vermögen eingehen, so dass

$$\mathbf{y} = a_1\mathbf{x}_1 + a_2\mathbf{x}_2$$

gilt, wobei $\mathbf{y}$ das poetische Vermögen repräsentiert, $\mathbf{x}_1$ das sprachliche Vrmögen und $\mathbf{x}_2$ die Phantasie, und a_1 und a_2 sind die Gewichte. Natürlich ist denkbar, dass das sprachliche Vermögen und die Phantasie in interaktiver, nichtlinearer Weise das poetische vermögen bestimmen. Ist y_i die Ausprägung des poetischen Vermögens der i-ten Person in einer Stichprobe und sind x_{i1} und x_{i2} die Ausprägungen des sprachlichen Vermögens bzw. der Phantasie der i-ten Person, so könnte das poetische Vermögen dieser Person durch

$$y_i = a_1x_{i1} + a_2x_{i2} + a_3x_{i1}x_{i2}$$

gegeben sein. Der Term $x_{i1}x_{i2}$ besagt, dass die Phantasie proportional zum sprachlichen Vermögen in das poetische Vermögen eingeht, und zwar mit dem Gewicht a_3. Ein anderes Modell bestünde darin, dass das sprachliche Vermögen nicht nur linear in das poetische Vermögen eingeht, sondern auch noch proportional zu x_{i1}^2 ist, so dass man

$$y_i = a_1x_{i1} + a_2x_{i2} + a_3x_{i1}x_{i2} + a_4x_{i1}^2$$

ansetzen kann. Man kann nun die "Prädiktoren" $\mathbf{x}_1$ und $\mathbf{x}_2$ umbenennen, so dass man einen Ausdruck der Form

$$\mathbf{y} = a_1\mathbf{u}_1 + a_2\mathbf{u}_2 + a_3\mathbf{u}_3 + a_4\mathbf{u}_4$$

erhält, wobei nun

$$u_{i1} = x_{i1},\ u_{i2} = x_{i2},\ u_{i3} = x_{i1}x_{i2},\ u_{i4} = x_{i1}^2$$

gilt. Man sieht, dass auf diese Weise eine ganze Menge möglicher Modelle formuliert werden können. Natürlich sind die Prädiktoren $\mathbf{u}_3$ und $\mathbf{u}_4$ unabhängig von $\mathbf{u}_1$ und $\mathbf{u}_2$. Obwohl diese Modelle nichtlineare Funktionen von bestimmten Variablen enthalten, sind sie immer noch *linear in den Koeffizienten* a_j, $j = 1, 2, 3, 4$. Modelle dieser Art werden im Folgenden nicht weiter diskutiert, sie werden im Zusammenhang mit den statistischen Verfahren wie multiple Regression besprochen werden[6].

Definition 1.3 *Die Summe*

$$\boldsymbol{x} = a_1\boldsymbol{x}_1 + a_2\boldsymbol{x}_2 + \cdots + a_p\boldsymbol{x}_p, \tag{1.15}$$

der p n-dimensionalen Vektoren, $a_j \in \mathbb{R}$ Skalare, $1 \leq j \leq p$ heißt Linearkombination *der $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p$.*

Der Begriff der Linearkombination charaktisiert einen Vektor als gewogene Summe anderer Vektoren, was nicht ausschließt, dass die Komponenten einiger Vektoren sich zB multiplikativ aus den Komponenten anderer Vektoren definiert sind, wie die obigen Beispiele zeigen. Wie oben ebenfalls schon angedeutet wurde, Vektoren $\mathbf{x}_1, \ldots, \mathbf{x}_p$ nicht mehr voneinander unabhängige IInformationen repräsentieren. Die Abhängigkeit von Vektoren wird seiter unten noch ausführlich diskutiert.

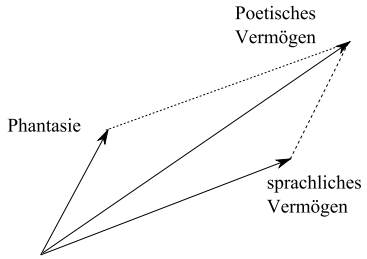
Beispiel 1.1 Multiple Regression Es wird angenommen, dass der Wert einer Variablen Y durch drei Prädiktorvariablen X_1, X_2 und X_3 bis auf einen zufälligen Fehler "vorhergesagt" werden kann:

$$Y_i = a_1X_{i1} + a_2X_{i2} + a_3X_{i3} + e_i, \quad i = 1, \ldots, m \tag{1.16}$$

Für die i-te Person sind also die Messungen Y_i, X_{i1}, X_{i2} und X_{i3} gegeben und die Y_i sollen anhand der X_{ij}, $j = 1, \ldots, 3$ vorhergesagt werden; e_i ist ein Fehlerterm; er repräsentiert alle Effekte in Y_i, die nicht durch die drei Prädiktorvariablen definiert werden. Die Koeffizienten a_1, a_2 und a_3 werden im Allgemeinen mit

[6] Auch das (Vektor-)Parallelogramm der Kräfte in der Physik ist keineswegs trivial: vergl. Lange (2009) *A Tale of Two Vectors*; dieser Artikel ist auf der Web-Seite http://www.uwe-mortensen.de/SkriptenStatistik.html verfügbar.

Abbildung 2: Addition von Vektoren, analog zur Addition von Kräften in der Physik



der Methode der Kleinsten Quadrate so geschätzt, dass die Summe $\sum_i e_i^2$ der Fehlerquadrate minimal wird. Ausgeschrieben erhält man

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_m \end{pmatrix} = a_1 \begin{pmatrix} X_{11} \\ X_{21} \\ \vdots \\ x_{m1} \end{pmatrix} + a_2 \begin{pmatrix} X_{12} \\ X_{22} \\ \vdots \\ x_{m2} \end{pmatrix} + a_3 \begin{pmatrix} X_{13} \\ X_{23} \\ \vdots \\ x_{m3} \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_m \end{pmatrix} \tag{1.17}$$

Bei dieser Schreibweise ist von dem Sachverhalt Gebrauch gemacht worden, dass die Koeffizienten a_j für alle X_{ij} identisch sind. Die X_{ij} können als Komponenten eines Vektors $\mathbf{x}_j$ aufgefasst werden, ebenso die e_i, so dass man abkürzend

$$\mathbf{y} = a_1\mathbf{x}_1 + a_2\mathbf{x}_2 + a_3\mathbf{x}_3 + \mathbf{e} \tag{1.18}$$

schreiben kann. $\mathbf{y}$ erscheint hier als Linearkombination der $\mathbf{x}_j$ und $\mathbf{e}$. Die Vektorschreibweise erscheint hier zunächst als vereinfachte Schreibweise; die "operativen Implikationen" dieser Schreibweise zeigen sich im Folgenden. Da die Komponenten von $\mathbf{x}_1$, $\mathbf{x}_2$ und $\mathbf{y}$ Messwerte sind, die üblicherweise mit zufälligen Fehlern behaftet sind, heißen diese Vektoren auch *zufällige Vektoren*. Natürlich ist $\mathbf{e}$ ebenfalls ein Zufallsvektor. Zufällige Vektoren werden in Abschnitt 2.3 ausführlicher behandelt.

Man bemerke, dass $\mathbf{y}$ als Addition von Vektoren analog zum in Abbildung 2 gezeigten Beispiel definiert ist; es handelt sich nicht um eine unziemliche Übertragung der Physik auf die Psychologie. Vielmehr sind die Addition von Kräften einerseits und die multiple Regression andererseits strukturgleiche Modelle. Dem *Modell* zufolge ist das "poetische Vermögen" eine additive Mischung von "verbalem Vermögen" und "Phantasie", die durch einen um einen Fehlervektor $\mathbf{e}$ erweiterten Regressionsansatz

$$\mathbf{y} = a_1\mathbf{x}_1 + a_2\mathbf{x}_2 + \mathbf{e}$$

repräsentiert wird. $\mathbf{e}$ repräsentiert zufällige Effekte und alle Aspekte von poetischem Vermögen, die nicht zu "verbalem Vermögen" oder "Phantasie" gehören.

Die Komponenten x_{ij} beziehen sich auf ein Koordinatensystem, dass in Abb. 2 nicht eingezeichnet wurde; die Wahl eines speziellen Koordinatensystems ist für die Vektoraddition nicht wesentlich. Ob dieser Ansatz vernünftig ist, ist eine andere Frage; man kann vermuten, dass der Fehleranteil in $\mathbf{y}$ verhältnismäßig groß relativ zu den Anteilen von $\mathbf{x}_1$ und $\mathbf{x}_2$ sein wird. □

Vektoren können auch miteinander "multipliziert" werden. Es gibt verschiedene Definitionen von Vektorprodukten. Hier werden nur die eingeführt, die in der Multivariaten Statistik verwendet werden.

Definition 1.4 *Es seien $\boldsymbol{x}$ und $\boldsymbol{y}$ n-dimensionale Vektoren. Dann heißt*

$$\boldsymbol{x}'\boldsymbol{y} = (x_1, x_2, \ldots, x_n) \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = x_1y_1 + \cdots + x_ny_n = \sum_{i=1}^{n} x_iy_i \tag{1.19}$$

das Skalarprodukt *oder* inneres Produkt *von $\boldsymbol{x}$ und $\boldsymbol{y}$. Nun sei $\boldsymbol{y}$ m-dimensional. Dann heißt*

$$\boldsymbol{x}\boldsymbol{y}' = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} (y_1, \ldots, y_m) = \begin{pmatrix} x_1y_1 & x_1y_2 & \cdots & x_1y_m \\ x_2y_1 & x_2y_2 & \cdots & x_2y_m \\ \vdots & \vdots & \ddots & \vdots \\ x_ny_1 & x_ny_2 & \cdots & x_ny_m \end{pmatrix} \tag{1.20}$$

dyadisches Produkt *oder* tensorielles Produkt *von $\boldsymbol{x}$ und $\boldsymbol{y}$; natürlich ist $m = n$ möglich.*

Anmerkung: Ein Vektor $\mathbf{u}$, dessen Komponenten durch $u_i = x_iy_i$ definiert sind, wobei x_i und y_i Komponenten zweier Vektoren $\mathbf{x}$ und $\mathbf{y}$ sind, bildet also *kein* Produkt der Vektoren $\mathbf{x}$ und $\mathbf{y}$. Vektoren wie $\mathbf{u}$ werden nur in speziellen Modellierungen von Zusammenhängen von Variablen betrachtet, sie sind kein Bestandteil der allgemeinen Vektoralgebra. □

Beispiel 1.2 Es seien $\mathbf{x}_1, \ldots, \mathbf{x}_n$ m-dimensionale Vektoren, und es sei $\mathbf{y} = a_1\mathbf{x}_1 + \cdots + a_n\mathbf{x}_n$ eine Linearkombination dieser Vektoren. Dann sind die Komponenten y_i von $\mathbf{y}$ durch das jeweilige Skalarprodukt

$$y_i = \mathbf{a}'\tilde{\mathbf{x}}_i, \quad i = 1, 2, \ldots, m \tag{1.21}$$

gegeben, wobei $\mathbf{a} = (a_1, a_2, \ldots, a_n)'$ und $\tilde{\mathbf{x}}_i = (x_{i1}, x_{i2}, \ldots, x_{in})'$ ist; die $x_{i1}, x_{i2}, \ldots, x_{in}$

sind die i-ten Komponenten der $\mathbf{x}_1, \ldots, \mathbf{x}_n$. Denn es ist ja

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix} = \sum_{j=1}^{n} a_j \mathbf{x}_j = \begin{pmatrix} a_1 x_{11} + a_2 x_{12} + \cdots + a_n x_{1n} \\ a_1 x_{21} + a_2 x_{22} + \cdots + a_n x_{2n} \\ \vdots \\ a_1 x_{m1} + a_2 x_{m2} + \cdots + a_m x_{mn} \end{pmatrix}$$
$$= \begin{pmatrix} \mathbf{a}'\tilde{\mathbf{x}}_1 \\ \mathbf{a}'\tilde{\mathbf{x}}_2 \\ \vdots \\ \mathbf{a}_n\tilde{\mathbf{x}}_m \end{pmatrix}. \tag{1.22}$$

Von diesem Zusammenhang zwischen Linearkombination und Skalarprodukt wird bei der Definition der Matrixmultiplikation Gebrauch gemacht. □

Im Rahmen der multivariaten Analyse treten sowohl das Skalarprodukt wie auch das dyadische Produkt im Zusammenhang mit der Betrachtung von Kovarianzen zwischen Variablen auf. Unter bestimmten Bedingungen entspricht das Skalarprodukt der Kovarianz zwischen zwei Variablen. Der Begriff des dyadischen Produkts erlaubt eine kompakte Schreibweise insbesondere für Kovarianz- oder Korrelationsmatrizen (Kovarianz- oder Korrelationsmatrizen), die als Summen von dyadischen Vektorprodukten dargestellt werden können; in Abschnitt 2.3 wird auf diesen Sachverhalt explizit eingegangen.

Anmerkungen: Man beachte, dass jeweils einer der Vektoren, deren Produkt bestimmt werden soll, als Zeilenvektor angeschrieben wurde. Diese Schreibweise ist weniger willkürlich, als es auf den ersten Blick scheinen mag: sie ergibt sich aus dem in Beispiel 1.2 illustrierten Zusammenhang zwischen Skalarprodukt und Linearkombination. Insbesondere für das Skalarprodukt gibt es andere Schreibweisen, die in bestimmten Zusammenhängen übersichtlicher sein können, etwa $\langle \mathbf{x}, \mathbf{y} \rangle$ oder $\mathbf{x} \cdot \mathbf{y}$. Der Name 'Skalarprodukt' erklärt sich aus dem Sachverhalt, dass $\mathbf{x}'\mathbf{y} \in \mathbb{R}$ ein *Skalar* ist. Skalare sind einzelne reelle Zahlen, der Ausdruck leitet sich aus der Bezeichnung 'Skala' für die Menge der reellen Zahlen zwischen $-\infty$ und $+\infty$ ab.

Während also das Skalarprodukt zweier Vektoren $\mathbf{x}$ und $\mathbf{y}$ eine einzelne reelle Zahl ist und voraussetzt, dass $\mathbf{x}$ und $\mathbf{y}$ dieselbe Anzahl von Komponenten haben, ist das dyadische Produkt eine Matrix und $\mathbf{x}$ und $\mathbf{y}$ müssen nicht dieselbe Anzahl von Komponenten haben.

Norm eines Vektors Für $\mathbf{y} = \mathbf{x}$ ergibt sich als Skalarprodukt

$$\mathbf{x}'\mathbf{x} = \sum_{i=1}^{n} x_i^2. \tag{1.23}$$

Für $\mathbf{x}'\mathbf{x}$ ist auch die Schreibweise $\|\mathbf{x}\|^2$ gebräuchlich. $\|\mathbf{x}\|^2$ ist das Quadrat der Länge von $\mathbf{x}$ (Satz des Pythagoras), d.h.

$$\sqrt{\mathbf{x}'\mathbf{x}} = \|\mathbf{x}\| \tag{1.24}$$

ist die Länge von $\mathbf{x}$. Für $\lambda \in \mathbb{R}$ folgt sofort $\|\lambda\mathbf{x}\| = |\lambda|\|\mathbf{x}\|$, d.h. die Multiplikation eines Vektors mit einem Skalar bedeutet die Skalierung der Länge des Vektors; für $\lambda < 1$ erhält man eine Stauchung, d.h. eine Verkürzung des Vektors, für $\lambda > 1$ eine Dehnung oder Verlängerung. Es werde λ so gewählt, dass mit $\mathbf{y} = \lambda\mathbf{x}$ die Bedingung

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

erfüllt ist. Dann folgt

$$|\lambda| = \frac{1}{\|\mathbf{x}\|} > 0 \tag{1.25}$$

Der Vektor

$$\mathbf{y} = \left(\frac{1}{\|\mathbf{x}\|}x_1, \frac{1}{\|\mathbf{x}\|}x_2, \ldots, \frac{1}{\|\mathbf{x}\|}x_n\right)' \tag{1.26}$$

heißt dann *normiert*, d.h. er hat die Länge 1. Die Länge $\|\mathbf{x}\|$ wird auch die *Norm* des Vektors $\mathbf{x}$ genannt. Der Begriff der Norm wird in Abschnitt 1.5 weiter spezifiziert.

Einheitsvektoren Vektoren der Länge 1 ($\|\mathbf{x}\| = 1$) heißen auch *Einheitsvektoren*. Ein Spezialfall sind die kanonischen Einheitsvektoren $\mathbf{e}_j$, deren Konponenten alle gleich Null sind mit Ausnahme der j-ten Komponente, die gleich 1 ist. Einheitsvektoren speziell aus $\mathbb{R}^2$ lassen sich in der Form $\mathbf{x} = (\sin\theta, \cos\theta)'$ für geeignet gewählten Winkel θ anschreiben, denn $\sin^2\theta + \cos^2\theta = 1$ für alle θ.

Zentrierte Vektoren Es sei $X = (X_1, X_2, \ldots, X_n)'$ und

$$\bar{x} = \frac{1}{n}X'\vec{1} = \frac{1}{n}\vec{1}'X = \frac{1}{n}\sum_{i=1}^{n} X_i$$

sei das arithmetische Mittel der Komponenten X_i von $\mathbf{X}$. Dann heißt

$$\mathbf{x} = X - \bar{x}\vec{1} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix} - \begin{pmatrix} \bar{x} \\ \bar{x} \\ \vdots \\ \bar{x} \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \tag{1.27}$$

zentrierter Vektor, mit $x_i = X_i - \bar{x}$. Offenbar ist

$$\sum_{i=1}^{n}(X_i - \bar{x}) = \sum_{i=1}^{n} x_i = \mathbf{x}'\vec{1} = 0,$$

und wenn $\mathbf{y}$ ein ebenfalls zentrierter n-dimensionaler Vektor ist, so ist

$$Kov = \frac{1}{n}\mathbf{x}'\mathbf{y} = \frac{1}{n}\sum_{i=1}^{n} x_i y_i = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{x})(Y_i - \bar{y}) \tag{1.28}$$

die Kovarianz der Messwerte X_i, Y_i, und für $\mathbf{x} = \mathbf{y}$ erhält man

$$s_x^2 = \frac{1}{n}\mathbf{x}'\mathbf{x} = \frac{1}{n}\|\mathbf{x}\|^2 \tag{1.29}$$

für die Varianz; s_y^2 ist analog definiert. Dann ist

$$r_{xy} = \frac{\mathbf{x}'\mathbf{y}}{\|\mathbf{x}\|\|\mathbf{y}\|} \tag{1.30}$$

die Produkt-Moment-Korrelation der Messwerte X_i und Y_i.

Der Kosinussatz: Es gilt die Beziehung

$$\|\mathbf{x} - \mathbf{y}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 - 2\|\mathbf{x}\|\|\mathbf{y}\|\cos\theta \tag{1.31}$$

(vergl. Abbildung 3 (b))[7] . Für $\theta = \pi/2$, also für einen Winkel von 90^0, folgt $\cos\theta = 0$ und es ergibt sich der Satz des Pythagoras in Vektorschreibweise. Hieraus folgt eine Beziehung zwischen dem Skalarprodukt $\mathbf{x}'\mathbf{y}$ und dem Kosinus des Winkels θ: es ist

$$\begin{aligned}\|\mathbf{x} - \mathbf{y}\|^2 = \sum_i (x_i - y_i)^2 &= \sum_i x_i^2 + \sum_i y_i^2 - 2\sum_i x_i y_i \\ &= \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 - 2\mathbf{x}'\mathbf{y}.\end{aligned}$$

Setzt man den Ausdruck auf der rechten Seite für $\|\mathbf{x} - \mathbf{y}\|^2$ in (1.31) ein, so wird man auf die Beziehung

$$\mathbf{x}'\mathbf{y} = \|\mathbf{x}\|\|\mathbf{y}\|\cos\theta, \tag{1.32}$$

oder

$$\cos\theta = \frac{\mathbf{x}'\mathbf{y}}{\|\mathbf{x}\|\|\mathbf{y}\|} \tag{1.33}$$

geführt. Mit Bezug auf (1.30) folgt daraus für zentrierte Vektoren

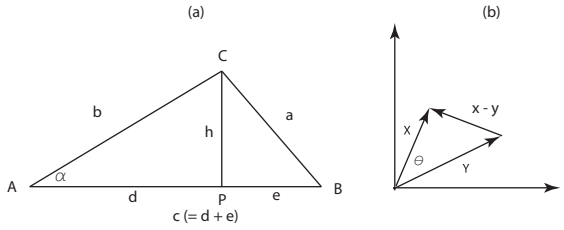
$$\cos\theta = r_{xy}. \tag{1.34}$$

[7] **Beweis:** h ist das von Punkt C auf die Verbindungslinie $c = \overline{AB}$ gefällte Lot (P). Es ist $d = \overline{AP}$, $e = \overline{PB}$. Nach dem Satz des Pythagoras ist $a^2 = h^2 + e^2$ und $b^2 = h^2 + d^2$, d.h. $h^2 = b^2 - d^2$, und nach Abb. 3 ist $e^2 = (c-d)^2$, so dass

$$a^2 = h^2 + e^2 = b^2 - d^2 + (c-d)^2 = b^2 + c^2 - 2cd$$

folgt. Weiter gilt $\cos\alpha = d/b$, dh $d = b\cos\alpha$. Damit erhält man $a^2 = b^2 + c^2 - 2bc\cos\alpha$. □

Abbildung 3: Zum Kosinussatz $a^2 = b^2 + c^2 - 2bc\cos\alpha$



Dieser Ausdruck für $\cos\theta$ wird auch *Kosinus-Ähnlichkeit* (*cosine similarity*) der Vektoren $\mathbf{x}$ und $\mathbf{y}$ genannt, vor allem in Programmen zur künstlichen Intelligenz, s. a. die Anmerkungen zur Ähnlichkeit in Punkt 1, Seite 21; die Kosinus-Ähnlichkeit ist also einfach das Skalarprodukt der normierten Vektoren $\mathbf{x}/\|\mathbf{x}\|$ und $\mathbf{y}/\|\mathbf{y}\|$.

Anmerkung: Für 2- oder 3-dimensionale Vektoren hat der Begriff des Winkels zwischen zwei Vektoren eine anschauliche Bedeutung, für mehr als drei Dimensionen ist diese Bedeutung nicht klar. Deshalb kann (1.34) als als *Definition* des Kosinus des Winkels und damit des Winkels selbst zwischen zwei Vektoren aufgefasst werden. Damit kann auch der Begriff der Orientierung eines Vektors schärfer gefasst werden: ist $\mathbf{x}$ ein n-dimensionaler Vektor und definiert $\mathbf{y}$ eine Koordinatenachse, so gibt (1.34) die Orientierung des Vektors $\mathbf{x}$ in Bezug auf diese Achse. Ist $\mathbf{x} = \vec{0}$ der Nullvektor, so ist $\mathbf{x}'\mathbf{y} = 0$ *für alle* θ, so dass nach (1.33) $\cos\theta = 0/0$; dieser Ausdruck ist bekanntlich nicht definiert, und damit ist eine Orientierung für $\vec{0}$ nicht definiert. □

Die Gleichung (1.33) impliziert die *Cauchy-Schwarzsche Ungleichung*:

Satz 1.2 *Es seien* $\boldsymbol{x}$ *und* $\boldsymbol{y}$ *zwei* n*-dimensionale Vektoren; dann gilt*

$$|\boldsymbol{x}'\boldsymbol{y}|^2 \leq \|\boldsymbol{x}\|^2\|\boldsymbol{y}\|^2. \tag{1.35}$$

Beweis: Nach (1.32) gilt

$$\mathbf{x}'\mathbf{y} = \|\mathbf{x}\|\|\mathbf{y}\|\cos\theta, \tag{1.36}$$

so dass auch

$$|\mathbf{x}'\mathbf{y}|^2 = \|\mathbf{x}\|^2\|\mathbf{y}\|^2\cos^2\theta$$

gilt. Wegen $-1 \leq \cos\theta \leq 1$ gilt $\cos^2\theta \leq 1$. Dann folgt (1.35). □

Korollar 1.1 *Es gilt*

$$\boldsymbol{x}'\boldsymbol{y} = \|\boldsymbol{y}\|\|\boldsymbol{x}\| \tag{1.37}$$

genau dann, wenn $\boldsymbol{y} = \lambda\boldsymbol{x}$*,* $\lambda \in \mathbb{R}$*, d.h. wenn* $y_i = \lambda x_i$*,* $i = 1, \ldots, n$*.*

Beweis: Offenbar ist $\mathbf{x}'\mathbf{y} = \|\mathbf{y}\|\|\mathbf{x}\|$ genau dann, wenn $\theta = 0$. Dann sind $\mathbf{x}$ und $\mathbf{y}$ parallel, d.h. sie unterscheiden sich nur durch ihre Längen, und dies bedeutet $\mathbf{y} = \lambda\mathbf{x}$, was gleichbedeutend mit $y_i = \lambda x_i$ für alle i ist. □

(1.35) ist die Cauchy-Schwarzsche Ungleichung. Sie wird oft in der Form

$$\left|\sum_{i=1}^{n} x_i y_i\right|^2 \leq \sum_{i=1}^{n} x_i^2 \sum_{i=1}^{n} y_i^2 \tag{1.38}$$

angeschrieben. Für den Fall $\mathbf{y} = \lambda\mathbf{x}$, $\lambda \in \mathbb{R}$, hat man wegen $\cos\theta = 1$ für $\theta = 0$

$$|\mathbf{x}'\mathbf{y}|^2 = \|\mathbf{x}\|\|\mathbf{y}\| = \lambda\|\mathbf{x}\|^2,$$

d.h.

$$\left|\sum_{i=1}^{n} x_i y_i\right|^2 = \sum_{i=1}^{n} x_i^2 \sum_{i=1}^{n} y_i^2 = \lambda\left(\sum_{i=1}^{n} x_i^2\right)^2. \tag{1.39}$$

(s.a. (1.36)).

Für $\theta = \pi/2$ (90^o) ist $\cos\theta = 0$; (1.32) impliziert dann $\mathbf{x}'\mathbf{y} = 0$; die Vektoren $\mathbf{x}$ und $\mathbf{y}$ sind dann *orthogonal*[8]. Variablen X und Y, deren Korrelation gleich Null ist, werden durch orthogonale Vektoren $\mathbf{x}$ und $\mathbf{y}$ repräsentiert.

Anmerkungen:

1. **Ähnlichkeit von Vektoren** Man kann das Skalarprodukt als ein Ähnlichkeitsmaß interpretieren: relativ zu den Längen $\|\mathbf{x}\|$ und $\|\mathbf{y}\|$ stimmen die Komponenten x_i und y_i um so mehr überein, je kleiner der Winkel θ zwischen $\mathbf{x}$ und $\mathbf{y}$ ist; für $\theta = 0$ unterscheiden sie sich allenfalls durch einen gemeinsamen Faktor λ. Für $\lambda \neq 1$ ein ein Vektor länger als der andere, d.h. der längere Vektor zeigt größere Ausprägungen der Merkmale, die von den Koordinatenachsen repräsentiert werden, als der kürzere, sofern größere Koordinatenwerte eine größere Ausprägung der Merkmale bedeuten. So habe eine Person eine Körperlänge von 170 cm und ein Gewicht von 70 kg, und eine zweite Person habe eine Köperlänge von 175 cm und ein Gewicht von 72.1 kg. Hier ist $\lambda \approx 1.03$ und $\theta = 0$ und man kann sagen, dass beide Personen dieselbe körperliche Konstitution haben (was Körperlänge und -gewicht angeht). Für $\theta = \pi/2$ ist $\cos\theta = 0$ und man hat eine maximale Unähnlichkeit: die eine Person ist klein und dürr und die zweite Person ist lang und dick, oder die eine ist klein und dick und die zweite ist lang und dürr. Zwei Personen können unterschiedliche Intelligenzquotienten haben, sich aber in Bezug auf ihr Intelligenzprofil (definiert durch die Scores in den Untertests) sehr ähnlich sein, wenn sich nämlich die Scores nur durch einen gemeinsamen Faktor λ unterscheiden). Zwei Personen können denselben Intelligenzquotienten haben, sich aber maximal in Bezug auf ihr

[8] Von griechisch *orthogonios*: ortho = rechts, *gon* = winklig

Profil unterscheiden, so dass die entsprechenden Vektoren (deren Komponenten die Scores der Untertests sind) orthogonal sind: die eine Person wird Dichter, die andere Physiker, etc.

2. **Nullkorrelation** Ist eine Korrelation $r_{xy} = 0$, so folgt *noch nicht*, dass die Variablen auch stochastisch unabhängig sind; es lassen sich Variablen definieren, die deterministisch miteinander verknüpft sind, deren Korrelationskoeffizient gleichwohl Null ist. Die Beziehung zwischen dem Korrelationskoeffizienten und der Abhängigkeit zwischen zwei Variablen hängt von der Wahrscheinlichkeitsverteilung bzw. - dichte ab. So seien X und Y 2-dimensional normalverteiltezufällige Veränderliche, so dass die Dichte für $X = x$ und $Y = y$ durch

$$f(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-r^2}} \exp\left[-\frac{1}{2(1-r^2)}(z_x^2 + z_y^2 - 2rz_xz_y)\right] \tag{1.40}$$

gegeben ist mit mit $z_x = (x-\mu_x)/\sigma_x$, $z_y = (y-\mu_y)/\sigma_y$. Für $r = 0$ folgt $f(x,y) = g(x)h(y)$, d.h. x und y sind stochastisch unabhängig. Für andere Dichten muß diese Folgerung nicht gelten.

1.4 Lineare Abhängigkeit und Unabhängigkeit von Vektoren

Der Begriff der linearen Abhängigkeit bzw. Unabhängigleit ist von zentraler Bedeutung für die lineare Algebra und deren Anwendung in der Multivariaten Statistik: eine Menge $\{\mathbf{x}_1, \ldots, \mathbf{x}_p\}$ von n-dimensionalen Vektoren ist linear abhängig, wenn sich jeweil einer der Vektoren als Linearkombination der übrigen darstellen läßt; dementsprechend sind sie linear unabhängig, wenn eine derartige Darstellung für keinen der Vektoren möglich ist; die bereits erwähnten latenten Variablen werden durch linear unabhängige Vektoren repräsentiert. Darüber hinaus ist der Begriff der linearen Unabhängigkeit von Vektoren eng mit der Frage nach der Existenz von Lösungen linearer Gleichungssysteme verbunden. Eine formale Definition der Abhängigkeit ist allerdings notwendig, um diesen Begriff für die jeweiligen Betrachtungen einbringen zu können.

Es seien $\mathbf{x}_j \in \mathbb{R}^n$, $1 \leq j \leq p$, $\mathbf{x}_j \neq \vec{0}$, und es werde die Darstellung des Nullvektors $\vec{0}$ als Linearkombination der $\mathbf{x}_j$ betrachtet:

$$\vec{0} = a_1\mathbf{x}_1 + \cdots + a_p\mathbf{x}_p. \tag{1.41}$$

Die Koeffizienten a_j, $j = 1, \ldots, n$ sind zunächst nicht bekannt; man kann sie als Unbekannte eines Gleichungssystems betrachten (s. unten). Generell kann man aber eine – von möglicherweise mehreren – Lösung für diese Gleichung sofort angeben:

$$a_1 = a_2 = \cdots = a_p = 0. \tag{1.42}$$

Weil diese Lösung *stets* eine Lösung ist, wird sie auch als *triviale Lösung* bezeichnet. Die Frage ist nun, ob weitere Lösungen existieren, bei denen mindestens ein

Koeffizient $a_j \neq 0$ ist. Es sei etwa $a_p \neq 0$, z.B. $a_p = -1$, und $a_{p-1}\mathbf{x}_{p-1} \neq \vec{0}$, so dass

$$\mathbf{x}_p = a_1\mathbf{x}_1 + \cdots + a_{p-1}\mathbf{x}_{p-1} \neq \vec{0}, \tag{1.43}$$

woraus folgt, dass für mindestens einen weiteren Koeffizienten $a_j \neq 0$ gelten muß. Man hat dann die

Definition 1.5 *Gilt (1.41) und sind nicht alle a_j gleich Null, so heißen die Vektoren $\boldsymbol{x}_j$* linear abhängig. *Ist dagegen (1.42) die einzige Lösung für die a_j, so heißen die $\boldsymbol{x}_j$* linear unabhängig.

Sind also die $\mathbf{x}_j$, $j = 1, \ldots, p$ linear unabhängig, so kann keiner von ihnen als Linearkombination der übrigen dargestellt werden.

Korollar 1.2 *Die Vektoren $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$, $\boldsymbol{x}_j \neq \vec{0}$ für alle j, seien paarweise orthogonal. Dann sind sie linear unabhängig.*

Beweis: Es sei

$$\vec{0} = a_1\mathbf{x}_1 + a_2\mathbf{x}_2 + \cdots + a_n\mathbf{x}_n.$$

Dann gilt

$$\mathbf{x}_j'\vec{0} = 0 = a_1\mathbf{x}_j'\mathbf{x}_1 + a_2\mathbf{x}_j'\mathbf{x}_2 + \cdots + a_j\mathbf{x}_j'\mathbf{x}_j + \cdots + a_n\mathbf{x}_j'\mathbf{x}_n$$

und es folgt $\mathbf{x}_j'\mathbf{x}_k = 0$ für alle $j \neq k$ wegen der vorausgesetzten Orthogonalität der $\mathbf{x}_j$ und $\mathbf{x}_k$, $j \neq k$. Dann muß aber auch

$$0 = a_j\mathbf{x}_j'\mathbf{x}_j = a_j\|\mathbf{x}_j\|^2$$

für alle j gelten. Wegen $\|\mathbf{x}_j\|^2 > 0$ folgt $a_j = 0$ für alle j, also sind die $\mathbf{x}_j$ linear unabhängig. □

Die Umkehrung – linear unabhängige Vektoren sind paarweise orthogonal – gilt *nicht*.

Korollar 1.3 *Die Einheitsvektoren $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ sind paarweise orthogonal.*

Beweis: Es gilt

$$\mathbf{e}_j'\mathbf{e}_k = \begin{cases} 1, & j = k \\ 0, & j \neq k \end{cases} \tag{1.44}$$

wie man unmittelbar verifziert. □

Nach Satz 1.2 sind die $\mathbf{e}_j$, $j = 1, \ldots, n$ linear unabhängig.

Satz 1.3 *Es sei $\boldsymbol{x}$ ein beliebiger n-dimensionaler Vektor. Dann ist $\boldsymbol{x}$ als Linearkombination der $\boldsymbol{e}_j$ darstellbar.*

Beweis: Es ist

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = x_1 \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + x_2 \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix} + \cdots + x_n \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}.$$

Da die $x_i \in \mathbb{R}$ beliebig gewählt werden können, kann jeder Vektor aus $\mathbb{R}^n$ auf diese Weise dargestellt werden. □

Satz 1.4 *Es seien $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p$ n-dimensionale Vektoren, die nicht alle gleich dem Nullvektor sind. Gilt $p > n$, so sind die $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p$ linear abhängig.*

Beweis: Es genügt, den Fall $p = n + 1$ zu betrachten. Es gelte

$$a_1\mathbf{x}_1 + \cdots + a_n\mathbf{x}_n + a_{n+1}\mathbf{x}_{n+1} = \vec{0}, \quad \mathbf{x}_{n+1} \neq \vec{0}.$$

Man kann

$$a_1\mathbf{x}_1 + \cdots + a_n\mathbf{x}_n = -a_{n+1}\mathbf{x}_{n+1}$$

schreiben; für $a_{n+1} \neq 0$ hat man dann

$$\frac{a_1}{a_{n+1}}\mathbf{x}_1 + \cdots + \frac{a_n}{a_{n+1}}\mathbf{x}_n = -\mathbf{x}_{n+1}.$$

$\mathbf{x}_{n+1} \neq \vec{0}$ impliziert, dass nicht $\mathbf{x}_j = \vec{0}$ für $j = 1, \ldots, n$ gilt und dass nicht alle $a_j = 0$, $1 \leq j \leq n$. Das heißt aber, dass $\mathbf{x}_{n+1}$ eine Linearkombination der von $\vec{0}$ verschiedenen $\mathbf{x}_j$, $j \leq n$, ist die $\mathbf{x}_1, \ldots, \mathbf{x}_{n+1}$ linear abhängig sind. □

Satz 1.5 *$\boldsymbol{x}$ und $\boldsymbol{y}$ seien zwei n-dimensionale Vektoren $\neq \vec{0}$. Der Winkel zwischen ihnen sei θ. $\boldsymbol{x}$ und $\boldsymbol{y}$ sind linear abhängig genau dann, wenn $\theta = 0$, d.h. wenn $\cos\theta = r_{xy} = 1$.*

Beweis: Es sei

$$\cos\theta = \frac{\mathbf{x}'\mathbf{y}}{\|\mathbf{x}\|\|\mathbf{y}\|} = 1.$$

Dann ist $\theta = 0$, d.h. $\mathbf{x}$ und $\mathbf{y}$ sind parallel (in Zeichen: $\mathbf{x} \| \mathbf{y}$), so dass ein $0 \neq a \in \mathbb{R}$ existiert derart, dass $\mathbf{y} = a\mathbf{x} \Rightarrow \vec{0} = a\mathbf{x} - \mathbf{y}$ und die Koeffizienten von $\mathbf{x}$ und $\mathbf{y}$ sind nicht alle gleich Null, d.h. die beiden Vektoren sind linear abhängig.

Umgekehrt sei $\mathbf{y} = a\mathbf{x}$, $a \in \mathbb{R}$. Dann sind $\mathbf{x}$ und $\mathbf{y}$ linear abhängig (vergl. Beispiel 1.4) und es ist

$$\frac{a\mathbf{x}'\mathbf{x}}{a\|\mathbf{x}\|^2} = 1,$$

d.h. $\cos\theta = 1$, also $\theta = 0$.

Nun sei $\mathbf{x} \nparallel \mathbf{y}$ ($\mathbf{x}$ und $\mathbf{y}$ seien nicht parallel). Dann kann $\mathbf{y}$ nicht als Linearkombination von $\mathbf{x}$ berechnet werden (und umgekehrt, $\mathbf{x}$ kann nicht als Linearkombination von $\mathbf{y}$ berechnet werden). Also sind die Vektoren linear unabhängig und es ist $\theta \neq 0$ und $|\cos\theta| < 1$. □

Lineare Abhängigkeit und Korrelationen: Gegeben seien zwei Merkmale, die durch die Vektoren $\mathbf{x} = \mathbf{x}_0$ und $\mathbf{y} = a\mathbf{x}_0$, $a \in \mathbb{R}$, repräsentiert werden, – *wenn* die Messungen der zu den Merkmalen korrespondierenden Variablen messfehlerfrei wären, was sie im Allgemeinen nicht sind. Den Messungen entsprechend hat man $\mathbf{x} = \mathbf{x}_0 + \vec{\varepsilon}_1$, $\mathbf{y} = a\mathbf{x}_0 + \vec{\varepsilon}_2$, wobei $\vec{\varepsilon}_1$ und $\vec{\varepsilon}_2$ Vektoren sind, deren Komponenten die Messfehler repräsentieren. Der Korrelationskoeffizient r_{xy} entspricht dann

$$-1 < r_{xy} = \cos\theta_{xy} = \frac{(\mathbf{x}_0 + \vec{\varepsilon}_1)'(a\mathbf{x}_0 + \vec{\varepsilon}_2)}{\|\mathbf{x}_0 + \vec{\varepsilon}_1\|\|a\mathbf{x}_0 + \vec{\varepsilon}_2\|} < 1, \quad \varepsilon_1, \varepsilon_2 \neq 0 \tag{1.45}$$

Nur für den Spezialfall $\vec{\varepsilon}_1 = \vec{\varepsilon}_2 = \vec{0}$ würde man den Fall $r_{xy} = \cos\theta_{xy} = 1$ erhalten, und der ist zwar nicht unmöglich, hat aber die Wahrscheinlichkeit 0; diese Aussage ergibt sich aus der impliziten Annahme, dass die Komponenten der Fehlervektoren stetige zufällige Veränderlichen sind und dem aus der Statistik bekannten Sachverhalt, dass die Wahrscheinlichkeit, einen *bestimmten* Wert einer stetigen zufälligen Veränderlichen zu messen stets gleich Null ist . Für die gemessenen Vektoren $\mathbf{x}_j$, $j = 1, \ldots, n$ heißt dies, dass *rechnerisch* alle mit Messfehlern behafteten Vektoren linear unabhängig sind. □

Lineare Gleichungssysteme: Die Begriffe der linearen Abhängigkeit bzw. Unabhängigkeit stehen in engem Zusammenhang mit der Frage nach den Lösungen linearer Gleichungssysteme. Gegeben sei die Vektorgleichung

$$\mathbf{y} = a_1\mathbf{x}_1 + \cdots + a_p\mathbf{x}_p, \tag{1.46}$$

wobei $\mathbf{y}, \mathbf{x}_1, \ldots, \mathbf{x}_p$ n-dimensionale Vektoren seien. Gibt man die Koeffizienten $a_1, \ldots, a_p$ vor, so ist $\mathbf{y}$ eine Linearkombination der $\mathbf{x}_j$, $1 \leq j \leq p$. Gibt man den Vektor $\mathbf{y}$ vor, so ergibt sich die Frage, ob Koeffizienten $a_1, \ldots, a_p$ existieren derart, dass $\mathbf{y}$ als Linearkombination der $\mathbf{x}_j$ dargestellt werden kann. Die Gleichung (1.46) entspricht dann einem linearen Gleichungssystem in den Unbekannten a_j, $j = 1, \ldots, p$; die Gleichungen sind

$$y_i = a_1x_{i1} + a_2x_{i2} + \ldots + a_px_{ip}, \quad i = 1, \ldots, n \tag{1.47}$$

Der folgende Satz gibt eine Teilauskunft zur Frage nach den a_j; auf ihn wird oft bei der Herleitung von Aussagen zurückgegriffen:

Satz 1.6 *Gegeben sei die Vektorgleichung (1.46). Sind die Vektoren $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p$ linear unabhängig, so sind die $a_1, \ldots, a_p$ eindeutig bestimmt.*

Beweis: Sind die $\mathbf{x}_j$ linear unabhängig, so gilt $a_1\mathbf{x}_1 + \cdots + a_p\mathbf{x}_p = \mathbf{y}$ mit $\mathbf{y} = \vec{0}$ nur dann, wenn $a_1 = \cdots = a_p = 0$. Für den Fall $\mathbf{y} \neq \vec{0}$ werde angenommen, dass

(mindestens) zwei Lösungen $a_1, \ldots, a_p$ und $b_1, \ldots, b_p$ existieren:

$$\begin{aligned} \mathbf{y} &= a_1\mathbf{x}_1 + \cdots + a_p\mathbf{x}_p \\ \mathbf{y} &= b_1\mathbf{x}_1 + \cdots + b_p\mathbf{x}_p \end{aligned}$$

Dann ist

$$\mathbf{y} - \mathbf{y} = \vec{0} = (a_1 - b_1)\mathbf{x}_1 + \cdots + (a_p - b_p)\mathbf{x}_p,$$

und wegen der linearen Unabhängigkeit der $\mathbf{x}_j$ muß $a_j - b_j = 0$ gelten für $j = 1, \ldots, p$. Das heißt aber $a_j = b_j$, d.h. die Koeffizienten sind eindeutig bestimmt.

□

Auf die Frage nach der Gesamtheit von Lösungen linearer Gleichungssysteme wird im Abschnitt 2.9 in größerer Allgemeinheit eingegangen; hier soll zur Illustration der Fall $n = 2$ behandelt werden. Gegeben seien drei 2-dimensionale Vektoren $\mathbf{y}$, $\mathbf{x}_1$ und $\mathbf{x}_2$, alle $\neq \vec{0}$. Die Frage sei, ob sich $\mathbf{y}$ als Linearkombination der Vektoren $\mathbf{x}_1$ und $\mathbf{x}_2$ darstellen läßt, d.h. ob Koeffizienten $a_1, a_2 \in \mathbb{R}$ existieren derart, dass

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = a_1\mathbf{x}_1 + a_2\mathbf{x}_2 = a_1 \begin{pmatrix} x_{11} \\ x_{21} \end{pmatrix} + a_2 \begin{pmatrix} x_{12} \\ x_{22} \end{pmatrix} \tag{1.48}$$

gilt. Schreibt man die rechte Seite aus, so ergibt sich das System von Gleichungen

$$y_1 = a_1x_{11} + a_2x_{12} \tag{1.49}$$

$$y_2 = a_1x_{21} + a_2x_{22} \tag{1.50}$$

mit $\mathbf{y} = (y_1, y_2)'$, $\mathbf{x}_1 = (x_{11}, x_{21})'$, $\mathbf{x}_2 = (x_{12}, x_{22})'$, und den Unbekannten a_1 und a_2 (die zu einem Vektor $\mathbf{a} = (a_1, a_2)'$ zusammengefasst werden können). Man findet

$$a_1 = \frac{y_1x_{22} - y_2x_{12}}{x_{11}x_{22} - x_{12}x_{21}} \tag{1.51}$$

$$a_2 = \frac{y_2x_{11} - y_1x_{21}}{x_{11}x_{22} - x_{12}x_{21}} \tag{1.52}$$

Es wird deutlich, dass eine *notwendige Bedingung* für die Existenz einer eindeutigen Lösung $\mathbf{a} = (a_1, a_2)'$ durch

$$x_{11}x_{22} - x_{12}x_{21} \neq 0 \tag{1.53}$$

gegeben ist; sie reflektiert die Bedingung der linearen Unabhängigkeit der $\mathbf{x}_1$ und $\mathbf{x}_2$, wie jetzt gezeigt wird. Der Fall $x_{11}x_{22} - x_{12}x_{21} = 0$ würde bedeuten, dass durch 0 dividiert werden muß, um eine Lösung zu erhalten, – und diese Operation macht bekanntlich keinen Sinn. Gleichwohl ist dieser Fall möglich, denn bezüglich der Wahl von $\mathbf{x}_1$ und $\mathbf{x}_2$ sind keinerlei einschränkende Annahmen gemacht worden, und die Implikationen dieses Falles sind für das Folgende von Bedeutung.

Es werde also der Fall

$$x_{11}x_{22} - x_{12}x_{21} = 0 \tag{1.54}$$

betrachtet. Er impliziert

$$\frac{x_{21}}{x_{11}} = \frac{x_{22}}{x_{12}} \tag{1.55}$$

Aber $x_{21}/x_{11} = \tan\theta$, θ der Winkel, der die Orientierung von $\mathbf{x}_1$ angibt[9], und (1.55) besagt, dass dieser Winkel identisch mit dem von $\mathbf{x}_2$ ist. Daraus folgt, dass (1.54) dann erfüllt ist, wenn $\mathbf{x}_1$ und $\mathbf{x}_2$ parallel (kollinear) sind. Dann gilt $\mathbf{x}_2 = \lambda\mathbf{x}_1$ mit $0 \neq \lambda \in \mathbb{R}$, und die Gleichung (1.48) geht über in die Gleichung

$$\mathbf{y} = a_1\mathbf{x}_1 + a_2\mathbf{x}_2 = a_1\mathbf{x}_1 + a_2\lambda\mathbf{x}_1 = (a_1 + a_2\lambda)\mathbf{x}_1, \tag{1.56}$$

aus der sofort hervorgeht, dass (1.48) nun keine eindeutige Lösung (a_1, a_2) hat, denn wegen (1.56) kann nur der Parameter $\alpha = a_1 + a_2\lambda$ bestimmt werden. Alle a_1 und a_2, die der Beziehung

$$a_1 = \alpha - a_2\lambda \tag{1.57}$$

genügen, sind nun eine Lösung des Gleichungssystems. Die Beziehung $\mathbf{x}_2 = \lambda\mathbf{x}_1$, $\lambda \neq 0$, bedeutet, dass $\mathbf{x}_1$ und $\mathbf{x}_2$ linear abhängig sind, denn $b_1\mathbf{x}_1 + b_2\mathbf{x}_1 = \vec{0}$ mit $b_1 = 1$ and $b_2 = -\lambda \neq 0$. Wenn, im 2-dimensionalen Fall, lineare Abhängigkeit die Parallelität zweier Vektoren bedeutet, so kann man vermuten, dass lineare Unabhängigkeit einen Winkel ungleich Null zwischen den Vektoren impliziert.

Anmerkung: Man kann die Vektoren $\mathbf{x}_1$ und $\mathbf{x}_2$ zu einer *Matrix*

$$X = [\mathbf{x}_1, \mathbf{x}_2] = \begin{pmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{pmatrix}$$

zusammenfassen: die erste Spalte von X ist der Vektor $\mathbf{x}_1$, die zweite Spalte ist der Vektor $\mathbf{x}_2$. Dann heißt der Nenner $x_{11}x_{22} - x_{12}x_{21}$ in den Ausdrücken (1.49) und (1.50) die *Determinante* der (2×2)-Matrix X; sie wird mit $|X|$ bezeichnet:

$$|X| = x_{11}x_{22} - x_{12}x_{21}, \tag{1.58}$$

s. Anhang, Abschnitt 4.7. Die Determinante einer Matrix ist stets gleich einer reellen Zahl und ist nur von Null verschieden, wenn die Spaltenvektoren der Matrix linear unabhängig sind, wie die vorangegangenen Betrachtungen zeigen. Diese Aussage gilt allgemein für (n, n)-Matrizen und damit für Systeme von linearen Gleichungen mit n Unbekannten. Wie die Inspektion der Gleichungen (1.51) und (1.52) nahelegt haben auch die Zähler in den Ausdrücken für a_1 und a_2 die Form einer Determinante. So ersetze man in der Matrix X die erste Spalte durch den Vektor $\mathbf{y}$ und berechne die Determinante der so veränderten Matrix:

$$X_1 = \begin{pmatrix} y_1 & x_{12} \\ y_2 & x_{22} \end{pmatrix}, \quad |X_1| = y_1x_{22} - y_2x_{12}.$$

[9] In Bezug auf das Koordinatensystem, in dem die x_{ij} die Komponenten der Vektoren $\mathbf{x}_1$ und $\mathbf{x}_2$ sind.

Dann kann der Ausdruck für a_1 in der Form

$$a_1 = \frac{|X_1|}{|X|} \tag{1.59}$$

geschrieben werden. Bildet man die Matrix X_2, indem man die zweite Spalte durch $\mathbf{y}$ ersetzt und berechnet die Determinante von X_2, so kann man den Ausdruck für a_2 in der Form

$$a_2 = \frac{|X_2|}{|X|} \tag{1.60}$$

schreiben. Man hat damit einen Spezialfall der *Cramerschen Regel*, nach der lineare Gleichungen auch für dfen Fall von n Gleichungen mit n Unbekanntenn gelöst werden können: man ersetzt in der (n, n)-Matrix X die i-te Spalte durch $\mathbf{y}$ und erhält für die i-te Unbekannte die Lösung

$$a_i = \frac{|X_i|}{|X|}, \quad i = 1, \ldots, n \tag{1.61}$$

für den Fall, dass $|X| \neq 0$, d.h. für den Fall, dass die Spaltenvektoren von X linear unabhängig sind.

Für die folgenden Betrachtungen ist aber der Begriff der Determinante nicht notwendig, weshalb an dieser Stelle nicht weiter auf ihn eingegangen wird.

□

Lineare Gleichungssysteme werden allgemein in Abschnitt 2.9 besprochen.

1.5 Vektorräume

Ein Vektorraum ist eine Menge von Vektoren, deren Linearkombinationen wiederum Elemente der Menge sind und für die für jedes Paar von Vektoren das Skalarprodukt definiert ist.

Man mag fragen, wozu man eine solche Spezifikation von Mengen benötigt. An dieser Stelle muß es genügen, darauf hinzuweisen, dass der Begriff des Vektorraumes im Rahmen der Entwicklung einer Theorie der Systeme linearer Gleichungen mit mehreren Unbekannten formuliert wurde (Dorier (1995)). Einige Systeme haben gar keine Lösung, andere haben genau eine und wieder andere Systeme haben Mengen von Lösungen; welcher dieser Fälle jeweils vorliegt, hängt von der Existenz bestimmter Abhängigkeiten zwischen den Gleichungen ab, und deren Charakterisierung führt zum Begriff des Vektorraums. Im Rahmen dieses Textes würde es zu weit führen, diese Sachverhalte hier zu elaborieren, der Sinn des Begriffs des Vektorraums wird sich im Zuge der folgenden Darstellung ergeben.

Definition 1.6 *Es sei V eine Menge von Vektoren, und für beliebige $\boldsymbol{x}_1, \boldsymbol{x}_2 \in V$ gelte*

1. $\boldsymbol{x} = a_1\boldsymbol{x}_1 + a_2\boldsymbol{x}_2 \in V$ *mit* $a_1, a_2 \in \mathbb{R}$,
2. Für $\boldsymbol{x}_1, \boldsymbol{x}_2 \in V$ *existiert das Skalarprodukt* $\boldsymbol{x}_1'\boldsymbol{x}_2$.
Dann heißt V Vektorraum.[10]

Anmerkungen:
1. Die in der Definition genannte Menge von Vektoren ist eine spezielle *algebraische Struktur*, d.h. eine Menge von Elementen, für die bestimmte Verknüpfungen definiert sind. Die genannten Verküpfungen (Addition und Bildung von Skalarprodukten) setzen voraus, dass die Vektoren der Menge dieselbe Anzahl von Komponenten haben.
2. Nicht jede Menge von Vektoren bildet einen Vektorraum. Es sei $\mathcal{M}$ eine Menge von $p \in \mathbb{N}$ Vektoren gleicher Dimensionalität mit der Länge 1, etwa $\mathcal{M} = \{\mathbf{x}_j | \mathbf{x}_j \in \mathbb{R}^n, \|\mathbf{x}_j\| = 1, j = 1, \ldots, p\}$. Diese Menge ist kein Vektorraum, denn die Linearkombination $\mathbf{x} = \sum_j a_j\mathbf{x}_j$ hat, für beliebige Koeffizienten a_j, nicht mehr notwendig die Länge 1, d.h. es gilt nicht notwendig $\mathbf{x} \in M$.
3. Die Definition 1.6 des Begriffs des Vektorraums ist minimalistisch insofern, als sie sich nur auf die hier behandelten Vektoren mit endlicher Anzahl von Komponenten beschränkt. Sie kann aber auf andere Objekte wie Matrizen oder Funktionen übertragen werden, allerdings muß die Definition dann ausführlicher werden. Die Definition 1.6 genügt zunächst, um sich an den Begriff zu gewöhnen. □

Auf Seite 18 wurde der Begriff der Norm eines Vektors eingeführt. Allgemein ist eine Norm[11] eine Abbildung eines Vektorraums in die Menge der nichtnegativen reellen Zahlen $\mathbb{R}_+$; sie wird mit $\|\cdot\|$ bezeichnet:

$$\|\cdot\| : V \to \mathbb{R}_+, \quad \mathbf{x} \mapsto \|\mathbf{x}\| \in \mathbb{R} \text{ für } \mathbf{x} \in V. \tag{1.62}$$

Normen haben die folgenden Eigenschaften: für alle $\mathbf{x}, \mathbf{y} \in V$ und alle $a \in \mathbb{R}$ gelten
1. *Definitheit*: $\|\mathbf{x}\| = 0 \to \mathbf{x} = \vec{0}$,
2. *absolute Homogenität*: $\|a\mathbf{x}\| = |a|\|\mathbf{x}\|$, $a \in \mathbb{R}$,
3. *Subadditivität*: $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$ (Dreiecksungleichung).

Während die absolute Homogenität sofort klar ist (s. Definition der Länge eines Vektors), mag die Subadditivität nicht sofort evident sein, aber sie läßt sich leicht herleiten. Es ist

$$\begin{aligned} \|\mathbf{x} + \mathbf{y}\|^2 &= (\mathbf{x} + \mathbf{y})'(\mathbf{x} + \mathbf{y}) \\ &= \mathbf{x}'\mathbf{x} + \mathbf{x}'\mathbf{y} + \mathbf{y}'\mathbf{x} + \mathbf{y}'\mathbf{y} \\ &= \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 + 2\mathbf{x}'\mathbf{y} \end{aligned}$$

[10] Diese Definition ist im Vergleich zur entsprechenden Definition in Lehrbüchern der Linearen Algebra stark vereinfacht; es wird nur die für die Zwecke dieses Skriptums wesentliche Eigenschaft von Vektorräumen genannt.
[11] Von lat. *norma* = Richtschnur

Es ist aber $\mathbf{x}'\mathbf{y} = \|\mathbf{x}\|\|\mathbf{y}\|\cos\theta \leq \|\mathbf{x}\|\|\mathbf{y}\|$, denn $\cos\theta \leq 1$, θ der Winkel zwischen den Vektoren $\mathbf{x}$ und $\mathbf{y}$. Es folgt

$$\|\mathbf{x}+\mathbf{y}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 + 2\mathbf{x}'\mathbf{y} \leq \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 + 2\|\mathbf{x}\|\|\mathbf{y}\| = (\|\mathbf{x}\| + \|\mathbf{y}\|)^2,$$

d.h. aber

$$\|\mathbf{x}+\mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|, \tag{1.63}$$

also die Dreiecksungleichung. □

Definition 1.7 *Es sei V ein Vektorraum, dessen Elemente n-dimensionale Vektoren seien, und es seien $\boldsymbol{x}, \boldsymbol{y} \in V$. Dann heißt die Norm der Differenz*

$$\|\boldsymbol{x}-\boldsymbol{y}\| = \left(\sum_{i=1}^{n}(x_i - y_i)^2\right)^{1/2} \tag{1.64}$$

die euklidische Distanz *zwischen $\boldsymbol{x}$ und $\boldsymbol{y}$, und V heißt auch* euklidischer Vektorraum *(vergl. (1.5)).*

$\|\mathbf{x}-\mathbf{y}\|$ ist ein Vektor, dessen Länge der Distanz $d_e(x,y)$ zwischen den Endpunkten x und y der Vektoren $\mathbf{x}$ und $\mathbf{y}$ entspricht; $d_e(x,y)$ ist euklidisch, weil der Vektor $\mathbf{x}-\mathbf{y}$ die Distanz zwischen den Endpunkten von $\mathbf{x}$ und $\mathbf{y}$ durch eine Gerade repräsentiert[12], und definiert die euklidische Metrix. Der Index e in der Bezeichnung d_e soll anzeigen, dass die euklidische Distanz gemeint ist, im Unterschied zur Distanz $d_p(x,y)$, der Minkowski-Distanz (1.6) (Seite 9). Es sei noch einmal darauf hingewiesen, dass eine Distanz über ein Skalarprodukt definiert ist: es sei

$$d_i = |x_i - y_i|, \quad i = 1,\ldots,n \tag{1.65}$$

der Betrag der Koordinatendifferenz $x_i - y_i$, und es sei $\mathbf{d} = (d_1, d_2, \ldots, d_n)'$ der Vektor mit den Komponenten d_i. Dann ist

$$\|\mathbf{d}\| = (\mathbf{d}'\mathbf{d})^{1/2} = \|\mathbf{x}-\mathbf{y}\| = d_e(x,y). \tag{1.66}$$

Der Vektor $\mathbf{d}$ kann als Spezialfall eines Vektors $\mathbf{d}_p = (d_1^{p/2}, d_2^{p/2}, \ldots, d_n^{p/2})'$ mit $p > 1, p \in \mathbb{R}$ angesehen werden. Für $p = 2$ erhält man gerade den Vektor $\mathbf{d}$. Dann ist

$$\mathbf{d}_p'\mathbf{d}_p = \sum_{i=1}^{n} d_i^{p/2} d_i^{p/2} = \sum_{i=1}^{n} d_i^p = \sum_{i=1}^{n} |x_i - y_i|^p, \quad p > 1 \tag{1.67}$$

und

$$d_p(x,y) = (\mathbf{d}_p'\mathbf{d}_p)^{1/p} = \left(\sum_{i=1}^{n} |x_i - y_i|^p\right)^{1/p}, \quad p > 1 \tag{1.68}$$

ist die Minkowski-Distanz zwischen x und y, die die Minkowski-Metrik definiert. Minkowski-Distanzen spielen in diesem Skript keine weitere Rolle, weshalb hier nicht weiter auf sie eingegangen wird.

[12] Nach Euklid ist die Gerade die kürzeste Verbindung zwischen zwei Punkten.

Definition 1.8 *Es sei V ein Vektorraum und $U \subset V$ sei eine Teilmenge von Vektoren aus V. Für beliebige Vektoren $\boldsymbol{x}_1, \boldsymbol{x}_2 \in U$ gelte $\boldsymbol{x} = a_1\boldsymbol{x}_1 + a_2\boldsymbol{x}_2 \in U$, mit $a_1, a_2 \in \mathbb{R}$, und es existiere das Skalarprodukt $\boldsymbol{x}_1'\boldsymbol{x}_2$. Dann heißt U* Teilvektorraum *(oder einfach* Teilraum*) von V. Die leere Menge $\{\emptyset\}$ und der Vektorraum V selbst heißen* triviale Teilräume.

Bemerkung 1.1 Für bestimmte multivariate Verfahren, etwa die Analyse von Zeitreihen, ist es notwendig, von einem allgemeineren Begriff des Vektorraums auszugehen als von dem, der in Definition 1.6 spezifiziert wurde. So kann es notwendig werden, Funktionen der Zeit zu betrachten und sie, analog zu den hier betrachteten Vektoren, als Linearkombinationen von Basisfunktionen zu repräsentieren, die sich als Eigenfunktionen einer Matrix von Autokorrelationen bestimmen lassen. Abschnitt 3 enthält eine erste Einführung. □

1.6 Erzeugendensysteme und Basen von Vektorräumen

Definition 1.9 *Es sei V ein Vektorraum und $M = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p\} \subseteq V$ eine Menge von Vektoren aus V.*
1. Es sei

$$\mathcal{L}(M) = \{\boldsymbol{x} | \boldsymbol{x} = a_1\boldsymbol{x}_1 + \cdots, a_p\boldsymbol{x}_p\}, \quad \mathcal{L}(\emptyset) = \vec{0}, \tag{1.69}$$

d.h. $\mathcal{L}(M)$ sei die Menge aller Linearkombinationen der Vektoren aus M. $\emptyset$ bezeichnet die leere Menge. $\mathcal{L}(M)$ heißt die lineare Hülle *von M.*
2. Es sei $V = \mathcal{L}(M)$. Dann bildet die Menge M ein Erzeugendensystem *für V.*
3. Ein Erzeugendensystem M heißt minimal*, wenn kein Vektor $\boldsymbol{x} \in M$ existiert derart, dass*[13] *$M \setminus \{\boldsymbol{x}\}$ kein Erzeugendensystem für V mehr ist.*

Satz 1.7 *Es sei V ein Vektorraum und $M = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p\} \subset V$. Dann gilt $\mathcal{L}(M) \subseteq V$. Für $\mathcal{L}(M) \subset V$ ist $\mathcal{L}(M)$ ein Teilvektorraum von V.*

Beweis: Es seien $\mathbf{x}$ und $\mathbf{y}$ Linearkombinationen von Vektoren aus M,

$$\mathbf{x} = \sum_{j=1}^{p} a_j\mathbf{x}_j, \quad \mathbf{y} = \sum_{j=1}^{p} b_j\mathbf{x}_j.$$

Dann folgt

$$\lambda\mathbf{x} + \mu\mathbf{y} = \lambda\sum_{j=1}^{p} a_j\mathbf{x}_j + \mu\sum_{j=1}^{p} b_j\mathbf{x}_j = \sum_{j=1}^{p} c_j\mathbf{x}_j, \;\; c_j = \lambda a_j + \mu b_j$$

d.h. $\lambda\mathbf{x} + \mu\mathbf{y}$ ist ebenfalls eine Linearkombination der $\mathbf{x}_j$ und damit Element von $\mathcal{L}(M)$. Gilt $\mathcal{L}(M) \subset V$, so existieren Vektoren $\mathbf{x} \in V$ mit $\mathbf{x} \notin \mathcal{L}(M)$ so dass $\mathcal{L}(M)$ nur ein Teilvektorraum (einfach: Teilraum) von V sein kann. □

[13] Sei $\mathbf{x} \in M$. $M \setminus \{\mathbf{x}\}$ ist die Menge M ohne den Vektor $\mathbf{x}$.

Die Elemente $\mathbf{x}_1, \ldots, \mathbf{x}_p$ müssen nicht linear unabhängig sein, um als Erzeugendensystem einen Vektorraum oder einen Teilraum davon aufzuspannen. Der Fall ihrer linearen Unabhängigkeit erweist sich allerdings als von großer Bedeutung:

Definition 1.10 *Es sei $V = \{\boldsymbol{x} | \boldsymbol{x} \in \mathbb{R}^n\}$ ein Vektorraum, d.h. die Elemente von V seien n-dimensionale Vektoren. Weiter sei $\mathcal{B} = \{\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n\} \subset V$. $\mathcal{B}$ heißt* Basis *von V, wenn gilt*

(i) die $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n$ sind linear unabhängig,

(ii) $V = \mathcal{L}(\mathcal{B})$, dh. V ist die lineare Hülle von $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n$.

Anmerkungen:
1. Eine Basis ist ein minimales Erzeugendensystem.
2. Sind die Bedingungen (i) und (ii) der Definition 1.10 erfüllt, so existieren Koeffizienten $a_1, \ldots a_n$ derart, dass für einen beliebigen Vektor $\mathbf{x}$ Vektor aus V die Gleichung

$$\mathbf{x} = a_1\mathbf{b}_1 + a_2\mathbf{b}_2 + \cdots + a_n\mathbf{b}_n \tag{1.70}$$

gilt. Wegen der linearen Unabhängigkeit der $\mathbf{b}_j$ sind die Koeffizienten a_j eindeutig bestimmt (vergl. Satz 1.6, Seite 25). Die Koeffizienten $a_1, \ldots, a_n$ heißen *Koordinaten* von $\mathbf{v}$ bezüglich $\mathcal{B}$. □

Satz 1.8 *Es*[14] *sei V ein Vektorraum und M sei eine Teilmenge von V. Dann existiert ein Teilraum U von V mit $M \subseteq U$ und es gilt: Ist W ein Teilraum von V mit $M \subseteq W$, so folgt $U \subseteq W$, und es gilt*

$$U = \mathcal{L}(M), \tag{1.71}$$

d.h. U ist der kleinste Teilraum von V, der M enthält.

Beweis: Es sei $U := \mathcal{L}(M)$. Nach Satz 1.7 ist U ein Teilraum von V. Sicherlich ist $M \subset \mathcal{L}(M)$, und $U = \{\vec{0}\}$ für $M = \emptyset$. Dann folgt für W ein Teilraum von V mit $M \subseteq W$, dass $U = \mathcal{L}(M) \subseteq W$.

Nun sei $M \neq \emptyset$ und W sei ein Teilraum von V mit $M \subseteq W$. Für jeden Vektor $\mathbf{x} \in U = \mathcal{L}(M)$ existiert eine Darstellung

$$\mathbf{x} = a_1\mathbf{x}_1 + \cdots + a_p\mathbf{x}_p,$$

und wegen $M \subseteq W$ folgt $\mathbf{x} \in W$. Mithin folgt $U \subseteq W$, wie behauptet.

Der Teilraum U ist eindeutig bestimmt. Denn sei $U' \subseteq M$ ein zweiter Teilraum, der der Behauptung des Satzes genügt. Dann folgt einerseits $U' \subseteq U$ und andererseits $U \subseteq U'$, so dass $U' = U$ folgt. □

[14] Lorenz (1988), Band I, p.33

Anmerkung: Die Redeweise vom kleinsten Teilraum von V, der M enthält, ergibt sich aus der Definition von W: man kann W wählen, wie man will, so lange nur $M \subseteq W$ gilt. Dann gilt stets $M \subseteq U = \mathcal{L}(M) \subseteq W$. □

Definition 1.11 *Es sei* $V = \{\boldsymbol{x}|\boldsymbol{x} \in \mathbb{R}^n\}$, $\mathcal{B}_r = \{\boldsymbol{b}_1, \ldots, \boldsymbol{b}_r\} \subset V$, *und die* $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_r$ *seien linear unabhängig. Für* $r < n$ *heißt* $\mathcal{L}(\mathcal{B}_r)$ r-dimensionaler Teilraum *von* V*; er auch mit* V_r *bezeichnet. Für* $r = n$ *heißt* V *ein* n*-dimensionaler Vektorraum; er wird oft mit* $\mathbb{R}^n$*, gelegentlich mit* V_n *bezeichnet.*

Die Dimensionalität von $V_r := \mathcal{L}(\mathcal{B}_r)$ wird also durch die Anzahl r der Basisvektoren, die den jeweiligen Vektorraum aufspannen angegeben.

Satz 1.9 *Es sei* V_n *ein* n*-dimensionaler Vektorraum. Es existiert keine Basis* $\mathcal{B}$ *für* V_n *mit mehr als* n *Basisvektoren.*

Beweis: Die Aussage ist eine unmittelbare Konsequenz von Satz 1.4, Seite 24. □

Beispiel 1.3 Es sei V ein Vektorraum und $\mathbf{x}_1 \in V$; $\mathcal{B} = \{\mathbf{x}_1\}$ ist die Basis für einen eindimensionalen Teilraum $V_1 = \{\mathbf{x}|\mathbf{x} = a\mathbf{x}_1, a \in \mathbb{R}\}$ von V.

V sei insbesondere ein n-dimensionaler Vektorraum mit $n = 2$; V wird dann geometrisch durch eine Ebene repräsentiert. Es seien $\mathbf{x}_1, \mathbf{x}_2 \in V$ irgendzwei nicht parallele, d.h. linear unabhängige Vektoren aus V. Sie bilden eine Basis $\mathcal{B} = \{\mathbf{x}_1, \mathbf{x}_2\}$ von V (also $\mathbf{b}_1 = \mathbf{x}_1, \mathbf{b}_2 = \mathbf{x}_2$ in der Notation von Definition 1.10). Jede Linearkombination $\mathbf{x} = a_1\mathbf{x}_1 + a_2\mathbf{x}_2$ mit $a_1, a_2 \in \mathbb{R}$ ist ein Element von V, und $\mathcal{B}_1 = \{\mathbf{x}_1\}$ und $\mathcal{B}_2 = \{\mathbf{x}_2\}$ sind Basen für 1-dimensionale Teilräume von V, die durch Geraden mit verschiedener Orientierung in dieser Ebene repräsentiert werden.

Nun sei $n > 2$. Die nicht parallelen Vektoren $\mathbf{x}_1$ und $\mathbf{x}_2$ sind linear unabhängig und bilden die Basis $\mathcal{B} = \{\mathbf{x}_1, \mathbf{x}_1\}$ eines 2-dimensionalen Teilraums, d.h. einer Ebene in V. Um dies zu sehen betrachte man eine beliebige Linearkombination $\mathbf{x} = a_1\mathbf{x}_1 + a_2\mathbf{x}_2$, $a_1, a_2 \in \mathbb{R}$. Da $n > 2$, existiert ein Vektor $\mathbf{n} \in V$, der orthogonal auf $\mathbf{x}_1$ und $\mathbf{x}_2$ steht; $\mathbf{n}$ ist der *Normalenvektor* für diese Vektoren. Dann folgt

$$\mathbf{n}'\mathbf{x} = a_1\mathbf{n}'\mathbf{x}_1 + a_2\mathbf{n}'\mathbf{x}_2 = 0, \tag{1.72}$$

wegen $\mathbf{n}'\mathbf{x}_1 = \mathbf{n}'\mathbf{x}_2 = 0$. Dies gilt üfr alle $a_1, a_2 \in \mathbb{R}$, d.h. $\mathbf{n}$ ist ein Normalenvektor für die gesamte Ebene. Da $\mathbf{n}$ eine bestimmte Orientierung hat, ist mit ihm auch die Orientierung der Ebene definiert. Diese Betrachtung gilt für einen beliebigen n-dimensionalen Vektorraum mit $n > 2$. □

Aufgabe: Es sei $n = 3$ und es sei die Basis $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3\}$ für den 3-dimensionalen Vektorraum V gegeben. Es seien $\mathbf{x}, \mathbf{y} \in V$ und es sei $\mathbf{z} = \alpha\mathbf{x} + \beta\mathbf{y}$. Welche Koordinaten hat $\mathbf{z}$ in Bezug auf die Basis $\mathcal{B}$?

Linear unabhängige Vektoren sind nicht notwendig auch paarweise orthogonal zueinander, aber paarweise orthogonale Vektoren sind notwendig linear unabhängig. Orthogonale Vektoren können demnach als Basisvektoren gewählt werden. Dieser Fall ist besonders für Anwendungen in der multivariaten Statistik wichtig, weshalb eine eigene Definition dafür eingeführt wird:

Definition 1.12 *Es sei V ein n-dimensionaler Vektorraum. Eine Basis*

$$\mathcal{B} = \{\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n\}$$

von V heißt Orthonormalbasis *(ONB) (oder* orthonormale Basis*), wenn die $\boldsymbol{b}_j$ auf die Länge 1 normiert und paarweise orthogonal sind, d.h. wenn*

$$\boldsymbol{b}_j' \boldsymbol{b}_k = \begin{cases} 0, & j \neq k \\ 1, & j = k \end{cases}, \quad j,k = 1, \ldots, n \tag{1.73}$$

gilt. Die Basis $\mathcal{B}_r = \{\boldsymbol{b}_1, \ldots, \boldsymbol{b}_r\}$ mit $r < n$ heißt orthonormale Teilbasis.

Die n-dimensionalen Einheitsvektoren sind ein Beispiel für eine orthonormale Basis:

Satz 1.10 *Die n-dimensionalen Einheitsvektoren $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ mit*

$$\boldsymbol{e}_i = (0, \ldots, 0, 1, 0, \ldots, 0)'$$

der i-te n-dimensionale Einheitsvektor, bilden eine orthonormale Basis des V_n.

Beweis: Die Einheitsvektoren sind linear unabhängig, denn $\vec{0} = \lambda_1 \mathbf{e}_1 + \lambda_2 \mathbf{e}_2 + \cdots + \lambda_n \mathbf{e}_n$ ist nur möglich für $\lambda_1 = \cdots = \lambda_n = 0$; für die i-te Komponente hat man nämlich $0 = \lambda_i 1$, woraus sofort $\lambda_i = 0$ folgt. Darüber hinaus sind die $\mathbf{e}_i$ orthonormal, vergl. (1.44). Die Vektoren $\mathbf{e}_1, \ldots, \mathbf{e}_n$ bilden deshalb eine orthonormale Basis des V_n. Da stets

$$\mathbf{x} = x_1 \mathbf{e}_1 + x_2 \mathbf{e}_2 + \cdots + x_n \mathbf{e}_n,$$

sind die Komponenten x_j von $\mathbf{x}$ auch stets die Koordinaten von $\mathbf{x}$ bezüglich der $\mathbf{e}_1, \ldots, \mathbf{e}_n$. □

Definition 1.13 *Die Basis $\{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n\}$ heißt* kanonische Basis *des V_n.*

Eine beliebige orthonormale Basis läßt sich als Rotation der kanonischen Basis herleiten, vergl. den Abschnitt 2.8.2 über Basiswechsel, Seite 100.

Beispiel 1.4 Es sei V ein n-dimensionaler Vektorraum und $\mathcal{B} = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ sei eine Basis für V, so dass für jeden Vektor $\mathbf{x} \in V$ die Darstellung

$$\mathbf{x} = a_1 \mathbf{x}_1 + \cdots + a_n \mathbf{x}_n$$

existiert, wobei $a_1, \ldots, a_n$ die Koordinaten von $\mathbf{x}$ in Bezug auf die Basis $\mathcal{B}$ sind. Gesucht sind die Koordinaten von $\mathbf{x}$ in Bezug auf die kanonische Basis $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$.

Nach Satz 1.3 kann jeder Vektor $\mathbf{x} = (x_1, \ldots, x_n)' \in V$ als Linearkombination der n-dimensionalen Einheitsvektoren $\mathbf{e}_1, \ldots, \mathbf{e}_n$ dargestellt werden, und damit können auch die Vektoren der Basis $\mathcal{B}$ als Linearkombinationen der $\mathbf{e}_1, \ldots, \mathbf{e}_n$ dargestellt werden. Es sei x_{kj} die k-te Komponente von $\mathbf{x}_j \in \mathcal{B}$. Dann hat man

$$\mathbf{x} = \sum_{j=1}^{n} a_j \mathbf{x}_j = \sum_{j=1}^{n} a_j \sum_{k=1}^{n} x_{kj} \mathbf{e}_k = \sum_{k=1}^{n} \underbrace{(\sum_{j=1}^{n} a_j x_{kj})}_{x_k} \mathbf{e}_k = \sum_{k=1}^{n} x_k \mathbf{e}_k.$$

Die gesuchten Koordinaten sind also durch

$$x_k = \sum_{j=1}^{n} a_j x_{kj}, \quad k = 1, \ldots, n$$

gegeben. □

Die Frage ist nun, ob jeder Vektorraum V auch eine Basis haben muß.

Satz 1.11 *Jeder endlich erzeugte Vektorraum V hat eine Basis.*

Beweis: "Endlich erzeugt" soll heißen, dass die Bedingung $n < \infty$ erfüllt ist. Für einen n-dimensionalen Vektorraum ist jede Menge von $n + 1$ Vektoren eines n-dimensionalen Vektorraums linear abhängig. Sei nun $\{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ eine Menge von n-dimensionalen Vektoren. Dann gilt

(i) Ein einzelner Vektor, etwa $\mathbf{b}_1 \neq \vec{0}$, ist linear unabhängig, denn $\lambda_1 \mathbf{b}_1 = \vec{0}$ nur dann, wenn $\lambda_1 = 0$,

(ii) $\{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ sei eine Menge von linear unabhängigen Vektoren. Diese Vektoren bilden eine Basis von V, da alle Vektoren aus V als Linearkombinationen der $\mathbf{b}_j$ dargestellt werden können.

(iv) Ist $\{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ linear abhängig, so gilt

$$\lambda_1 \mathbf{b}_1 + \lambda_2 \mathbf{b}_2 + \cdots + \lambda_n \mathbf{b}_n = \vec{0}$$

und nicht alle λ_j sind gleich Null. Es sei $\lambda_1 \neq 0$; dann gilt

$$\mathbf{b}_1 = -\frac{\lambda_2}{\lambda 1} \mathbf{b}_2 - \cdots - \frac{\lambda_n}{\lambda_1} \mathbf{b}_n.$$

Sind die $\mathbf{b}_2, \ldots, \mathbf{b}_n$ linear unabhängig, so bilden sie eine Basis des Vektorraums, und $\mathcal{L}(\mathbf{b}_2, \ldots, \mathbf{b}_n) = V$. Ist $\{\mathbf{b}_2, \ldots, \mathbf{b}_n\}$ linear abhängig, so kann man die Betrachtung wiederholen; ist $\{\mathbf{b}_3, \ldots, \mathbf{b}_n\}$ linear unabhängig, so bildet diese Menge eine Basis und $\mathcal{L}(\mathbf{b}_3, \ldots, \mathbf{b}_n\} = V$, etc. Dieser Prozess kann im Prinzip fortgesetzt werden, bis man bei $\{\mathbf{b}_n\}$ angelangt ist. Da $\mathbf{b}_n$ notwendig linear unabhängig ist, ist in diesem Fall V ein eindimensionaler Vektorraum mit $\mathbf{b}_n$ als Basis. Man findet also in jedem Falle eine Basis, d.h. für jeden Vektorraum existiert eine Basis.

□

Orthonormale Basisentwicklung eines Vektors: Die zur Darstellung eines beliebigen Vektors $\mathbf{v} \in \mathcal{L}$ benötigten Koeffizienten a_j ergeben sich besonders einfach, wenn Orthonormalbasen gewählt werden: Es sei $\mathbf{x} \in V_n$ ($\mathbf{x}$ sei ein n-dimensionaler Vektor) und die $\mathbf{b}_1, \ldots, \mathbf{b}_n$ seien orthonormale Basisvektoren. Dann existieren Koordinaten $a_1, \ldots, a_n$ derart, dass

$$\mathbf{x} = a_1\mathbf{b}_1 + \cdots + a_n\mathbf{b}_n = \sum_{k=1}^{n} a_k\mathbf{b}_k. \tag{1.74}$$

Man betrachte nun das Skalarprodukt $\mathbf{b}_j'\mathbf{x}$:

$$\mathbf{b}_j'\mathbf{x} = \sum_{k=1}^{n} a_k\mathbf{b}_j'\mathbf{b}_k = a_j, \quad j = 1, \ldots, n \tag{1.75}$$

denn

$$\mathbf{b}_j'\mathbf{b}_k = \begin{cases} 0, & j \neq k \\ 1, & j = k \end{cases} \tag{1.76}$$

(1.74) kann dann in der Form

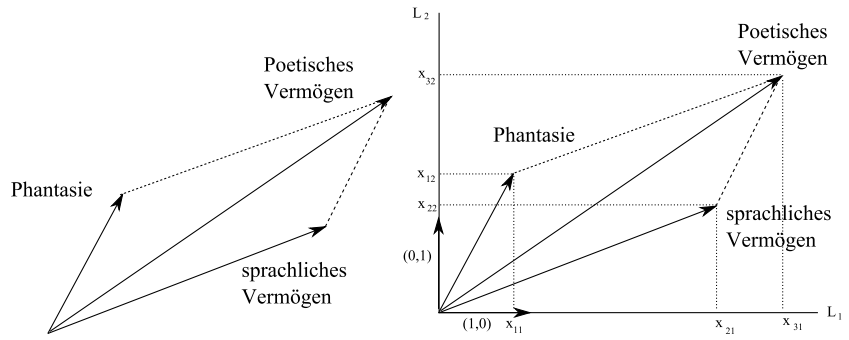
$$\mathbf{x} = \sum_{k=1}^{n} (\mathbf{b}_k'\mathbf{x})\mathbf{b}_k. \tag{1.77}$$

dargestellt werden. Dieser Ausdruck heißt auch *orthonormale Basisentwicklung* des Vektors $\mathbf{x}$.

Anmerkung: Bekanntlich kann unter bestimmten Normierungsbedingungen ein Skalarprodukt als Korrelation interpretiert werden. Dann bedeutet $\mathbf{b}_k'\mathbf{x}$ in Gleichung (1.77) die Korrelation zwischen dem Vektor $\mathbf{x}$ und dem Basisvektor $\mathbf{b}_k$. In der Faktorenanalyse und in Approximationen der Faktorenanalyse wird die *Ladung* eines Items (d.h. die Koordinate des Items) auf einer latenten Dimension als Korrelation zwischen dem Item und der latenten Dimension interpretiert. Diese Interpretation beruht auf (1.77). □

In Abbildung 4 wird noch einmal die Vektorrepräsentation bestimmter kognitiver Fähigkeiten gezeigt: links ergibt sich das poetische Vermögen als Vektoraddition (Linearkombination) der zwei Fähigkeiten 'sprachliches Vermögen' und

Abbildung 4: Poesie als Addition von Vektoren (links), und als Linearkombination von Vektoren der kanonischen Basis (rechts); aber: die Merkmale Phantasie und sprachliches Vermögen sind selbst komplexe, d.h. multivariate Konzepte.



'Phantasie'; die repräsentierenden Vektoren für diese beiden Kompetenzen sind nicht parallel und deshalb linear unabhängig, sie bilden eine Basis im $\mathbb{R}^2$. Rechts ist noch ein durch die beiden Einheitsvektoren $\mathbf{e}_1$ und $\mathbf{e}_2$ definiertes Koordinatensystem eingezeichnet worden. Alle Vektoren können als Linearkombinationen der orthonormalen Basis $\{\mathbf{e}_1, \mathbf{e}_2\}$ dargestellt werden. Das poetische Vermögen erscheint jetzt als Linearkombinationen der kognitiven Grundfunktionen L_1 und L_2, die wegen ihrer Repräsentation durch orthogonale Vektoren als unkorreliert angenommen werden[15].

Satz 1.12 *Es sei $\mathcal{B} = \{\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n\}$ eine beliebige Basis eines n-dimensionalen Vektorraums. Dann läßt sich aus $\mathcal{B}$ eine Orthonormalbasis konstruieren.*

Beweis:[16] Zunächst wird $\mathbf{b}_1$ normiert, d.h. es sei

$$\mathbf{c}_1 = \frac{1}{\|\mathbf{b}_1\|}\mathbf{b}_1.$$

Dann wird aus $\mathcal{L}(\mathbf{c}_1, \mathbf{b}_2)$ ein Vektor $\mathbf{d}_2$ gewählt, der der Bedingung $\mathbf{c}_1'\mathbf{d}_2 = 0$ genügt, d.h. es gelte $\mathbf{c}_1 \perp \mathbf{d}_2$. Dann mache man den Ansatz

$$\mathbf{d}_2 = \mathbf{b}_2 + \alpha\mathbf{c}_1,$$

[15]Derartige Repräsentationen von Kompetenzen bzw. Eigenschaften sind einerseits Standard, andererseits nicht unproblematisch, denn Ausdrücke wie 'Phantasie', 'sprachliches Vermögen' etc. bezeichnen komplexe Prozesse und 'poetisches Vermögen' steht für eine Interaktion dieser Prozesse, und der Ansatz, diese Interaktion durch einfache Addition darzustellen, kann eine zu große Vereinfachung bedeuten. Dieses Problem kann hier nicht diskutiert werden. Es sei aber angemerkt, dass die Vektoralgebra ebenfalls bei der Modellierung dynamischer Prozesse benötigt wird.

[16]Nach Ehrhard Schmidt (1876 – 1956). s. Koecher (1997), p. 157

und es folgt $\mathbf{c}_1'\mathbf{d}_2 = \mathbf{c}_1'\mathbf{b}_2 + \alpha\mathbf{c}_1'\mathbf{c}_1 = 0$, d.h. es folgt $\alpha = -\mathbf{c}_1'\mathbf{b}_2$. $\mathbf{c}_1$ und $\mathbf{b}_2$ sind linear unabhängig, daher ist $\mathbf{d}_2 \neq 0$, und somit kann

$$\mathbf{c}_2 = \frac{1}{\|\mathbf{d}_2\|}\mathbf{d}_2$$

definiert werden, und es ist $\mathbf{c}_i'\mathbf{c}_j = 0$ für $i, j = 1, 2$. Man fährt in dieser Weise fort und erhält eine orthonormale Basis. □

Orthogonale Teilräume

Satz 1.13 *Es sei $\mathcal{B} = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ eine orthonormale Basis eines n-dimensionalen Vektorraums V, und es sei $\mathcal{B}_k = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\}$, $k < n$, und die lineare Hülle $\mathcal{L}_k = \mathcal{L}(\mathcal{B}_k)$ definiert einen Teilraum von V (Satz 1.7), Seite 31). Dann existiert ein Vektor $\boldsymbol{n} \in V$, der orthogonal zu allen Vektoren aus $\mathcal{L}_k$ ist.*

Beweis: Es ist $V = \mathcal{L}(\mathbf{v}_1, \ldots, \mathbf{v}_n)$, trivialerweise ist $\mathbf{v}_{k+1} \in V$. Die $\mathbf{v}_1, \ldots, \mathbf{v}_n$ sind nach Voraussetzung paarweise orthogonal. Es sei $\mathbf{y} \in \mathcal{L}_k$, so dass

$$\mathbf{y} = a_1\mathbf{v}_1 + \cdots + a_k\mathbf{v}_k.$$

Dann folgt

$$\mathbf{v}_{k+1}'\mathbf{y} = a_1\mathbf{v}_{k+1}'\mathbf{v}_1 + \cdots + a_n\mathbf{v}_{k+1}'\mathbf{v}_k = 0,$$

da $\mathbf{v}_{k+1}'\mathbf{v}_j = 0$ für $j = 1, \ldots, k$. Also ist $\mathbf{v}_{k+1}$ ein Normalenvektor $\mathbf{n}$ für alle Vektoren aus $\mathcal{L}_k$. □

Damit ist die Annahme der Existenz eines Normalenvektors für eine Ebene auf Seite 33 gerechtfertigt.

Satz 1.14 *V, $\mathcal{B}$ und $\mathcal{L}_k$ seien wie im vorangegangenen Satz definiert. Es sei $\mathcal{B}_c = \{\boldsymbol{v}_{k+1}, \ldots, \boldsymbol{v}_n\}$ und $\mathcal{L}_c = \mathcal{L}(\mathcal{B}_c)$ sei der zu $\mathcal{B}_c$ korrespondierende Teilraum von V. Weiter seinen $\boldsymbol{x} \in \mathcal{L}_k$ und $\boldsymbol{y} \in \mathcal{L}_c$. Dann gilt $\boldsymbol{x} \perp \boldsymbol{y}$, d.h. $\boldsymbol{x}$ und $\boldsymbol{y}$ sind orthogonal.*

Beweis: Es gilt $\mathbf{x} = a_1\mathbf{v}_1 + \cdots + a_k\mathbf{v}_k$, $\mathbf{y} = b_{k+1}\mathbf{v}_{k+1} + \cdots + b_n\mathbf{v}_n$. Dann ist

$$\begin{aligned}
\mathbf{x}'\mathbf{y} &= (a_1\mathbf{v}_1 + \cdots + a_k\mathbf{v}_k)'(b_{k+1}\mathbf{v}_{k+1} + \cdots + b_n\mathbf{v}_n) \\
&= a_1b_{k+1}\mathbf{v}_1'\mathbf{v}_{k+1} + \cdots + a_kb_{k+1}\mathbf{v}_k'\mathbf{v}_{k+1} + \cdots \\
&\quad + a_1b_{k+1}\mathbf{v}_1'\mathbf{v}_{k+1} + \cdots + a_kb_{k+1}\mathbf{v}_k'\mathbf{v}_{k+1} = 0
\end{aligned}$$

da wegen der Orthogonalität der $\mathbf{v}_j$ die Skalarprodukte zwischen den $\mathbf{v}_j \in \mathcal{B}_k$ und den $\mathbf{v}_k \in \mathcal{B}_c$ aller gleich Null sind. □

Anmerkung: Man sagt kurz, die Teilräume $\mathcal{L}_k$ und $\mathcal{L}_c$ seien orthogonal; $\mathcal{L}_c$ heißt auch das orthogonale Komplement von $\mathcal{L}_k$. Eine formale Definition orthogonaler Teilräume ist

Definition 1.14 *Es sei V ein n-dimensionaler Vektorraum und $U \subset V$ sei ein Teilraum on V. Weiter sei*[17]

$$U^{\perp} = \{\boldsymbol{x} \in V | \boldsymbol{x}'\boldsymbol{u} = 0, \forall \boldsymbol{u} \in U\} \tag{1.78}$$

Dann heißt $U^{\perp}$ das orthogonale Komplement *von U.*

Beispiel 1.5 Es sei $V = \mathbb{R}^2$ ein 2-dimensionaler Vektorraum und U sei ein 1-dimensionaler Teilraum von V, d.h. U sei eine Gerade durch den Ursprung des Koordinatensystems mit der Orientierung des Vektors $\mathbf{a} \in V$; die Gerade werde mit $\mathbb{R}\mathbf{a}$ bezeichnet. Dann ist das zu U orthogonale Komplement die Gerade $\mathbb{R}\mathbf{b}$ durch den Ursprung, wobei $\mathbf{b} \perp \mathbf{a}$, $\mathbf{b} \in V$. □

Beispiel 1.6 Es sei $V = \mathbb{R}^3$ und U sei ein 2-dimensionaler Teilraum von V, d.h. eine Ebene in V. $\mathbf{n} \in V$ sei der Normalenvektor für U, d.h. $\mathbf{n}$ stehe senkrecht auf U. Dann ist $U^{\perp} = \mathbf{n}$ das orthogonale Komplement zur Ebene. □

Kombination von Teilräumen:

Definition 1.15 *Es sei W ein n-dimensionler Vektorraum und U und V seien zwei Teilräume von W. W heißt* orthogonale Summe *von U und V, in Zeichen $W = U \perp V$, wenn*
(i) Für jeden Vektor $\boldsymbol{w} \in W$, $\boldsymbol{w} = \boldsymbol{u} + \boldsymbol{v}$, $\boldsymbol{u} \in U$, $\boldsymbol{v} \in V$, und
(ii) $\boldsymbol{u}'\boldsymbol{v} = 0$ für alle $\boldsymbol{u} \in U$, $\boldsymbol{v} \in V$ gilt.

Satz 1.15 *(Dimensionalitätssatz)Es seien $U, V \subset W$, W ein n-dimensionaler Vektorraum. Die folgenden Aussagen (i) und (ii) sind äquivalent:*
(i) V ist das orthogonale Komplement von U, d.h. $V = U^{\perp}$,
(ii) $W = U \perp V$, d.h. W ist die orthogonale Summe von U und V. Dann gilt

$$\dim W = \dim U + \dim V^{\perp}. \tag{1.79}$$

Beweis: (i) ⇒(ii): Zu zeigen ist $W = U \perp U^{\perp}$. Es sei $\{\mathbf{b}_1, \ldots, \mathbf{b}_r\}$ eine orthonormale Basis des Teilraums U; sie kann nach Satz 1.12 zu einer orthonormalen Basis $\{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ von W erweitert werden. Nach Definition von $U^{\perp}$ ist $U^{\perp} = \mathcal{L}(\mathbf{b}_{r+1}, \ldots, \mathbf{b}_n)$, und da $\{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ eine Basis von W ist, folgt Aussage (i). Da die $\mathbf{b}_j$, $j = 1, \ldots, n$ paarweise orthonormal sind, folgt die Aussage (ii) und damit (1.79).

(ii) ⇒ (i): Nach (ii) der Definition 1.15 gilt $\mathbf{u}'\mathbf{v} = 0$ für $W = U \perp V$; Sei $\mathbf{x} \in W$. Dann $\mathbf{x} = \mathbf{u} + \mathbf{v}$ mit $\mathbf{u} \in U$ und $\mathbf{v} \in V$. Dann $\mathbf{x}^{\perp}$ genau dann, wenn $\mathbf{x}'\mathbf{w} = 0$ für alle $\mathbf{w} \in U$, also $\mathbf{u}'\mathbf{w} = 0$ für alle $\mathbf{u} \in U$, d.h $\mathbf{u} = \vec{0}$. □

Es gilt der

[17] ∀ steht für "für alle"

Satz 1.16 *Es sei W ein Vektorraum und $U \subset W, V \subset W$ seien Teilräume von W. Dann gilt*
1. Der Durchschnitt $U \cap V$ ist wieder ein Teilraum von W,
2. Die Vereinigung $U \cup V$ ist dann und nur dann ein Teilraum von W, wenn entweder $U \subseteq V$ oder $V \subseteq U$.

Beweis: Zu 1.: Es sei $U \cap V = \emptyset$; dann ist $U \cap V$ gerade der triviale Teilraum (vergl. Definition 1.8, Seite 31). Nun sei $U \cap V \neq \emptyset$. Es seien $\mathbf{u}, \mathbf{v} \in U \cap V$. Es seien $a_1, a_2 \in \mathbb{R}$ beliebige Koeffizienten, und es sei $\mathbf{w} = a_1\mathbf{u} + a_2\mathbf{v} \in W$, denn $\mathbf{u}, \mathbf{v} \in W$ und da W ein Vektorraum ist, folgt $\mathbf{w} \in W$. Da aber U und V Teilräume sind, sind sie abgeschlossen gegenüber der Multiplikation mit einem Skalar und der Summenbildung, also ist $\mathbf{w} \in U$ und $\mathbf{w} \in V$,d.h. $\mathbf{w} \in U \cap V$. Also ist $U \cap V$ ein Teilraum von W.

Zu 2.: Es seien U und V Teilräume von W und es gelte weder $U \subseteq V$ noch $V \subseteq U$. Nach Voraussetzung existiert ein Vektor $\mathbf{w}_1 \in U$ und $\mathbf{w}_1 \notin V$, und ein Vektor $\mathbf{w}_2 \notin U$, $\mathbf{w}_2 \in V$. Es werde angenommen, dass $U \cup V$ ein Teilraum von W ist. Dann muß $\mathbf{w} = \mathbf{w}_1 + \mathbf{w}_2 \in U \cup V$ gelten. Mit $\mathbf{w}_1 \in U$ muß aber auch $-\mathbf{w}_1 \in U$ gelten (da U ein Teilraum von W ist, muß mit $\mathbf{w}_1 \in U$ auch $\lambda\mathbf{w}_1 \in U$ gelten, $\lambda \in \mathbb{R}$, und es kann insbesondere $\lambda = -1$ gewählt werden). Nun muß $\mathbf{w} \in U$ oder $\mathbf{w} \in V$ sein. Sei $\mathbf{w} \in V$. Aber dann muß auch $\mathbf{w} - \mathbf{w}_2 \in V$ gelten, da ja $-\mathbf{w}_2 \in V$. Aber dann $(\mathbf{w}_1 + \mathbf{w}_2) - \mathbf{w}2 = \mathbf{w}_1 \in V$, entgegen der Voraussetzung $\mathbf{w}_1 \notin V$. Analog dazu sollte $\mathbf{w} - \mathbf{w}_1 \in U$ sein, da ja $-\mathbf{w}_1 \in U$. Aber nun folgt $\mathbf{w}_2 \in U$, entgegen der Voraussetzung $\mathbf{w}_2 \notin U$. Man erhält also einen Widerspruch zur Annahme, dass $U \cup V$ ein Teilraum von W ist. Damit W ein Teilraum ist, muß entweder $U \subset V$ oder $V \subset U$ gelten. □

Beispiel 1.7 Als Beispiel für eine Vereinigung zweier Teilräume betrachte man den 2-dimensionalen Vektorraum, wobei U und V zwei nicht parallelele Geraden seien. Sicherlich sind U und V Teilräume, aber ihre Vereinigung ist kein Teilraum: für $\mathbf{x}_1 \in U$ und $\mathbf{x}_2 \in V$ ist $\mathbf{x}_1 + \mathbf{x}_2 \notin U \cup V$. □

Definition 1.16 *Es seien U und V Teilräume eines Vektorraums W. Dann heißt*

$$U + V = \{\boldsymbol{u} + \boldsymbol{v} | \boldsymbol{u} \in U,\ \boldsymbol{v} \in V\} \tag{1.80}$$

die Summe *der Teilräume U und V.*

Satz 1.17 *Die Summe zweier Teilräume eines Vektorraums W ist wieder ein Teilraum von W.*

Beweis: Es seien $\mathbf{w}_1, \mathbf{w}_2 \in W_0 = U + V$. Dann ist auch $a_1\mathbf{w}_1 + a_2\mathbf{w}_2 \in W_0$. Denn $a_1\mathbf{w}_1 = a_1(\mathbf{u}_1 + bv_1)$, $\mathbf{u}_1 \in U$, $\mathbf{v}_1 \in V$, und $a_2\mathbf{w}_2 = a_2(\mathbf{u}_2 + \mathbf{v}_2)$, so dass

$$\mathbf{w}_1 + \mathbf{w}_2 = a_1\mathbf{u}_1 + a_2\mathbf{u}_2 + a_1\mathbf{v}_1 + a_2\mathbf{v}_2 \in W_0,$$

da $a_1\mathbf{u}_1 + a_2\mathbf{u}_2 \in U$ und $a_1\mathbf{v}_1 + a_2\mathbf{v}_2 \in V$, da U und V Teilräume sind, und es folgt $\mathbf{w}_1 + \mathbf{w}_2 \in W_0$, da W_0 ja alle Summen $\mathbf{u} + \mathbf{v}$ enthält. □

Die Summe zweier Teilräume ist also von der Vereinigung zweier Teilräume zu unterscheiden.

Die folgende Definition führt im Wesentlichen die Redeweise vom 'Rang eines Vektorraumes' ein:

Definition 1.17 *Es sei V ein Vektorraum und S eine Teilmenge von V. Dann ist der* Rang *von S gleich der Dimension des von S erzeugten Teilraums $\mathcal{L}(S)$. Ist $V = V_n$ ein n-dimensionaler Vektorraum und ist $S \subset V_n$, so hat S den Rang $r < n$, wenn S von r linear unabhängigen Vektoren aufgespannt wird; für $r = n$ hat S den* vollen Rang.

Anmerkung: r heißt auch die *Dimension* des Teilraums Teilraums $\mathcal{L}(S)$, und $n - r$ heißt die *Kodimension* des Teilraums $\mathcal{L}(S)$. Die Dimension eines Unter- oder Teilraums eines Vektorraums ist also nicht notwendig gleich der Dimension, d.h. der Anzahl der Komponenten der Vektoren, die die Elemente des Teilraums sind. □

Satz 1.18 *Für die Summe der Teilräume gilt die* Dimensionsformel

$$\dim(U + V) = \dim(U) + \dim(V) - \dim(U \cap V). \tag{1.81}$$

Beweis:[18] Die Dimension eines (Teil-)Vektorraums ist gleich der Anzahl der Basisvektoren des Raums. $U \cap V$ ist ein Teilraum und habe die Basis $\mathbf{a}_1, \ldots, \mathbf{a}_m$, also $m = \dim(U \cap V)$. Diese Basis kann zu Basen von einerseits U, andererseits V ergänzt werden: $\mathbf{a}_1, \ldots, \mathbf{a}_m, \mathbf{b}_1, \ldots, \mathbf{b}_r$ sei die Basis von U, also $\dim(U) = m + r$, und $\mathbf{a}_1, \ldots, \mathbf{a}_m, \mathbf{c}_1, \ldots, \mathbf{c}_s$ sei die Basis von V, also $\dim(V) = m + 2$, und $\dim(U \cap V) = m$. Dann bildet

$$\mathbf{a}_1, \ldots, \mathbf{a}_m, \mathbf{b}_1, \ldots, \mathbf{b}_r, \mathbf{c}_1, \ldots, \mathbf{c}_s$$

sicherlich ein Erzeugendensystem von $U + V$, d.h.

$$\mathcal{L}(\mathbf{a}_1, \ldots, \mathbf{a}_m, \mathbf{b}_1, \ldots, \mathbf{b}_r, \mathbf{c}_1, \ldots, \mathbf{c}_s) = U + V.$$

Zunächst muß gezeigt werden, dass auch

$$\mathrm{rg}(\mathbf{a}_1, \ldots, \mathbf{a}_m, \mathbf{b}_1, \ldots, \mathbf{b}_r, \mathbf{c}_1, \ldots, \mathbf{c}_s) = r + s + m$$

gilt, d.h. dass die $\mathbf{a}_1, \ldots, \mathbf{c}_s$ linear unabhängig sind, also

$$\underbrace{\sum_{i=1}^{m} \lambda_i \mathbf{a}_i}_{\mathbf{a}} + \underbrace{\sum_{i=1}^{r} \mu_i \mathbf{b}_i}_{\mathbf{b}} + \underbrace{\sum_{i=1}^{s} \nu_i \mathbf{c}_i}_{\mathbf{c}} = \mathbf{a} + \mathbf{b} + \mathbf{c} = \vec{0}$$

[18] Nach Lorenz I, p. 56

nur dann, wenn die Koeffizienten λ_i, μ_i, ν_i alle gleich Null sind. Es ist $\mathbf{a} \in U \cap V$, $\mathbf{b} \in U$, $\mathbf{c} \in V$. Dann folgt aus der vorangehenden Gleichung $\mathbf{b} = -\mathbf{a} - \mathbf{c} \in U \cap V$. Aus der Wahl der $\mathbf{b}_1, \ldots, \mathbf{b}_r$ folgt, dass alle $\lambda_i = 0$. Analog dazu folgt, dass alle $\mu_i = 0$, so dass $\mathbf{b} = \mathbf{c} = 0$, woraus wiederum $\mathbf{a} = 0$, d.h. alle $\lambda_i = 0$. Schließlich folgt

$$m + r + s = (m + r) + (m + s) - m = \dim(U) + \dim(V) - \dim(U \cap V),$$

und das war zu zeigen. □

Korollar 1.4 *Es seien U und V Teilräume eines Vektorraums W. Aus (1.81) folgt*

$$\dim(U + V) \leq \dim(U) + \dim(V), \tag{1.82}$$

da $\dim(U \cap V) \geq 0$.

2 Matrizen

2.1 Definitionen

Gegeben sei eine Menge $X = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ von m-dimensionalen Vektoren

$$\mathbf{x}_j = (x_{1j}, x_{2j}, \ldots, x_{mj})'.$$

Schreibt man diese Vektoren spaltenweise nebeneinander, so entsteht die Matrix

$$X = \begin{pmatrix} x_{11} & x_{12} & \ldots & x_{1n} \\ x_{21} & x_{22} & \ldots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{pmatrix} \tag{2.1}$$

X heißt auch (m, n)-Matrix; gelegentlich wird einfach $X_{m,n}$ dafür geschrieben, oder $X = (x_{ij})_{m,n}$ oder $X = (x_{ij})$, wenn klar ist, dass $1 \leq i \leq m$, $1 \leq j \leq n$. Eine andere Schreibweise ist $X \in \mathbb{R}^{m,n}$, womit angedeutet wird, dass die Elemente von X reelle Zahlen sind, denn man kann auch Matrizen betrachten, deren Elemente komplexe Zahlen sind. Derartige Matrizen werden aber in diesem Skript nicht behandelt. Auf analoge Weise kann eine Matrix entstehen, indem man eine Anzahl m-dimensionaler Zeilenvektoren untereinander schreibt. X heißt *quadratisch*, wenn $m = n$. Die $\mathbf{x}_1, \ldots, \mathbf{x}_n$ heißen die *Spaltenvektoren* von X. Die Schreibweise

$$X = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n] \tag{2.2}$$

erweist sich oft als nützlich. Die Zeilen $(x_{i1}, x_{i2}, \ldots, x_{in})$ heißen die *Zeilenvektoren* von X, $i = 1, \ldots, m$. Eine Matrix wird *gestürzt* oder *transponiert*, indem die

Zeilenvektoren als Spaltenvektoren angeschrieben werden; man schreibt X' dafür:

$$X' = \begin{pmatrix} x_{11} & x_{21} & \cdots & x_{m1} \\ x_{12} & x_{22} & \cdots & x_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1n} & x_{2n} & \cdots & X_{mn} \end{pmatrix} = [\tilde{\mathbf{x}}_1.\tilde{\mathbf{x}}_2, \ldots, \tilde{\mathbf{x}}_m] \tag{2.3}$$

X' ist also eine $(n \times m)$-Matrix. Die $\tilde{\mathbf{x}}_i$, $i = 1, \ldots, m$ sind die n-dimensionalen Spaltenvektoren von X', d.h. die Zeilenvektoren von X. Generell werden die Zeilenvektoren einer Matrix X oder allgemein A stets als Spaltenvektoren $\tilde{\mathbf{x}}_i$ bzw. $\tilde{\mathbf{a}}_i$ eben der transponierten Matrix X' bzw. A' angeschrieben.

Die Matrix X heißt *symmetrisch*, wenn

$$X' = X, \quad x_{ij} = x_{ji}, \;\; 1 \leq i, j \leq n \tag{2.4}$$

Symmetrische Matrizen sind notwendig quadratisch. Eine Matrix heißt *Diagonalmatrix*, wenn alle Elemente gleich Null sind bis auf r Diagonalelemente x_{ii}.

2.2 Operationen mit Matrizen

2.2.1 Addition und Multiplikation mit einem Skalar

Mit Matrizen können eine Reihe von Operationen durchgeführt werden; die beiden folgenden Operationen sind elementar. Die Multiplikation einer Matrix mit einem Vektor und mit einer Matrix erfordern eine etwas längere Elaboration und werden in den folgenden Unterabschnitten vorgestellt.

1. Multiplikation mit einem Skalar: $\lambda X = (\lambda x_{ij})$, $\lambda \in \mathbb{R}$, d.h. die Multiplikation von X mit einem Skalar bedeutet, dass jedes Element x_{ij} von X mit diesem Skalar multipliziert wird.

2. X und Y seinen zwei (m, n)-Matrizen. Dann ist die Summe $X + Y$ durch

$$X + Y = (x_{ij} + y_{ij}) \tag{2.5}$$

   definiert, d.h. die Elemente von $X + Y$ sind die Summen der korrespondierenen Elemente x_{ij} und y_{ij}.

2.2.2 Multiplikation einer Matrix mit einem Vektor

Gegeben seien die n m-dimensionalen Vektoren $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_n$, und der Vektor $\mathbf{y}$ sei eine Linearkombination dieser Vektoren mit den Koeffizienten $x_1, \ldots, x_n$, dh. es möge

$$\mathbf{y} = x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n. \tag{2.6}$$

Für gegebene Vektoren $\mathbf{a}_k$ wird $\mathbf{y}$ durch die Koeffizienten $x_1, \ldots, x_n$ bestimmt. Man kann sie zu einem Vektor $\mathbf{x} = (x_1, \ldots, x_n)'$ zusammenfassen. Ebenso kann man die $\mathbf{a}_k$ spaltenweise zu einer Matrix $A = [ba_1, \ldots, \mathbf{a}_n]$, d.h.

$$A = \begin{pmatrix} a_{11} & a_{12}, & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

zusammenfassen. Dann kann die Linearkombination (2.6) in der Form

$$\mathbf{y} = A\mathbf{x} \tag{2.7}$$

geschrieben werden. Diese Gleichung definiert die *Multiplikation einer Matrix mit einem Vektor*. Sie soll mit der Gleichung (2.6) äquivalent sein. Als Regel formuliert heißt dies, dass man den ersten Spaltenvektor von A mit der ersten Komponente von $\mathbf{x}$ multiplizieren soll, den so entstehenden Vektor zum zweiten Vektor $\mathbf{a}_2$ addieren soll, nachdem dieser mit der zweiten Komponente von $\mathbf{x}$ multipliziert wurde, etc. Wie schon in Beispiel 1.2, Seite 16 gezeigt wurde ist diese Regel gleichbedeutend damit, die erste Komponente y_1 von $\mathbf{y}$ als Skalarprodukt des ersten Zeilenvektors von A mit $\mathbf{x}$ zu berechnen, die zweite Komponente y_2 als Skalarprodukt des zweiten Zeilenvektors von A mit $\mathbf{x}$, etc. Man erhält

$$\mathbf{y} = A\mathbf{x} = \begin{pmatrix} \tilde{\mathbf{a}}_1'\mathbf{x} \\ \tilde{\mathbf{a}}_2'\mathbf{x} \\ \vdots \\ \tilde{\mathbf{a}}_m'\mathbf{x} \end{pmatrix} \tag{2.8}$$

Hält man die Matrix A fix, so wird $\mathbf{y}$ durch $\mathbf{x}$ bestimmt. Man kann das so ausdrücken, dass man sagt, dass für gegebene (m, n)-Matrix A der n-dimensionale Vektor $\mathbf{x}$ dem m-dimensionalen Vektor $\mathbf{y}$ zugeordnet wird, oder dass $\mathbf{x}$ in den Vektor $\mathbf{y}$ transformiert wird, oder dass $\mathbf{x}$ auf $\mathbf{y}$ abgebildet wird. Diese Abbildung ist durch die Matrix A definiert. Insofern kann man die Schreibweise von Funktionen verwenden:

$$A\colon \mathbf{x} \to \mathbf{y} \tag{2.9}$$

Diese Interpretation wird in Abschnitt 4.6 des Anhangs vertieft. Im Folgenden wird in erster Linie von der Interpretation der Gleichung $\mathbf{y} = A\mathbf{x}$ als Linearkombination der Spalten von A Gebrauch gemacht.

Beispiel 2.1 Ein Beispiel ist

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{pmatrix}, \quad \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \quad \mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix},$$

und

$$A\mathbf{x} = x_1\mathbf{a}_1 + x_2\mathbf{a}_2 = x_1 \begin{pmatrix} a_{11} \\ a_{21} \\ a_{31} \end{pmatrix} + x_2 \begin{pmatrix} a_{12} \\ a_{22} \\ a_{32} \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \mathbf{y}. \qquad (2.10)$$

Inspektion dieser Gleichung zeigt, dass die Komponenten y_i von $\mathbf{y}$ durch

$$\begin{aligned} y_1 &= x_1a_{11} + x_2a_{12} \\ y_2 &= x_1a_{21} + x_2a_{22} \\ y_3 &= x_1a_{31} + x_2a_{32} \end{aligned}$$

oder

$$y_i = x_1a_{i1} + x_2a_{i2} + x_3a_{i3}, \quad i = 1, \dots m$$

gegeben sind, d.h. die Ausdrücke für y_i sind durch das Skalarprodukt des Vektors $\mathbf{x}$ mit dem i-ten *Zeilen*vektor $\tilde{\mathbf{a}}_i' = (a_{i1}, a_{i2}, a_{i3})$ gegeben, d.h.

$$y_i = \tilde{\mathbf{a}}_i'\mathbf{x} = \mathbf{x}'\tilde{\mathbf{a}}_i, \quad i = 1, 2, 3 \qquad (2.11)$$

□

Es gilt nun der

Satz 2.1 *Es sei $\boldsymbol{y} = A\boldsymbol{u}$, A eine (m, n)-Matrix, $\boldsymbol{u}$ ein n-dimensionaler Vektor und $\boldsymbol{y}$ ein m-dimensionaler Vektor. Dann gilt*

$$\boldsymbol{y}' = (A\boldsymbol{u})' = \boldsymbol{u}'A'. \qquad (2.12)$$

Beweis: Für die i-te Komponente y_i von $\mathbf{y} = A\mathbf{u}$ gilt nach (2.11) $y_i = \tilde{\mathbf{a}}_i'\mathbf{u} = \mathbf{u}'\tilde{\mathbf{a}}_i$ d.h. es ist

$$\mathbf{y}' = (\tilde{\mathbf{u}}'\tilde{\mathbf{a}}_1, \tilde{\mathbf{u}}'\tilde{\mathbf{a}}_2, \dots, \tilde{\mathbf{u}}'\tilde{\mathbf{a}}_n) = \mathbf{u}'[\tilde{\mathbf{a}}_1, \dots, \tilde{\mathbf{a}}_n] = \mathbf{u}'A'.$$

□

(2.12) bedeutet, dass $\mathbf{y}'$ eine Linearkombination der Zeilenvektoren von A' ist. Statt A' kann man eine beliebige (n, m)-Matrix B einsetzen, so dss man eine Gleichung der Form $\mathbf{y}' = \mathbf{x}'B$ erhält, und so ergibt sich die Zusammenfassung

1. $\mathbf{y} = A\mathbf{u}$ ist eine Linearkombination der Spaltenvektoren von A;
2. $\mathbf{z}' = \mathbf{v}'B$ ist eine Linearkombination der Zeilenvektoren von B.

Beispiel 2.2 Ist A wie im obigen Beispiel eine (3×2)-Matrix, so hat man

$$\mathbf{z}' = \mathbf{v}'A = (v_1, v_2, v_3) \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{pmatrix} = (v_1a_{11}+v_2a_{21}+v_3a_{31}, v_1a_{12}+v_2a_{22}+v_3a_{32}),$$

also

$$\mathbf{z} = \begin{pmatrix} v_1 a_{11} + v_2 a_{21} + v_3 a_{31} \\ v_1 a_{12} + v_2 a_{22} + v_3 a_{32} \end{pmatrix},$$

und $z_j = \mathbf{v}'\mathbf{a}_j, \quad j = 1, 2.$ □

Beispiel 2.3 Es sei $D = (d_{ij})$ eine (m, n)-Datenmatrix: d_{ij} ist die Messung des j-ten Merkmals ("Variable") beim i-ten Fall. Die Spaltenvektoren seien $\mathbf{d}_j$, $j = 1, \ldots, n$, d.h. die Komponenten d_{ij}, $i = 1, \ldots, m$ sind die Messwerte für die j-te Variable. Man sucht "latente" Variablen, die sich durch Vektoren $\mathbf{L}_k$ ($k = 1, \ldots, r \leq \min(m, n)$) repräsentieren lassen, mit denen man die Messdaten d_{ij} "erklären" kann, indem man die Vektoren $\mathbf{d}_j$ als Linearkombinationen der latenten Vektoren darstellt:

$$\mathbf{d}_j = b_{1j}\mathbf{L}_1 + \cdots + b_{rj}\mathbf{L}_r, \quad j = 1, \ldots, n, \tag{2.13}$$

mit $\mathbf{L}_k = (\ell_{1k}, \ell_{2k}, \ldots, \ell_{mk})'$. Da man aber nur die Daten hat, muß man dazu zuerst die latenten Vektoren als Linearkombinationen der $\mathbf{d}_j$ darstellen: Es sei $\mathbf{a}_1 = (a_{11}, a_{21}, \ldots, a_{n1})'$ ein n-dimensionaler Vektor, und es sei

$$D\mathbf{a}_1 = a_{11}\mathbf{d}_1 + \cdots + a_{n1}\mathbf{d}_n = \mathbf{L}_1, \tag{2.14}$$

oder, ausführlicher geschrieben,

$$\begin{aligned} D\mathbf{a}_1 &= \begin{pmatrix} d_{11} & d_{12} & \cdots & d_{1n} \\ d_{21} & d_{22} & \cdots & d_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ d_{m1} & d_{m2} & \cdots & d_{mn} \end{pmatrix} \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{n1} \end{pmatrix} \\ &= a_{11} \begin{pmatrix} d_{11} \\ d_{21} \\ \vdots \\ d_{m1} \end{pmatrix} + \cdots + a_{n1} \begin{pmatrix} d_{1n} \\ d_{2n} \\ \vdots \\ d_{mn} \end{pmatrix} = \begin{pmatrix} \ell_{11} \\ \ell_{21} \\ \vdots \\ \ell_{m1} \end{pmatrix} = \mathbf{L}_1 \in \mathbb{R}^m \end{aligned}$$

Jede Komponente ℓ_{i1} von $\mathbf{L}_1$, $i = 1, \ldots, m$, korrespondiert zu einem Fall.

Nach den Regeln der Multiplikation einer Matrix mit einem Vektor ist ℓ_{i1} das Skalarprodukt des Zeilenvektors $\tilde{\mathbf{d}}_i$ mit $\mathbf{a}_1$, d.h.

$$\tilde{\mathbf{d}}_i'\mathbf{a}_1 = \ell_{i1}, \quad 1 \leq i \leq m \tag{2.15}$$

Die Komponenten ℓ_{i1} von $\mathbf{L}_1$ können als Koordinaten der Fälle auf einer durch $\mathbf{L}_1$ definierten Achse interpretiert werden.

Es sei $\mathbf{a}_2 = (a_{21}, \ldots, a_{n2})'$ ein zweiter Vektor, dessen Komponenten wiederum als "Gewichte" aufgefasst werden können, um einen zweiten Vektor $\mathbf{L}_2$ als Linearkombination der $\mathbf{d}_j$ zu definieren:

$$D\mathbf{a}_2 = a_{12}\mathbf{d}_1 + \cdots + a_{n2}\mathbf{d}_n = \mathbf{L}_2 \tag{2.16}$$

Für die Komponente ℓ_{i2} von $\mathbf{L}_2$ gilt demnach

$$\tilde{\mathbf{d}}_i'\mathbf{a}_2 = \ell_{i2}, \quad 1 \leq i \leq m \tag{2.17}$$

$\mathbf{L}_2$ repräsentiert wie $\mathbf{L}_1$ ein Merkmal, das sich durch eine Linearkombination der $\mathbf{d}_j$, als Mischung der gemessenen Variablen darstellen oder charakterisieren läßt. Es seien L_1 und L_2 Geraden, deren Orientierung die $\mathbf{L}_1$ und $\mathbf{L}_2$ gegeben ist. Man kann ℓ_{i1} und ℓ_{i2} sind Koordinaten im Koordinatensystem (L_1, L_2) eines Punktes mit den Koordinaten (d_{i1}, d_{i2}) im ursprünglichen Koordinatensystem.

Es lassen sich nun Koeffizientenvektoren $\mathbf{b}_1$ und $\mathbf{b}_2$ finden derart, dass mit $L = [\mathbf{L}_1, \mathbf{L}_2]$

$$\begin{aligned} \mathbf{d}_1 &= L\mathbf{b}_1 \\ \mathbf{d}_2 &= L\mathbf{b}_2 \end{aligned}$$

so dass die Datenvektoren durch die "Latenten" Vektoren $\mathbf{L}_1$ und $\mathbf{L}_2$ erklärt werden (vergl. (2.13)). In Abschnitt 2.6 wird die Bestimmung der Matrizen $A = [\mathbf{a}_1, \mathbf{a}_2]$ und $B = [\mathbf{b}_1, \mathbf{b}_2]$ und die Beziehung zwischen diesen beiden Matrizen diskutiert werden. □

2.2.3 Multiplikation einer Matrix mit einer Matrix

Es seien A eine (m, n)- und B eine $(n \times p)$-Matrix. $\mathbf{b}_j$ sei der j-te Spaltenvektor von B. Dann ist

$$\mathbf{c}_j = A\mathbf{b}_j$$

eine Linearkombination der Spaltenvektoren von A mit den Komponenten von $\mathbf{b}_j$ als Koeffizienten. Schreibt man die Vektoren $\mathbf{c}_j$ spaltenweise nebeneinander, so entsteht eine $(m \times p)$-Matrix C, so so dass man insgesamt die Matrixgleichung

$$C = AB \tag{2.18}$$

erhält. Da die Komponenten von $\mathbf{c}_j$ die Skalarprodukte der Zeilenvektoren von A mit den Spaltenvektoren von B sind, hat man für die Komponenten c_{ij} von C die Beziehung

$$c_{ij} = \tilde{\mathbf{a}}_i'\mathbf{b}_j, \quad i = 1, \ldots, m;\ j = 1, \ldots, p \tag{2.19}$$

wobei $\tilde{\mathbf{a}}_i$ der i-te Spaltenvektor von A', also der i-te Zeilenvektor von A ist.

Es gilt:

1. Die Spaltenvektoren von C sind Linearkombinationen der Spaltenvektoren von A,

2. Die Zeilenvektoren von C sind Linearkombinationen der Zeilenvektoren von B; diese Aussage folgt aus Satz 2.12, p. 45.

Satz 2.2 *Es sei A eine (m,n)-Matrix, B sei eine $(n \times p)$, und C eine $(m \times p)$-Matrix, und es sei $C = AB$. Dann gilt*

$$C' = (AB)' = B'A' \tag{2.20}$$

Ist C eine $(p \times q)$-Matrix, so gilt

$$(AB)C = A(BC) \quad \textit{(Assoziativität)} \tag{2.21}$$

Beweis: Zu (2.20): Gilt $C = AB$, so sind die Zeilenvektoren von C Linearkombinationen der Zeilenvektoren von B, d.h. die Spaltenvektoren von C' sind Linearkombinationen der Spalten von B'. Die Spaltenvektoren von C sind Linearkombinationen der Spaltenvektoren von A, also sind die Zeilenvektoren von C' Linearkombinationen der Zeilenvektoren von A'. Damit muß $C' = B'A'$ gelten.

Zu (2.21): Es sei $U = AB$, $V = BC$, so dass $F = (AB)C = UC$, $G = A(BC) = AV$. U ist eine $(m \times p)$-Mastrix, V ist eine $(n \times q)$-Matrix, und F und G sind beide $(m \times q)$-Matrizen. Es seien $\mathbf{f}_j$ und $\mathbf{g}_j$ die j-ten Spaltenvektoren von F bzw. G. Offenbar gilt

$$\mathbf{f}_j = U\mathbf{c}_j, \quad \mathbf{g}_j = A\mathbf{v}_j, \quad j = 1, \ldots, q$$

Nun ist $\mathbf{v}_j = B\mathbf{c}_j$, so dass $\mathbf{g}_j = AB\mathbf{c}_j = U\mathbf{c}_j$ folgt, was wiederum $\mathbf{f}_j = \mathbf{g}_j$ und damit $F = G$ bedeutet, – und das ist (2.21). □

Die in (2.21) ausgedrückte Eigenschaft der Matrixmultiplikation wird als *assoziativ* bezeichnet. Sie ist aus der Multiplikation reeller Zahlen bekannt: Sind $0 \neq a, b, c \in \mathbb{R}$, so gilt $(ab)c = a(bc)$. Die Multiplikation reeller Zahlen ist auch *kommmutativ*, d.h. es gilt stets auch $ab = ba$.

Die Matrixmultiplikation ist aber im Allgemeinen *nicht* kommutativ, d.h. *im Allgemeinen* gilt für irgendzwei Matrizen A und B, wobei A eine (m,n)- und B eine $(n \times p)$-Matrix ist,

$$AB \neq BA. \tag{2.22}$$

Damit $AB = BA$ gilt müssen einige *notwendige*, wenn auch noch nicht *hinreichende* Bedingungen erfüllt sein. Zunächst einmal muß $p = m$ gelten, damit überhaupt das Produkt BA gebildet werden kann, für $p \neq m$ kann BA gar nicht berechnet werden und die Aussage $AB = BA$ ist sinnlos. Nun sei $p = m$, und es werde $U = AB$ und $V = BA$ gesetzt. U ist eine $(m \times p)$-Matrix und V ist eine (m,n)-Matrix, so dass $AB = BA$ impliziert, dass auch $n = p$ ist, d.h. $p \neq n$ impliziert, dass BA nicht gebildet werden kann. Eine notwendige Voraussetzung für (2.22) ist also $m = n = p$, d.h. A und B müssen quadratische Matrizen mit identischer Zeilen- und Spaltenzahl sein. Dann gilt $\mathbf{u}_j = A\mathbf{b}_j$ und $\mathbf{v}_j = B\mathbf{a}_j$, und die Beziehung $AB = BA$ impliziert $\mathbf{u}_j = \mathbf{v}_j$ für alle $j = 1, \ldots$. Diese Gleichheit *kann* für spezielle Matrizen gegeben sein, etwa falls $B = A^{-1}$ die Inverse von A ist (s. Abschnitt 2.5), oder wenn $B = A'$ und A eine orthonormale Matrix (Rotationsmatrix) ist, s. Abschnitt 2.6.1. Irgendzwei (m,n) Diagonalmatrizen A und B sind ein weiterer Spezialfall, für den $AB = BA$ gilt, wie man leicht nachrechnet.

Aber für beliebige Matrizen mit $m = n = p$ *muß* die Bedingung $AB = BA$ *nicht* erfüllt sein.

Längenskalierung: Multiplikation mit einer Diagonalmatrix. Es sei $X \in \mathbb{R}^{m,n}$, und es seien Λ und $\tilde{\Lambda}$ Diagonalmatrizen, also Matrizen, die nur in den Diagonalzellen von Null verschiedene Elemente haben:

$$\Lambda = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}, \quad \tilde{\Lambda} = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_m \end{pmatrix} \tag{2.23}$$

Man schreibt dafür auch kurz

$$\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n), \quad \tilde{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_m) \tag{2.24}$$

Dann bedeutet
(i) das Produkt $X\Lambda$ eine Längenskalierung der Spaltenvektoren von X,
(ii) das Produkt $\tilde{\Lambda}X$ eine Längenskalierung der Zeilenvektoren von X:

$$X\Lambda = \begin{pmatrix} \lambda_1 x_{11} & \lambda_2 x_{12} & \cdots & \lambda_n x_{1n} \\ \lambda_1 x_{21} & \lambda_2 x_{22} & \cdots & \lambda_n x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_1 x_{m1} & \lambda_2 x_{m2} & \cdots & \lambda_n x_{mn} \end{pmatrix}, \tag{2.25}$$

bzw.

$$\tilde{\Lambda}X = \begin{pmatrix} \lambda_1 x_{11} & \lambda_1 x_{12} & \cdots & \lambda_1 x_{1n} \\ \lambda_2 x_{21} & \lambda_2 x_{22} & \cdots & \lambda_2 x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_m x_{m1} & \lambda_m x_{m2} & \cdots & \lambda_m x_{mn} \end{pmatrix}. \tag{2.26}$$

Wie man sieht, ist der j-te Spaltenvektor von $X\Lambda$ gleich dem j-ten Spaltenvektor $\mathbf{x}_j$ von X, multipliziert mit λ_j, und der i-te Zeilenvektor von $\tilde{\lambda}X$ ist der i-te Zeilenvektor $\tilde{\mathbf{x}}_i$ von X multipliziert mit λ_i. Dies bedeutet eine Längenskalierung der $\mathbf{x}_j$ bzw. der $\tilde{\mathbf{x}}_i$.

Beispiel 2.4 In Beispiel 2.3, Seite 46 wurde die Multiplikation einer Matrix mit einem Vektor betrachtet: eine Datenmatrix D wurde mit jeweils einem Vektor $\mathbf{a}_k$ multipliziert, so dass ein Vektor $\mathbf{L}_k$ als Linearkombination der Spaltenvektoren von D erzeugt wurde. Man kann nun die Vektoren $\mathbf{a}_k$ zu einer Matrix

$$A = [\mathbf{a}_1, \mathbf{a}_2] = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \tag{2.27}$$

zusammenfassen. Dann gilt

$$DA = L = [\mathbf{L}_1, \mathbf{L}_2] \tag{2.28}$$

Die Spalten von A – also die Vektoren $\mathbf{a}_1$ und $\mathbf{a}_2$ sind spezifisch für die Spaltenvektoren von L, also von $\mathbf{L}_1$ und $\mathbf{L}_2$, und dass A durch nur zwei Vektoren definiert wurde, ist ein Resultat des in Beispiel 2.3 betrachteten Spezialfalls von nur zwei gemessenen Variablen, Körperlänge und Körpergewicht. Bei $n > 2$ gemessenen Variablen wird man eine (n, n)-Matrix A zur Bildung der Linearkombinationen $\mathbf{u}_k$ finden müssen; man hat ja

$$a_{1k}\mathbf{d}_1 + a_{2k}\mathbf{d}_2 + \cdots + a_{nk}\mathbf{d}_n = \mathbf{L}_k,$$

wobei die Komponenten u_{ik} die Koordinaten der Fälle auf dem k-ten durch eine Linearkombination definierten Merkmals M_k sind. Die Komponenten a_{jk} sind spezifisch für das j-te *gemessene* Merkmal und können als Koordinaten von Punkten interpretiert werden, die die j-te gemessene Variable repräsentieren. Die folgenden Abschnitte zielen u.a. auf die Frage, wie die Matrix A zu bestimmen ist.

□

2.3 Zufällige Vektoren und Matrizen

In der multivariaten Analyse werden die Beziehungen zwischen mehreren Variablen untersucht; sofern diese Variablen gemessen werden, werden sie als zufällige Veränderliche interpretiert. Ein Vektor, dessen Komponenten zufällige Veränderliche sind, heißt *zufälliger Vektor* (auch: *Zufallsvektor*), und eine Matrix, deren Elemente zufällige Veränderliche sind, heißt dementsprechend *zufällige Matrix* (auch: *Zufallsmatrix*). Algebraisch wird kein Unterschied zwischen zufälligen Vektoren und Matrizen und nichtzufälligen Vektoren und Matrizen gemacht; das Adjektiv "zufällig" bezieht sich eher auf die Interpretation der Vektoren und Matrizen.

Ein zufälliger Vektor wird etwa mit

$$\vec{X} = \mathbf{X} = (X_1, X_2, \ldots, X_n)' \tag{2.29}$$

bezeichnet, wobei die X_i eben zufällige Veränderliche bezeichnen. Analog dazu ist

$$X = \begin{pmatrix} X_{11} & X_{12} & \cdots & X_{1n} \\ X_{21} & X_{22} & \cdots & X_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ X_{m1} & X_{m2} & \cdots & X_{mn} \end{pmatrix} = (X_{ij}), \quad i = 1, \ldots, m; j = 1, \ldots, n \tag{2.30}$$

eine zufällige Matrix. Datenmatrizen sind Beispiele für zufällige Matrizen. Zufäälige Vektoren werden mit einer multivariaten Verteilung bzw. Dichtefunktion $f(x_1, \ldots, sx_n)$ assoziiert; in Abschnitt 4.7.3 wird insbesondere die multivariate Normalverteilung besprochen. Datenmatrizen sind *zeilen-* oder *spaltenzentriert*, wenn entweder die Zeilen- oder die Spaltenvektoren zentriert wurden. Repräsentieren die Spalten der Matrix Variablen und repräsentieren die Zeilen der Matrix

"Fälle" (Objekte oder Personen), an denen Messungen dieser Variablen vorgenommnen wurden, so sind die Skalarprodukte der Spaltenvektoren proportional zu Kovarianzen der entsprechenden Variablen. So sei die (m, n)-Datenmatrix X spaltenzentriert; im Allgemeinen werden die Elemente dieser Matrix dann mit kleinen Buchstaben bezeichnet: x_{ij} statt X_{ij}. $\mathbf{x}_j$ und $\mathbf{x}_k$ irgendzwei Spaltenvektoren von X. Dann ist

$$\mathbf{x}_j' \mathbf{x}_k = \sum_{i=1}^{m} x_{ij} x_{ik} \tag{2.31}$$

proportional zur geschätzten Kovarianz der Variablen V_j und V_k; der Proportionalitätsfaktor ist $1/m$ oder $1/(m-1)$.

Man betrachte nun das dyadische Produkt des *Zeilen*vektors $\tilde{\mathbf{x}}_i$ mit sich selbst:

$$\tilde{\mathbf{x}}_i \tilde{\mathbf{x}}_i' = \begin{pmatrix} x_{i1}x_{i1} & x_{i1}x_{i2} & \cdots & x_{i1}x_{in} \\ x_{21}x_{i1} & x_{i2}x_{i2} & \cdots & x_{i2}x_{in} \\ \vdots & \vdots & \ddots & \vdots \\ x_{in}x_{i1} & x_{ni}x_{i2} & \cdots & x_{ni}x_{in} \end{pmatrix} \tag{2.32}$$

Die Elemente dieser Matrix entsprechen den Produkten $x_{ij}x_{ik}$ in Gleichung (2.31). Dementsprechend ist die Summe über die dyadischen Produkte

$$\sum_{i=1}^{m} \tilde{\mathbf{x}}_i \tilde{\mathbf{x}}_i' = m\mathbf{\Sigma} = (\mathbf{x}_j' \mathbf{x}_k), \quad 1 \leq j, k \leq n \tag{2.33}$$

proportional zu $\mathbf{\Sigma}$, der Matrix der Skalarprodukte zwischen den Variablen, oder

$$\mathbf{\Sigma} = \frac{1}{m} \sum_{i=1}^{m} \tilde{\mathbf{x}}_i \tilde{\mathbf{x}}_i', \tag{2.34}$$

– bei kleinerem Stichprobenumfang m wird man durch $m-1$ dividieren.

Zufällige Veränderliche werden durch Verteilungsfunktionen und deren Ableitungen, den Dichtefunktionen beschrieben, sofern sie stetige Variable repräsentieren; im diskreten Fall sind die Dichtefunktionen Wahrscheinlichkeitsverteilungen. Man kann also annehmen, dass ein zufälliger Vektor durch eine n-dimensionale Verteilungsfunktion

$$F(x_1, \ldots, x_n) = P(X_1 \leq x_1, X \leq x_2, \ldots, X_n \leq x_n) \tag{2.35}$$

charakterisiert.

Im Folgenden sind insbesondere die Erwartungswerte, Varianzen und Kovarianzen von zufälligen Veränderlichen von Interesse. Der Erwartungswert einer zufälligen Veränderlichen ist der Mittelwert über *alle möglichen Realisationen* der Veränderlichen; er wird it $\mathbb{E}$ bezeichnet:

$$\mathbb{E}(X) = \begin{cases} \int_{-\infty}^{\infty} x f(x) dx, & X \text{ ist stetig} \\ \\ \sum_i p_i X_i, & X \text{ ist diskret} \end{cases} \tag{2.36}$$

$\mathbb{E}$ ist ein *linearer Operator*, d.h. es gilt

$$\mathbb{E}(\sum_{i=1}^{n} a_i X_i) = \sum_{i=1}^{n} a_i \mathbb{E}(X_i); \tag{2.37}$$

diese Aussage folgt sofort aus der Definition von $\mathbb{E}$.

Der Erwartungswert eines Zufallsvektors ist ein Vektor, dessen Komponenten die Erwartungswerte der Komponenten des Vektors sind:

$$\mathbb{E}(\mathbf{X}) = (\mathbb{E}(X_1), \mathbb{E}(X_2), \ldots, \mathbb{E}(X_n))'. \tag{2.38}$$

Oft wird $\mathbb{E}(\mathbf{X})$ mit $\boldsymbol{\mu}$ bezeichnet:

$$\boldsymbol{\mu} = \mathbb{E}(\mathbf{X}) = (\mu_1, \mu_2, \ldots, \mu_n)', \tag{2.39}$$

wobei $\mu_i = \mathbb{E}(X_i)$ ist.

Mit

$$\mathbf{X} - \boldsymbol{\mu} = (X_1 - \mu_1, X_2 - \mu_2, \ldots, X_n - \mu_n)' \tag{2.40}$$

wird der Vektor der Abweichungen vom jeweiligen Erwartungswert bezeichnet. Der Erwartungswert des dyadischen Produkts

$$\boldsymbol{\Sigma} = \mathbb{E}[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})'] = (\mathbb{E}[(X_i - \mu_i)(X_j - \mu_j)']) = (\sigma_{ij}) \tag{2.41}$$

ist die Matrix der Kovarianzen zwischen den Komponenten eines zufälligen Vektors; die Diagonalelemente σ_{ii} von $\boldsymbol{\Sigma}$ sind die Varianzen der Komponenten von $\mathbf{X}$:

$$\sigma_{ii} = \sigma_i^2 = \mathbb{E}[(X_i - \mu_i)^2] \tag{2.42}$$

Satz 2.3 *Es sei $\boldsymbol{Z} = (Z_1, \ldots, Z_n)'$ ein zufälliger Vektor, dessen Komponenten paarweise unabhängig sind mit $\mathbb{E}(Z_i) = 0$, $Var(Z_i) = 1$, $1 \leq i \leq n$. Weiter sei A eine (n, n)-Matrix und $\boldsymbol{X} = A\boldsymbol{Z} + \boldsymbol{\mu}$. Dann ist $\boldsymbol{\Sigma} := AA'$ die Matrix der Kovarianzen zwischen den Komponenten von $\boldsymbol{X}$.*

Beweis: Es ist $A\mathbf{Z} = \mathbf{X} - \boldsymbol{\mu}$ und

$$\mathbb{E}[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})'] = \boldsymbol{\Sigma} = \mathbb{E}[AZZ'A'] = A\mathbb{E}(ZZ')A' = AA',$$

wegen

$$\mathbb{E}(ZZ') = (\mathbb{E}(Z_i Z_j)) = I,$$

I die Einheitsmatrix, denn wegen der postulierten Unabhängigkeit der Komponenten von $\mathbf{Z}$ ist $\mathbb{E}(Z_i Z_j) = 0$ für $i \neq j$ und $\mathbb{E}(Z_i Z_i) = 1$, ebenfalls nach Voraussetzung. □

2.4 Der Rang einer Matrix

Auf Seite 41 ist der Begriff des Ranges eines Vektorraums bzw. eines Teilraums eines Vektorraums eingeführt worden. Es sei nun $X = [\mathbf{x}_1, \ldots, \mathbf{x}_n]$ eine (m, n)-Matrix. Die linearen Hüllen $\mathcal{L}_s = \mathcal{L}(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ und $\mathcal{L}_r(\tilde{\mathbf{x}}_1, \ldots, \tilde{\mathbf{x}}_m)$ der Spalten- bzw. Zeilenvektoren sind Teilvektorräume mit Basen, die $s \leq n$ bzw. $r \leq m$ Vektoren enthalten (vergl. Definition 1.11, Seite 33). s heißt *Spaltenrang* von X, und r heißt der *Zeilenrang* von X.

Es sei $U = [\mathbf{u}_1, \ldots, \mathbf{u}_s]$ eine Matrix, deren m-dimensionale Spaltenvektoren $\mathbf{u}_j$ eine Basis von $\mathcal{L}_s$ bilden, und es sei $V = [\mathbf{v}_1, \ldots, \mathbf{v}_r]$ eine Matrix, deren n-dimensionale Vektoren $\mathbf{v}_k$ eine Basis von $\mathcal{L}_r$ bilden[19]. Die Spaltenvektoren $\mathbf{x}_j$ von X können dann als Linearkombinationen der Spaltenvektoren von U und die Zeilenvektoren $\tilde{\mathbf{x}}_i$ können als Linearkombinationen der Spaltenvektoren von V dargestellt werden:

$$\mathbf{x}_j = a_{1j}\mathbf{u}_1 + a_{2j}\mathbf{u}_2 + \cdots + a_{sj}\mathbf{u}_s. \tag{2.43}$$

$$\tilde{\mathbf{x}}_i = b_{1i}\mathbf{v}_1 + b_{2i}\mathbf{v}_2 + \cdots + b_{ri}\mathbf{v}_r \tag{2.44}$$

Die Koeffizienten $a_{1j}, \ldots, a_{sj}$ können zu s-dimensionalen Vektoren $\mathbf{a}_j$ zusammengeffasst werden und die $b_{1i}, \ldots, b_{ri}$ zu r-dimensionalen Vektoren $\mathbf{b}_i$. Nach Satz 1.6, S. 25 liegen die Koeffizienten a_{kj} und b_{ji} wegen der linearen Unabhängigkeit der $\mathbf{u}_k$ und $\mathbf{v}_k$ eindeutig fest. Fasst man die $\mathbf{a}_j$ und $\mathbf{b}_i$ jeweils zu einer (s, n)-dimensionalen Matrix A und zu einer (r, m)-dimensionalen Matrix B zusammen, so entsprechen den Gleichungen (2.43) und (2.44) die Matrixgleichungen

$$X = UA \tag{2.45}$$

$$X' = VB \tag{2.46}$$

Aus der Definition der Matrixmultiplikation (vergl. die Kommentare zur Matrixmultiplikation 1. und 2.) folgt, dass die Zeilenvektoren von X als Linearkombinationen der s Zeilenvektoren von A aufgefasst werden können, und die Zeilenvektoren von X' – also die Spalten von X – sind als Linearkombinationen der r Zeilenvektoren von B repräsentierbar. Dieser Sachverhalt legt nahe, dass es einen engen Zusammenhang zwischen dem Zeilen- und dem Spaltenrang von X gibt. Tatsächlich gilt der folgende

Satz 2.4 *Es sei X eine (m, n)-Matrix mit dem Zeilenrang r und dem Spaltenrang s. Dann gilt*

$$r = s \tag{2.47}$$

d.h. der Zeilenrang ist stets gleich dem Spaltenrang, sowie

$$X = UV', \tag{2.48}$$

[19] Zur Abkürzung wird im Folgenden eine Matrix, deren Spalten- oder Zeilenvektoren die Rolle einer Basis haben, selbst als Basis bezeichnet.

wobei $U \in \mathbb{R}^{m,r}$ und $V \in \mathbb{R}^{n,r}$ Matrizen mit dem Rang r sind. Weiter gilt

$$r \leq \min(m, n) \tag{2.49}$$

Beweis: Nach Voraussetzung gelten die Gleichungen (2.45) und (2.46), also $X = UA$ und $X' = VB$, sobei U eine (m, s)- und A eine (s, n)-Matrix ist, und V ist eine (n, r)- und B eine (r, m)-Matrix. Es gilt aber auch, dass die Zeilenvektoren von X (also die Spalten von X') Linearkombinationen der s Zeilenvektoren von A sind. Aber der Zeilenrang von X ist r, also kann s nicht kleiner als r sein, d.h. es muß $r \leq s$ gelten.

Umgekehrt gilt $X' = VB$, d.h. die Spalten von X' (die Zeilen von X) sind Linearkombinationen von r linear unabhängigen Spaltenvektoren von V, und die Zeilen von X' (die Spalten von X) sind Linearkombinationen der r Zeilenvektoren von B, aber sie sind auch Linearkombinationen der s linear unabhängigen Spaltenvektoren von U. s kann dann nicht kleiner als r sein, d.h. es muß $r \geq s$ gelten. Zusammen mit dem Befund $r \leq s$ folgt dann $r = s$, – der Zeilenrang ist gleich dem Spaltenrang.

Also existiert eine Matrix U mit dem Spaltenrang $s = r$ und eine (r, n)-Matrix A, so dass $X = UA$. Da $X' = A'U'$ sind die Spaltenvektoren von X' Linearkombinationen der r Spaltenvektoren von A', und es folgt, dass die Spaltenvektoren von A' linear unabhängig sind, denn wären sie linear abhängig, könnten nicht alle Spalten von X' als Linearkombinationen der Spalten von A' rerpräsentiert werden. Wegen $X' = VB$ kann man also $V = A'$ setzen, bzw. $V' = A$, so dass man zu der Gleichung $X = UV'$ gelangt, wobei U und V denselben Rabg haben.

Es sei $m > n$. Da der Zeilenrang gleich dem Spaltenrang ist kann der Rang der Matrix nicht größer als n sein, so dass $r \leq \min(m, n)$. Die Argumentation für den Fall $m \leq n$ ist analog, d.h. es gilt (2.49). □

Anmerkungen:
1.Der Beweis wird üblicherweise durch Anwendung elementarer Umformungen auf X geführt, s. Abschnitt 4.1.
2. Die (m, n)-Matrix A hat den Rang $r = 0$ dann und nur dann, wenn $A = 0$ die Nullmatrix ist, d.h. wenn $a_{ij} = 0$ für alle i und j ist. □

Definition 2.1 *Es seien $\boldsymbol{x}_j$, $j = 1, \ldots, n$ die Spaltenvektoren von X, und $\tilde{\boldsymbol{x}}_i$, $i = 1, \ldots, m$, seien die Zeilenvektoren von X (als Spaltenvektoren von X' aufgefasst). Dann heißt die lineare Hülle $\mathcal{L}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$ der* Spaltenraum *von X, und die lineare Hülle $\mathcal{L}(\tilde{\boldsymbol{x}}_1, \ldots, \tilde{\boldsymbol{x}}_m)$ heißt* Zeilenraum *von X.*

Anmerkungen:
1. Die jeweils r Spaltenvektoren von U und V sind linear unabhängig; sie sind die Basis des Spaltenraums bzw. des Zeilenraums von X.
2. Für $r = \min(m, n)$ sagt man, X habe den *vollen Rang.*

3. Aus dem Beweis zu Satz 2.4 ergibt sich, dass sich *jede* (m, n)-Matrix X als Produkt der Form (2.48) darstellen läßt.
4. Da mit der Wahl von U die Matrix V' eindeutig festliegt und umgekehrt mit der Wahl von V' die Wahl von U eindeutig festliegt stellt sich die Frage, ob eine der beiden Matrizen berechnet werden kann, wenn man die andere gewählt hat. Das ist möglich, wenn man etwa U so wählt, dass die Spaltenvektoren von U orthonormal sind: dann gilt $U'U = I_r$, I_r die (r, r)-Einheitsmatrix, und man hat

$$U'X = U'X = U'UV' = V', \tag{2.50}$$

d.h. die Zeilenvektoren von V' (Spaltenvektoren von V, eine (n, r)-Matrix) ergeben sich als Linearkombinationen der Zeilenvektoren von X. Oder man habe die Spaltenektoren von V als orthonormal gewählt, so dass $V'V = I_r$. Dann folgt

$$XV = UV'V = U, \tag{2.51}$$

d.h. die Spaltenvektoren von U ergeben sich als Linearkombinationen der Spaltenvektoren von X. Die in Abschnitt 2.6.6 eingeführte Singularwertzerlegung von X ist ein Spezialfall von (2.51) bzw. (2.51), der vielen faktorenanalytischen Ansätzen zur Interpretation von Datenmatrizen zugrunde liegt. □

Der folgende Satz über den Rang eines dyadischen Produkts erweist sich als nützlich:

Satz 2.5 *Es sei $\boldsymbol{x}$ ein n-dimensionaler und $\boldsymbol{y}$ ein m-dimensionaler Vektor. Das dyadische Produkt $\boldsymbol{xy'}$ ist eine (m, n)-Matrix mit dem Rang 1.*

Beweis: Die Inspektion der Matrix $\mathbf{xy'}$ (vergl. (1.20), (S. 16) zeigt, dass die Zeilen von $\mathbf{xy'}$ alle proportional zu $\mathbf{y}$ sind, d.h. Elemente eines 1-dimensionalen Teilraums des $\mathbb{R}^m$ sind. Analog dazu sind die Spaltenvektoren alle proportional zu $\mathbf{x}$, d.h. sie sind Elemente eines 1-dimensionalen Teilraums des $\mathbb{R}^n$. Somit hat $\mathbf{xy'}$ den Rang 1. □

Korollar 2.1 *Aus (2.48) folgt, dass jedes Element x_{ij} von X als Skalarprodukt*

$$x_{ij} = \tilde{\boldsymbol{u}}_i' \tilde{\boldsymbol{v}}_j \tag{2.52}$$

dargestellt werden kann, wobei $\tilde{\boldsymbol{u}}_i$ der i-te Spaltenvektor von U' (Zeilenvektor von U) und $\tilde{\boldsymbol{v}}_j$ der j-te Spaltenvektor von V' (Zeilenvektor von V) ist.

Beweis: (2.52) ist eine unmittelbare Folgerung aus der Gleichung $X = UV'$ (vergl. auch Gleichung (2.19), Seite 47). □

Satz 2.6 *Es sei A eine (m, n)-Matrix. Es gelten die folgenden Aussagen:*
1. B sei eine $(n \times p)$-Matrix. Dann gilt

$$rg(AB) \leq \min[rg(A), rg(B)]. \tag{2.53}$$

2. $Q(m,m)$ und $P(n,n)$ haben jeweils vollen Rang, d.h. $rg(Q) = m$, $rg(P) = n$. Dann gilt

$$rg(PAQ) = rg(A), \tag{2.54}$$

d.h. das Produkt von A von links mit einer Matrix mit vollem Rang und die Multiplikation von A von rechts mit einer Matrix mit vollem Rang hat denselben Rang wie A.
3. A und B seien zwei (m,n)-Matrizen. Dann gilt

$$rg(A+B) \leq rg(A) + rg(B) \tag{2.55}$$

Beweis: Zu 1. A habe den Rang s. Die Spaltenvektoren von $C = AB$ sind Linearkombinationen der Spaltenvektoren von A, d.h. $C \subseteq \mathcal{L}(A) = \mathcal{L}(\mathbf{a}_1, \ldots, \mathbf{a}_s)$, $\mathbf{a}_1, \ldots, \mathbf{a}_s$ eine Basis der linearen Hülle $\mathcal{L}(A)$ von A, so dass C höchstens den Spaltenrang s hat. Die Matrix B habe den Rang r, und die Zeilenvektoren von C sind Linearkombinationen der Zeilenvektoren von B, so dass $C' \subseteq \mathcal{L}(B') = \mathcal{L}(\mathbf{b}_1, \ldots, \mathbf{b}_r)$, $\mathbf{b}_1, \ldots, \mathbf{b}_r$ eine Basis von $\mathcal{L}(B')$, so dass C höchstens den Rang r haben kann. Es sei $r \leq s$; dann hat C höchstens den Rang r, denn nach Satz 2.47 sind dann auch die Spaltenvektoren von C als Linearkombinationen von maximal r linear unabhängigen Vektoren darstellbar. Sei umgekehrt $s \leq r$; analog zur vorangegehenden Argumentation ist dann ist der Rang von C höchstens gleich s, d.h. $\mathrm{rg}(C) \leq \min[\mathrm{rg}(A), \mathrm{rg}(B)]$.

Zu 2. Es ist $\mathrm{rg}(A) = r \leq \min(m,n)$. Es sei $B = PA$. Aus 1. folgt

$$\mathrm{rg}(B) \leq \min[\mathrm{rg}(P), \mathrm{rg}(A)] = \min(m,r) = r \Rightarrow \mathrm{rg}(PA) \leq \mathrm{rg}(A).$$

Andererseits ist $P^{-1}B = A$, und wieder nach 1. folgt

$$\mathrm{rg}(P^{-1}B) = \mathrm{rg}(A) \leq \min(m, \mathrm{rg}(B)) = \mathrm{rg}(B),$$

d.h.

$$\mathrm{rg}(PA) \leq \mathrm{rg}(A) \leq \mathrm{rg}(PA) \Rightarrow \mathrm{rg}(PA) = \mathrm{rg}(A).$$

Es sei $C = AQ$; auf analoge Weise folgt $\mathrm{rg}(C) = \mathrm{rg}(AQ) = \mathrm{rg}(A)$. Dann folgt aber auch $\mathrm{rg}(A) = \mathrm{rg}(PAQ)$.
Zu 3. Der Rang einer Matrix ist gleich der Dimension des von den Spalten der Matrix aufgespannten Teilraums, d.h. $\mathrm{rg}(A) = \dim \mathcal{L}(A)$. Es sei $C = A + B$; dann ist die j-te Spalte $\mathbf{c}_j$ von C durch $\mathbf{a}_j + \mathbf{b}_j$ gegeben, $\mathbf{a}_j$ die j-te Spalte von A und $\mathbf{b}_j$ die j-te Spalte von B. $\mathbf{c}_j$ ist ein Element der Summe der Teilräume $\mathcal{L}(A)$ und $\mathcal{L}(B)$. Für die Dimensionalität der Summe zweier Teilräume gilt nach (2.55)

$$\dim(\mathcal{L}(A) + \mathcal{L}(B)) \leq \dim(\mathcal{L}(A)) + \dim(\mathcal{L}(B)),$$

und das heißt (2.55). □

Korollar 2.2 *Es sei A eine (m,n)-Matrix mit dem Rang $rg(A) \leq \min(m,n)$, und P sei eine (m,m)-Matrix mit $rg(P) = m$. Dann gilt $rg(PA) = rg(A)$. Ist Q eine (m,n)-Matrix mit dem Rang $rg(Q) = n$, so folgt $rg(AQ) = rg(A)$.*

Beweis: Die Aussagen folgen direkt aus (2.54) für entweder $Q = I_n$ oder $P = I_m$, (vergl. auch den Beweis zu Satz 2.6, 3.). □

Definition 2.2 *Es sei A eine (m,n)-Matrix. Die Menge $\mathcal{N}(A) = \{\boldsymbol{x}|A\boldsymbol{x} = 0, \boldsymbol{x} \in \mathbb{R}^n\}$ heißt* Nullraum *oder* Kern *von A. Die Dimensionalität des Nullraums wird auch* Nullität *(nullity) genannt.*

$\mathcal{N}$ ist ein Vektorraum, denn für $\mathbf{x}_1$ und $\mathbf{x}_2$ aus $\mathcal{N}$ folgt

$$A\alpha_1\mathbf{x}_1 + A\alpha_2\mathbf{x}_2 = \alpha_1 A\mathbf{x}_1 + \alpha_2 A\mathbf{x}_2 = 0 + 0 = 0.$$

Beispiel 2.5 Es sei A eine(m,n)-Matrix mit dem Rang n, d.h, die Spaltenvektoren von A seien linear unabhängig. Dann gilt $A\mathbf{x} = \vec{0}$ genau dann, wenn $\mathbf{x} = \vec{0}$, und $\mathcal{N}(A) = \{\vec{0}\}$. Die Gleichung $A\mathbf{x} = \mathbf{y}$ hat dann eine (falls $\mathbf{y} \in \mathcal{L}(A)$) oder gar keine Lösung (falls $\mathbf{y} \notin \mathcal{L}(A)$). □

Satz 2.7 *Es sei A eine (m,n)-Matrix vom Rang $r \leq \min(m,n)$. Es sei s der Rang von $\mathcal{N}(A)$. Dann gilt*

$$r + s = \min(m,n). \tag{2.56}$$

Beweis: Die Behauptung des Satzes kann auf die Aussage über orthogonale Komplemente (Definition 1.14, Seite 39) und Satz 1.15, Seite 39, zurückgeführt werden. Denn sei V das orthogonale Komplement von $\mathcal{N}(A)$. Gemäß (1.79), Seite 39, gilt dann die Beziehung (2.56). □

Anmerkung: Es gelte inbesondere $r = \min(m,n)$; dann folgt aus (2.56) $s = 0$, d.h. der Nullraum hat den Rang 0. Der Nullraum enthält dann nur den Nullvektor $\vec{0}$ (s. Anmerkung 3. auf Seite 54). Für $\min(m,n) = n$ heißt dies, dass $A\mathbf{x} = \vec{0}$ nur die Lösung $\mathbf{x} = \vec{0}$ hat, und dies ist der Fall, wenn die Spaltenvektoren linear unabhängig sind. Für den Fall $m = \min(m,n)$ gilt die analoge Argumentation für $\mathbf{x}'A = \vec{0}'$, $\mathbf{x} \in \mathbb{R}^m$. □

Kovarianz- und Korrelationsmatrizen spielen in der multivariaten Statistik eine zentrale Rolle; sie werden oft auch *Kreuzproduktmatrizen* genannt. Ist X eine (m,n)-Matrix, bei der (wie üblich) die Zeilen "Fälle" repräsentieren und die Spalten die gemessenen Variablen, so ist $(1/m)X'X$ gleich der Kovarianzmatrix oder der Korrelationsmatrix für die Variablen, je nachdem, ob nur eine Spaltenzentrierung oder eine Spaltenstandardisierung der Datenmatrix vorgenommen wurde; eine analoge Aussage gilt für XX'. Die folgende Aussage gilt für beliebige (m,n)-Matrizen X.

Satz 2.8 *Es sei X eine (m,n)-Matrix mit dem Rang $rg(X) = r \leq \min(m,n)$. Dann gilt für den Rang der Kreuzproduktmatrizen*

$$rg(X'X) = rg(XX') = rg(X). \tag{2.57}$$

Beweis:[20] (i) $\mathrm{rg}(X) = \mathrm{rg}(X'X)$. Es sei $\mathbf{u} \in \mathbb{R}^n$ und es gelte $X\mathbf{u} = \vec{0}$, so dass $\mathbf{u} \in \mathcal{N}(X)$, $\mathcal{N}(X)$ der Nullraum von X. Dann gilt auch $X'X\mathbf{u} = \vec{0}$, so dass ebenfalls $\mathbf{u} \in \mathcal{N}(X'X)$. Sei nun $\mathbf{v} \in \mathcal{N}(X'X)$, d.h. $X'X\mathbf{v} = \vec{0}$. Dann folgt $\mathbf{v}'X'X\mathbf{v} = \|X\mathbf{v}\|^2 = 0$, woraus $X\mathbf{v} = \vec{0}$ und damit $\mathbf{v} \in \mathcal{N}(X)$ folgt. Also haben X und $X'X$ denselben Nullraum und damit, nach Satz 2.7, denselben Rang.

(ii) $\mathrm{rg}(X) = \mathrm{rg}(XX')$: analog. □

2.5 Die Inverse einer Matrix

Es sei A eine (m,n)-Matrix, und $\mathbf{x} \in \mathbb{R}^n$; betrachtet werde die Linearkombination $\mathbf{y} = A\mathbf{x}$. Sie kann als Transformation $\mathbf{x} \to \mathbf{y}$ des Vektors $\mathbf{x}$ in den Vektor $\mathbf{y}$ gedeutet werden. Bestimmte Transformationen wie etwa die Rotation eines Vektors um einen bestimmten Winkel können invertiert werden, d.h. man kann für die Transformation $\mathbf{x} \to \mathbf{y}$ eine Transformation $\mathbf{y} \to \mathbf{x}$ finden. Dies bedeutet, eine Matrix B zu finden derart, dass $\mathbf{x} = B\mathbf{y}$. Die Frage nach der zu $A\mathbf{x} = \mathbf{y}$ inversen Transformation ist dann die Frage nach der Existenz einer Matrix B derart, dass $B\mathbf{y} = \mathbf{x}$. Substituiert man hier für $\mathbf{y}$ den Ausdruck $A\mathbf{x}$, so erhält man die Beziehung $BA\mathbf{x} = \mathbf{x}$, so dass man $BA = I_n$ erhält, I_n die (n,n)-Einheitsmatrix. Existiert B, so heißt B eine *Linksinverse*. Analog dazu heißt C eine *Rechtsinverse*

Satz 2.9 *Es sei A eine (m,n)-Matrix. Notwendig für die Existenz der Linksinversen B von A ist, dass A den vollen Spaltenrang n hat, und notwendig für die Existenz einer Rechtsinversen C von A ist, dass A vollen Zeilenrang m hat.*

Beweis: Es gelte $A\mathbf{x} = \mathbf{y}$, $\mathbf{x} = B\mathbf{y}$, $BA = I_n$. Notwendig für die Eindeutigkeit von $\mathbf{x}$ ist nach Satz 1.6, Seite 25)), dass die Spaltenvektoren von A linear unabhängig sind, d.h. A muß vollen Spaltenrang haben. Weiter sei $A'\mathbf{u} = \mathbf{v}$, $\mathbf{u} \in \mathbb{R}^m$, $\mathbf{v} \in \mathbb{R}^b$. Notwendig für die Existenz von $\mathbf{u}$ ist der volle Spaltenrang von A', d.h. der volle Zeilenrang von A. □

Ist $m > n$, so sind die Zeilenvektoren von A notwendig linear abhängig und es existiert keine Rechtsinverse für A. Ist $m < n$, so sind die Spaltenvektoren von A notwendig linear abhängig und es existiert keine Linksinverse für A. Für den Fall $m = n$ ergibt sich der Satz

Satz 2.10 *Existieren für eine Matrix A sowohl die Linksinverse B wie auch die Rechtsinverse C, so folgt, dass A eine quadratische Matrix ist, d.h. es ist $m = n$, und es gilt*

$$B = C. \tag{2.58}$$

[20] Seber, p. 385

Beweis: Es sei $m \neq n$; dann gilt $BA = I_n$ und die Rechtsinverse C existiert nicht, oder es gilt $AC = I_m$ und die Linksinverrse existiert nicht, je nachdem, ob $m > n$ oder $m < n$. Existieren sowohl B wie auch C, so folgt, dass die Bedingung $m \neq n$ nicht erfüllt sein kann, also muß $m = n$ gelten[21].

Weiter gilt

$$B = BI_m = B\underbrace{AC}_{I_m} = \underbrace{BA}_{I_n}C = C,$$

d.h. es gilt (2.58). □

Definition 2.3 *Es sei A eine (n, n)-Matrix mit vollem Rang n. Dann heißt $A^{-1} = B = C$ die zu A* inverse Matrix*, wobei B die Links- und C die Rechtsinverse von A ist.*

Folgerung: Wegen $A^{-1} = B = C$ folgt sofort

$$A^{-1}A = AA^{-1} = I_n, \tag{2.59}$$

I_n die (n, n)-Einheitsmatrix. □

Im Beweis für die Aussage (2.58) wurden keine expliziten Ausdrücke für B und C gegeben, weshalb der folgende alternative Beweis von Interesse sein mag:

> Es gelte $BA = I_n$, B eine (m, n)-Matrix. Die Multiplikation von rechts mit A' liefert $BAA' = A'$. Da A vollen Spaltenrang hat, hat die (m, m)-Matrix AA' den Rang m, so dass die Inverse $(AA')^{-1}$ existiert, so dass die Multiplikation von rechts mit $(AA')^{-1}$ auf
>
> $$B = A'(AA')^{-1} \tag{2.60}$$
>
> führt. Weiter gelte $AC = I_m$, C eine (n, m)-Matrix. Multiplikation von links mit A' liefert $A'AC = A'$. Da voller Zeilenrang von A vorausgesetzt wird, hat die (n, n)-Matrix $A'A$ den Rang n, so dass die Inverse $(A'A)^{-1}$ existiert, und die Multiplikation von links mit $(A'A)^{-1}$ liefert
>
> $$C = (A'A)^{-1}A'. \tag{2.61}$$
>
> Für den Fall, dass A eine (n, n)-Matrix ist, existieren sowohl B wie auch C. Es soll gezeigt werden, dass $B = C$. Dazu werde von $A' = A'$ ausgegangen. Dann folgt
>
> $$A'(AA')(AA')^{-1} = A'A(A'(AA')^{-1}) = A'$$
>
> Multplikation mit $(A'A)^{-1}$ liefert
>
> $$A'(AA')^{-1} = (A'A)^{-1}A',$$
>
> also $B = C$. □

[21] Dieser Schluß ist eine Anwendung des *modus tollens*: sind p und q Aussagen und gilt $p \to q$, so folgt $\neg q \to \neg p$, $\neg p$ = nicht-p, $\neg q$ = nicht-q.

Spezialfall: Die (n,n)-Matrix A sei orthonormal. Dann gilt $A'A = AA' = I$, und mithin $A^{-1} = A'$. □

Beispiel 2.6 Es sei

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

und gesucht ist

$$A^{-1} = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}.$$

Es ist

$$AA^{-1} = \begin{pmatrix} a\alpha + b\gamma & a\beta + b\delta \\ c\alpha + d\gamma & c\beta + d\delta \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Man hat also das Gleichungssystem

$$a\alpha + b\gamma = 1 \Rightarrow \alpha = \frac{1 - b\gamma}{a} \tag{2.62}$$

$$a\beta + b\delta = 0 \Rightarrow \beta = -\frac{b\delta}{a} \tag{2.63}$$

$$c\alpha + d\gamma = 0 \Rightarrow \gamma = -\frac{c\alpha}{d} \tag{2.64}$$

$$c\beta + d\delta = 1 \Rightarrow \delta = \frac{1 - c\beta}{d} \tag{2.65}$$

Durch Einsetzen etwa des in Gleichung (2.64) gegebenen Ausdrucks für γ in die Gleichung (2.62) etc findet man schließlich

$$A^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}. \tag{2.66}$$

Man überprüft durch Nachrechnen, dass in der Tat $AA^{-1} = A^{-1}A = I$ gilt , – vorausgesetzt, dass $ad - bc \neq 0$ ist.

Es werde nun angenommen, dass $ad - bc = 0$ ist, d.h. es gelte $ad = bc$. Dann ist der Ausdruck $1/(ad - bc)$ in (2.66) nicht definiert, d.h. A^{-1} existiert nicht. Man hat dann einerseits

$$b = a\frac{d}{c}, \quad d = c\frac{b}{a},$$

andererseits folgt auch

$$\frac{d}{c} = \frac{b}{a} =: \lambda,$$

d.h. für den Spaltenvektor $(b,d)'$ von A gilt

$$\begin{pmatrix} b \\ d \end{pmatrix} = \lambda \begin{pmatrix} a \\ c \end{pmatrix},$$

die Spaltenvektoren von A sind linear abhängig. Man rechnet leicht nach, dass umgekehrt die lineare Abhängigkeit der Spaltenvektoren die Gleichung $ad - bc =$

0 impliziert. Die Voraussetzung $ad - bc \neq 0$ gilt also genau dann, wenn die Spalten- und damit auch die Zeilenvektoren von A linear unabhängig sind. Es sei angemerkt, dass $ad - bc = |A|$ die Determinante der Matrix A ist (vergl. (1.58), S. 27) ist; allgemein gilt, dass die Inverse einer quadratischen Matrix nur dann existiert, wenn die Determinante der Matrix ungleich Null ist, und dies ist der Fall, wenn A vollen Rang hat. □

Beispiel 2.7 Es sei R eine (2×2)-Korrelationsmatrix,

$$R = \begin{pmatrix} 1 & r \\ r & 1 \end{pmatrix}. \tag{2.67}$$

Die Inverse R^{-1} ergibt sich aus (2.66)

$$R^{-1} = \begin{pmatrix} \frac{1}{1-r^2} & -\frac{r}{1-r^2} \\ -\frac{r}{1-r^2} & \frac{1}{1-r^2} \end{pmatrix} = \frac{1}{1-r^2} \begin{pmatrix} 1 & -r \\ -r & 1 \end{pmatrix} \tag{2.68}$$

R^{-1} heißt auch *Präzisionsmatrix*. Der Ausdruck wird klar, wenn man R^{-1} für $r \to 0$ und $r \to 1$ bzw $r \to -1$ betrachtet. Offenbar ist

$$\lim_{r \to 0} R^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \ \lim_{r \to 1} R^{-1} = \begin{pmatrix} \infty & -\infty \\ -\infty & \infty \end{pmatrix} \tag{2.69}$$

Für $r \to -1$ ändert sich das Vorzeichen der Elemente neben der Diagonalen. Aus der Regressionsrechnung ist bekannt, dass für $y = bx + a + e$

$$r_{xy}^2 = r^2 = 1 - \frac{s_e^2}{s_y^2},$$

d.h. $r \to 1$, wenn $s_e^2 \to 0$. Ein kleiner Wert für die Fehlervarianz bedeutet größere Präzision der Vorhersage von y, und dies drückt sich in einem betragsmäßig großen Wert der Elemente von R^{-1} aus. Für $s_e^2 \to s_y^2$ folgt $r \to 0$ und die Präzision der Vorhersage geht gegen Null. □

Beispiel 2.8 Es werde der Fall einer Matrix A mit den Zeilen $(1, 3)$ und $(2, 1)$ betrachtet; gesucht ist die zu A inverse Matrix A^{-1}:

$$AA^{-1} = \begin{pmatrix} 1 & 3 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Die Spalten von A sind linear unabhängig, denn die Spaltenvektoren sind offenbar nicht parallel. Es müssen die Elemente a, b, c und d von A^{-1} bestimmt werden. Man erhält zwei Gleichungssysteme:

$$\begin{array}{ll} 1 \cdot a + 3 \cdot c = 1 & 1 \cdot b + 3 \cdot d = 0 \\ 2 \cdot a + 1 \cdot c = 0 \ , & 2 \cdot b + 1 \cdot d = 1 \end{array}$$

Dann ist

$$A^{-1} = \begin{pmatrix} -1/5 & 3/5 \\ 2/5 & -1/5 \end{pmatrix} = -\frac{1}{5} \begin{pmatrix} 1 & -3 \\ -2 & 1 \end{pmatrix},$$

und man rechnet leicht nach, dass $AA^{-1} = A^{-1}A = I$ ist. □

Anmerkung: Es sei A eine (n, r)-Matrix mit $r < n$, und die Spaltenvektoren von A seien orthonormal. Dann ist $AA' = I_n$, d.h. AA' ist eine $(n \times n)$-Matrix. $A'A$ dagegen ist eine (r, r)-Matrix, so dass sicherlich $A'A \neq AA'$. Die inverse Matrix existiert in diesem Fall nicht, weil A nicht quadratisch ist und AA' überdies singulär ist, denn $\text{rg}(AA') < n$. □

Satz 2.11 *Es seien A und B zwei (n, n)-Matrizen mit vollem Rang, so dass die Inversen Matrizen A^{-1} und B^{-1} existieren. Dann gilt*

$$(AB)^{-1} = B^{-1}A^{-1}. \tag{2.70}$$

Beweis: Sicherlich gilt $AB(AB)^{-1} = I_n$, I_n die (n, n)-Einheitsmatrix. Multiplikation von links mit A^{-1} liefert $B(AB)^{-1} = A^{-1}$; nochmalige Multiplikation mit B^{-1} von links führt auf $(AB)^{-1} = B^{-1}A^{-1}$. □

2.6 Eigenvektoren symmetrischer Matrizen

Gesucht ist eine Methode, für die Zeilen- bzw. Spaltenvektoren einer $(m \times n)$ Datenmatrix X latente Vektoren zu finden derart, dass die Vektoren von X als Linearkombinationen von m- bzw. n-dimensionalen latenten Vektoren dargestellt werden können. Es zeigt sich, dass diese latenten Vektoren durch eine Rotation der Zeilen- bzw. Spaltenvektoren von X gefunden werden können. Die Rotation eines Vektors $\mathbf{x}$ ist ein Spezialfall einer Transformation des Vektors, die stets eine Linearkombination ist, die wiederum stets als Multiplikation einer Matrix mit einem Vektor dargestellt werden kann:

$$\mathbf{y} = T\mathbf{x} = x_1\mathbf{t}_1 + x_2\mathbf{t}_2 + \cdots + x_n\mathbf{t}_n, \tag{2.71}$$

wobei $T = [\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_n]$ die *Transformationsmatrix* ist. Die Aufgabe ist also, eine derartige Matrix T zu bestimmen.

Es sei A eine beliebige (m, n)-Matrix, $\mathbf{x}$ sei ein n-dimensionaler Vektor und es gelte $A\mathbf{x} = \mathbf{y}$. $\mathbf{y}$ ist dann m-dimensional. A bildet Vektoren aus einem $\mathbb{R}^n$ auf Vektoren aus einem $\mathbb{R}^m$ ab. Für den Fall $m = n$ (A ist quadratisch) sind sowohl $\mathbf{x}$ also auch $\mathbf{y}$ Elemente des $\mathbb{R}^n$. Für das Folgende sind zwei Spezialfälle: von Bedeutung:

1. Es seien $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, $\mathbf{y} = A\mathbf{x}$, und es gelte einerseits $\mathbf{x} \neq \mathbf{y}$, und andererseits $\|\mathbf{x}\| = \|\mathbf{y}\|$, d.h. $\mathbf{x}$ und $\mathbf{y}$ haben identische Längen. Man sagt, die Transformation A ist *längeninvariant*. $\mathbf{x}$ und $\mathbf{y}$ unterscheiden sich nur durch ihre Orientierungen. A heißt dann auch *Rotationsmatrix*.

2. Für $\mathbf{x} \in \mathbb{R}^n$ gelte $A\mathbf{x} = \mathbf{y} = \lambda\mathbf{x}$. $\mathbf{x}$ und $\mathbf{y}$ haben also dieselbe Orientierung, unterscheiden sich aber im Falle $\lambda \neq 1$ durch ihre Länge. Für eine gegebene

Matrix A können die Vektoren $\mathbf{x}$ *nicht* beliebig gewählt werden, die *Orientierungsinvarianz* kann nur für spezielle, für A charakteristische Vektoren $\mathbf{x}$ gelten, die deswegen auch *charakteristische Vektoren* oder *Eigenvektoren* von A genannt werden. Dabei kann der wiederum für die Anwendungen sehr wichtige Fall eintreten, dass die zu einer Matrix T zusammengefassten Eigenvektoren einer Matrix A die Eigenschaft einer Rotationsmatrix haben.

In den folgenden Abschnitten wird die Rolle von Rotationsmatrizen und Matrizen von Eigenvektoren elaboriert.

2.6.1 Rotationen

Es seien $\mathbf{x}, \mathbf{y}$ zwei verschiedene n-dimensionale Vektoren und T sei eine Matrix derart, dass $\mathbf{y} = T\mathbf{x}$. T muß eine (n, n)-Matrix sein, da andernfalls die Vektoren $\mathbf{x}$ und $\mathbf{y}$ nicht beide n-dimensional sein können. T lasse die Länge von $\mathbf{x}$ invariant, so dass sich $\mathbf{y}$ von $\mathbf{x}$ nur in Bezug auf die Orientierung unterscheidet. Dementsprechend soll

$$\mathbf{x} \neq \mathbf{y}, \quad \|\mathbf{y}\|^2 = \mathbf{y}'\mathbf{y} = \mathbf{x}'T'T\mathbf{x} = \mathbf{x}'\mathbf{x} = \|\mathbf{x}\|^2 \tag{2.72}$$

gelten.

Satz 2.12 *Die Beziehung (2.72) gilt genau dann, wenn die Spaltenvektoren von T orthonormal sind, so dass $T'T = I$ gilt, wobei I die (n, n)-Einheitsmatrix ist. Darüber hinaus gilt $T'T = I$ genau dann, wenn $TT' = I$, d.h. die Orthonormalität der Spalten von T impliziert die Orthonormalität der Zeilen und umgekehrt.*

Beweis: Der Fall $\mathbf{x} = \mathbf{y}$ und $T = I$ die Einheitsmatrix ist trivial. Für $\mathbf{x} \neq \mathbf{y}$ ist die Bedingung $T'T = I$ sicher hinreichend dafür, dass $\mathbf{x}'T'T\mathbf{x} = \|\mathbf{x}\|^2$ erfüllt ist, denn dann gilt $\mathbf{x}'T'T\mathbf{x} = \mathbf{x}'I\mathbf{x} = \mathbf{x}'\mathbf{x} = \|\mathbf{x}\|^2$.

Die Beziehung $T'T = I$ ist auch notwendig für die Gültigkeit von (2.72). Um diese Behauptung einzusehen, werde $U = T'T$ gesetzt, wobei zunächst noch offen sei, ob $U = T'T = I$ gilt. Nach (2.72) soll $\mathbf{x}'U\mathbf{x} = \mathbf{x}'\mathbf{x}$ für alle $\mathbf{x}$ gelten. Dann liefert (4.27) aus dem Anhang, Abschnitt 4.4.2, Seite 148

$$U\mathbf{x} = I\mathbf{x}, \quad \text{für alle } \mathbf{x},$$

und die nochmalige Ableitung liefert $U = I$ (s. Gleichung (4.24), Seite 147), d.h. $T'T = I$. Dies bedeutet, dass die Spaltenvektoren von T orthonormal sind.

$T'T = I \Rightarrow TT' = I$: Da der Zeilenrang einer Matrix gleich ihrem Spaltenrang ist, muß $\mathrm{rg}(T') = \mathrm{rg}(T) = n$ gelten, d.h. die Spaltenvektoren $\tilde{\mathbf{t}}_i$ von T' (also die Zeilenvektoren von T) sind linear unabhängig, so dass nur noch ihre Orthonormalität gezeigt werden muß. Es gilt

$$T'T = I \Rightarrow T'TT' = T' \Rightarrow T' \underbrace{(TT' - I)}_{B=[\mathbf{b}_1,\ldots,\mathbf{b}_n]} = 0 \Rightarrow T'\mathbf{b}_j = \vec{0}_j \Rightarrow \mathbf{b}_j = \vec{0}_j$$

für $j = 1, \ldots, n$ wegen der linearen Unabhängigkeit der Spalten von T', also folgt $B = TT' - I = 0$, d.h. $TT' = I$.

$TT' = I \Rightarrow T'T = I$: diese Aussage folgt analog zur Aussage $T'T = I \Rightarrow TT' = I$.

□

Hinweis: Setzt man den Begriff der Determinante voraus, läßt sich ein sehr kurzer Beweis für die Aussage $T'T = I \Rightarrow TT' = I$ finden. Nach dem Produktsatz für Determinanten gilt $|AB| = |A||B|$ (vergl. (4.72), Seite 163), und nach (4.87), Seite 167, gilt für eine orthonormale Matrix T die Aussage $|T| = |I| = 1$, I die Einheitsmatrix. Dann hat man $|T'T| = |I| = |T'||T| = |T||T'| = |TT'| = 1$, also $T'T = TT' = I$.

□

Anmerkung: Die Beziehung $T'T = TT' = I$ für orthonormale Matrizen bedeutet, dass im Falle von orthonormalen Matrizen T die jeweilige Inverse durch

$$T^{-1} = T' \tag{2.73}$$

gegeben ist. Die Beziehung ist damit auch ein weiteres Beispiel für eine kommutative Matrixmultiplikation: mit $A = T$ und $B = T'$ hat man $AB = BA$.

□

Korollar 2.3 *Eine Rotation läßt die Skalarprodukte zwischen den rotierten Vektoren invariant.*

Beweis: Für $\mathbf{u} = T\mathbf{x}$, $\mathbf{v} = T\mathbf{y}$ folgt sofort $\mathbf{u}'\mathbf{v} = \mathbf{x}'T'T\mathbf{y} = \mathbf{x}'\mathbf{y}, \quad T'T = I.$ □

T läßt sich durch trigonometrische Betrachtungen zur Rotation (etwa von Koordinatensystemen) herleiten; im 2-dimensionalen Fall erhält man für T den Ausdruck

$$T = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}, \tag{2.74}$$

wobei θ der Rotationswinkel ist. Eine gegebene (m, n)-Rotationsmatrix T rotiert alle n-dimensionalen Vektoren $\mathbf{y}$ um einen bestimmten, fixen Winkel θ, so dass man auch $T(\theta)$ schreiben könnte, um diesen Sachverhalt auszudrücken. Davon wird im Folgenden kein Gebrauch gemacht, weil θ nicht explizit in die Betrachtungen eingeht. Dem Spezialfall auf Seite 60 entsprechend hat man $T' = T^{-1}$.
.

□

2.6.2 Rotationen und Koordinatentransformationen

Es sei X eine (m, n)-Matrix. Die m-Zeilen repräsentieren "Fälle", d.h. Personen oder Objekte, an denen jeweils n Messungen vorgenommen wurden. Die Messungen sind Messungen von n Variablen. X_{ij} ist die Messung der j-ten Variablen

beim i-ten Fall. Die Matrix X heißt *spaltenzentriert*, wenn von den Messwerten X_{ij} der j-ten Variablen (den Messwerten in der j-ten Spalte von X) der Mittelwert $\bar{x}_j$ subtrahiert wurde: $x_{ij} = X_{ij} - \bar{x}_j$. Benennt man X um in die Matrix der x_{ij}-Werte, so heißt X *spaltenzentriert*. Werden die Messwerte X_{ij} *spaltenstandardisiert*, so gehen die X_{ij}, über in $z_{ij} = (X_{ij} - \bar{x}_j)/s_j$, s_j die Streuung der Messwerte in der j-ten Spalte der Datenmatrix. Offenbar [22] ist

$$C = \frac{1}{m} X'X, \quad R = \frac{1}{m} Z'Z, \tag{2.75}$$

wenn X die spaltenzentrierte Datenmatrix ist und $Z = (z_{ij})$ die Matrix der spaltenstandardisierten Messwerte ist. C ist die Matrix der Varianzen und Kovarianzen, R ist die Matrix der Korrelationen zwischen den Variablen

Die Betrachtungen für eine zeilenzentrierte bzw. zeilenstandardisierte Matrix sind analog. Man muß allerdings bedenken, dass eine Zeilenzentrierung oder eine Zeilenstandardisierung nur Sinn macht, wenn die damit verbundene Mittelung über die Variablen Sinn macht.

Es sei X spaltenzentriert. Die Spalten von X können als (Spalten-)Vektoren $\mathbf{x}_j$, $j = 1, \ldots, n$ betrachtet werden. Die Spaltenvektoren von X' werden mit $\tilde{\mathbf{x}}_i$ bezeichnet; die $\tilde{\mathbf{x}}_i$ sind die Zeilenvektoren von X. Sie werden aber im Folgenden stets als Spaltenvektoren betrachtet, – eben als Spaltenvektoren von X'. Auf diese Weise kann u.a. die Schreibweise von Skalarprodukten aufrecht erhalten werden.

In Regressionsproblemen werden üblicherweise die Fälle als Punktekonfigurationen ("Punktwolken") dargestellt: im 2-dimensionalen Fall ist jeder Fall durch zwei Messungen X_1 und X_2 definiert, so dass der i-te Fall durch den Punkt (x_{i1}, x_{i2}) repräsentiert wird. Sind die Spaltenvektoren von X zentriert worden. fällt der Schwerpunkt der Punktekonfiguration mit dem Ursprung des Koordinatensystems zusammen, s. die Abbildung 5. Im Folgenden wird angenommen, dass die Daten spaltenzentriert sind, insbesondere können sie spaltenstandardisiert sein. Die Transformation von den Koordinaten (X_1, X_2) auf die Koordinaten (L_1, L_2) ist eine Rotation. Diese Koordinatentransformation kann dann in der Form

$$\tilde{\mathbf{y}}_i = T\tilde{\mathbf{x}}_i, \quad i = 1, \ldots, m \tag{2.76}$$

angeschrieben werden; die inverse Transformation ist

$$\tilde{\mathbf{x}}_i = T'\tilde{\mathbf{y}}_i, \quad i = 1, \ldots, m \tag{2.77}$$

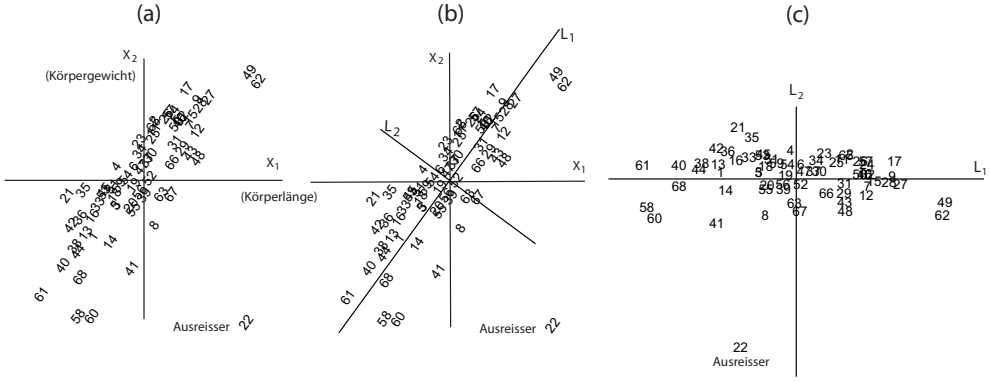
Die Schreibweisen T und T' können vertauscht werden; man wählt diejenige, die am Ende zu einer übersichtlicheren Formel führt.

2.6.3 Quadratische Formen und Eigenvektoren

Die folgenden Betrachtungen bereiten die Beantwortung der Frage vor, wie für einen gegebenen Datensatz latente Variable bestimmt werden können. Es wird

[22] Natürlich kann durch $m - 1$ geteilt werden.

Abbildung 5: Konfiguration von Fällen im ursprünglichen Koordinatensystem: Gewicht versus Körpergröße (a). In (b) sind mögliche latente Variable eingezeichnet worden: L_1 hat die Orientierung der maximalen Ausdehnung der Konfiguration, L_2 ist orthogonal zu L_1 und repräsentiert die Orientierung mit im allgemeinen zweitgrößter Ausdehnung der Konfiguration. (c) zeigt die Konfiguration im Koordinatensystem (L_1, L_2); die Koordinaten in diesem System sind die Projektionen der Punkte im ursprünglichen System auf die Achsen L_1 und L_2. Die Punkte werden durch Zahlen repräsentiert, um die Identifikation der Punkte im rotierten System zu erleichtern. Der Ausreisser 22 wurde bei der Bestimmung von L_1 und L_2 *nicht* berücksichtigt, weil er wegen seiner *Hebelwirkung* (leverage) die optimale Bestimmung dieser Achsen verhindert hätte (vergl. Beispiel 2.10, Seite 129).



zunächst gezeigt, dass bestimmten symmetrischen Matrizen M – insbesondere also Varianz-Kovarianzmatrizen – Ellipsoide zugeordnet werden können. Dabei sind die Orientierungen der Ellipsoide identisch und durch die im Folgenden hergeleiteten "Eigenvektoren" von M gegeben. Zu jedem Eigenvektor von M korrespondiert ein "Eigenwert". Den Hauptachsen der Ellipsoide entsprechen mögliche latente Dimensionen, und die Eigenwerte enthalten Informationen über die Anzahl der latenten Dimensionen.

Die Idee der latenten Variablen ist, dass diese durch (Basis-)Vektoren $\mathbf{L}_1, \ldots, \mathbf{L}_r$ repräsentiert werden, so dass die Spaltenvektoren $\mathbf{x}_j$ von X als Linearkombinationen

$$\mathbf{x}_j = a_{1j}\mathbf{L}_1 + a_{2j}\mathbf{L}_2 + \cdots + a_{jr}\mathbf{L}_r$$

dargestellt werden können. Da aber nur die Matrix X gegeben ist, müssen die $\mathbf{L}_k$ $(k = 1, \ldots, r)$ aus den Daten errechnet werden. Formal bedeutet dies, dass man Vektoren $\mathbf{t}_k$ bestimmt derart, dass $X\mathbf{t}_k = \mathbf{L}_k$. Wie Betrachtungen zur Abbildung 5 zeigen, kann die Beziehung zwischen den $\mathbf{x}_j$ und den $\mathbf{L}_k$ als Rotation der Koordinatenachsen bzw. der Punktekonfiguration aufgefasst werden, so dass

$T = [\mathbf{t}_1 \ldots, \mathbf{t}_n]$ eine Rotationsmatrix ist, woraus sofort folgt, dass die Darstellung der $\mathbf{x}_j$ als Linearkombinationen der $\mathbf{L}_k$ ebenfalls eine Rotation der $\mathbf{L}_k$ sein muß. Diese Rotation invertiert die durch T gegebene Rotation und ist deshalb durch $T^{-1} = T'$ gegeben.

Ist X (spalten-)zentriert, so gilt $\vec{1}'X = (0, \ldots, 0) = \vec{0}'$, d.h. die Spaltensummen von X sind gleich Null; alle Betrachtungen übertragen sich auf zeilenzentrierte Datenmatrizen). Dann folgt $\vec{1}'X\mathbf{t}_k = \vec{1}'\mathbf{L}_l = 0$, d.h. die Summen der Komponenten der $\mathbf{L}_k$ sind ebenfalls gleich Null. Dann ist $\mathbf{L}_k'\mathbf{L}_k = \|\mathbf{L}_k\|^2$ proportional zur Varianz der Komponenten von $\mathbf{L}_k$; der Proportionalitätsfaktor ist $1/m$ bzw. $1/(m-1)$. Es ist $\mathbf{L}_k'\mathbf{L}_k = \mathbf{t}_k'X'X\mathbf{t}_k$. Hier ist $X'X$ eine symmetrische Matrix, weshalb der Ausdruck $\mathbf{t}_k'X'X\mathbf{t}_k$ ein Beispiel für den allgemeinen Begriff der *quadratischen Form* ist, der im Folgenden allgemein eingeführt wird. Die Bedeutung dieses Begriffs wird deutlich, wenn man bedenkt, dass die $\mathbf{L}_k$ so bestimmt werden können, dass die Varianz $\mathbf{L}_1'\mathbf{L}_1$ maximal wird. Dies bedeutet die Maximierung der quadratischen Form $\mathbf{t}_1'X'X\mathbf{t}_1$ als Funktion des Vektors $\mathbf{t}_1$. Die Details dieses Ansatzes werden weiter unten gegeben.

Die folgenden Betrachtungen gelten für beliebige symmetrische Matrizen M und Vektoren $\mathbf{x}$, ein spezieller Bezug auf Zeilenvektoren $\tilde{\mathbf{x}}_i$ oder Spaltenvektoren $\mathbf{x}_j$ einer Matrix X ist hier nicht nötig.

Definition 2.4 *Es sei M eine symmetrische (n,n)-Matrix, und es gelte*

$$Q_M(\boldsymbol{x}) = \boldsymbol{x}'M\boldsymbol{x}, \quad \textit{für alle}\, \boldsymbol{x} \in \mathbb{R}^n \tag{2.78}$$

Dann heißt $Q_M(\boldsymbol{x})$ quadratische Form*, und*
1. M heißt positiv semidefinit*, wenn $\boldsymbol{x}'M\boldsymbol{x} \geq 0$*
2. M heißt negativ semidefinit*, wenn $\boldsymbol{x}'M\boldsymbol{x} \leq 0$*
3. M heißt positiv definit *bzw.* elliptisch*, wenn $\boldsymbol{x}'M\boldsymbol{x} > 0$, und*
4. M heißt negativ definit *bzw.* hyperbolisch*, wenn $\boldsymbol{x}'M\boldsymbol{x} < 0$*
jeweils für alle $\boldsymbol{x} \in \mathbb{R}^n$ gilt.

Satz 2.13 *Es sei $X \in \mathbb{R}^{m,n}$ eine zentrierte Matrix. Die Matrizen $X'X$ und XX' sind positiv semidefinit.*

Beweis: Es sei $\mathbf{x} \in \mathbb{R}^n$, d.h. $\mathbf{x}$ sei ein n-dimensionaler Vektor, und es sei $X\mathbf{x} = \mathbf{y}$. Dann ist $\mathbf{x}'X'X\mathbf{x} = \mathbf{y}'\mathbf{y} = \|\mathbf{y}\|^2 \geq 0$, also ist $X'X$ positiv-semidefinit. Nun sei $\mathbf{x} \in \mathbb{R}^m$ und es sei $\mathbf{x}'X = \mathbf{y}'$. Dann folgt $\mathbf{x}'XX'\mathbf{x} = \mathbf{y}'\mathbf{y} = \|\mathbf{x}\|^2 \geq 0$, d.h. XX' ist positiv semidefinit. □

Folgerung: Varianz-Kovarianzmatrizen sind von der Form $X'X$ bzw. XX', also sind sie positiv semidefinit. □

Definition 2.5 *Es seien $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ eine Menge von Vektoren mit gleicher Dimensionalität. Es sei A die (n,n)-Matrix der Skalarprodukte $\boldsymbol{x}_j'\boldsymbol{x}_k$, $j,k = 1,\ldots,n$ der Vektoren. Dann heißt A* Gram-Matrix *oder* Gramsche Matrix[23]*.*

Varianz-Kovarianzmatrizen sind also Gramsche Matrizen, und aus dem Vorangegangenen folgt, dass Gramsche Matrizen positiv-semidefinit sind.

Satz 2.14 *Es sei $M \in \mathbb{R}^{n,n}$ eine symmetrische, positiv semidefinite Matrix. Dann definiert die Menge $\mathcal{E}_x = \{\boldsymbol{x} | \boldsymbol{x}'M\boldsymbol{x} = k, \boldsymbol{x} \in \mathbb{R}^n, k \in \mathbb{R}$ eine Konstante$\}$ ein n-dimensionales Ellipsoid, wobei die Anfangspunkte der Vektoren $\boldsymbol{x}$ im Nullpunkt des Koordinatensystems und die Endpunkte auf dem jeweiligen Ellipsoid liegen.*

Beweis: Die Aussage folgt sofort aus der Definition von $Q_M(\mathbf{x})$: multipliziert man (2.78) aus, so erhält man

$$\mathbf{x}'M\mathbf{x} = \sum_{i=1}^{n} m_{ii}x_i^2 + 2\sum_{i<j} m_{ij}x_ix_j = k \tag{2.79}$$

Für $\mathbf{x}'M\mathbf{x} = k > 0$ definiert $\mathbf{x}'M\mathbf{x}$ ein n-dimensionales Ellipsoid. □

Der Ausdruck 'quadratische Form' ergibt sich aus dem Sachverhalt, dass die Summe der Exponenten der Komponenten x_i stets gleich 2 ist. Für den Spezialfall $n = 2$ hat man

$$\mathbf{x}'M\mathbf{x} = m_{11}x_1^2 + m_{22}x_2^2 + 2m_{12}x_1x_2 = k > 0. \tag{2.80}$$

Die Menge der 2-dimensionalen Vektoren $\mathbf{x} = (x_1, x_2)'$, die dieser Gleichung genügen, definiert eine Ellipse.

Anmerkung: Die in Satz 2.13 eingeführte Varianz-Kovarianzmatrix S ist symmetrisch und positiv-definit, also definiert sie eine Menge von Ellipsoiden mit identischer Orientierung, wobei jedes Ellipsoid durch einen bestimmten Wert von $k \in \mathbb{R}$ charakterisiert wird. □

Spezialfall: Insbesondere sei $M = \Lambda = \text{diag}(\lambda_1, \ldots, \lambda_n)$ eine Diagonalmatrix. Dann sind die Ellipsoide $\mathbf{x}'\Lambda\mathbf{x} = k$ für verschiedene Werte von k *achsenparallel*, d.h. die Hauptachsen der Ellipsoide sind parallel zu den Achsen des Koordinatensystems; diese Aussage folgt sofort aus (2.79) bzw. (2.80), denn für $M = \Lambda$ sind alle $m_{ij} = 0$ für $i \neq j$.

Es seien nun $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ Vektoren und M sei eine symmetrische (n,n)-Matrix, und es gelte

$$\mathbf{x}'M\mathbf{x} = \mathbf{y}'\Lambda\mathbf{y} = k > 0, \quad \Lambda = \text{diag}(\lambda_1, \ldots, \lambda_n). \tag{2.81}$$

[23] Nach Jörgen Pedersen Gram (1850 – 1916), dänischer Mathematiker.

$\mathcal{E}_x = \{\mathbf{x}|\mathbf{x}'M\mathbf{x} = k > 0\}$ und $\mathcal{E}_y = \{\mathbf{y}|\mathbf{y}'\Lambda\mathbf{y} = k > 0\}$ sind Ellipsoide, $\mathcal{E}_Y$ ist insbesondere achsenparallel. Weiter gelte $\mathbf{x} = T\mathbf{y}$, wobei die (m, n)-Matrix T eine Rotation repräsentiere. Offenbar gilt

$$\mathbf{x}'M\mathbf{x} = \mathbf{y}'T'MT\mathbf{y} = \mathbf{y}'\Lambda\mathbf{y} = k, \tag{2.82}$$

woraus

$$T'MT = \Lambda \tag{2.83}$$

folgt[24]. Denn $\mathcal{E}_y$ ist ein achsenparalleles Ellipsoid, so dass auch $\mathbf{y}'T'MT\mathbf{y} = k$ ein achsenparalleles Ellipsoid sein muß, d.h. $T'MT = \Lambda$ muß eine Diagonalmatrix sein. □

Da T als Rotationsmatrix angenommen wurde, folgt, dass T orthonormal ist. Deshalb folgt durch Multiplikation der Gleichung (2.83) von links mit T die Gleichung

$$MT = T\Lambda. \tag{2.84}$$

Diese Gleichung besagt, dass für die Spaltenvektoren $\mathbf{t}_k$, $k = 1, \ldots, n$, von T die Beziehung $M\mathbf{t}_k = \lambda_k\mathbf{t}_k$ gilt, d.h. die $\mathbf{t}_k$ werden durch M so transformiert, dass sich nur ihre Länge, nicht aber ihre Orientierung verändert (vergl. (2.25), S. 49). Diese Aussage gilt natürlich nicht für beliebige Vektoren $\mathbf{x}$, sondern nur für spezielle Vektoren $\mathbf{t}$, die charakteristisch für die Matrix M sind.

Definition 2.6 *Es sei M eine beliebige (n, n)-Matrix und $\boldsymbol{t} \in \mathbb{R}^n$ sei ein Vektor, der der Beziehung*

$$M\boldsymbol{t} = \lambda\boldsymbol{t}, \quad \boldsymbol{t} \neq \vec{0}, \tag{2.85}$$

genügt. Dann heißt $\boldsymbol{t}$ Eigenvektor[25] *von M und λ heißt der zu $\boldsymbol{t}$ gehörende* Eigenwert *von M.*

Anmerkungen:

1. In der Definition 2.6 wurde *nicht* vorausgesetzt, dass M symmetrisch ist, d.h. es wurde dem Sachverhalt, dass Eigenvektoren auch für nicht-symmetrische Matrizen existieren Rechnung getragen. Die folgenden Betrachtungen beschränken sich aber auf symmetrische Matrizen M.

2. In der Definition 2.6 wurde T nicht als Rotationsmatrix spezifiziert, die Beziehung (2.84), die T als Matrix von Eigenvektoren ausweist, ergab sich aber durch Ausnutzung der Orthonormaliät der Rotationsmatrix T. Die Frage

[24] Zwei quadratische Matrizen A und B heißen *ähnlich*, wenn eine reguläre Matrix S existiert – also eine Matrix, für die eine Inverse existiert – derart, dass $A = S^{-1}BS$. Die Matrizen M und Λ in Gleichung (2.83) sind offenbar ähnliche Matrizen, da $T' = T^{-1}$. S.a. Definition 4.89, Seite 167. Die Beziehung (2.82) läßt sich auch "rechnerisch" durch Anwendung der Regeln für die Differentiation von Vektoren (s. Anhang, Abschnitt 4.4) zeigen.

[25] Synonym sind auch die Ausdrücke 'latenter Vektor' und 'latenter Wert' und 'charakteristischer Vektor' und 'charakteristischer Wert'.

ist nun, ob Orthonormalität eine charakteristische Eigenschaft von Eigenvektoren der symmetrischen Matrix ist, unabhängig davon, ob T vorher als Rotationsmatrix spezfiziert wurde oder nicht. Dies ist in der Tat der Fall, wie in den Sätzen 2.18 und 2.19, Seite 77, gezeigt wird.

3. Der Nullvektor $\vec{0}$ ist *kein* Eigenvektor. Wäre $\vec{0}$ ein Eigenvektor, könnte man $M\vec{0} = \vec{0} = \lambda\vec{0}$ schreiben. Diese Gleichung ist aber für beliebige $\lambda \in \mathbb{R}$ erfüllt, d.h. man hätte beliebig viele Eigenwerte, die zu $\vec{0}$ korrespondieren. Darüber hinaus hat $\vec{0}$ keine Orientierung. Die Orientierung ist aber ein charakteristisches Merkmal von Eigenvektoren; da sie alle die Länge 1 haben, unterscheiden sie sich nur durch ihre Orientierung. Der Nullvektor als Eigenvektor ist daher kein sinnvoller Begriff.

4. Nach (2.85) kann man für irgendeinen Eigenvektor $\mathbf{t}$ mit zugehörigem Eigenwert λ $M\mathbf{t} = \lambda\mathbf{t}$ schreiben, woraus

$$M\mathbf{t} - \lambda\mathbf{t} = (M - \Lambda I)\mathbf{t} = 0 \tag{2.86}$$

folgt, wobei I die (n, n)-Einheitsmatrix ist. Dies ist ein Gleichungssystem mit der Matrix $M - \lambda I$ und den Komponenten von $\mathbf{t}$ als Unbekannten. Hätte $M - \lambda I$ vollen Rang, müsste $\mathbf{t} = \vec{0}$ sein, was aber nicht möglich ist, da $\mathbf{t}$ ja ein Eigenvektor ist. Also hat $M - \lambda I$ nicht vollen Rang, d.h. $M - \lambda I$ ist *singulär*. Dies bedeutet, dass für die Determinante (s. Abschnitt 4.7) $|M - \lambda I| = 0$ gilt. Entwickelt man diese Determinante (vergl. Punkt 19, Seite 168, in Abschnitt 4.7), so ergibt sich ein Polynom n-ten Grades in λ, d.h.

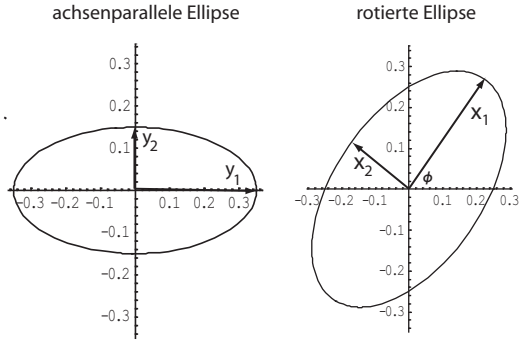
$$f(\lambda) = |M - \lambda I| = \alpha_0 + \alpha_1\lambda + \alpha_2\lambda^2 + \cdots + \alpha_n\lambda^n, \tag{2.87}$$

wobei die α_j reelle Zahlen sind, die durch die Elemente der Matrix $M - \lambda I$ definiert sind. Die Nullstellen dieses Polynoms, also diejenigen Werte für λ, für die $f(\lambda) = 0$ gilt, sind die Eigenwerte von M. $f(\lambda)$ heißt *charakteristisches Polynom* der Matrix $A - \lambda I$, I die Einheitsmatrix, das im Anhang, Abschnitt 4.7 vorgestellt wird, worauf aber im Folgenden kein Bezug genommen wird. Der Begriff des charakteristischen Polynoms wird an dieser Stelle nur der Vollständigkeit wegen erwähnt. □

Die Gleichung (2.84) besagt also, dass alle Spaltenvektoren $\mathbf{t}_j$ von T Eigenvektoren von M sind, und die Diagonalmatrix Λ enthält in der Diagonalen die zugehörigen Eigenwerte von M.

Zur Bedeutung der Eigenvektoren: Es sei M eine symmetrische, positivsemidefinite Matrix mit T als Matrix der zugehörigen Eigenvektoren, so dass $MT = T\Lambda$, also $T'MT = \Lambda$ gilt, Λ die Diagonalmatrix der zugehörigen Eigenwerte von M. Die Menge $\mathcal{E}$ der Vektoren $\mathbf{x}$ mit $\mathbf{x}'M\mathbf{x} = k \in \mathbb{R}$ beschreibt ein Ellipsoid (Ellipse) der Art "rotiert" in Abbildung 6, sofern M keine Diagonalmatrix ist. Die Menge $\mathcal{Y}$ mit $\mathbf{y}'\Lambda\mathbf{y} = k$ dagegen beschreibt eine achsenparallele

Abbildung 6: Ellipsen in verschiedenen Orientierungen; die Hauptachsen sind die skalierten Eigenvektoren der zugehörigen Matrix M.



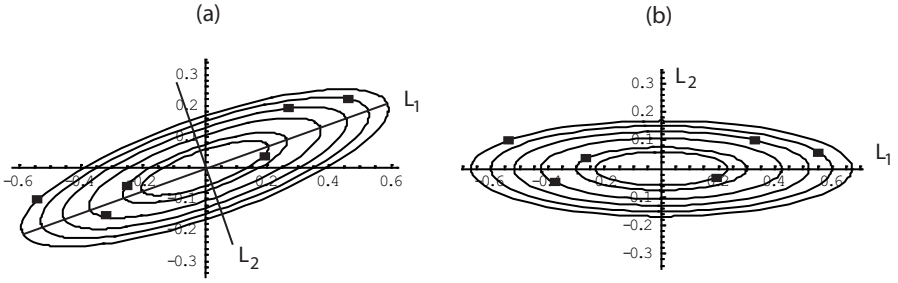
Ellipse. Da T orthonormal ist, ist T eine Rotationsmatrix, d.h. T rotiert $\mathbf{y} \in \mathcal{Y}$ in einen Vektor $\mathbf{x} \in \mathcal{E}$. Insbesondere definiere der Vektor $\mathbf{y}_1 \in \mathcal{Y}$ die Orientierung der ersten Halbachse des durch Λ definierten achsenparallelen Ellipsoids (d.h. $\mathbf{y}_1$ liege auf der ersten Achse des Koordinatensystems, in dem das Ellipsoid liegt). Dann ist $\mathbf{y}_1 = y_1\mathbf{e}_1$, $\mathbf{e}_1 = (1, 0, \ldots, 0)'$, $y_1 \in \mathbb{R}$, so dass $\|\mathbf{y}_1\| = y_1\|\mathbf{e}_1\|$. Dann transformiert T $\mathbf{y}_1$ in den Vektor $\mathbf{x}_1$, der die erste Halbachse der rotierten, durch M definierten Ellipse (allgemein: des Ellipsoids) repräsentiert, und man hat

$$\mathbf{x}_1 = T\mathbf{y}_1 = T(y_1\mathbf{e}_1) = y_1 T\mathbf{e}_1 = y_1 \begin{pmatrix} t_{11} \\ t_{21} \\ \vdots \\ t_{n1} \end{pmatrix} = y_1\mathbf{t}_1, \quad y_1 \in \mathbb{R}, \tag{2.88}$$

d.h. $\mathbf{x}_1$ ist proportional zum ersten Eigenvektor von M, der in der ersten Spalte von T steht. $\mathbf{t}_1$ definiert demnach die Orientierung der ersten Hauptachse des durch M definierten Ellipsoids. Da die $\mathbf{t}_j$ orthogonal sind, definieren die restlichen Vektoren $\mathbf{t}_j$, $j \neq 1$, die Orientierungen der restlichen Hauptachsen des Ellpsoids.

T rotiert alle Vektoren $\mathbf{y}$, für die $\mathbf{y}'\Lambda\mathbf{y} = k$ gilt. $\mathbf{x}_1$ definiert dann die erste Halbachse des Ellipsoids $\mathbf{x}'M\mathbf{x} = k$. Nach (2.88) ist die Orientierung dieser Halbachse durch den ersten Eigenvektor $\mathbf{t}_1$ von M gegeben. Analoge Interpretationen

Abbildung 7: Punktekonfiguration und Ellipsen. Im Anhang, Abschnitt 4.3 wird die Konstruktion dieser Ellipsen näher erläutert.



ergeben sich für die übrigen Halbachsen des durch M definierten Ellipsoids:

$$\mathbf{x}_j = Ty_j \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = y_j \begin{pmatrix} t_{1j} \\ t_{2j} \\ \vdots \\ t_{nj} \end{pmatrix}, \quad \mathbf{x}_j \in \mathcal{E}_x,\ \mathbf{y}_j = y_j \mathbf{e}_j \in \mathcal{E}_y \tag{2.89}$$

y_j ist die Länge der jeweiligen Halbachse.

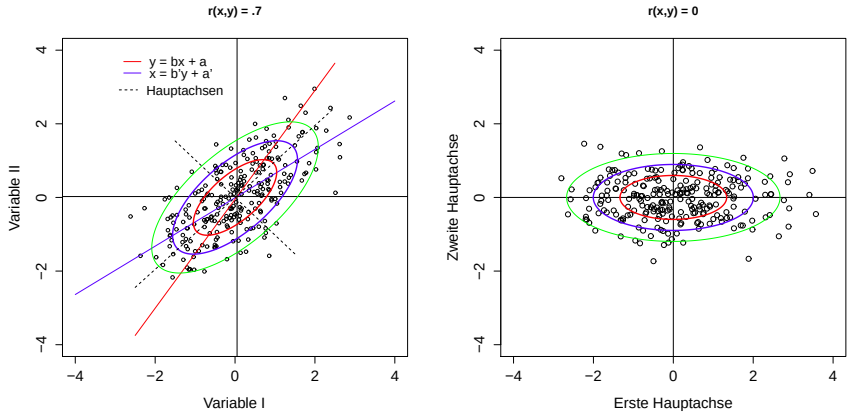
Definition 2.7 *Die orthonormale Transformationsmatrix T rotiert das achsenparallele Ellipsoid $\mathcal{E}_y$ in das orientierte Ellipsoid $\mathcal{E}_x$, und wegen $T^{-1} = T'$ rotiert das Ellipsoid $\mathcal{E}_x$ in das Ellpsoid $\mathcal{E}_y$. T und T' heißen deshalb* Hauptachsentransformationen.

Beispiel 2.9 Rotation der Fälle Es sei X eine spaltenzentrierte (m, n)-Datenmatrix; die Spaltenvektoren $\mathbf{x}_j$, $1 \leq j \leq m$, repräsentieren Variablen, die Zeilenvektoren $\tilde{\mathbf{x}}_i$, $1 \leq i \leq n$, repräsentieren die Fälle. Für jeden Fall existiert ein Ellipsoid derart, dass der Endpunkt von $\tilde{\mathbf{x}}_i$ auf der Oberfläche des Ellipsoids liegt, vergl. Abbildung 7. Denn es sei $C = X'X$, $\tilde{\mathbf{x}}_i' C \tilde{\mathbf{x}}_i = k_i$. Dann ist

$$\mathcal{E}_{x,i} = \{\tilde{\mathbf{x}} | \tilde{\mathbf{x}} \in \mathbb{R}^n, \tilde{\mathbf{x}}' C \tilde{\mathbf{x}} = k_i > 0\} \tag{2.90}$$

das Ellipsoid für den i-ten Fall in den ursprünglichen Koordinaten. Die Orientierung der Ellipsoide ist für alle i identisch, da sie von der Matrix C festgelegt wird. Nach Satz 2.4, Seite 53, kann X stets als Produkt $X = UV'$ zweier Matrizen U und V' dargestellt werden, wobei die Spaltenvektoren von V linear unabhängig

Abbildung 8: Links: Punktekonfiguration für $r_{xy} = .7$ mit Regressionsgeraden, Ellipsen und deren Hauptachsen; Die Orientierungen der Regressionsgeraden sind im Allgemeinen nicht identisch mit der Orientierung der ersten Hauptachse. Rechts: Die Hauptachsen als neue Koordinatenachsen für die Punktekonfiguration. Die Komponenten der $\mathbf{y}_i$ sind die Koordinaten in Bezug auf diese Achsen. Zur Berechnung der Ellipsen s. den Anhang 4.3, Seite 145.



sind. Insbesondere kann dann für V eine orthonormale Matrix gewählt werden, weil dann die Spalten von V auf jeden Fall linear unabhängig sind. Somit sei V die orthonormale Matrix der Eigenvektoren von C: $CV = V\Lambda$, Λ die Diagonalmatrix der Eigenwerte von C. Dann folgt $XV = U$, und wegen

$$U'U = V'X'XV = V'CV = V'V\Lambda V'V = \Lambda$$

folgt dann, dass die Spaltenvektoren von U orthogonal sind. Der i-te Zeilenvektor $\tilde{\mathbf{u}}_i$ von U ist dann die Transformation des i-ten Zeilenvektors von X: aus $XV = U$ folgt $X' = VU'$, d.h.

$$\tilde{\mathbf{x}}_i = V\tilde{\mathbf{u}}_i, \quad \tilde{\mathbf{u}}_i = V'\tilde{\mathbf{x}}_i, \tag{2.91}$$

d.h. $\tilde{\mathbf{x}}_i$ und $\tilde{\mathbf{u}}_i$ gehen durch Rotation auseinander hervor. $\tilde{\mathbf{u}}_i$ definiert das achsenpallele Ellipsoid

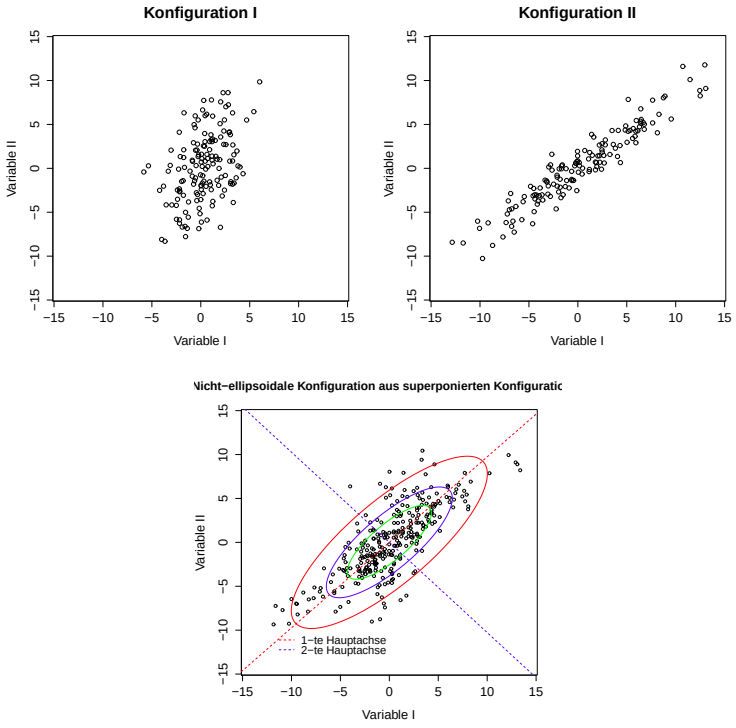
$$\mathcal{E}_{u,i} = \{\tilde{\mathbf{u}} | \tilde{\mathbf{u}}'\Lambda\tilde{\mathbf{u}} = k_i > 0\} \tag{2.92}$$

wobei $k_i = \tilde{\mathbf{u}}_i'\Lambda\tilde{\mathbf{u}}_i$. Die Ellipsoide $\mathcal{E}_{u,i}$ haben für die verschiedenen i alle dieselbe Orientierung, da Λ für alle Ellipsoide identisch ist, und da Λ eine Diagonalmatrix ist sind die Ellipsoide $\mathcal{E}_{u,i}$ achsenparallel. Abbildung 8 (S. 73) illustriert die Rotation der Punktekonfigurationen. □

Anmerkung zur Punktekonfiguration: Es ist gezeigt worden, dass für jeden Punkt der Konfiguration der Fälle ein Ellipsoid existiert, auf dem der Punkt liegt, aber dies bedeutet *nicht*, dass die Konfiguration auch tatsächlich ellipsoid

sein muß, – d.h. die unterliegende Verteilung muß nicht die multivariate Normalverteilung sein(vergl. die Dichte der Normalverteilung in Beispiel 4.7, Gleichung (4.132), Seite 182), – die Dichtefunktion ist durch eine qudratische Form definiert, aber die Beschreibung eines Datensatzes durch eeine quadratische Form bedeutet noch nicht, dass die Daten auch normalverteilt sind). Auch wenn die Konfiguration selbst nicht ellipsoid ist kann doch stets eine Menge von Ellipsoiden gefunden werden derart, dass jeder Fall auf einem Ellipsoid liegt, – einfach weil $X'X$ stets eine Menge von Ellipsoiden definiert. Dieser Fall kann z.B. dann eintreten, wenn sich die Stichprobe der Fälle aus Stichproben aus verschiedenen Populationen zusammensetzt. In Abbildung 9 wird dieser Sachverhalt illustriert.

Abbildung 9: Superponierte Punktekonfigurationen und Ellipsen



Die folgenden Aussagen beziehen sich auf die Anzahl der Dimensionen, d.h. auf die Anzahl der Hauptachsen der Ellipsoide.

Satz 2.15 *Es sei M eine reelle, symmetrische (n,n)-Matrix. M ist positiv semidefinit dann und nur dann, wenn die Eigenwerte λ_j größer als bzw. mindestens gleich Null sind.*

Beweis: Es sei $M = T'\Lambda T$, wobei T die orthonormalen Eigenvektoren von M

seien. Multiplikation von links mit T impliziert $TM = \Lambda T$, nochmalige Multiplikation von rechts mit T' impliziert $TMT' = \Lambda$. Weiter sei $q = \mathbf{x}'M\mathbf{x} \in \mathbb{R}$, $\mathbf{x}$ ein beliebiger n-dimensionaler Vektor. Für einen geeignet gewählten Vektor $\mathbf{y}$ ist $\mathbf{x} = T\mathbf{y}$, so dass

$$q = \mathbf{y}'T'MT\mathbf{y} = \mathbf{y}'\Lambda\mathbf{y} = \sum_{i=1}^{n} \lambda_i y_i^2$$

Da $\mathbf{x}$ beliebig gewählt werden kann, kann insbesondere $\mathbf{x} = T\mathbf{e}_j$ gewählt werden, also $\mathbf{y}_j = \mathbf{e}_j$ der j-te Einheitsvektor, $j = 1, \ldots, n$. Dann ist $q = \lambda_j$, und $q \geq 0$ genau dann, wenn $\lambda_j \geq 0$.

Man zeigt auf analoge Weise, dass für eine negativ (semi-)definite Matrix $\lambda_j \leq 0$ für alle j gilt. □

Spektraldarstellung von M: Die Orthonormalität von T bedeutet, dass man aus $MT = T\Lambda$ (Gleichungen (2.99) und (2.100)) durch Multiplikation von rechts mit T' die Beziehung

$$M = T\Lambda T' = \sum_{k=1}^{n} \lambda_k \mathbf{t}_k \mathbf{t}_k'. \tag{2.93}$$

erhält. Der Ausdruck $\sum_k \lambda_k \mathbf{t}_k \mathbf{t}_k'$ drückt $T\Lambda T'$ über die dyadischen Produkte $\mathbf{t}_k \mathbf{t}_k'$ aus und erweist sich bei bestimmten Betrachtungen als nützlich. Man macht sich leicht klar, wie dieser Ausdruck zustande kommt. Es ist ja $T = [\mathbf{t}_1, \ldots, \mathbf{t}_n]$ und $T\Lambda = [\lambda_1\mathbf{t}_1, \ldots, \lambda_n\mathbf{t}_n]$, so dass

$$T\Lambda T' = [\lambda_1\mathbf{t}_1, \ldots, \lambda_n\mathbf{t}_n] \begin{bmatrix} \mathbf{t}_1' \\ \mathbf{t}_2' \\ \vdots \\ \mathbf{t}_n' \end{bmatrix} = \lambda_1\mathbf{t}_1\mathbf{t}_1' + \lambda_2\mathbf{t}_2\mathbf{t}_2' + \cdots + \lambda_n\mathbf{t}_n\mathbf{t}_n' = \sum_{k=1}^{n} \lambda_k \mathbf{t}_k \mathbf{t}_k'.$$

Definition 2.8 *Die Darstellung (2.93) von M heißt* Spektraldarstellung *von M.*

Satz 2.16 *Es sei M eine beliebige relle, symmetrische (n, n)-Matrix vom Rang $r \leq n$, T sei die $n \times n$-Matrix der Eigenvektoren, Λ sei die (n, n)-Diagonalmatrix der Eigenvektoren von M, so dass $M = T\Lambda T'$. Dann gilt*

$$rg(M) = rg(T\Lambda T') = rg(\Lambda) = r \tag{2.94}$$

und $rg(\Lambda)$ ist gleich der Anzahl von Null verschiedenen Eigenwerte.

Beweis: Da T vollen Rang hat ($T'T = I$) folgt $\mathrm{rg}(M) = \mathrm{rg}(T\Lambda T') = \mathrm{rg}(\Lambda)$ aus Satz 2.6, Gleichung (2.54) (Seite 56) mit $P = Q = T$. $\mathrm{rg}(\Lambda) = r$ folgt aus (2.93), denn es muß nun

$$\sum_{k=1}^{n} \lambda_k \mathbf{t}_k \mathbf{t}_k' = \sum_{k=1}^{r} \lambda_k \mathbf{t}_k \mathbf{t}_k'$$

gelten, d.h. es folgt $\lambda_{r+1} = \cdots = \lambda_n = 0$. □

Korollar 2.4 *Es sei X eine (m,n)-Matrix. Dann haben $X'X$ und XX' dieselben von Null verschiedenen Eigenwerte.*

Beweis: Allgemein gilt $X = UV'$ (vergl. (2.48), Seite 53), wobei U eine (m,r)- und V eine (n,r)-Matrix ist, $r = \mathrm{rg}(X)$. Dann ist $X' = VU'$, d.h. die Spaltenvektoren von X' sind Linearkombinationen der Spaltenvektoren von V. Es sei insbesondere $V = T$, T die Matrix der Eigenvektoren von $X'X$; T ist eine orthonormale Basis für die Spaltenvektoren von X'. Dann folgt $XT = U$ und

$$U'U = T'X'XT = T'T\Lambda T'T = \Lambda,$$

Λ die Diagonalmatrix der r von Null verschiedenen Eigenwerte, d.h. U ist orthogonal und die Eigenwerte sind gleich den Quadraten der Längen der Spaltenvektoren von U. Dann enthält $Q = U\Lambda^{-1/2}$ die normiertem Vektoren, d.h. $U = Q\Lambda^{1/2}$. Man hat nun

$$XX' = UT'TU' = UU' = Q\Lambda^{1/2}\Lambda^{1/2}Q' = Q\Lambda Q' \Rightarrow XX'Q = Q\Lambda,$$

d.h. Q ist die Matrix der zu Eigenwerten ungleich Null korrespondierenden Eigenvektoren von XX', und offenbar sind diese Eigenwerte identisch mit denen von $X'X$. □

Anmerkung: Man kann also $T\Lambda T' = T_r\Lambda_r T_r'$ schreiben, wobei Λ_r eine (r,r)-Diagonalmatrix ist, deren Diagonalelemente nur die von Null verschiedenen Eigenwerte entält, und T_r ist eine (n,r)-Matrix mit den Eigenvektoren von M, die zu den von Null verschiedenen Eigenwerten korrespondieren. □

Der folgende Satz gibt weitere Auskunft über die Eigenschaften einer reellen, symmetrischen Matrix; er spezifiert die Bedingungen, die eine symmetrische Matrix M erfüllen muß, um eine Ellipse bzw. ein Ellipsoid zu definieren.

Satz 2.17 *Es sei M eine symmetrische (n,n)- Matrix vom Rang $r \leq n$. Dann ist M genau dann positiv semidefinit, wenn eine (n,r)-Matrix G existiert derart, dass*

$$M = GG'. \tag{2.95}$$

Beweis: (1) $\Rightarrow$: Es gelte $M = GG'$. Dann folgt

$$\mathbf{x}'GG'\mathbf{x} = (G\mathbf{x})'G\mathbf{x} = \|G\mathbf{x}\|^2 \geq 0,$$

so dass M positiv semidefinit ist.

(2) $\Leftarrow$: Aus der Symmetrie von M folgt die Existenz der orthonormalen Matrix T und der Diagonalmatrix $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$, $\lambda_j \geq 0$ für $j = 1, \ldots, n$ (s. Satz 2.15), mit $M = T\Lambda T'$. Es sei

$$\Lambda^{1/2} = \mathrm{diag}(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_r}, \underbrace{0, \ldots, 0}_{n-r}).$$

Dann kann man

$$M = T\Lambda^{1/2}\Lambda^{1/2}T' = (T\Lambda^{1/2})(T\Lambda^{1/2})'$$

schreiben. Streicht man in $T\Lambda^{1/2}$ alle Spalten, die nur Nullen enthalten, so erhält man eine Matrix $G = T_r\Lambda_r^{1/2}$ und M ist in der Form $M = GG'$ darstellbar. □

In der Anmerkung 2 auf Seite 69 wurde die Frage gestellt, ob die Matrix T der Eigenvektoren einer symmetrischen Matrix charakteristischerweise orthonormal ist oder diese Eigenschaft nur folgt, wennn T vorher als Rotationsmatrix definiert wurde. Dazu wird der folgende Satz bewiesen:

Satz 2.18 *$M \in \mathbb{R}^{n,n}$ sei symmetrisch und habe die Eigenvektoren $\boldsymbol{t}_1, \ldots, \boldsymbol{t}_n$. Für irgendzwei Eigenvektoren $\boldsymbol{t}_j$ und $\boldsymbol{t}_k$ mit zugehörigen Eigenwerten $\lambda_j \neq \lambda_k$ gilt*

$$\boldsymbol{t}_j'\boldsymbol{t}_k = \begin{cases} 0, & j \neq k \\ \|\boldsymbol{t}_j\| \neq 0, & j = k \end{cases} \tag{2.96}$$

d.h. Eigenvektoren mit verschiedenen zugehörigen Eigenwerten sind orthonormal.

Beweis: Ist (i) T eine Rotationsmatrix, ist (ii) M symmetrisch und gilt (iii) $MT = T\Lambda$, so ist T eine Matrix von Eigenvektoren und aus der Orthonormalität von Rotationsmatrizen folgt die Orthonormalität der Eigenvektoren. Es sei umgekehrt $T = [\mathbf{t}_1, \ldots, \mathbf{t}_n]$ eine Matrix von Eigenvektoren von M, d.h. für $\mathbf{t}_j, \mathbf{t}_k \in T$ mit $\lambda_j \neq \lambda_k$ gilt

$$M\mathbf{t}_j = \lambda_j\mathbf{t}_j \tag{2.97}$$
$$M\mathbf{t}_k = \lambda_k\mathbf{t}_k \tag{2.98}$$

mit$\lambda_j \neq \lambda_k$. Die Gleichung (2.97) werde von links mit $\mathbf{t}_k'$, die Gleichung (2.98) von links mit $\mathbf{t}_j$ multipliziert. Es entstehen die Gleichungen

$$\mathbf{t}_k'M\mathbf{t}_j = \lambda_j\mathbf{t}_k'\mathbf{t}_j \tag{2.99}$$
$$\mathbf{t}_j'M\mathbf{t}_k = \lambda_k\mathbf{t}_j'\mathbf{t}_k. \tag{2.100}$$

Nun ist einerseits $\mathbf{t}_j'\mathbf{t}_k = \mathbf{t}_k'\mathbf{t}_j$, und andererseits $(\mathbf{t}_k'M\mathbf{t}_j)' = \mathbf{t}_j'M\mathbf{t}_k$, da ja $M' = M$. Subtrahiert man also die zweite Gleichung von der ersten, ergibt sich

$$0 = (\lambda_j - \lambda_k)\mathbf{t}_j'\mathbf{t}_k,$$

woraus wegen $\lambda_j - \lambda_k \neq 0$ die Behauptung $\mathbf{t}_j'\mathbf{t}_k = 0$ folgt. □

Anmerkung: In (2.96) ist nicht gefordert worden, dass $\|\mathbf{t}_j\| = 1$ ist; $M\mathbf{t}_j = \lambda_j\mathbf{t}_j$ bedeutet, dass sich die Längen der Vektoren $M\mathbf{t}_j$ und $\mathbf{t}_j$ um den Faktor λ_j unterscheiden, unabhängig von der Länge von $\mathbf{t}_j$. Insofern ist die Länge eines Eigenvektors irrelevant und deswegen kann $\|\mathbf{t}_j\| = 1$ gesetzt werden, womit man die für weitere Betrachtungen nützliche Eigenschaft der Orthonormalität der Matrix T der Eigenvektoren erhält. Ist bereits bekannt, dass T auch eine Rotationsmatrix ist, so wird die Normiertheit der $\mathbf{t}_j$ gewissermaßen gleich mitgeliefert. □

Die Frage ist nun, welche Aussage über die Eigenvektoren einer symmetrischen Matrix gemacht werden kann, wenn nicht alle Eigenwerte voneinander verschieden sind. Es gilt der folgende Satz:

Satz 2.19 *Es sei λ_j ein Eigenwert der symmetrischen Matrix M mit der Mehrfachheit m, d.h. es gelte $\lambda_j = \lambda_{j+1} = \cdots = \lambda_{j+m}$. Dann existieren m orthogonale, zu diesen Eigenwerten korrespondierende Eigenvektoren.*

Beweis: Der Beweis wird hier nicht gegeben, da er auf Resultate zurückgreift, die in dieser Kurzfassung nicht dargestellt werden, weil sie im Folgenden nicht weiter gebraucht werden. Andererseits ist der Satz aber intuitiv klar: Die Matrix T der Eigenvektoren einer symmetrischen Matrix ergab sich aus der Eigenschaft der Orthonormalität von Rotationsmatrizen one Bezug auf die Eigenwerte der Eigenvektoren. n-dimensionale Vektoren $\mathbf{x} = T\mathbf{y}$, so dass T auf jeden Fall eine orthonormale (m, n)-Matrix ist, d.h. sie enthält in jedem Fall n orthonormale Spalten- und damit auch n orthonormale Zeilenvektoren, unabhängig davon, ob es identische Eigenwerte gibt oder nicht. Damit ist auch der Spezialfall $\lambda_{r+1} = \cdots = \lambda_n = 0$ abgedeckt, wenn also die (m, n)-Matrix M einen Rang $r < n$ hat. Es gilt $M = T\Lambda T' = AT'$, $A = T\Lambda$, und die Spaltenvektoren von M sind Linearkombinationen der Spalten von A. Dabei gehen nur die ersten r Spaltenvektoren von A in die Berechnung der Spalten von M ein, weil nur die ersten Spaltenvektoren von A ungleich Null sind (man überzeuge sich durch Nachrechnen). Gleichwohl enthält T insgesamt n paarweise orthogonale Vektoren. □

Bemerkung 2.1 Eine Matrix muß nicht symmetrisch sein, damit Eigenvektoren für sie existieren, und andererseits existieren nicht für jede symmetrische Matrix Eigenvektoren mit *reellwertigen* Komponenten (d.h. es ist möglich, dass Eigenvektoren mit komplexwertigen Komponenten existieren, vergl. Satz 2.22, 88). Dazu betrachte man die Matrix (2.74), Seite 64, d.h.

$$T = \begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix}.$$

Dann ist

$$T\mathbf{x} = x_1 \begin{pmatrix} \cos\phi \\ \sin\phi \end{pmatrix} + x_2 \begin{pmatrix} -\sin\phi \\ \cos\phi \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \mathbf{y}.$$

$\mathbf{x}$ ist ein Eigenvektor von T, wenn $\mathbf{y} = \lambda\mathbf{x}$, d.h. wenn sie dieselbe Orientierung haben, also wenn

$$\frac{x_2}{x_1} = \frac{y_2}{y_1}$$

gilt. Es ist aber

$$\frac{y_2}{y_1} = \frac{x_1 \sin\phi + x_2 \cos\phi}{x_1 \cos\phi - x_2 \sin\phi},$$

so dass $\mathbf{y}$ parallel zu $\mathbf{x}$ nur für spezielle Werte von ϕ ist, für die $\cos\phi = 1$ und $\sin\phi = 0$ gilt, also z.B. für $\phi = 0$, so dass $T = I$ mit den Spaltenvektoren

$(1,0)'$ und $(0,1)'$. Dies ist der gewissermaßen triviale Fall, bei dem gar keine Rotation erzeugt wird. Man findet allerdings komplexwertige Eigenvektoren mit zugehörigen komplexwertigen Eigenwerten, – für $\phi = \pi/4$ etwa findet man die Eigenvektoren $(i,1)'$ und $(-i,1)'$ mit den Eigenwerten $(1+i)/\sqrt{2}$ und $(1-i)/\sqrt{2}$, mit $i = \sqrt{-1}$, wie man durch Nachrechnen bestätigt. Komplexe Eigenvektoren und - werte werden allerdings im Folgenden keine Rolle spielen.

2.6.4 Inverse und Wurzel einer symmetrischen Matrix

Die Inverse von M: Es sei M eine symmetrische (n,n)-Matrix und M habe vollen Rang, so dass $\mathrm{rg}(M) = n$. Die Spektraldarstellung von M sei $M = T\Lambda T'$. Da M vollen Rang hat, existiert die Inverse M^{-1} von M. Es ist (vergl. Satz 2.70, Seite 62)

$$M^{-1} = (T\Lambda T')^{-1} = (T')^{-1}\Lambda^{-1}T^{-1},$$

wobei $\Lambda^{-1} = \mathrm{diag}(\lambda_1^{-1}, \ldots, \lambda_N^{-1})$. Die Orthonormalität von T impliziert $T' = T^{-1}$, so dass $(T')^{-1} = (T^{-1})^{-1} = T$, so dass die Inverse von M durch

$$M^{-1} = T\Lambda^{-1}T' = \sum_{k=1}^{n} \frac{1}{\lambda_k}\mathbf{t}_k\mathbf{t}_k' \tag{2.101}$$

gegeben ist. Die $\mathbf{t}_k\mathbf{t}_k'$ sind die dyadischen Produkte der Eigenvektoren von M.

Anmerkung: Die Beziehung (2.101) ist von Bedeutung u.a. bei der Interpretation von Regressionsparametern: die Stichprobenvarianzen der Schätzungen der Regressionsparameter $\mathbf{b} = (b_1, \ldots, b_n)'$ einer multiplen Regression $\mathbf{y} = X\mathbf{b} + \mathbf{e}$, X die zentrierte Matrix der Prädiktorwerte, hängen von den Eigenwerten λ_k der Kovarianzmatrix $M = X'X$ ab, und die Stichprobenvarianzen sind proportional zu $(X'X)^{-1}$. Gleichung (2.104) zeigt, dass diese Varianzen groß werden, wenn es kleine Eigenwerte λ_k gibt.

Die Wurzel von M: Es sei $\Lambda^{1/2} = \mathrm{diag}(\lambda_1^{1/2}, \ldots, \lambda_n^{1/2})$. Dann gilt sicherlich

$$M = T\Lambda^{1/2}\Lambda^{1/2}T'.$$

Es sei nun[26]

$$M^{1/2} =_{\mathrm{Def}} T\Lambda^{1/2}T' = \sum_{k=1}^{n} \sqrt{\lambda_k}\mathbf{t}_k\mathbf{t}_k' \tag{2.102}$$

$\Lambda^{1/2}T'$ kann sicherlich berechnet werden, so dass der Ausdruck $M^{1/2}$ einer berechenbaren Größe entspricht. Darüber hinaus entspricht sie der üblichen Schreibweise $a^{1/2}a^{1/2} = a$ für $a \in \mathbb{R}$, denn

$$M^{1/2}M^{1/2} = T\Lambda^{1/2}T'T\Lambda^{1/2}T' = T\Lambda T'.$$

[26]Mit dem Zeichen $=_{\mathrm{Def}}$ soll ausgedrückt werden, dass der Ausdruck auf der linken Seite durch den auf der rechten Seite definiert wird.

Weiter folgt

$$(M^{1/2})' = (T\Lambda^{1/2}T')' = T\Lambda^{1/2}T', \tag{2.103}$$

d.h. $M^{1/2}$ ist symmetrisch, und

$$(M^{1/2})^{-1} = M^{-1/2} = (T\Lambda^{1/2}T')^{-1} = T\Lambda^{-1/2}T' = \sum_{k=1}^{n} \frac{1}{\sqrt{\lambda_k}} \mathbf{t}_k \mathbf{t}_k'. \tag{2.104}$$

2.6.5 Der Rayleigh-Quotient

In Abschnitt 2.6.3 wurde gezeigt, dass die Transformationsmatrix T, die ein achsenparalleles Ellipsoid in ein nicht achsenparalleles Ellipsoid überführt (rotiert), die Matrix der Eigenvektoren der (n, n)-Matrix $M = X'X$ ist. Der erste, d.h. der zum größten Eigenwert korrespondierende Eigenvektor entspricht dabei der Orientierung der ersten Hauptachse des nicht-achsenparallelen Ellipsoids. Dieser Sachverhalt legt nahe, dass diese Orientierung die der größten Ausdehnung der Punktekonfiguration der Fälle ist. Das ist in der Tat der Fall.

Man betrachte die Projektion der Fälle auf eine durch die Konfiguration der Fälle verlaufende Achse. Die Projektionen sind eine Funktion der Orientierung der Achse. Es sei $\mathbf{y}$ der Vektor, dessen Komponenten durch die Projektionen, die als Koordinaten der Fälle auf dieser Achse gesehen werden können, definiert sind, und der Einfachheit halber werde angenommen, dass die Komponenten zentriert worden seien. Das Quadrat der Länge $\mathbf{y}'\mathbf{y} = \|\mathbf{y}\|^2$ von $\mathbf{y}$ ist proportional dem Quadrat der Länge von $\mathbf{y}$; der Proportionalitätsfaktor ist $1/m$, der aber für die folgenden Betrachtungen vernachlässigt werden kann. Sicherlich hängt der Wert von $\|\mathbf{y}\|^2$ von der Orientierung der Geraden, also von $\mathbf{y}$ ab. Entspricht ihre Orientierung der Richtung der maximalen Ausdehnung der Konfiguration, so wird $\|\mathbf{y}\|^2$ und damit die Varianz maximal. Da nur die Datenmatrix X gegeben ist und $\mathbf{y}$ aus den Daten zu errechnen ist, folgt, dass $\mathbf{y}$ nur als Linearkombination der Spaltenvektoren $\mathbf{x}_j$ der Datenmatrix X gegeben sein kann. Also muß $\mathbf{y} = X\mathbf{s}$ gelten, wobei $\mathbf{s}$ ein Koeffizientenvektor ist. Die triviale Lösung für das Maximumproblem ist $\|\mathbf{y}\|^2 = \infty$, die aber keine Lösung ist, weil nur endliche $\|\mathbf{y}\|$-Werte sinnvoll interpretiert werden können. $\|\mathbf{y}\|^2 < \infty$ setzt $\|\mathbf{s}\| < \infty$ voraus, d.h. man muß eine endliche Länge für $\mathbf{s}$ vorgeben. Da es nur auf die Orientierung von $\mathbf{s}$ ankommt, kann dieser endliche Wert für $\|\mathbf{s}\|$ beliebig gewählt werden, so dass man insbesondere $\|\mathbf{s}\| = 1$ setzen kann. Es ist $\mathbf{y}'\mathbf{y} = \mathbf{s}'X'X\mathbf{s} = \mathbf{s}'M\mathbf{s}$, d.h. $M = X'X$. Das Ziel ist, den Vektor $\mathbf{s}_{\max}$ zu bestimmen, für den $\mathbf{y} = \mathbf{y}_{\max} = \mathbf{L}$ gilt; die Bezeichnung $\mathbf{L}$ soll daran erinnern, dass es sich um die Bestimmung von "latenten" Vektoren handelt. Man kann für $\mathbf{s}$ zunächst einen beliebigen Vektor $\mathbf{x}$ betrachten und der Forderung nach der Normiertheit von $\mathbf{s}$ (also der Forderung $\|\mathbf{s}\| = 1$) dadurch gerecht werden, dass man für $\mathbf{s}$ den Vektor $\frac{\mathbf{x}}{\|\mathbf{x}\|}$ einsetzt. $\mathbf{s}'M\mathbf{s}$ wird dann zum *Rayleigh-Quotienten*:

Definition 2.9 *Es sei M eine symmetrische Matrix. Dann heißt*

$$R(\boldsymbol{x}) = \frac{\boldsymbol{x}'}{\|\boldsymbol{x}\|} M \frac{\boldsymbol{x}}{\|\boldsymbol{x}\|} = \frac{\boldsymbol{x}' M \boldsymbol{x}}{\|\boldsymbol{x}\|^2} = \frac{\boldsymbol{x}' M \boldsymbol{x}}{\boldsymbol{x}' \boldsymbol{x}} \tag{2.105}$$

Rayleigh-Quotient[27] *oder* Rayleigh-Koeffizient.

Eine Methode zur Maximierung von $R(\mathbf{x})$ besteht darin, diesen Quotienten durch Anwendung der Differentialrechung zu maximieren. Dieses Vorgehen wird im Anhang, Abschnitt 4.4 besprochen. In diesem Abschnitt wird ein anderer Weg beschritten, der ohne die Anwendung der Differentialrechnung auskommt.

Satz 2.20 *(Satz von Courant-Fischer) Es sei M eine symmetrische, positiv definite Matrix mit $M = T\Lambda T'$, T die Matrix der Eigenvektoren und*

$$\Lambda = diag(\lambda_1, \ldots, \lambda_n), \quad \lambda_1 \geq \cdots \geq \lambda_n$$

die Diagonalmatrix der zugehörigen Eigenwerte von M. Dann ist

$$\max_{\boldsymbol{x} \neq \vec{0}} R(\boldsymbol{x}) = \max_{\boldsymbol{x} \neq \vec{0}} \frac{\boldsymbol{x}' M \boldsymbol{x}}{\boldsymbol{x}' \boldsymbol{x}} = \max_j \lambda_j = \lambda_1, \tag{2.106}$$

und der Vektor $\boldsymbol{x}$, für den das Maximum angenommen wird, ist der zu λ_1 korrespondierende Eigenvektor $\boldsymbol{t}_1$, so dass $\max_{\boldsymbol{x} \neq \vec{0}} R(\boldsymbol{x}) = \boldsymbol{t}_1' M \boldsymbol{t}_1 = \lambda_1$. Weiter gilt

$$\max_{\boldsymbol{x} \perp \boldsymbol{t}_1, \ldots, \boldsymbol{t}_k} \frac{\boldsymbol{x}' M \boldsymbol{x}}{\boldsymbol{x}' \boldsymbol{x}} = \lambda_{k+1}, \quad k < n, \tag{2.107}$$

für $\boldsymbol{x} = \boldsymbol{t}_{k+1}$ der $(k+1)$-te Eigenvektor ($\perp$ steht für "ist orthogonal zu".), und

$$\min_{\boldsymbol{x} \neq \vec{0}} R(\boldsymbol{x}) = \frac{\boldsymbol{x}' M \boldsymbol{x}}{\boldsymbol{x}' \boldsymbol{x}} = \min_j \lambda_j, \tag{2.108}$$

mit dem zugehörigen Eigenvektor $\boldsymbol{t}_{\min}$, d.h. $\min_{\boldsymbol{x} \neq \vec{0}} R(\boldsymbol{x}) = \boldsymbol{t}_{\min} M \boldsymbol{t}_{\min} = \lambda_{\min}$.

Beweis: Sei $T = [\mathbf{t}_1, \ldots, \mathbf{t}_n]$ die Matrix der Eigenvektoren von M mit den zugehörigen Eigenwerten $\lambda_1 \geq \cdots, \geq \lambda_n$, so dass $MT = T\Lambda$, $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$. Die Eigenvektoren $\mathbf{t}_j$, $j = 1, \ldots, n$ bilden eine orthonormale Basis des $V_n = \mathcal{L}(\mathbf{t}_1, \ldots, \mathbf{t}_n)$. Es sei nun $\mathbf{x}$ ein beliebiger n-dimensionaler Vektor. Er kann stets als Linearkombination einer Basis. Es gilt also insbesondere

$$\mathbf{x} = c_1 \mathbf{t}_1 + \cdots + c_n \mathbf{t}_n$$

für für geeignet gewählte Koeffizienten $c_1, \ldots, c_n$. Dann ist

$$\mathbf{x}' M \mathbf{x} = (c_1 \mathbf{t}_1 + \cdots + c_n \mathbf{t}_n)' M (c_1 \mathbf{t}_1 + \cdots + c_n \mathbf{t}_n) = \sum_{j=1}^{n} c_j^2 \mathbf{t}_j' M \mathbf{t}_j = \sum_{j=1}^{n} c_j^2 \lambda_j.$$

[27] Nach dem britischen Physiker John William Strutt, Dritter Baron Rayleigh (1842–1919)

Alle Kreuzprodukte $c_j c_k \mathbf{t}_j' \mathbf{t}_k$ verschwinden, weil $\mathbf{t}_j' \mathbf{t}_k = 0$ für alle $j \neq k$, so dass

$$\frac{\mathbf{x}' M \mathbf{x}}{\mathbf{x}' \mathbf{x}} = \frac{\sum_{j=1}^n c_j^2 \lambda_j}{\sum_{j=1}^n c_j^2}$$

und $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. Ersetzt man also auf der rechten Seite alle $\lambda_j \neq \lambda_1$ durch λ_1, so folgt

$$\frac{\mathbf{x}' M \mathbf{x}}{\mathbf{x}' \mathbf{x}} = \frac{\sum_{j=1}^n c_j^2 \lambda_j}{\sum_{j=1}^n c_j^2} \leq \frac{\lambda_1 \sum_{j=1}^n c_j^2}{\sum_{j=1}^n c_j^2} = \lambda_1,$$

so dass $\max_{\mathbf{x}} R(\mathbf{x}) = \lambda_1$, λ_1 der größte Eigenwert. Jetzt muß noch gezeigt werden, für welchen Vektor $\mathbf{x}$ das Maximum angenommen wird. Es sei $\mathbf{t}_1$ der zu $\lambda_{\max} = \lambda_1$ korrespondierende Eigenvektor. Für $\mathbf{x} = \mathbf{t}_1$ folgt $\mathbf{t}_1' M \mathbf{t}_1 = \lambda_1$ und wegen $\mathbf{t}_1' \mathbf{t}_1 = 1$ sieht man, dass $R(\mathbf{x}) = \max$ für $\mathbf{x} = \mathbf{t}_1$.

Die Aussage (2.108) ergibt sich aus (2.107). Es sei wieder $T = [\mathbf{t}_1, \cdots, \mathbf{t}_n]$ die Matrix der Eigenvektoren von M. Dann existieren relle Zahlen $y_1, \ldots, y_n$ derart, dass ein Vektor $\mathbf{x}$ in der Form $\mathbf{x} = T\mathbf{y}$ dargestellt werden kann, wobei die Komponenten von $\mathbf{y}$ durch die y_j gegeben sind. Nun soll speziell $\mathbf{x} \perp \mathbf{t}_1, \ldots, \mathbf{t}_k$ gelten. Dann folgt aber

$$\mathbf{t}_k' \mathbf{x} = y_1 \mathbf{t}_k' \mathbf{t}_1 + y_2 \mathbf{t}_k' \mathbf{t}_2 + \cdots + y_n \mathbf{t}_k' \mathbf{t}_n = y_k = 0,$$

denn $\mathbf{t}_k' \mathbf{t}_k = 1$, so dass $y_k \mathbf{t}_k' \mathbf{t}_k = y_k$. Die Forderung der Orthogonalität von $\mathbf{x}$ zu den ersten k Eigenvektoren impliziert also $y_1 = \cdots = y_k = 0$. Dann folgt aus (2.106)

$$\frac{\mathbf{x}' M \mathbf{x}}{\mathbf{x}' \mathbf{x}} = \frac{\sum_{j=k+1}^n \lambda_j y_j^2}{\sum_{j=k+1}^n y_j^2},$$

und analog zur Argumentation im Beweis zu Satz 2.20, Seite 81, folgt (2.107). □

Anmerkungen:

1. Der Beweis geht davon aus, dass die Eigenvektoren $\mathbf{t}_1, \ldots, \mathbf{t}_n$ als Basis des V_n gewählt werden können. Diese Wahl ist keine Einschränkung der Allgemeinheit, da man von einer beliebigen Basis V stets zu einer orthonormalen Basis übergehen kann, vergl. die Singularwertzerlegung (Abschnitt 2.6.6), angewendet auf V. Deswegen bedeutet dieser Übergang auch nur, einen Umweg zu gehen: man kann gleich mit den Eigenvektoren als Basis beginnen.

2. Für $k = 0$ betrachtet man $\max_{\mathbf{x} \neq \vec{0}} \frac{\mathbf{x}' M \mathbf{x}}{\mathbf{x}' \mathbf{x}} = \lambda_1$, und für $k = n-1$ erhält man

$$\max_{\mathbf{x} \perp \mathbf{t}_1, \ldots, \mathbf{t}_{n-1}} \frac{\mathbf{x}' M \mathbf{x}}{\mathbf{x}' \mathbf{x}} = \lambda_n,$$

d.h.

$$\min_{\mathbf{x}} R(\mathbf{x}) = \lambda_n \tag{2.109}$$

so dass

$$\lambda_n \leq R(\mathbf{x}) \leq \lambda_1. \tag{2.110}$$

3. Es sei M eine beliebige symmetrische (m, n)-Matrix und $\mathbf{x} \in \mathbb{R}^n$. Weiter sei $\lambda = \mathbf{x}'M\mathbf{x}/\mathbf{x}'\mathbf{x}$. Die Werte λ heißen *maximal*, wenn sie den Rayleigh-Quotienten $\mathbf{x}'M\mathbf{x}/\mathbf{x}'\mathbf{x}$ maximieren.

4. Es sei T die Matrix der Eigenvektoren von $M = X'X$; für irgendzwei Eigenvektoren $\mathbf{t}_j$ und $\mathbf{t}_k$, $j \neq k$, gilt also $\mathbf{t}_j'\mathbf{t}_k = 0$. Es sei $L = XT$, so dass $\mathbf{L}_j = X\mathbf{t}_j$, $j = 1, \ldots, n$. Dann gilt

$$\mathbf{L}_j'\mathbf{L}_k = \mathbf{t}_j'X'X\mathbf{t}_k = \mathbf{t}_j'M\mathbf{t}_k = 0, \tag{2.111}$$

d.h. die Matrix L ist orthogonal. Denn $M = T\Lambda T'$, Λ die Diagonalmatrix der Eigenwerte on M. Dann hat man

$$\mathbf{t}_j'M\mathbf{t}_k = \mathbf{t}_j'T\Lambda T'\mathbf{t}_k = \mathbf{e}_j'\Lambda\mathbf{e}_k = \lambda_j\mathbf{e}_j'\mathbf{e}_k = 0, \quad j \neq k.$$

□

2.6.6 Die Singularwertzerlegung einer Matrix

Die im Folgenden dargestellte Zerlegung einer Matrix gilt für beliebige (m, n)-Matrizen, und die Betrachtung von Datenmatrizen erleichtert möglicherweise den intuitiven Zugang. Es geht um die Frage, wie latente Variablen zur Erklärung von beobachteten Daten bestimmt werden können. Die Daten seien in einer (m, n)-Matrix X zusammengefasst worden, bei der die Komponenten der Spaltenvektoren $\mathbf{x}_j$ die Messungen x_{ij} der j-ten Variablen $(1 \leq j \leq n)$ bei $i = 1, \ldots, m$ Fällen repräsentieren. Die Komponenten der Zeilenvektoren $\tilde{\mathbf{x}}_i$ sind die Messungen x_{ij}, $j = 1, \ldots, n$, beim i-ten Fall für alle n Variablen.

Die Betrachtungen der vorangegangenen Abschnitte können in einem Satz[28] über die *Singularwertzerlegung* der Matrix X zusammengefasst werden, der für die multivariate Analyse von Daten von zentraler Bedeutung ist:

Satz 2.21 *Es sei X eine (m, n)-Matrix mit dem Rang $r \leq \min(m, n)$. Dann gilt die* Singularwertzerlegung *(SVD[29])*

$$X = Q\Sigma T', \quad \Sigma = \Lambda^{1/2}, \tag{2.112}$$

[28] Der Satz ist einfach, hat aber eine längere Vorgeschichte, vergl. Stewart (1993).

[29] Übliche Abkürzung für Singularwertzerlegung, von engl. Singular Value Decomposition.

wobei Λ die (m,n)-Matrix der Form

$$\Lambda = \left(\begin{array}{ccc|ccc} \lambda_1 & & & & \vdots & \\ & \ddots & & \cdots & 0 & \cdots \\ & & \lambda_r & & \vdots & \\ \hline \vdots & & & & \vdots & \\ 0 & \cdots & & \cdots & 0 & \cdots \\ \vdots & & & & \vdots & \end{array}\right) \tag{2.113}$$

ist, d.h. Λ enthält nur Nullen mit Ausnahme der ersten r Diagonalzellen, in denen die Eigenwerte von XX' und $X'X$ ungleich Null stehen; diese Eigenwerte sind für XX' und $X'X$ identisch. Q ist die (m,n)-Matrix der Eigenvektoren von XX' und T ist die (n,n)-Matrix der Eigenvektoren von $X'X$, die jeweils zu Eigenwerten ungleich Null korrespondieren.

Anmerkungen: Die Dimensionen der Matrizen Q und T sind so gewählt worden, dass allen möglichen Fällen ($m > n$, $m = n$ und $m < n$) Genüge getan wird. Die Diagonalmatrix Σ enthält offenbar die Wurzeln aus den Eigenwerten; sie heißen auch *Singularwerte*; $\sigma_j = \sqrt{\lambda_j}$, $j = 1, \ldots, r$. Die Spaltenvektoren von Q heißen *Linkseigenvektoren*, die von T heißen *Rechtseigenvektoren* von X. Ein anderer, häufig gebrauchter Ausdruck für die SVD ist 'Grundstruktur' einer Matrix (engl. basic structure). □

Beweis: Es sei T eine orthonormale (n,n)-Matrix und es sei $L = XT$[30], L eine (m,n)-Matrix, wobei T so gewählt werde, dass

$$T'X'XT = L'L = \Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$$

d.h. Λ sei eine Diagonalmatrix, die Spaltenvektoren $\mathbf{L}_k$, $k = 1, \ldots, n$ von L seien also orthogonal[31]. Multiplikation von $XT = L$ von rechts mit T' liefert die Gleichung $X = LT'$. Die Spaltenvektoren von L werden durch Multiplikation von rechts mit $\Lambda^{-1/2}$ normalisiert: es entsteht die Matrix $Q = L\Lambda^{-1/2}$ mit orthonormalen Spaltenvektoren $\mathbf{q}_k$ (s. Längenskalierung, Seite 49). Dann kann $L = Q\Lambda^{1/2}$ geschrieben werden und es resultiert

$$X = Q\Lambda^{1/2}T',$$

also (2.112). Man hat weiter

$$X'X = TL'LT' = T\Lambda T'.$$

[30] d.h. $T'X' = L'$ und T' rotiert die $\tilde{\mathbf{x}}_i$, also die die Fälle repräsentierenden Zeilenvektoren von X, in die Vektoren $\tilde{\mathbf{L}}_i = (\ell_{i1}, \ldots, \ell_{in})'$, die die Fälle in einem neuen Koordinatensystem abbilden.

[31] Die Notation ist im Vergleich zu der in Abschnitt 2.6.2 verändert worden, um der üblichen Schreibweise für die SVD zu entsprechen: die Matrix T heißt dort T'.

Die Multiplikation von rechts mit T liefert $X'XT = T\Lambda$, d.h. T ist die Matrix der Eigenvektoren von $X'X$ und Λ ist die Diagonalmatrix der zugehörigen Eigenwerte von $X'X$. Ebenso hat man

$$XX' = LT'TL' = LL' = Q\Lambda^{1/2}\Lambda^{1/2}Q' = Q\Lambda Q',$$

so dass nach Multiplikation von rechts mit Q die Beziehung $XX'Q = Q\Lambda$ folgt, d.h. Q ist die Matrix der ersten n Eigenvektoren von XX' und Λ enthält die von Null verschiedenen Eigenwerte von XX'; die von Null verschiedenen Eigenwerte von $X'X$ und XX' sind also identisch. □

Anmerkungen:

1. Die SVD $X = Q\Lambda^{1/2}T'$ repräsentiert offenbar eine spezielle Wahl für die Matrizen U und V' für die Beziehung $X = UV'$: man kann $U = Q$ und $V = T\Lambda^{1/2}$ setzen, oder $U = Q\Lambda^{1/2}$ und $V = T$. Noch allgemeiner kann man $X = Q\Lambda^{\frac{p}{2}}\Lambda^{\frac{1-p}{2}}T'$ setzen, mit $0 \leq p \leq 1$, um dann $U = Q\Lambda^{\frac{p}{2}}$ und $V = T\Lambda^{\frac{1-p}{2}}$ anzunehmen; üblicherweise werden allerdings nur die Fälle $p = 1$ oder $p = 0$ betrachtet.

2. Aus $X'X = T\Lambda T'$ folgt $T'X'XT = \Lambda$. Es sei $\mathbf{v}$ ein n-dimensionaler Vektor und es werde die quadratische Form $\mathbf{v}'X'X\mathbf{v} = \mu \in \mathbb{R}$ betrachtet. Nach dem Satz von Courant-Fischer nimmt μ den maximal möglichen Wert λ_1 an, wenn $\mathbf{v} = \mathbf{t}_1$ (es wird die übliche Annahme gemacht, dass die λ_k in λ der Größe nach geordnet angeschrieben werden und die Anordnung der Eigenvektoren $\mathbf{t}_k$ in T der der Eigenwerte in λ entspricht. Da $\mathbf{L}_1'\mathbf{L}_1 = \lambda_1$ ist klar, dass die Projektionen der Fälle auf die erste Hauptachse, also die Komponenten von $\mathbf{L}_1$, die größte Ausdehnung der Konfiguration definieren. $\mathbf{L}_2$ repräsentiert dann die zweitgrößte Ausdehnung, etc. Die SVD entspricht einer Hauptachsentransformation.

3. Die SVD kann auch in der Form $X = QA'$ geschrieben werden, wobei $A = T\Lambda^{1/2}$. Die Spaltenvektoren $\mathbf{x}_j$ von X können als Linearkombinationen einerseits der $\mathbf{L}_k$ geschrieben werden:

$$\mathbf{x}_j = t_{1j}\mathbf{L}_1 + \cdots + t_{nj}\mathbf{L}_n = L\tilde{\mathbf{t}}_j, \tag{2.114}$$

   $\tilde{\mathbf{t}}_j$ der j-te Spaltenvektor von T' (Zeilenvektor von T): oder als Linearkombinationen der $\mathbf{q}_k$, der Spaltenvektoren von Q:

$$\mathbf{x}_j = a_{1j}\mathbf{q}_1 + \cdots + a_{nj}\mathbf{q}_n = Q\tilde{\mathbf{a}}_j, \tag{2.115}$$

   $\tilde{\mathbf{a}}_j$ der j-te Spaltenvektor von A' (Zeilenvektor von A). Die t_{kj} bzw. die a_{kj} beziehen sich offenbar auf die j-te gemessene Variable; sie indizieren das Gewicht, mit dem die k-te latente Variable in die j-te gemessene Variable eingeht. Eine analoge Aussage gilt für die t_{kj}. Die t_{kj} und die a_{kj} repräsentieren verschiedenen Skalierungen der Variablen, vergl. Abschnitt 2.10.1.

4. Die SVD ist nicht an eine Spaltenzentrierung oder Standardisierung der Matrix X gebunden; eine SVD kann für eine beliebige Matrix X bestimmt werden.

Eine Anwendung der SVD wird in Abschnitt 2.10.1 vorgestellt.

Zur Beziehung zwischen den Eigenvektoren von XX' und $X'X$: Nach Multiplikation der SVD $X = Q\Sigma T'$ von rechts mit T erhält man

$$XT = Q\Sigma, \tag{2.116}$$

und die Multiplikation der SVD von links mit Q' ergibt $Q'X = \Sigma T'$, oder

$$X'Q = T\Sigma, \tag{2.117}$$

so dass man für die Eigenvektoren $\mathbf{q}_k$ aus Q und $\mathbf{t}_k$ aus T die Beziehungen

$$X\mathbf{t}_k = \sigma_k \mathbf{q}_k \tag{2.118}$$

$$X'\mathbf{q}_k = \sigma_k \mathbf{t}_k, \quad k = 1, \ldots, n \tag{2.119}$$

erhält. Diese beiden Gleichungen zeigen die Beziehungen zwischen den $\mathbf{t}_k$ und den $\mathbf{q}_k$: $\sigma_k \mathbf{q}_k$ ist eine Linearkombination der Spalten von X mit den Komponenten von $\mathbf{t}_k$ als Koeffizienten, und $\sigma_k \mathbf{q}_k$ ist eine Linearkombination der Zeilen von X (Spalten von X') mit den Komponenten von $\mathbf{q}_k$ als Koeffizienten.

Darstellung der SVD über das dyadische Produkt: Die Zerlegung $X = Q\Sigma T'$ kann in der Form

$$X = \sum_{k=1}^{n} \sigma_k \mathbf{q}_k \mathbf{t}_k', \quad \sigma_k = \sqrt{\lambda_k},\ m \geq n \tag{2.120}$$

dargestellt werden, wobei $\mathbf{q}_j$ die Spaltenvektoren von Q und $\mathbf{t}_j$ die Spaltenvektoren von T sind. $\mathbf{q}_j \mathbf{t}_j'$ ist das dyadische Produkt dieser Vektoren. X kann also als Summe von Matrizen aufgefasst werden, die jeweils eine Dimension repräsentieren. Das Element x_{ij} ist demnach durch die Summe

$$x_{ij} = \sigma_1 q_{i1} t_{j1} + \cdots + \sigma_n q_{in} t_{jn} \tag{2.121}$$

gegeben, $\sigma_1 q_{i1} t_{j1}$ ist der Beitrag der ersten latenten Dimension, etc.

Üblicherweise werden die Eigenwerte λ_k in Λ so angordnet, dass λ_1 der größte und λ_n der kleinste Eigenwert ist, – numerisch sind üblicherweise schon aufgrund von Rundungsfehlern alle $\lambda_k \neq 0$, auch wenn der "wahre" Rang von X kleiner als $\min(m, n)$ ist. Man hat also

$$\sigma_1 > \sigma_2 > \cdots > \sigma_n.$$

Die Matrix X wird in (2.120) als Summe der als dyadische Produkte der Eigenvektoren $\mathbf{q}_k$ von XX' und der Eigenvektoren $\mathbf{t}_k$ von $X'X$ definierten Matrizen,

jeweils gewichtet mit $\sigma_k = \sqrt{\lambda_k}$, dargestellt. Betrachtet man die σ_k für einen bestimmten Wert $r < k$ als "hinreichend" klein, kann man X durch eine Matrix $\hat{X}_r$ approximieren:

$$\hat{X}_r = \sum_{k=1}^{r} \sigma_k \mathbf{q}_k \mathbf{t}_k' \approx X, \quad \sigma_k = \sqrt{\lambda_k}, \quad r < n \tag{2.122}$$

"Hinreichend" klein bedeutet, dass man die σ_k mit $k > r$ als eben nur "zufällig" von Null abweichend betrachtet. Darauf wird in Abschnitt 2.10.1 noch eingegangen.

Anmerkung: In Abschnitt 4.4.4, Seite 151, wird der Satz von Courant-Fischer durch Differentiation der quadratischen Form (2.107), Seite 81 (hier also $T'X'XT = L'L = \Lambda$) unter der Nebenbedingung $\mathbf{t}'\mathbf{t} = 1$ bewiesen; dieser Herleitung entnimmt man leicht, dass es nur eine Lösung für T gibt, eben die Matrix der Eigenvektoren von $X'X$, denn die Ableitung $Q(\mathbf{t}) = d(\mathbf{t}'X'X\mathbf{t})/d\mathbf{t} = 0$ hat nur eine Lösung für $\mathbf{t}$. Es gibt also keine Rotation $T_1 \neq T$ derart, dass $XT_1 = L_1$ mit $L_1'L_1 = \Lambda_1$, Λ_1 eine Diagonalmatrix. Wählt man demnach eine von T verschiedene Matrix T_1, um die Vektoren von X zu rotieren, so repräsentieren die zu T_1 korrespondierenden $\mathbf{L}_k^{(1)}$ ein Koordinatensystem, in Bezug auf das die Konfiguration der Fälle nicht mehr achsenparallel ist, d.h. die latenten Variablen sind nicht mehr unkorreliert. □

2.7 Eigenvektoren und Eigenwerte nichtsymmetrischer Matrizen

2.7.1 Der allgemeine Fall

Der Begriff des Eigenvektors und der des zugehörigen Eigenwerts ergab sich in Abschnitt 2.6.3 bei der Betrachtung einer Koordinatentransformation auf eine natürliche Art und Weise für den Spezialfall symmetrischer Matrizen. Für nichtsymmetrische quadratische Matrizen können ebenfalls Eigenvektoren existieren, die aber nicht notwendig reell sind; so sei A eine orthonormale Matrix. Das Produkt von A mit einem Vektor $\mathbf{x}$ liefert einen Vektor $\mathbf{y}$, der sich von $\mathbf{x}$ möglicherweise durch eine Rotation unterscheidet: so sei

$$A = \begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix}.$$

Dann ist

$$A\mathbf{x} = x_1 \begin{pmatrix} \cos\phi \\ \sin\phi \end{pmatrix} + x_2 \begin{pmatrix} -\sin\phi \\ \cos\phi \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \mathbf{y},$$

und $\mathbf{y}$ ist nur parallel zu $\mathbf{x}$ für diejenigen Werte von ϕ, für die $\cos\phi = 1$ und $\sin\phi = 0$ ist, also z.B. für $\phi = 0$, so dass $A = I$ mit den Spaltenvektoren $(1, 0)'$ und $(0, 1)$. Dies ist der gewissermaßen triviale Fall, bei dem gar keine Rotation erzeugt wird. Man findet allerdings komplexwertige Eigenvektoren mit zugehörigen komplexwertigen Eigenwerten, – für $\phi = \pi/4$ etwa findet man die Eigenvektoren

$(i,1)'$ und $(-i,1)'$ mit den Eigenwerten $(1+i)/\sqrt{2}$ und $(1-i)/\sqrt{2}$, mit $i=\sqrt{-1}$, wie man durch Nachrechnen bestätigt. Wie komplexe Eigenvektoren und - werte zu deuten sind, wird später noch besprochen werden.

Charakteristische Gleichung einer Matrix: Es sei A eine quadratische Matrix. Aus $A\mathbf{u} = \lambda\mathbf{u}$ folgt $A\mathbf{u} - \lambda\mathbf{u} = \lambda I\mathbf{u} = \vec{0}$, I die Einheitsmatrix; die Dimensionen von I entsprechen denen von A. Diese Gleichung kann in der Form

$$(A - \lambda I)\mathbf{u} = \vec{0} \tag{2.123}$$

geschrieben werden. Diese Gleichung beschreibt ein homogenes Gleichungssystem: In ausgeschriebener Form hat man

$$\begin{aligned}
(a_{11}-\lambda)u_1 + a_{12}u_2 + \cdots + a_{1m}u_n &= 0 && (2.124)\\
a_{21}u_1 + (a_{22}-\lambda)u_2 + \cdots + a_{2n}u_n &= 0 && (2.125)\\
&\vdots && (2.126)\\
a_{n1}u_1 + a_{n2}u_2 + \cdots + (a_{nn}-\lambda)u_n &= 0 && (2.127)
\end{aligned}$$

Solche Gleichungssysteme haben nur dann mindestens eine nicht-triviale Lösung (d.h. eine Lösung, die nicht gleich dem Nullvektor $\vec{0}$ ist), wenn die Koeffizientenmatrix nicht vollen Rang hat, d.h. wenn ihre Determinante verschwindet, so dass

$$|A - \lambda I| = \begin{vmatrix} a_{11}-\lambda & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22}-\lambda & \cdots & a_{2n} \\ & & \ddots & \\ a_{n1} & a_{n2} & \cdots & a_{nn}-\lambda \end{vmatrix} = 0 \tag{2.128}$$

Entwickelt man die Determinante, so ergibt sich ein Polynom $P(\lambda)$ in λ vom Grad n:

$$|A - \lambda I| = (-1)^n[\lambda^n - \beta_1\lambda^{n-1} + \beta_2\lambda^{n-2} + \cdots + (-1)^{n-1}\beta_{n-1}\lambda + (-1)^n\beta_n] = 0. \tag{2.129}$$

Diese Gleichung heißt *charakteristische Gleichung* der Matrix A, wenn man den Faktor $(-1)^n$ wegläßt, und das Polynom auf der rechten Seite heißt *charakteristisches Polynom* von A. Die Gleichung hat, wie aus der Theorie der Polynome bekannt ist, insgesamt n Lösungen – also mögliche Eigenwerte von A –, von denen aber nicht alle identisch sein müssen. Die Nullstellen von $P(\lambda)$ sind die möglichen Eigenwerte von A. Bekanntlich existieren für ein Polnom n-ten Grades genau n Nullstellen $\lambda_1, \ldots, \lambda_n$, die allerdings nicht alle verschieden sein müssen und die durch komplexe Zahlen $\lambda = \alpha + i\beta$, $i = \sqrt{-1}$, gegeben sein können. Es gilt dabei

Satz 2.22 *Ist ein Eigenwert λ der quadratischen Matrix mit reellen Elementen komplex, so existiert ein zweiter Eigenwert $\bar{\lambda}$, der zu λ konjugiert komplex ist, dh gilt $\lambda = \alpha + i\beta$, so ist auch $\bar{\lambda} = \alpha - i\beta$ ein Eigenwert von A.*

Beweis: Es ist $(A - \lambda I)\mathbf{u} = 0$. Der Übergang zu konjugiert komplexen Zahlen führt zu $(\bar{A} - \bar{\lambda} I)\bar{\mathbf{u}} = 0$. Aber A ist als reell vorausgesetzt worden, also folgt

$$(A - \bar{\lambda} I)\bar{\mathbf{u}} = 0, \tag{2.130}$$

und dies heißt, dass $\bar{\lambda}$ ebenfalls ein Eigenwert von A ist. □

Links- und Rechtseigenvektoren: Es sei A eine nicht notwendig symmetrische (m, n)-Matrix, und für einen n-dimensionalen Vektor $\mathbf{u}$ gelte die Beziehung

$$A\mathbf{u} = \lambda\mathbf{u}. \tag{2.131}$$

Dann ist $\mathbf{u}$ ein Eigenvektor von A, und λ ist der zugehörige Eigenwert.

Es sei $B = A'$; gilt

$$B\mathbf{v} = \mu\mathbf{v}, \tag{2.132}$$

so ist $\mathbf{v}$ ein Eigenvektor von B und μ der zughörige Eigenwert. Es ist

$$(B\mathbf{v})' = \mathbf{v}'B' = \mathbf{v}'A = \mu\mathbf{v}'.$$

$\mathbf{v}$ heißt auch *Linkseigenvektor* von A; $\mathbf{u}$ in (2.131) heißt dementsprechend *Rechtseigenvektor*. Wegen (2.132) übertragen sich alle Aussagen über Rechtseigenvektoren auf Linkseigenvektoren, was allerding nicht bedeutet, dass Links- und Rechtseigenvektoren notwendig identisch sind. Notwendig identisch sind sie nur für den Spezialfall symmetrischer Matrizen. Denn wenn $A' = B = A$ gilt, so folgt aus (2.132) $B\mathbf{v} = A\mathbf{v} = \mu\mathbf{v}$, d.h. ein gegebener Linkseigenvektor entspricht einem Rechtseigenvektor. Im Falle $A' \neq A$ gilt der

Satz 2.23 *Es sei A eine quadratische, nicht-symmetrische Matrix. Es gelte einerseits $\boldsymbol{v}'A = \mu\boldsymbol{v}'$, andererseits $A\boldsymbol{u} = \lambda\boldsymbol{u}$ mit $\lambda \neq \mu$. Dann folgt $\boldsymbol{u}'\boldsymbol{v} = 0$, d.h. die Links- und Rechtseigenvektoren sind orthogonal zueinander.*

Beweis: Multiplikation von $\mathbf{v}'A = \mu\mathbf{v}'$ von rechts mit $\mathbf{u}$ und von $A\mathbf{u} = \lambda\mathbf{u}$ von links mit $\mathbf{v}'$ liefert

$$\begin{aligned} \mathbf{v}'A\mathbf{u} &= \mu\mathbf{v}'\mathbf{u} = \lambda\mathbf{v}'\mathbf{u} \\ \mathbf{v}'A\mathbf{u} &= \lambda\mathbf{v}'\mathbf{u} = \mu\mathbf{v}'\mathbf{u} \end{aligned}$$

Da $\mathbf{v}'A\mathbf{u} - \mathbf{v}'A\mathbf{u} = 0$ folgt $\lambda\mathbf{v}'\mathbf{u} - \mu\mathbf{v}'\mathbf{u} = (\lambda - \mu)\mathbf{v}'\mathbf{u} = 0$, woraus wegen $\lambda - \mu \neq 0$ die Behauptung $\mathbf{v}'\mathbf{u} = 0$ folgt, d.h. $\mathbf{u}$ und $\mathbf{v}$ sind orthogonal. □

Im Falle nicht-symmetrischer Matrizen sind Links- und Rechtseigenvektoren also verschieden, da sie ja orthogonal zueinander sind. Dieses Resultat bedeutet nicht, dass auch die Rechts- und Linkseigenvektoren untereinander orthogonal zueinander sind. Aber die Gültigkeit des folgenden Satzes läßt sich zeigen:

Satz 2.24 *Es sei A eine nicht-symmetrische, quadratische Matrix mit mehr als einem Rechtseigenvektor. Die Rechtseigenvektoren sind linear unabhängig, sofern die zugehörigen Eigenwerte verschieden sind.*

Beweis: Es seien $\mathbf{u}_1$ und $\mathbf{u}_1$ zwei Rechtseigenvektoren von A mit zugehörigen Eigenwerten $\mu \neq \lambda$, für die $\mu \neq \lambda$ gelte. Angenommen, sie seien linear abhängig; dann existieren Koeffizienten a_1 und a_2 ungleich Null derart, dass

$$a_1\mathbf{u}_1 + a_2\mathbf{u}_2 = 0 \tag{2.133}$$

Multiplikation von links mit A führt dann auf $a_1A\mathbf{u}_1 + a_2A\mathbf{u}_2 = 0$, d.h. auf

$$a_1\mu\mathbf{u}_1 + a_2\lambda\mathbf{u}_2 = 0. \tag{2.134}$$

Multipliziert man (2.133) mit λ und subtrahiert (2.133) dann von (2.134), so erhält man

$$a_1(\lambda - \mu)\mathbf{u}_1 + a_2(\lambda - \lambda)\mathbf{u}_2 = 0,$$

woraus entgegen der Annahme wegen $\lambda - \mu \neq 0$ sofort $a_1 = 0$ folgt. Auf analoge Weise fogt $a_2 = 0$, d.h. $\mathbf{u}_1$ und $\mathbf{u}_2$ sind linear unabhängig.

Diese Aussage gilt für irgendzwei Rechtseigenvektoren von A. Hat man also insgesamt drei Eigenvektoren, so sind sie paarweise linear unabhängig, so dass man sagen könnte, sie seien insgesamt linear unabhängig. Das Argument ist aber intuitiv, und ein strenger Beweis ist einer intuitiven Betrachtung stets vorzuziehen. Dieser ergibt sich durch das Prinzip der vollständigen Induktion. Es gebe also $r > 2$ linear unabhängige Eigenvektoren, so dass

$$a_1\mathbf{u}_1 + a_2\mathbf{u}_2 + \cdots + a_r\mathbf{u}_r = 0 \text{ genau dann, wenn } a_1 = \cdots = a_r = 0.$$

Es ist zu zeigen, dass dann auch $r + 1$ Eigenvektoren linear unabhängig sind, so dass

$$a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_r\mathbf{v}_r + a_{p+1}\mathbf{v}_{p+1} = 0 \tag{2.135}$$

gilt mit $a_1 = a_2 = \cdots = a_{p+1} = 0$ als einziger Lösung. Da die $\mathbf{u}_j$ Eigenvektoren sind, gilt $A\mathbf{u}_j = \lambda_j\mathbf{u}_j$. Multiplikation von (2.135) mit A führt dann unter Berücksichtigung dieser Beziehung auf die Gleichung

$$a_1\lambda_1\mathbf{u}_1 + a_2\lambda_2\mathbf{u}_2 + \cdots + a_r\lambda_r\mathbf{u}_r + a_{p+1}\lambda_{p+1}\mathbf{u}_{p+1} = 0. \tag{2.136}$$

Multipliziert man (2.135) mit λ_{p+1} und subtrahiert die Gleichung dann von (2.136), so erhält man

$$a_1(\lambda_1 - \lambda_{p+1})\mathbf{u}_1 + \cdots + a_p(\lambda_p - \lambda_{p+1})u_p = 0,$$

und wegen der vorausgesetzten linearen Unabhängigkeit der $\mathbf{v}_j$, $1 \leq j \leq r$ hat man einerseits $a_1 = \cdots = a_p = 0$ und wegen der ebenso vorausgesetzten Ungleichheit der λ_j folgt dann aus (2.136) $a_{p+1}\lambda_{p+1}\mathbf{u}_{p+1} = 0$. Daraus folgt wegen

$\lambda_{p+1} \neq 0$ dann $a_{p+1} = 0$, so dass die $\mathbf{u}_1, \ldots, \mathbf{u}_{p+1}$ ebenfalls linear unabhängig sind. □

Der Beweis gilt für eine beliebige quadratische Matrix, also auch für A' und damit für die Rechtseigenvektoren von A', die aber die Linkseigenvektoren von A sind, so dass deren lineare Unabhängigkeit ebenfalls nachgewiesen ist. Gilt der Spezialfall $A' = A$, ist A also symmetrisch, so folgt sofort, dass in diesem Fall die Linkseigenvektoren gleich den Rechtseigenvektoren sind, und wie bereits gezeigt wurde gilt dann nicht nur die lineare Unabhängigkeit der Eigenvektoren, sondern darüber hinaus auch die Orthogonalität der Eigenvektoren.

Im Folgenden werden nur die Rechtseigenvektoren betrachtet und es wird der Kürze wegen nur von Eigenvektoren geredet; alle Aussagen übertragen sich auf die Linkseigenvektoren. Zunächst soll die Beziehung zwischen einer quadratischen Matrix A und ihren Eigenvektoren und Eigenwerten auf eine andere Art dargestellt werden, die Aufschluss über die Anzahl und Art der Eigenvektoren und -eigenwerte gibt.

Ähnliche Matrizen: Es sei V die Matrix der Eigenvektoren einer beliebigen quadratischen Matrix A. Da die Spaltenvektoren von V linear unabhängig sind, folgt die Existenz der zu V inversen Matrix V^{-1}. Aus $AV = V\Lambda$, Λ die Matrix der zugehörigen Eigenwerte, folgt dann durch Multiplikation von rechts mit V^{-1}

$$A = V\Lambda V^{-1}. \tag{2.137}$$

Durch Multiplikation von rechts mit V und von links mit V^{-1} erhält man hieraus

$$V^{-1}AV = \Lambda. \tag{2.138}$$

Mit diesen beiden Gleichungen hat man einen Spezialfall einer Beziehung zwischen Matrizen, die durch die folgende Definition charakterisiert wird:

Definition 2.10 *Es seien A und B zwei (n, n)-Matrizen und es existiere eine nichtsinguläre Matrix C derart, dass*

$$B = C^{-1}AC \tag{2.139}$$

gilt. Dann heißen A und B ähnlich.

Offenbar bedeuten (2.137) und (2.138), dass A und Λ ähnlich sind. Da Λ eine Diagonalmatrix ist, heißt A auch *diagonalisierbar*. Damit eine (n, n)-Matrix diagonalisierbar ist, muß also die Matrix V^{-1} existieren, und diese Matrix existiert, wenn A vollen Rang hat, denn dann hat A n linear unabhängige Eigenvektoren.

Komplexe Eigenwerte und - vektoren: Es ist bisher stets vorausgesetzt worden, dass für eine gegebene quadratische Matrix Eigenwerte und - vektoren existieren. Die Frage ist aber, ob für eine beliebige quadratische Matrix überhaupt

Eigenvektoren existieren müssen. Gegeben sei etwa die Matrix

$$A(\phi) = \begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix} \tag{2.140}$$

Ein Eigenvektor $\mathbf{v}$ von A muß die Bedingung $A\mathbf{v} = \mathbf{w} = \lambda\mathbf{v}$ erfüllen, d.h. der Vektor $\mathbf{w}$ muß parallel zu $\mathbf{v}$ sein, er darf sich nur in der Länge von $\mathbf{v}$ unterscheiden. Aber für $\phi \neq 0$ und $\phi \neq \pi$ bewirkt A eine Rotation des Vektors $\mathbf{v}$, $\mathbf{w}$ kann also nicht parallel zu $\mathbf{v}$ sein. A hat mit Ausnahme spezieller ϕ-Werte zumindest keinen reellen Eigenvektor. Um die Situation allgemein zu klären, geht man noch einmal auf die Gleichung (2.131) zurück: so dass sich die Eigenwerte von A als Nullstellen des Polynoms ergeben. Speziell für die Matrix (2.140) erhält man

$$|A - \lambda I| = 4\lambda^2 - 4\lambda\cos\phi + 1 = 0; \tag{2.141}$$

auf die Herleitung des Polynoms wird hier verzichtet, da es an dieser Stelle nur nur auf die Implikationen von (2.141) ankommt. Man findet die Nullstellen

$$\lambda_1 = \frac{1}{2}(\cos\phi + i\sin\phi), \quad \lambda_2 = \frac{1}{2}(\cos\phi - i\sin\phi), \;\; i = \sqrt{-1} \tag{2.142}$$

Die in (2.140) definierte Matrix A hat also zwei komplexe Eigenwerte, relle Eigenwerte ergeben sich nur für solche ϕ-Werte, für die $\sin\phi = 0$ ist, also etwa für $\phi = 0$, wenn gar keine Rotation der Vektoren stattfindet, oder für $\phi = \pi/2$, wenn eine Rotation um 90^o stattfindet.

Es ist also möglich, dass für eine beliebig gewählte quadratische Matrix keine rellen Eigenwerte existieren, dass man aber komplexwertige Eigenwerte finden kann, die als Paare konjugiert komplexer Zahlen auftreten[32]. Nun hätte man noch gerne die zugehörigen Eigenvektoren bestimmt. Für A findet man zwei:

$$\mathbf{u}_1 = \begin{pmatrix} -i \\ 1 \end{pmatrix}, \quad \mathbf{u}_2 = \begin{pmatrix} i \\ 1 \end{pmatrix}. \tag{2.143}$$

Natürlich ergibt sich die Frage der Deutung von komplexen Eigenwerten und Eigenvektoren. Diese treten etwa bei der Analyse dynamischer Systeme und dementsprechend bei allgemeinen Diskussionen von Zeitreihenproblemen auf. Hier sollen zunächst noch bestimmte Typen von Matrizen eingeführt werden.

Typen von Matrizen: In der multivariaten Statistik spielen symmetrische Matrizen mit reellen Elementen eine zentrale Rolle, es ist aber trotzdem sinnvoll, auch den allgemeinen Fall einer Matrix mit möglicherweise komplexwertigen Elementen zu betrachten.

Sind die Elemente einer Matrix A komplex, d.h. von der Form $z = x + iy$ mit $i = \sqrt{-1}$, so heißt $\bar{A}$ die zu A konjugierte Matrix; die Elemente von A enthalten

[32]Zwei komplexe Zahlen z und $\bar{z}$ heißen konjugiert komplex, wenn sie sich nur im Vorzeichen des Imaginärteils unterscheiden, wenn also $z = x + iy$ und $\bar{z} = x - iy$ gilt.

die zu z konjugiert komplexen Elemente $\bar{z} = x - iy$. Sind nur die Imaginärteile iy der Elemente einer Matrix A von Null verschieden, so heißt A *imaginär*; in diesem Fall gilt $A = -\bar{A}$. Die Transponierte $\bar{A}'$ einer Matrix A heißt die mit A *assoziierte* Matrix.

Für symmetrische Matrizen gilt $A' = A$, d.h. $a_{ij} = a_{ji}$ für alle i, j. Gilt für eine Matrix die Ausssage $a_{ij} = -a_{ji}$, so heißt A *schief-symmetrisch.*

Ein wichtiger Fall ist durch die Gleichung

$$A = \bar{A}' \tag{2.144}$$

definiert; in diesem Fall heißt A *hermitesch*[33]. Ist $A = \bar{A}$, so sind die Elemente von A alle reell und (2.144) bedeutet einfach, dass A symmetrisch ist. Da der reelle Fall ein Spezialfall ist, gelten alle Aussagen über hermitesche Matrizen auch für reelle symmetrische Matrizen, so dass es Sinn macht, bestimmte Aussagen allgemein für hermitesche Matrizen zu machen.

2.7.2 Mehrfache Eigenwerte

Es sei A eine (n, n)-Matrix und es werde die Gleichung $A\mathbf{v} = \lambda\mathbf{v}$ betrachtet: λ ist ein Eigenwert von A und $\mathbf{v}$ der zugehörige Eigenvektor. Es gibt maximal n verschiedene Eigenwerte, d.h. es ist möglich, dass einige Eigenwerte mehrfach vorkommen (multiple Eigenwerte, *multiplicity*, *repeated eigenvalues*). Ein einfaches Beispiel ist die (m, n)-Identitätsmatrix I: für jeden n-dimensionalen Vektor $\mathbf{x}$ gilt $I\mathbf{x} = \mathbf{x}$, d.h. jeder Vektor $\mathbf{x}$ ist ein Eigenvektor von I, und alle haben den Eigenwert $\lambda = 1$.

Definition 2.11 *Es sei V ein Vektorraum und es sei $V_\lambda = \{\boldsymbol{v} \in V | A\boldsymbol{v} = \lambda\boldsymbol{v}\}$. Dann heißt V_λ der* Eigenraum *von A zum Eigenwert λ.*

Bemerkung: Aus $A\mathbf{v} = \lambda\mathbf{v}$ folgt $(A - \lambda I)\mathbf{v} = \vec{0}$. Diese Gleichung ist ein lineares Gleichungssystem in $\mathbf{v}$, d.h. in den Komponenten von $\mathbf{v}$ als Unbekannten. Bekanntlich heißt die Menge der Vektoren $\mathbf{x}$, die der Gleichung $A\mathbf{x} = \vec{0}$ genügen, der Kern von A: $\text{kern}(A) = \{\mathbf{x} | A\mathbf{x} = \vec{0}\}$. Dementsprechend ist $\text{kern}(A - \lambda I) = V_\lambda$, d.h. der Eigenraum $V\lambda$ ist der Kern von $(A - \lambda I)$. □

Da zu jedem Eigenwert λ ein Eigenvektor $\mathbf{v}$ korrespondiert, enthält V_λ zumindest ein Element. Da mit $\mathbf{v}$ auch $a\mathbf{v}$, $a \neq 0$ ein Eigenvektor ist, ist V_λ zumindest ein 1-dimensionaler Teilraum von V. Die Frage ist, ob V_λ stets ein Teilraum von V ist. Man sieht dies leicht ein: sind $\mathbf{v} \neq \mathbf{w}$ aus V_λ, so ist mit $a \in \mathbb{R}$, $b \in \mathbb{R}$ auch $\mathbf{u} = a\mathbf{v} + b\mathbf{w} \in V_\lambda$. Denn wegen $A\mathbf{v} = \lambda\mathbf{v}$, $A\mathbf{w} = \lambda\mathbf{w}$ hat man auch

$$A(a\mathbf{v} + b\mathbf{w}) = aA\mathbf{v} + bA\mathbf{w} = a\lambda\mathbf{v} + b\lambda\mathbf{w} = \lambda\mathbf{u}.$$

[33] Nach dem französischen Mathematiker Charles Hermite (1822 – 1901)

Die Eigenwerte von A sind die Nullstellen des Polynoms, das durch die Determinante

$$P_A = |A - \lambda I| = 0$$

definiert ist. Mehrfache Eigenwerte gibt es demnach dann, wenn dieses Polynom mehrfache Nullstellen hat. Man kann nun zeigen, dass, wenn λ eine m-fache Nullstelle von P_A ist, dann die Dimension des Eigenraums V_λ kleiner, höchstens gleich m ist, d.h. es gibt maximal m linear unabhängige Vektoren in V_λ (der Beweis für diese Aussage wird hier übergangen (vergl. Fischer (1984), Kapitel 4).

Der Begriff des Hauptraums ist eine Verallgemeinerung des Begriffs des Eigenraums:

Definition 2.12 *Die Matrix A definiere eine Abbildung f des Vektorraums V in sich selbst, d.h. $f : V \to V$, und λ sei ein Eigenwert von A (d.h. von f), und $r(\lambda)$ sei die algebraische Vielfachheit von λ. Der Kern der r-fachen Hintereinanderschaltung von $A - \lambda I$ heißt* Hauptraum zu λ $H(A, \lambda)$

$$H(A, \lambda) = \{\boldsymbol{v} \in V | (A - \lambda I)^r(\boldsymbol{v}) = 0\}. \tag{2.145}$$

Die Elemente von $H(A, \lambda)$ heißen die Hauptvektoren. *$v \in V$ ist* Hauptvektor der Stufe *p, wenn $(A - \lambda I)^p \boldsymbol{v} \neq \vec{0}$.*

Anmerkung: Alle Eigenvektoren sind Hauptvektoren der Stufe $p = 1$. □

Der in der folgenden Definition eingeführte Begriff des invarianten Teilraums ist eine weitere Verallgemeinerung des Begriffs des Eigenraums:

Definition 2.13 *Die Matrix A definiere eine Abbildung eines Vektorraums in sich selbst: f: $V \to V$, und es sei $U \subseteq V$. Gilt $f(U) \subseteq U$, d.h. ist die Menge der Vektoren $A\boldsymbol{u}$ wieder eine Teilmenge von U, so heißt U* invarianter Teilraum von V*, oder einfach* f-invariant[34].

Anmerkung: Alle Eigenräume sowie alle Haupträume sind invariante Teilräume. □

2.7.3 Das generalisierte Eigenvektorproblem

Eine Reihe von statistischen Fragestellungen führt auf das *generalisierte Eigenvektorproblem*, so etwa die Frage, ob zwei, an m "Fällen" erhobene Datensätze die gleiche oder eine ähnliche latente Struktur haben oder nicht. So kann man an m Personen (Patienten, etc) Messungen von n Variablen vor und nach einer Intervention (etwa einer Therapie) erheben. Die Frage nach einer Veränderung durch die Intervention (Therapieerfolg) führt auf die Frage, ob sich Vorher- und

[34] oder invariant unter f

Nachhermessungen systematisch voneinander unterscheiden. Die Berechnung der Kanonischen Korrelationen kann hier zu Antworten führen. An dieser Stelle kann nur auf die rein formalen Aspekte derartiger Methoden eingegangen werden.

Definition 2.14 *Es seien A und B symmetrische, positiv semidefinite Matrizen. Dann repräsentiert*

$$A\boldsymbol{w} = \lambda B\boldsymbol{w} \tag{2.146}$$

das generalisierte Eigenvektorproblem*.*

Der generalisierte Rayleigh-Quotient: Der *generalisierte Rayleigh-Quotient* ist durch

$$\rho(\mathbf{w}) = \frac{\mathbf{w}'A\mathbf{w}}{\mathbf{w}'B\mathbf{w}} \tag{2.147}$$

definiert. Wie beim schon behandelten Fall $B = I$ die Einheitsmatrix ergibt sich die Frage nach dem maximalen Wert von $\rho(\mathbf{w})$. In Abschnitt 2.10.2 wird eine Anwendung dieses Quotienten vorgestellt.

Dazu werde vorausgesetzt, dass die Inverse B^{-1} existiert. Da B als symmetrisch und positiv semidefinit vorausgesetzt worden ist, kann man die Wurzel $B^{1/2} = P\Lambda^{1/2}P'$ von B bestimmen, – offenbar ist

$$B^{1/2}B^{1/2} = B = P\Lambda^{1/2}P'P\Lambda^{1/2}P' = P\Lambda P',$$

denn $P'P = I$ die Einheitsmatrix, und P die Matrix der Eigenvektoren von B.. Es sei $\mathbf{v} = B^{1/2}\mathbf{w}$. Dann ist $\mathbf{w} = B^{-1/2}\mathbf{v}$ und man erhält

$$\rho(\mathbf{w}) = \frac{\mathbf{w}'A\mathbf{w}}{\mathbf{w}'B\mathbf{w}} = \frac{\mathbf{v}'B^{-1/2}AB^{-1/2}\mathbf{v}}{\mathbf{v}'\mathbf{v}} = \frac{\mathbf{v}'M\mathbf{v}}{\mathbf{v}'\mathbf{v}}. \tag{2.148}$$

$M = B^{-1/2}AB^{-1/2}$ ist symmetrisch (warum?) und die Maximierung des generalisierten Rayleigh-Quotienten ist auf den einfachen Fall zurückgeführt worden. Der Vektor $\mathbf{w}$ des ursprünglichen Problems ergibt sich aus der Lösung für (2.148) gemäß $\mathbf{w} = B^{-1/2}\mathbf{v}$.

Die Gleichung (2.146) führt durch Multiplikation von links mit B^{-1} auf

$$B^{-1}A\mathbf{w} = \lambda\mathbf{w}, \tag{2.149}$$

d.h. $\mathbf{w}$ ist ein Eigenvektor der nicht-symmetrischen Matrix $BV^{-1}A$, und λ ist der zugehörige Eigenwert. Multipliziert man diese Gleichung von links mit $B^{1/2}$, so erhält man

$$B^{1/2}B^{-1}A\mathbf{w} = B^{-1/2}A\mathbf{w} = \lambda B^{1/2}\mathbf{w}.$$

und nochmalige Multiplikation von links mit $B^{-1/2}$ führt zu

$$B^{-1/2}AB^{-1/2}\mathbf{v} = \lambda\mathbf{v}. \tag{2.150}$$

Damit hat man mit (2.150) ein Eigenwert- und Eigenvektorproblem der bekannten Art für symmetrische Matrizen.

Die Eigenvektoren $\mathbf{v}_j$ symmetrischer Matrizen sind bekanntlich orthonormal. Die Lösung für den generaliserten Rayleigh-Quotienten durch die Lösung für den gewöhnlichen Rayleigh-Quotienten in einem *transformierten Raum* gegeben ist. Man kommt damit zu der Aussage (Shaw-Taylor & Christianini (2004), p. 162)

Satz 2.25 *Ein beliebiger Vektor $\boldsymbol{v}$ kann als Linearkombination der $\boldsymbol{w}_j$, $j = 1, \ldots, k$ angeschrieben werden. Für die Eigenvektoren des generalisierten Eigenvektorproblems $A\boldsymbol{w} = \lambda B\boldsymbol{w}$ gelten die Relationen*

$$\boldsymbol{w}_i' B \boldsymbol{w}_j = \delta_{ij}, \quad \boldsymbol{w}_i' A \boldsymbol{w}_j = \delta_{ij}\lambda_i, \quad \delta_{ij} = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases} \tag{2.151}$$

Beweis: Es war $\mathbf{v} = B^{1/2}\mathbf{w}$, und als Lösungen von (2.150) sind die $\mathbf{v}_j$ orthonormal. Es $i \neq j$ und $\lambda_j \neq 0$. Dann folgt, wegen $A\mathbf{w} = \lambda B\mathbf{w}$ (vergl. (2.149)) und damit $B\mathbf{w} = (1/\lambda)A\mathbf{w}$,

$$0 = \mathbf{v}_i'\mathbf{v}_j = \mathbf{w}_i' B^{1/2} B^{1/2} \mathbf{w}_j = \mathbf{w}_i' B \mathbf{w}_j = \frac{1}{\lambda_j} \mathbf{w}_i A \mathbf{w}_j;$$

nach (2.146) gilt ja $A\mathbf{w} = \lambda B\mathbf{w}$ und deshalb $(1/\lambda_i)A\mathbf{w} = B\mathbf{w}$. Damit gilt (2.151) für den Fall $i \neq j$.

Nun sei $i = j$; es ist $1 = \mathbf{v}_i'\mathbf{v}_i = \mathbf{w}_i' B^{1/2} B^{1/2} \mathbf{w}_i$, also

$$\lambda_i = \lambda_i \mathbf{v}_i'\mathbf{v}_i = \lambda_i \mathbf{w}_i' B^{1/2} B^{1/2} \mathbf{w}_i = \lambda_i \mathbf{w}_i' B \mathbf{w}_i = \mathbf{w}_i A \mathbf{w}_i,$$

und dies ist (2.151) für den Fall $i = j$. □

Die Maximierung von (2.151) (Maximierung unter Nebenbedingungen, S. Anhang) führt auf die Gleichung (2.149).

Satz 2.26 *Für den generalisierten Rayleigh-Quotienten gilt*

$$\rho_1 \leq \rho \leq \rho_2, \tag{2.152}$$

und ρ_1, ρ_2 sind durch die Eigenvektoren definiert, die zum kleinsten bzw. größten Eigenwert korrespondieren.

Der Beweis ergibt sich analog zum Beweis für den Rayleigh-Quotienten für symmetrische Matrizen (Satz von Courant-Fisher, Seite 81).

Satz 2.27 *Gilt $A\boldsymbol{v} = \lambda B\boldsymbol{v}$ und sind λ und $\boldsymbol{v}$ die Eigenwerte und Eigenvektoren für den generalisierten Rayleigh-Quotienten, so kann A gemäß*

$$A = \sum_{j=1}^{r} \lambda_j B\boldsymbol{v}_j (B\boldsymbol{v}_j)' \tag{2.153}$$

zerlegt werden.

Beweis: Für eine beliebige symmetrische Matrix C mit der Matrix P der Eigenvektoren und der Diagonalmatrix $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ der Eigenwerte gilt bekanntlich $C = \sum_j \lambda_j \mathbf{p}_j \mathbf{p}_j'$. Die Matrix $B^{-1/2}AB^{-1/2}$ ist symmetrisch, mithin gilt

$$B^{-1/2}AB^{-1/2} = \sum_{j=1}^{r} \lambda_j \mathbf{v}_j \mathbf{v}_j'.$$

Multipliziert man von links mit $B^{1/2}$ und von rechts ebenfalls mit $B^{1/2}$, so folgt

$$A = \sum_{j=1}^{r} \lambda_j B^{1/2}(B^{1/2}\mathbf{v}_j\mathbf{v}_j)' = \sum_j \lambda_j B\mathbf{w}_j(B\mathbf{w}_j)',$$

und das war zu zeigen. □

2.8 Weitere Befunde

2.8.1 Die Zentrierungsmatrix

Die Zentrierung einer Datenmatrix kann durch eine Matrixmultiplikation bewerkstelligt werden:

Definition 2.15 *Es sei $\vec{1} = (1, 1, \ldots, 1)'$ ein m-dimensionaler Vektor, dessen Komponenten alle gleich 1 sind, und $\vec{1}\vec{1}'$ ist das dyadische Produkt von $\vec{1}$ mit sich selbst. Dann heißt*

$$H_m = I - \frac{1}{m}\vec{1}\vec{1}'. \tag{2.154}$$

Zentrierungsmatrix.

H_M hat die Form

$$H_m = \begin{pmatrix} 1-\frac{1}{m}, & -\frac{1}{m}, & -\frac{1}{m}, & \cdots, & -\frac{1}{m} \\ -\frac{1}{m}, & 1-\frac{1}{m}, & -\frac{1}{m}, & \cdots, & -\frac{1}{m} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -\frac{1}{m}, & -\frac{1}{m}, & -\frac{1}{m}, & \cdots, & 1-\frac{1}{m} \end{pmatrix} \tag{2.155}$$

Ist X eine nichtzentrierte (m, n)-Matrix und ist X_{sc} die korrespondierende spaltenzentrierte Matrix, X_{zc} die korrespondierende zeilenzentrierte Matrix so gilt

$$X_{sc} = H_m X, \quad X_{zc} = X H_n. \tag{2.156}$$

Es gilt der

Satz 2.28 *Die Zentrierungsmatrix hat die Eigenschaften:*
1. H_m ist idempotent (eine Matrix M ist idempotent*, wenn $MM = M$),*

2. H_m ist symmetrisch und positiv-semidefinit[35]*,*
3. H_m hat einen Eigenwert $\lambda = 0$ und den Eigenwert 1 mit der Multiplizität $m-1$, d.h. H_m hat den Rang $rg(H_m) = m-1$,

Beweis: Es gilt

$$\begin{aligned} H_m H_m &= (I - \frac{1}{m}\vec{1}\vec{1}')(I - \frac{1}{m}\vec{1}\vec{1}') \\ &= I - \frac{1}{m}I\vec{1}\vec{1}' - \frac{1}{m}I\vec{1}\vec{1}' + \frac{1}{m^2}\vec{1}\vec{1}'\vec{1}\vec{1}' \\ &= I - \frac{1}{m}I\vec{1}\vec{1}' = H_m, \end{aligned}$$

denn

$$\vec{1}\vec{1}'\vec{1}\vec{1}' = \vec{1}(\vec{1}'\vec{1})\vec{1}' = m\vec{1}\vec{1}',$$

d.h. H_m ist idempotent. Weiter gilt

$$H_m' = (I - \frac{1}{m}\vec{1}\vec{1}')' = I' - (\frac{1}{m}\vec{1}\vec{1}')' = I - \frac{1}{m}\vec{1}\vec{1}' = H_m, \tag{2.157}$$

d.h. H_m ist symmetrisch. Damit ist H_m gemäß Definition 2.21 eine Projektionsmatrix. Nach Satz 2.37 (Seite 128) hat H_m dann die Eigenwerte 0 und 1. Da H_m reell und symmetrisch ist, ist der Rang von H_m gleich der Anzahl von von Null verschiedenen Eigenwerte, mithin ist der Rang von H_m gleich $\mathrm{rg}(H_m) = m-1$, d.h. H_m hat keinen vollen Rang und ist damit singulär. Nach Definition 2.4 ist H_m positiv-semidefinit, wenn für $\mathbf{x} \in \mathbb{R}^m$ die Relation $\mathbf{x}'H_m\mathbf{x} = k \geq 0$ gilt. $\mathbf{x}'H_m$ ist ein Zeilenvektor; die j-te Komponente von $\mathbf{x}'H_m$ ist

$$-\frac{x_1}{m} - \cdots - \frac{x_{j-1}}{m} + x_j(1 - \frac{1}{m}) - \frac{x_{j+1}}{m} - \cdots - \frac{x_m}{m} = x_j - \frac{1}{m}\sum_{i=1} x_i = x_j - \bar{x}$$

Dann ist

$$\mathbf{x}'H_m\mathbf{x} = x_1^2 - x_1\bar{x} + x_2^2 - x_2\bar{x} + \cdots - x_m^2 - x_m\bar{x} = \sum_{i=1}^{m} x_i^2 - m\bar{x}^2 = k,$$

so dass

$$\frac{1}{m}\mathbf{x}'H_m\mathbf{x} = \frac{k}{m} = \frac{1}{m}\sum_{i=1}^{m} x_i^2 - \bar{x}^2 = \frac{1}{m}\sum_{i=1}^{m}(x_i - \bar{x})^2 \geq 0,$$

also ist H_m positiv-semidefinit.

Der Kern von H_m ist $\ker(H_m) = \{\mathbf{x}|H_m\mathbf{x} = \vec{0}\}$. Man hat

$$\begin{pmatrix} 1-\frac{1}{m}, & -\frac{1}{m}, & -\frac{1}{m}, & \cdots, & -\frac{1}{m} \\ -\frac{1}{m}, & 1-\frac{1}{m}, & -\frac{1}{m}, & \cdots, & -\frac{1}{m} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -\frac{1}{m}, & -\frac{1}{m}, & -\frac{1}{m}, & \cdots, & 1-\frac{1}{m} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

[35] d.h. H_m ist eine Projektionsmatrix, s. Abschnitt Projektionsmatrix

Für die i-te Komponente von $\vec{0}$ findet man dann

$$x_i - \bar{x} = 0 \Rightarrow x_i = \bar{x}$$

für alle i, d.h. die Komponenten von $\mathbf{x}$ sind identisch, $\mathbf{x} = (x, x, \ldots, x)'$. Da sich die Orientierung eines Vektors nicht ändert, wenn er mit einem Skalar multipliziert wird, ist

$$\frac{1}{x}\mathbf{x} = (1, 1, \ldots, 1)',$$

d.h. die Orientierung der $\mathbf{x}$ aus dem Kern von H_m ist identisch mit der von $\vec{1} = (1, 1, \ldots, 1)'$. Der Kern ist damit ein 1-dimensionaler Teilraum des $\mathbb{R}^m$. $\square$

Es sei X eine unzentrierte Datenmatrix, X_c sei die zugehörige spaltenzentrierte Matrix. Dann gilt

$$X_c = H_m X, \quad X_c' X_c = X H_m' H_m X = X' H_m^2 X = X' H_m X, \tag{2.158}$$

wegen der Symmetrie und Idempotenz von H_m.

Der Rang zentrierter Datenmatrizen: Nach (2.156) ist $X_{sc} = H_m X$, so dass wegen (2.53)

$$\mathrm{rg}(X_{sc}) \leq \min[\mathrm{rg}(H_m), \mathrm{rg}(X)]. \tag{2.159}$$

Satz 2.29 *Es ist* $rg(H_m) = m$.

Beweis: Zu zeigen ist, dass die Darstellung des Nullvektors $\vec{0}$ als Linearkombination der Spaltenvektoren von H_m nur möglich ist, wenn alle Koeffizienten $a_i = 0$ gilt. Allgemein hat man

$$a_1 \begin{pmatrix} 1 - 1/m \\ -1/m \\ \vdots \\ -1/m \end{pmatrix} + a_2 \begin{pmatrix} -1/m \\ 1 - 1/m \\ \vdots \\ -1/m \end{pmatrix} + \cdots + a_m \begin{pmatrix} -1/m \\ -1/m \\ \vdots \\ 1 - 1/m \end{pmatrix} = \vec{0}$$

Mit $S = \sum_i a_i$ gilt dann speziell für die i-te Komponente von $\vec{0}$

$$0 = a_i(1 - \frac{1}{m}) - \frac{m-1}{m}(S - a_i),$$

d.h.

$$\frac{m-1}{m} a_i = \frac{m-1}{m}(S - a_i) \Rightarrow a_i = S - a_i,$$

so dass

$$a_i = \frac{1}{2}S, \quad i = 1, \ldots, m$$

Andererseits hat man

$$\frac{m-1}{m}\sum_i a_i = (m-1)S + \frac{m-1}{m}\sum_i a_i,$$

woraus $0 = (m-1)S$, d.h. $S = 0$ folgt. Dann folgt $a_i = \frac{1}{2}S = 0$ für alle i. Damit folgt, dass die Spaltenvektoren von H_m linear unabhängig sind, d.h. H_m hat den vollen Rang m. □

Bei Datenmatrizen stehen die m Zeilen für die m Fälle und die n Spalten für die gemessenen n Variablen, und man hat $m \geq n$. Die Gleichung (2.159) impliziert dann

$$\mathrm{rg}(X_{sc}) = \mathrm{rg}(X) \leq n. \tag{2.160}$$

2.8.2 Die Beziehungen zwischen verschiedenen Basen

Es sei V ein n-dimensionaler Vektorraum. Jede Menge von n linear unabhängigen, n-dimensionalen Vektoren aus V ist eine Basis von V, und es fragt sich, welche Basis man wählen soll, wenn man eine Menge von Datenvektoren $\mathbf{x}_j$, $j = 1, \ldots, n$ als Linearkombination von Basisvektoren repräsentieren will. Bevor auf diese Frage näher eingegangen wird, sollen ein paar grundsätzliche Sachverhalte geklärt werden.

Es seien $\mathcal{B} = (\mathbf{b}_1, \ldots, \mathbf{b}_n)$ und $\mathcal{C} = (\mathbf{c}_1, \ldots, \mathbf{c}_n)$ irgend zwei Basen des Vektorraums V. Dann gilt

$$\mathcal{L}(\mathcal{B}) = \mathcal{L}(\mathcal{C}) = V$$

und damit sind die $\mathbf{b}_j \in \mathcal{L}(\mathcal{C})$ und die $\mathbf{c}_j \in \mathcal{L}(\mathcal{B})$. Jeder Basisvektor $\mathbf{c}_j$ läßt sich als Linearkombination der Basisvektoren $\mathbf{b}_j$ darstellen und umgekehrt, d.h. es gelten die Gleichungen

$$\mathbf{c}_j = t_{1j}\mathbf{b}_1 + \cdots + t_{nj}\mathbf{b}_n = B\mathbf{t}_j, \quad j = 1, \ldots, n \tag{2.161}$$

$$\mathbf{b}_j = s_{1j}\mathbf{c}_1 + \cdots + s_{nj}\mathbf{c}_n = C\mathbf{s}_j \tag{2.162}$$

wobei B eine Matrix ist, die entsteht, wenn man die $\mathbf{b}_j$ spaltenweise zu einer Matrix zusammenfasst, und analog dazu ist C die Matrix, deren Spaltenvektoren die $\mathbf{c}_j$ sind. Die Koeffizienten t_{kj} und s_{kj}, $1 \leq k \leq n$, können zu Vektoren $\mathbf{t}_j = (t_{1j}, \ldots, t_{nj})'$ und $\mathbf{s}_j = (s_{1j}, \ldots, s_{nj})'$ zusammengefasst werden. Ist T die Matrix mit den $\mathbf{t}_j$ als Spaltenvektoren und S die Matrix mit den $\mathbf{s}_j$ als Spaltenvektoren, so sind (2.161) und (2.162) äquivalent zu

$$C = BT \tag{2.163}$$

$$B = CS. \tag{2.164}$$

Diese Gleichungen repräsentieren die Möglichkeit eines *Basiswechsels*, d.h. des Übergangs von einer Basis zu einer anderen, wobei sich zeigt, dass zwischen den

Matrizen S und T eine enge Beziehung besteht. Setzt man nämlich die rechte Seite von (2.164) für B in (2.163) ein, so erhält man $C = CST$. C hat – da die Spaltenvektoren eine Basis repräsentieren – den vollen Rang n, so dass C^{-1} existiert, so dass

$$C^{-1}C = C^{-1}CST = ST = I$$

folgt; aus demselben Grund hat auch B vollen Rang n, so dass weiter $\mathrm{rg}(C) = \mathrm{rg}(BT)$ folgt. Nach Satz 2.6, Seite 55, gilt

$$\mathrm{rg}(C) \leq \min(\mathrm{rg}(B), \mathrm{rg}(T)) \leq \mathrm{rg}(B), \quad \mathrm{rg}(B) \leq \min(\mathrm{rg}(C), \mathrm{rg}(S),$$

und man kann

$$\mathrm{rg}(S) = \mathrm{rg}(T) = n$$

schließen; wäre $\mathrm{rg}(T) < n$, würde $\mathrm{rg}(C) < \mathrm{rg}(B)$ folgen, was nicht möglich ist, da ja $\mathrm{rg}(B) = \mathrm{rg}(C)$ gelten muß; eine analoge Argumentation gilt für den Rang von S. Dann folgt aus $ST = I$ die Beziehung

$$S = T^{-1}, \tag{2.165}$$

und analog dazu $S^{-1} = T$.

Natürlich ist der Fall möglich, dass die Vektoren $\mathbf{x}_j$ linear abhängig sind, so dass die aus den Spaltenvektoren $\mathbf{x}_j$ bestehende Matrix X einen Rang r kleiner als $\min(m, n)$ hat. Dann sind B_r und C_r (r, r)-Matrizen und die vorangegangenen Betrachtungen übertragen sich auf diesen Fall.

Formal sind die verschiedenen Basen eines Vektorraums insofern äquivalent, als sie es ermöglichen, alle Vektoren des (Teil-)Vektorraumes als Linearkombinationen der Basisvektoren darzustellen. Will man zu einer Matrix X zusammengefasste Daten interpretieren, so muß man sich für eine möglichen der Basen entscheiden. Die Entscheidung muß anhand von Kriterien geschehen, die sich aus dem Begriff des Vektorraums oder dem der Basis eines Vektoraumes selbst nicht herleiten lassen. Ein Ausgangspunkt für die Suche nach einer (Teil-)Basis kann stets die SVD gewählt werden. Setzt man $L = Q\Lambda^{1/2}$, so hat man $X = LT'$, T die Matrix der Eigenvektoren von $X'X$, und $L'L = \Lambda$ eine Diagonalmatrix, d.h. die Spaltenvektoren $\mathbf{L}_k$, $k = 1, \ldots, r \leq \min(m, n)$ sind orthogonal, weshalb man die $\mathbf{L}_k$ unabhängig voneinander interpretieren kann. Diese Möglichkeit ist allerdings mit Vorsicht zu betrachten: eine Korrelation gleich Null bedeutet ja nicht notwendig auch stochastische Unabhängigkeit; diese Interpretation setzt gewisse Annahmen übner die Verteilung der Daten voraus, worauf an dieser Stelle nicht weiter eingegangen werden kann. Jedenfalls gilt wegen der Orthonormalität von T die Beziehung

$$XT = L,$$

d.h. die Basisvektoren $\mathbf{L}_k$ ergeben sich als Linearkombinationen der Datenvekgtoren $\mathbf{x}_j$. Will man auf die Variablen (repräsentiert durch die Spalten von X)

fokussieren, so betrachte man den Fall $X = QA'$, mit $A = T\Lambda^{1/2}$, und wegen der Othonormalität on Q hat man $Q'X = A'$, oder

$$X'Q = A,$$

d.h. die n-dimensionalen Spalten von A ergeben sich als Linearkombinationen der Zeilenvektoren von X.

Der Übergang zu einer anderen Basis bedeutet nun die Multiplikation etwa von L mit einer orthogonalen Matrix U. Man erhält

$$XTU = LU,$$

und setzt man $V = LU$, so erhält man (Multiplikation von rechts (i) mit U^{-1}, (ii) mit T')

$$X = LUU^{-1}T' = VW', \quad W = TU^{-1}, \tag{2.166}$$

d.h. die Transformation von L mit U geht mit einer Transformation von T mit der Transformationsmatrix U^{-1} einher.

2.8.3 Die Pseudoinverse einer Matrix

Es werde das Gleichungssystem $A\mathbf{x} = \mathbf{y}$ betrachtet, wobei A eine (m, n)-Matrix sei mit $m > n$. $\mathbf{x}$ ist ein n-dimensionaler Vektor, dessen Komponenten die Unbekannten sind. Man hat jetzt mehr Gleichungen als Unbekannte. $\mathbf{y}$ ist ein m-dimensionaler Vektor. Ist $\mathbf{y} \in \mathcal{L}(A)$, d.h. ist $\mathbf{y}$ eine Linearkombination der Spaltenvektoren von A ($\mathcal{L}(A)$ ist die lineare Hülle der Spaltenvektoren von A), so existiert $\mathbf{x}$, andernfalls – wenn $\mathbf{y}$ keine Linearkombination der Spaltenvektoren von A ist ($\mathbf{y} \notin \mathcal{L}(A)$), so existiert $\mathbf{x}$ nicht.

Man kann nun eine Pseudoinverse (auch: *generalisierte Inverse*) A^+ definieren:

Definition 2.16 *Die Matrix A^+ heißt* Pseudoinverse *oder* Moore-Penrose-Inverse*, wenn sie die folgenden Bedingungen (Moore-Penrose-Bedingungen) erfüllt:*
1. $AA^+A = A$,
2. $A^+AA^+ = A^+$
3. $(AA^+)^* = AA^+$
4. $(A^+A)^* = A^+A$.

Anmerkung: Die Bedingungen 3. und 4. beziehen sich auf Matrizen mit komplexwertigen Elementen; der Stern definiert die konjugiert komplexe Zahl einer komplexen Zahl. In diesem Skript werden keine komplexen Matrizen und Vektoren betrachtet, so dass 3. und 4. nicht weiter berücksichtigt werden müssen, die beiden Punkte sind nur der Vollständigkeit halber mit aufgeführt worden. □

Für das Gleichungssystem $A\mathbf{x} = \mathbf{y}$ kann nun leicht eine Pseudoinverse gefunden werden, wenn $\mathrm{rg}(A) = n$ ist. Dann hat $A'A$ ebenfalls den Rang n, d.h.

$A'A$ hat vollen Rang, so dass die Inverse $(A'A)^{-1}$ existiert. Man hat dann nach Multiplikation von $A\mathbf{x} = \mathbf{y}$ von links mit A' die Gleichung

$$A'A\mathbf{x} = A'\mathbf{y} \Rightarrow \mathbf{x} = (A'A)^{-1}A'\mathbf{y}. \tag{2.167}$$

Für $\mathbf{y} \in \mathcal{L}(A)$ ist $\mathbf{x}$ die exakte Lösung für das Gleichungssystem, und für $\mathbf{y} \notin \mathcal{L}(A)$ liefert $(A'A)^{-1}A'\mathbf{y}$ die Kleinste-Quadrate-Approximation $\hat{\mathbf{x}}$ für $\mathbf{x}$, d.h. man hat

$$(A'A)^{-1}A'\mathbf{y} = \begin{cases} \mathbf{x}, & \mathbf{y} \in \mathcal{L}(A), \\ \hat{\mathbf{x}}, & \mathbf{y} \notin \mathcal{L}(A). \end{cases} \tag{2.168}$$

(vergl. Abschnitt 4.4.3, Gleichung (4.34), Seite 149).

Die Matrix $(A'A)^{-1}A'$ ist eine Pseudoinverse für A. Um diese Behauptung einzusehen, genügt es, die Bedingungen 1. und 2. zu überprüfen. In Bezug auf 1. hat man

$$A((A'A)^{-1}A')A = A(A'A)^{-1}A'A = A,$$

und in Bezug auf 2. hat man

$$((A'A)^{-1}A')A((A'A)^{-1}A') = (A'A)^{-1}A'A(A'A)^{-1}A' = (A'A)^{-1}A',$$

d.h. $(A'A)^{-1}A'$ ist eine Pseudoinverse für A. □

Der folgende Ansatz, eine Pseudoinverse zu definieren, gilt auch für den Fall $\mathrm{rg}(A) = r < \min(m, n)$ (Stewart (1973)). Es sei

$$A = Q\Sigma T', \quad \Sigma = \begin{pmatrix} \Lambda_r^{1/2} & 0 \\ 0 & 0 \end{pmatrix} \tag{2.169}$$

die Darstellung von A durch die SVD. Q ist die (m, m)-Matrix der orthonormalen Eigenvektoren von AA', T ist die (n, n)-Matrix der orthonormalen Eigenvektoren von $A'A$, und Λ_r ist eine (r, r)-Matrix der von Null verschiedenen Eigenwerte von AA' bzw. $A'A$. Σ ist eine (m, n)-Matrix, deren Elemente bis auf die Diagonalzellen von Λ_r gleich Null sind. Dann ist

$$A^+ = A' = T \begin{pmatrix} \Lambda_r^{-1/2} & 0 \\ 0 & 0 \end{pmatrix} Q' \tag{2.170}$$

eine Pseudoinverse für A. Denn nach 1. muß $AA^+A = A$ gelten, und man findet, indem man die SVD für A einsetzt,

$$Q \begin{pmatrix} \Lambda_r & 0 \\ 0 & 0 \end{pmatrix} T'T \begin{pmatrix} \Lambda_r^{-1/2} & 0 \\ 0 & 0 \end{pmatrix} Q'Q \begin{pmatrix} \Lambda_r^{1/2} & 0 \\ 0 & 0 \end{pmatrix} T' = Q \begin{pmatrix} \Lambda_r^{1/2} & 0 \\ 0 & 0 \end{pmatrix} T' = A,$$

und nach 2. muß $A^+AA^+ = A^+$ gelten:

$$T \begin{pmatrix} \Lambda_r^{-1/2} & 0 \\ 0 & 0 \end{pmatrix} Q'Q \begin{pmatrix} \Lambda_r^{1/2} & 0 \\ 0 & 0 \end{pmatrix} T'T \begin{pmatrix} \Lambda_r^{-1/2} & 0 \\ 0 & 0 \end{pmatrix} Q' = T \begin{pmatrix} \Lambda_r^{-1/2} & 0 \\ 0 & 0 \end{pmatrix} Q' = A^+,$$

d.h. (2.170) definiert tatsächlich eine Pseudoinverse.

2.8.4 Vektor- und Matrixnormen

Es ist immer wieder von normierten Vektoren die Rede gewesen: ein Vektor $\mathbf{x}$ ist normiert, wenn $\|\mathbf{x}\| = 1$, wobei $\|\mathbf{x}\|$ die Länge im Sinne des Satzes von Pythagoras ist, man spricht auch von Euklidischer Norm. Dies ist ein Spezialfall, die Norm eines Vektors kan allgemeiner definiert werden.

Die Norm eines Vektors definiert, in welchem Sinne von der "Gröÿe" eines Vektors gesprochen werden soll, – die übliche euklidische Norm $\|\mathbf{x}\| = (\sum_i x_i^2)^{1/2}$ definiert die Länge des Vektors $\mathbf{x}$ als seine "Größe"[36]. Ebenso kann eine *Matrixnorm* definiert werden. Dieser Begriff erweist sich als nützlich, wenn bestimmte Maxima oder Minima gefunden werden sollen, etwa die Varianzen von Projektionen einer Punktekonfiguration auf bestimmte Dimensionen, oder die Güte der Approximation an eine Datenmatrix. Es wird zuerst der Begriff der Vektornorm spezifiziert:

Definition 2.17 *Es seien $\boldsymbol{x} \in \mathbb{R}^n$ n-dimensionale Vektoren. Eine Vektornorm ist eine Abildung $f : \mathbb{R}^n \to \mathbb{R}$ (d.h. es wird einem Vektor $\boldsymbol{x}$ eine bestimmte reelle Zahl zugeordnet), die den Bedingungen*
1. $f(\boldsymbol{x}) \geq 0$,
2. $f(\boldsymbol{x} + \boldsymbol{y}) \leq f(\boldsymbol{x}) + f(\boldsymbol{y})$,
3. $f(a\boldsymbol{x}) = af(\boldsymbol{x})$, für $a \in \mathbb{R}$
genügt. Dann heißt f eine Vektornorm. *f wird durch $f(\boldsymbol{x}) = \|\boldsymbol{x}\|$ notiert. Der* Einheitsvektor *in Bezug auf eine Norm $\|\cdot\|$ ist derjenige Vektor, für den $\|\boldsymbol{x}\| = 1$ gilt.*

Von besonderem Interesse sind die p-Normen

$$\|\mathbf{x}\|_p = (|x_1|^p + \cdots + |x_n|^p)^{1/p} = \left(\sum_{j=1}^{n} |x_j|^p \right)^{1/p}. \tag{2.171}$$

Für $p = 1$ erhält man die 1-Norm

$$\|\mathbf{x}\|_1 = (|x_1| + \cdots + |x_n|) = \sum_{j=1}^{n} |x_j|. \tag{2.172}$$

und für $p = 2$ die euklidische Norm

$$\|\mathbf{x}\|_2 = (|x_1|^2 + \cdots + |x_n|^2)^{1/2} = \left(\sum_{j=1}^{n} |x_j|^2 \right)^{1/2} = \sqrt{\mathbf{x}'\mathbf{x}}. \tag{2.173}$$

[36] Um sich eine inhaltliche Vorstellung zu machen, stelle man sich vor dass die Komponenten x_i von $\mathbf{x}$ Maße für Begabungen $M_1, \ldots, M_n$ repräsentieren. Dann ist $\|\mathbf{x}\|$ ein *mögliches* Maß für die Gesamtbegabung einer Person.

Für $p = \infty$ schließlich findet man die Maximum-Norm

$$|\mathbf{x}\|_\infty = \max_{1\le i\le n} |x_i|. \tag{2.174}$$

Die Maximum-Norm ergibt sich aus der p-Norm für $p \to \infty$. Es sei $x_k = x_{\max}$ die maximale Komponente von $\mathbf{x}$. Dann ist

$$\|\mathbf{x}\|_p = \left(|x_k|^p \sum_{j=1}^{n} \frac{|x_j|^p}{|x_k|^p} \right)^{1/p}.$$

Wegen $|x_j|/|x_k| \le 1$ für alle $j \neq k$ folgt $\lim_{p\to\infty} |x_j|^p/|x_k|^p \to 0$ für $j \neq k$ und $|x_j|^p/|x_k|^p = 1$ für $j = k$, so dass

$$\lim_{p\to\infty} \|\mathbf{x}\|_p = x_k = x_{\max}.$$

Matrixnormen: Der Begriff der Norm kann auch auf Matrizen angewendet werden:

Definition 2.18 *Es sei $\mathbb{R}^{m\times n}$ die Menge der rellen (m,n)-Matrizen*[37]*. Eine* Matrixnorm *ist eine Abbildung $\|\cdot\| : \mathbb{R}^{m\times n} \to \mathbb{R}_+$, $\mathbb{R}_+$ die Menge der reellen Zahlen größer oder gleich Null, und $A \mapsto \|A\|$ derart, dass*

1. $\|A\| = 0$ genau dann, wenn $A = 0$ die Nullmatrix ist,
2. $\|\lambda X\| = \lambda\|A\|$,
3. $\|A + B\| \le \|A\| + \|B\|$
gilt. Zusammen mit der Norm $\|\cdot\|$ wird der Vektorraum der (m,n)-Matrizen dann zu einem normierten Vektorraum[38] *$(\mathbb{R}^{m\times n}, \|\cdot\|)$.*

Es gibt verschiedene Normen, von denen hier einige als Beispiel genannt werden:

1. Die *Frobenius-Norm.*

$$\|A\|_F = \left(\sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|^2 \right)^{1/2}. \tag{2.175}$$

   Für diese Norm wird auch der Name *Schur-Norm* oder *Hilbert-Schmidt-Norm* verwendet.

[37] Diese Definition ist etwas vereinfacht formuliert, eigentlich muß es heißen: es sei $\mathbb{K} = \mathbb{R}$ der Körper der rellen Zahlen und $\mathbb{K}^{m\times n} = \mathbb{R}^{m\times n}$ die Menge der reellen (m,n)-Matrizen, etc

[38] Hier wird vom *allgemeinen* Begriff des Vektorraums Gebrauch gemacht, demzufolge auch Matrizen als "Vektoren" aufgefasst werden können, so dass auch Mengen von Matrizen einen Vektorraum bilden können. Der Begriff des Vektorraums bezieht sich ja eigentlich nur auf die Kombination bzw. Verknüpfungen von Elementen einer Menge!

2. Die p-Norm: sie ist definiert durch

$$\|A\|_p = \max_{\mathbf{x}\neq\vec{0}} \frac{\|A\mathbf{x}\|_p}{\|\mathbf{x}\|_p} \tag{2.176}$$

Da $A\mathbf{x} = \mathbf{y}$ ein Vektor ist, ist $\|A\mathbf{x}\|_p$ eigentlich eine Vektornorm; allgemein heißen Normen der Form

$$\|A\| = \max_{\mathbf{x}\neq\vec{0}} \frac{\|A\mathbf{x}\|}{\|\mathbf{x}\|}. \tag{2.177}$$

durch eine Vektornorm induzierte Normen.

Die Frobenius- und die p-Norm sind die am häufigsten vorkommenden Matrixnormen. Für $\|A\|_p$ gilt, wenn A eine (n,n)-Matrix ist,

$$\|A\|_p = \sup_{\mathbf{x}\neq\vec{0}} \left\| \left(A\frac{\mathbf{x}}{\|\mathbf{x}\|} \right) \right\|_p = \max_{\|\mathbf{x}\|_p=1} \|A\mathbf{x}\|_p. \tag{2.178}$$

Speziell für $p = 2$ ist mit $\mathbf{y} = A\mathbf{x}$ die Norm $\|A\mathbf{x}\|_2$ durch die Norm $\|\mathbf{y}\|_2 = (\mathbf{y}'\mathbf{y})^{1/2} = \|\mathbf{y}\|$ gegeben, und nach dem Courant-Fischer Theorem 2.20 findet man

$$\|A\|_2 = \max_{\|\mathbf{x}\|_2=1} \|A\mathbf{x}\|_2 = \sqrt{\lambda_{\max}}, \tag{2.179}$$

wobei $\lambda_{\max}$ der maximale Eigenwert von $A'A$ ist. Für die Frobenius-Norm findet man

Satz 2.30 *Es sei A eine (m,n)-Matrix. Für die Frobenius-Norm $\|A\|_F$ gilt*

$$\|A\|_F^2 = spur(AA') = \sum_{i=1}^{n} \lambda_j \tag{2.180}$$

wobei $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ die Eigenwerte von $A'A$ sind.

.

Beweis: Auf A kann die SVD angewendet werden: $A = Q\Lambda^{1/2}P'$, $\lambda_j \geq 0$ für $j = 1,\ldots,n$. Dann ist $AA' = Q\Lambda^{1/2}P'P\Lambda^{1/2}Q' = Q\Lambda Q'$, und die Diagonalelemente von $Q\Lambda Q'$ sind von der Form $\sum_i \lambda_i q_{ji}^2$ für $j = 1,\ldots,n$ Die Spur von $A'A$ ist die Summe dieser Diagonalelemente, d.h.

$$\text{spur}(AA') = \sum_{j=1}^{n}\sum_{i=1}^{n} \lambda_i q_{ji}^2 = \sum_{i=1}^{n} \lambda_i \sum_{j=1}^{n} q_{ij}^2.$$

Aber $\sum_{j=1}^{n} q_{ij}^2 = 1$ für alle i, da Q orthonormal ist, d.h. die Eigenvektoren haben alle die Länge 1. Damit ist (2.180) gezeigt. □

Anmerkung: $A = Q\Lambda^{1/2}P'$ impliziert $A'A = P\Lambda P'$ und wegen der Orthonormalität der Spaltenvektoren von P folgt in analoger Weise

$$\text{spur}(A'A) = \sum_{i=1}^{n} \lambda_j. \tag{2.181}$$

2.8.5 Die Approximation von Matrizen

Der folgende Satz macht eine Aussage über die Güte der Approximation einer Matrix A durch eine Matrix mit kleinerem Rang. So sei etwa $A = X$ eine Datenmatrix mit dem Rang n und man will versuchen, X durch eine Matrix X_r mit dem Rang $r < n$ zu approximieren, d.h. durch möglichst wenige latente Variable zu "erklären".

Satz 2.31 *Es seien A und A_k (m, n)-Matrizen, $m \geq n$, und es seien die Matrizen A und A_k durch*

$$A = Q\Lambda^{1/2}P' = \sum_{j=1}^{n} \sqrt{\lambda_j}\boldsymbol{q}_j\boldsymbol{p}_j', \quad A_k = Q\Lambda_k^{1/2}P' = \sum_{j=1}^{k} \sqrt{\lambda_j}\boldsymbol{q}_j\boldsymbol{p}_j' \tag{2.182}$$

definiert, wobei $\Lambda = diag(\lambda_1, \lambda_2, \cdots, \lambda_n)$ mit $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ und $\Lambda_k = diag(\lambda_1, \lambda_2, \cdots, \lambda_k)$ mit $k < n$ sei. Dann gilt

$$\|A - A_k\|_2 = \sqrt{\lambda_{k+1}}. \tag{2.183}$$

Beweis: Es ist $A = Q\Sigma P'$, $A_k = Q\Sigma_k P'$, wobei $\Sigma = \Lambda^{1/2}$, $\Sigma_k = \Lambda_k^{1/2}$, Λ die Diagonalmatrix der Eigenwerte von $A'A$, Λ_k die Diagonalmatrix der ersten k Eigenwerte. Dann ist

$$A - A_k = Q\Sigma P' - Q\Sigma_k P' = Q(\Sigma - \Sigma_k)P' = Q\Sigma^* P',$$

$\Sigma^* = \text{diag}(\underbrace{0, \ldots, 0}_{k}, \sigma_{k+1}, \ldots, \sigma_n)$, $\sigma_j = \sqrt{\lambda_j}$. Da $\|A\|_2 = \sigma_{\max} = \sigma_1$ im Falle geordneter Singularwerte $\sigma_{k+1} \geq \cdots \geq \sigma_n$, folgt

$$\|A - A_k\|_2 = \sigma_{k+1} = \sqrt{\lambda_{k+1}},$$

da nun σ_{k+1} der maximale Singularwert ist. □

Anmerkungen:

1. Die Approximation wird trivialerweise immer besser, je größer der Wert von k ist, da ja der Wert von λ_{k+1} mit größer werdendem k immer kleiner wird. Der nichttriviale Teil der Aussage ist, dass $\|A - A_k\|_2$ gerade dem Wert von $\sqrt{\lambda_{k+1}}$ entspricht.

2. Bei der Approximation von A durch A_k wurde von der SVD von A Gebrauch gemacht. Die Gleichung $A = Q\Lambda^{1/2}P'$ ist insofern trivial, als die SVD stets gilt. Dass man A durch A_k approximiert, wobei A_k nur durch die ersten k Terme der SVD definiert ist, kann zur Frage führen, ob es eine andere Repräsentation für A_k gibt, die nicht auf der SVD beruht, aber besser ist in dem Sinne, dass $\|A - A_k\|_2 < \sqrt{\lambda_{k+1}}$. Eine solche gibt es nicht, wie noch gezeigt werden wird.

3. Man vergleiche die Aussage (2.183) mit der Aussage (2.182) von Satz 2.20. Wie die Gleichung (2.121), also die SVD von A, zeigt, ist A additiv durch Matrizen aufgebaut, die jeweils als dyadisches Produkt der Singularvektoren $\mathbf{q}_j$ und $\mathbf{p}_j$ definiert sind und die jeweils den Rang 1 haben (vergl. Satz 2.5, Seite 55). Der Rang $\text{rg}(A) \leq \min(m, n)$ ist durch die Anzahl der von Null verschiedenen Eigenwerte λ_j und damit durch die Anzahl der von Null verschiedenen $\sigma_k \mathbf{q}_k \mathbf{p}_k$ gegeben. Da die ersten k Eigenwerte von A und A_k identisch sind, enthält die Differenz $\Lambda^{1/2} - \Lambda_k^{1/2}$ nur Nullen, und der erste von Null verschiedene Wert in der Diagonalen ist $\sigma_{k+1} = \sqrt{\lambda_{k+1}}$. Da die Eigenwerte λ_j in Λ der Größe nach angeordnet sind, ist σ_{k+1} nun der größte Singularwert für $A - A_k$. □

Im Folgenden bedeutet $\min_{\text{rg}(B)=k} \|X - B\|$ bzw $\min_{\text{rg}(B)=k} \|X - B\|_F$ diejenige Matrix B, die (i) den Rang $\text{rg}(B) = k$ hat und die (ii) den Wert für die Norm $\|X-B\|$ bzw. $\|X-B\|_F$ minimiert. Es kann nun der folgende Satz bewiesen werden:

Satz 2.32 *Es seien A und B (m, n)-Matrizen mit $m > n$, wobei A den Rang r und B den Rang $k < r$ habe. Weiter sei*

$$A_k = Q\Lambda_k^{1/2} P' = \sum_{j=1}^{k} \sqrt{\lambda_j}\, \boldsymbol{q}_j \boldsymbol{p}_j', \tag{2.184}$$

Λ_k die zu den ersten[39] k Eigenvektoren korrespondierenden Eigenwerte von X enthält. Dann gilt

$$\min_{B \in \mathbb{R}^{m,n}, rg(B)=k} \|A - B\|_2 = \|A - A_k\|_2 = \sigma_{k+1} = \sqrt{\lambda_{k+1}} \tag{2.185}$$

Anmerkung: Dieser Satz wird gelegentlich als Satz von Eckart & Young bezeichnet, weil Eckart & Young (1936) eine derartige Aussage vorgestellt haben, allerdings nicht mit diesem Beweis. Tatsächlich hat schon Schmidt (1907) diese Aussage vorgestellt, und Mirsky (1960) hat diesen und den folgenden Satz 2.33 in allgemeiner Weise bewiesen, so dass auch zusammenfassend vom Schmidt-Mirsky-Theorem gesprochen wird.

Beweis: Zur Vereinfachung werde

$$\|A - B_{\min}\| = \min_{B \in \mathbb{R}^{m,n}, \text{rg}(B)=k} \|A - B\|_2$$

gesetzt. Zu zeigen ist, dass

$$\min_{B \in \mathbb{R}^{m,n}, \text{rg}(B)=k} \|A - B\|_2 = \|A - A_k\|_2.$$

[39] Es wird angenommen, dass die Eigenwerte der Größe nach geordnet sind, $\lambda_1 > \lambda_2 > \cdots > \lambda_k$

Dazu werde angenommen, dass

$$\|A - B_{\min}\| < \|A - A_k\|_2.$$

Die Ungleichung bleibt bestehen, wenn beide Seiten mit dem gleichen Faktor (>0) multipliziert werden. Für alle n-dimensionalen Vektoren $\mathbf{b}$ gilt dann

$$\|A - B_{\min}\|\|\mathbf{b}\| < \|A - A_k\|_2\|\mathbf{b}\| = \sigma_{k+1}\|\mathbf{b}\|.$$

Dann gilt auch

$$\|(A - B_{\min})\mathbf{b}\| \leq \|(A - A_k)\mathbf{b}\| \leq \sigma_{k+1}\|\mathbf{b}\|.$$

Insbesondere kann dann $\mathbf{b}$ als Linearkombination der ersten $k+1$ (Eigen-)Vektoren von P gewählt werden: sind also gerade die ersten $k+1$ Spalten von P die Spalten von P_{k+1}, so sei $\mathbf{b} = P_{k+1}\mathbf{x} = x_1\mathbf{p}_1 + \cdots + x_{k+1}\mathbf{p}_{k+1}$. Es ist

$$B_{\min}\mathbf{b} = B_{\min}P_{k+1}\mathbf{x},$$

$B_{\min}P_{k+1}$ ist also eine $(m\times(k+1))$-Matrix. Nach Satz 2.6, Gleichung (2.53) (Seite 55) ist aber $\mathrm{rg}(B_{\min}P_{k+1}) \leq \min(\mathrm{rg}(B_{\min})), \mathrm{rg}(P_{k+1})) = k$, da ja $\mathrm{rg}(B_{\min}) = k$ nach Voraussetzung. Dann folgt aber

$$\mathrm{rg}(B_{\min}P_{k+1}) + \dim(\mathrm{kern}(B_{\min}P_{k+1})) = k+1,$$

d.h.

$$\dim(\mathrm{kern}(B_{\min}P_{k+1})) \geq k+1-k = 1.$$

Also enthält $\mathrm{kern}(B_{\min}P_{k+1})$ mindestens einen Vektor $\mathbf{x}$ mit $B_{\min}P_{k+1}\mathbf{x} = \vec{0}$. Es sei also $\mathbf{x} \in \mathrm{kern}(B_{\min}P_{k+1})$; dann folgt

$$\|A\mathbf{b} - B_{\min}\mathbf{b}\| = \|AP_{k+1}\mathbf{x}\| < \|(AP_{k+1}\mathbf{x} - A_kP_{k+1}\mathbf{x}\| \leq \sigma_{k+1}\|\mathbf{b}\|.$$

Es ist aber

$$\|A\|\|P_{k+1}\mathbf{x}\| < \|(A - A_kP_{k+1}\|\|\mathbf{x}\| \leq \sigma_{k+1}\|\mathbf{b}\|,$$

d.h.

$$\sigma_1 < \sigma_{k+1},$$

im Widerspruch zu $\sigma_1 \geq \sigma_{k+1}$. Damit gilt (2.185). □

Satz 2.33 *(Satz von Schmidt-Mirsky) Es sei A und $B_{\min}$ (m,n)-Matrizen mit $m > n$, wobei A den Rang r und B den Rang $k < r$ habe, $B_{\min}$ sei wie in Satz 2.32 definiert und A_k sei wie in (2.184) definiert. Dann gilt*

$$\|A - B_{\min}\|_F = \sqrt{\sum_{j=k+1}^{n} \lambda_j}, \tag{2.186}$$

wobei $\|\cdot\|_F$ die Frobenius-Norm ist.

Beweis: Die Anwendung der SVD auf $A - A_k$ liefert

$$\|A - A_k\|^2 = \|Q(\Lambda^{1/2} - \Lambda_k^{1/2})P'\|^2 = \sum_{j=1}^{n} \lambda_j - \sum_{j=k+1}^{n} \lambda_j = \|A\|_F^2 - \sum_{j=k+1}^{n} \lambda_j.$$

$B_{\min}$ kann als Summe von durch dyadische Produkte definierte Matrizen definiert werden, also

$$B_{\min} = \sum_{j=1}^{n} \mathbf{x}_j \mathbf{y}_j',$$

wobei die $\mathbf{x}_j$ m-dimensionale und die $\mathbf{y}_j$ n-dimensionale Vektoren sind. Da auch für $B_{\min}$ eine Singularwertzerlegung gilt, können für $\mathbf{x}_j$ und $\mathbf{y}_j$ die jeweils mit $\sqrt{\sigma_j}$ multiplizierten Links- und Rechtssingulärvektoren von $B_{\min}$ gewählt werden, d.h. man kann orthogonale Vektoren wählen. Zu zeigen ist dann, dass

$$\|A - \sum_{j=1}^{k} \mathbf{x}_j \mathbf{y}_j\| \geq \|A\|^2 - \sum_{j=1}^{k} \lambda_j.$$

Nach Definition der Frobenius-Norm hat man

$$\begin{aligned} \|A - \sum_{j=1}^{k} \mathbf{x}_j \mathbf{y}_j\|_F^2 &= \text{spur}\left((A - \sum_{j=1}^{k} \mathbf{x}_j \mathbf{y}_j)'(A - \sum_{j=1}^{k} \mathbf{x}_j \mathbf{y}_j) \right) \\ &= \text{spur}\left(A'A + \sum_{j=1}^{k} (\mathbf{y}_j - A'\mathbf{x}_j)(\mathbf{y}_j - A'\mathbf{x}_j)' - \sum_{j=1}^{k} A'\mathbf{x}_j \mathbf{x}_j A \right) \end{aligned}$$

Es ist $\text{spur}((\mathbf{y}_j - A'\mathbf{x}_j)(\mathbf{y}_j - A'\mathbf{x}_j)) \geq 0$, $\text{spur}(A'\mathbf{x}_j\mathbf{x}_j'A) = \|A'\mathbf{x}_j\|$ und es ist zu zeigen, dass

$$\sum_{j=1}^{k} \|A'\mathbf{x}_j\|^2 \leq \sum_{j=1}^{k} \lambda_j.$$

Die SVD von A sei $A = Q\Sigma P'$, und es sei $P_1 = [|\mathbf{p}_1| \ldots |\mathbf{p}_k|0]$, $P_2 = [|0|\mathbf{p}_{k+1}| \ldots |\mathbf{P}_n|]$, so dass $P = [P_1|P_2]$. Analog dazu sei $\Sigma_1 = \text{diag}(\sigma_1, \ldots, \sigma_k)$, $\Sigma_2 = \text{diag}(\sigma_{k+1}, \ldots, \sigma_n)$. Dann hat man

$$\begin{aligned} \|A'\mathbf{x}_j\| &= \|Q\Sigma P'\|_F^2 = \|\Sigma P'\mathbf{x}_j\|_F^2 = \\ &= \|\Sigma_1 P_1'\mathbf{x}_j\|_F^2 + \|\Sigma_2 P_2'\mathbf{x}_j\|_F^2 + \lambda_k - \lambda_k + \lambda_k(\|P'\mathbf{x}_j\| - \|P_1'\mathbf{x}_j\|_F^2 - \|P_2'\mathbf{x}_j\|_F^2) \\ &= \lambda_k + (\|\Sigma_1 \mathbf{P}_1'\mathbf{x}_j\|_F^2 - \underbrace{\lambda_k \|P_1'\mathbf{x}_j\|_F^2)}_{(1)} - \underbrace{\lambda_k(1 - \|P'\mathbf{x}_j\|_F^2)}_{(2)} \end{aligned}$$

Der Term (1) ist positiv, ebenso (2), da P orthonormal, und $\mathbf{x}_j$ ist ebenfalls orthonormal. Dann folgt

$$\begin{aligned}\sum_{j=1}^{k}\|A'\mathbf{x}_j\|^2 &\leq k\lambda_k + \sum_{j=1}^{k}(\|\Sigma_1 P_1'\|^2 - \lambda_k\|P_1'\mathbf{x}_j\|^2)\\ &= k\lambda_k + \sum_{j=1}^{k}\sum_{i=1}^{k}(\lambda_i-\lambda_j)|\mathbf{v}_j'\mathbf{x}_j|^2\\ &\leq \sum_{i=1}^{k}(\lambda_k+(\lambda_i-\lambda_k)) = \sum_{j=1}^{k}\lambda_j.\end{aligned}$$

□

2.9 Lineare Gleichungssysteme

In Abschnitt 1.4, Seite 25 wurde zur Illustration der Darstellbarkeit von Vektoren als Linearkombination anderer Vektoren bereits ein System von zwei Gleichungen mit zwei Unbekannten betrachtet; man kann das System in der Form $X\mathbf{a} = \mathbf{y}$ schreiben, mit $X = [\mathbf{x}_1, \mathbf{x}_2]$ und $\mathbf{a} = [a_1, a_2]$. Die Lösung für die unbekannten Koeffizienten a_1 und a_2 ist eindeutig, wenn die Vektoren $\mathbf{x}_1$ und $\mathbf{x}_2$ linear unabhängig sind; sind sie linear abhängig, so ist die Lösung nicht mehr eindeutig. Nicht betrachtet wurde der Fall $\mathbf{y} = a_1\mathbf{x}_1 + a_2\mathbf{x}_2 + a_3\mathbf{x}_3$, wobei $\mathbf{y}, \mathbf{x}_1$ und $\mathbf{x}_3$ 2-dimensionale Vektoren sind; allgemein ist dies der Fall, wenn man weniger Gleichungen als Unbekannte hat. Die Lösungsgesamtheit hängt dann vom Rang der Matrix $X = [\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3]$ ab. Im Folgenden wird ein allgemeiner Lösungsansatz vorgestellt, wobei allerdings die Notation geändert wird. Die Matrix X wird in A umbenannt, und der Vektor $\mathbf{a}$ der Unbekannten wird in $\mathbf{x}$ umbenannt, – damit wird die allgemein übliche Schreibweise verwendet.

Es sei also A eine (m, n)-Matrix, $\mathbf{x}$ ein n-dimensionaler Vektor, und $\mathbf{y}$ ein m-dimensionaler Vektor, und es gelte

$$A\mathbf{x} = \mathbf{y}, \quad \mathbf{x} \in \mathbb{R}^n, \ \ \mathbf{y} \in \mathbb{R}^m \tag{2.187}$$

Formal bedeutet diese Gleichung, dass $\mathbf{y}$ eine Linearkombination der Spaltenvektoren von A ist oder sein soll. $\mathbf{x}$ ist dann der Vektor der Koeffizienten: gilt $A = [\mathbf{a}_1, \ldots, \mathbf{a}_n]$, $\mathbf{x} = (x_1, \ldots, x_n)'$, so hat man

$$A\mathbf{x} = x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots x_n\mathbf{a}_n = \mathbf{y} \tag{2.188}$$

Die Komponenten $x_1, \ldots, x_n$ seien nicht bekannt. Die Gleichung $A\mathbf{x} = \mathbf{y}$ kann als ein System von m linearen Gleichungen mit n Unbekannten, nämlich den Komponenten von $\mathbf{x}$, gesehen werden. $\mathbf{x}$ ist n-dimensional, $\mathbf{y}$ ist m-dimensional. Sind

A und $\mathbf{y}$ vorgegeben, so stellt sich die Frage, ob es überhaupt einen Lösungsvektor $\mathbf{x}$ gibt, und wenn ja, ob der Lösungsvektor eindeutig ist oder ob es mehrere Lösungsvektoren gibt.

Zunächst werde zwischen zwischen zwei Arten von Gleichungssystemen unterschieden:

1. $\mathbf{y} = \vec{0}$. Dann heißt das Gleichungssystem *homogen*,
2. $\mathbf{y} \neq \vec{0}$. Dann heißt das Gleichungssystem *inhomogen*.

Definition 2.19 *Es sei A eine (m,n)-Matrix mit dem Rang $r = rg(A) \leq \min(m,n)$. und $A\boldsymbol{x} = \boldsymbol{y}$ sei ein System von Gleichungen. Weiter sei*

$$kern(A) = \{\boldsymbol{x} \in \mathbb{R}^n | A\boldsymbol{x} = \vec{0}\} \quad (2.189)$$

$$\mathcal{L}(A) = \{\boldsymbol{y} \in \mathbb{R}^m | A\boldsymbol{x} = \boldsymbol{y}\}. \quad (2.190)$$

$kern(A)$ heißt Kern *von A, und $\mathcal{L}(A)$ ist die* lineare Hülle *der Spaltenvektoren von A.*

Anmerkungen:

1. kern(A) ist ein (Teil-)Vektorraum: sind die Vektoren $\mathbf{x}_1$ und $\mathbf{x}_2$ Elemente aus kern(A), so rechnet man leicht nach, dass dann auch $b_1\mathbf{x}_1 + b_2\mathbf{x}_2 \in$ kern(A) $(b_1, b_2 \in \mathbb{R})$ gilt[40].

2. Damit eine Lösung $\mathbf{x}$ existiert, müssen die Matrizen A und $(A, \mathbf{y})$ (die um die Spalte $\mathbf{y}$ erweiterte Matrix A) denselben Rang haben; dies folgt sofort aus der Tatsache, dass $\mathbf{y}$ eine Linearkombination der Spalten von A sein muß, damit eine Lösung $\mathbf{x}$ existiert. □

Der folgende Satz gilt für beliebige (m,n)-Matrizen X; er wird hier für $X = A$ angeschrieben, weil er in Bezug auf das Gleichungssystem $A\mathbf{x} = \mathbf{y}$ interpretiert werden soll:

Satz 2.34 *Es sei A eine (m,n)-Matrix mit der SVD $A = Q\Sigma T'$, wobei Q aus den Spaltenvektoren $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_n$ und T aus den Spaltenvektoren $\boldsymbol{t}_1, \ldots, \boldsymbol{t}_n$ bestehe; ist $rg(A) = r \leq \min(m,n)$, so sind r Singularwerte σ_k größer als Null und die restlichen sind gleich Null. Dann gilt*

$$\mathcal{L}(A) = \mathcal{L}(\boldsymbol{q}_1, \ldots, \boldsymbol{q}_r) \quad (2.191)$$

$$kern(A) = \mathcal{L}(\boldsymbol{t}_{r+1}, \ldots, \boldsymbol{t}_s), \quad s = \min(m,n) \quad (2.192)$$

$$rg(kern(A)) + rg(\mathcal{L}(A)) = \min(m,n) \quad (2.193)$$

[40] In Definition 4.3, Punkt 4., Seite 156 wird der Begriff des Kerns einer Abbildung eingeführt. Die Matrix A definiert eine Abbildung, und (2.189) definiert damit den Kern einer Abbildung.

Beweis: Der Beweis wird für den Fall $n \leq m$ (höchstens so viele Unbekannte wie Gleichungen) geführt, der Fall $m < n$ (weniger Gleichungen als Unbekannte) ist analog. Wegen $A = Q\Sigma T'$ sind die $\mathbf{a}_j$ Linearkombinationen der $r \leq \min(m,n)$ Spaltenvektoren $\mathbf{q}_k$ von Q; als Eigenvektoren von AA' sind die $\mathbf{q}_k$ paarweise orthogonal und damit linear unabhängig; sie bilden eine r-dimensionale Teilbasis des $\mathbb{R}^m$. Damit sind auch alle Linearkombinationen der $\mathbf{a}_j$ als Linearkombinationen der $\mathbf{q}_k$ darstellbar, so dass (2.191) gelten muß.

Der Kern von A sind alle n-dimensionalen Vektoren $\mathbf{x}$, für die $A\mathbf{x} = \vec{0}$ gilt. Die Spaltenvektoren von T bilden eine Basis des $\mathbb{R}^n$, so dass allgemein

$$\mathbf{x} = c_1\mathbf{t}_1 + \cdots + c_r\mathbf{t}_r + c_{r+1}\mathbf{t}_{r+1} + \cdots + c_n\mathbf{t}_n$$

geschrieben werden kann, wobei die $c_j \in \mathbb{R}$ geeignet gewählte Koeffizienten sind. Dann gilt

$$A\mathbf{x} = A\sum_{j=1}^{n} c_j\mathbf{t}_j = \sum_{j=1}^{n} c_j A\mathbf{t}_j = \sum_{j=1}^{r} c_j\sigma_j\mathbf{q}_j = \vec{0},$$

(vergl. (2.118), Seite 86), denn $\sigma_j = 0$ für $j > r$ (falls $r < \min(m,n)$). Wegen der linearen Unabhängigkeit der $\mathbf{q}_j$ kann diese Gleichung nur gelten, wenn $c_1 = \cdots = c_r = 0$. Dann kann $\mathbf{x} \neq \vec{0}$ kein Element des durch die $\mathbf{t}_1, \ldots, \mathbf{t}_r$ aufgespannten Vektorraums sein, sondern muß ein Element des $(n-r)$-dimensionalen Komplementäraums sein. Die $\mathbf{t}_{r+1}, \ldots, \mathbf{t}_n$ sind eine Basis für diesen Komplementärraum, so dass man

$$\mathbf{x} = c_{r+1}\mathbf{t}_{r+1} + \cdots + c_n\mathbf{t}_n,$$

ansetzen kann, und

$$A\mathbf{x} = A\sum_{j=r+1}^{n} c_j\mathbf{t}_j = \sum_{j=r+1}^{n} c_j A\mathbf{t}_j = \sum_{j=r+1}^{n} c_j\sigma_j\mathbf{q}_j = \vec{0},$$

wegen (2.118), Seite 86, und weil $\sigma_j = 0$ für $j > r$, falls $r < \min(m,n)$. Alle Linearkombinationen von Vektoren $\mathbf{x} \neq \vec{0}$ mit $A\mathbf{x} = \vec{0}$ sind Linearkombinationen der $\mathbf{t}_{r+1}, \ldots, \mathbf{t}_n$, und dies ist die Aussage von (2.192).

Die Gleichung (2.193) ist eine unmittelbare Folge der vorangegangenen Argumente: $\mathcal{L}(A)$ hat den Rang r und $\ker(A)$ hat den Rang $n-r$, so dass die Summe der Ränge gleich n sein muß. □

Anmerkung: Der Satz 2.34 ergab sich als Folgerung aus der SVD für die Matrix A. Die Eigenvektoren $\mathbf{t}_j$ von $A'A$ und $\mathbf{q}_k$ von AA' sind natürlich nicht die einzigen Basisvektoren, mit denen sich die Teilräume $\ker(A)$ und $\mathcal{L}(A)$ darstellen lassen. Einen alternativen, wenn auch etwas länglichen alternativen Beweis, in dem ein anderer Satz von Basisvektoren verwendet wird, findet man im Anhang, Abschnitt 4.5. □

Die allgemeine Lösungsmenge wird im folgenden Satz spezifiziert:

Satz 2.35 *Es sei $A\boldsymbol{x} = \boldsymbol{y} \in \mathcal{L}(A)$, $rg(A) = r$, und insbesondere sei $\boldsymbol{x} = \boldsymbol{x}_0$ eine bestimmte Lösung, so dass $A\boldsymbol{x}_0 = \boldsymbol{y}$ gilt. Der Kern kern(A) besteht aus dem $(n-r)$-dimensionalen Teilraum $\mathcal{L}_{n-r} = \mathcal{L}(\boldsymbol{t}_{r+1}, \ldots, \boldsymbol{t}_n)$ des V_n. Dann ist die Menge der Lösungsvektoren durch*

$$\mathcal{L} = \{\boldsymbol{x}_0 + \boldsymbol{x} | \boldsymbol{x} \in \mathcal{L}_{n-r}\} \tag{2.194}$$

gegeben.

Beweis: Tatsächlich ist $\mathbf{x}_0 + \mathbf{x}$ eine Lösung, denn

$$A(\mathbf{x}_0 + \mathbf{x}) = A\mathbf{x}_0 + A\mathbf{x} = A\mathbf{x}_0,$$

denn $A\mathbf{x} = \vec{0}$ ist nach Voraussetzung eine Lösung, und $\mathbf{x}_0$ war als Lösungsvektor vorausgesetzt worden. Umgekehrt sei $\mathbf{x}_1$ ein Lösungsvektor. Es muß gezeigt werden, dass $\mathbf{x}_1 \in \mathcal{L}$ ist. Nach Voraussetzung muß $A\mathbf{x}_1 = \mathbf{y}$ gelten. Für irgendeinen Vektor $\mathbf{x} \in \mathcal{L}_{n-r}$ muß $A\mathbf{x} = \vec{0}$ gelten. Dann muß aber auch

$$A(\mathbf{x}_1 + \mathbf{x}) = A\mathbf{x}_1 = \mathbf{y}$$

gelten, so dass $\mathbf{x}_1 + \mathbf{x} \in \mathcal{L}$ liegt. □

Der Fall $m = n = r$: Ist $m = n = r$, r der Rang von A, so existiert die Inverse A^{-1} und aus $A\mathbf{x} = \mathbf{y} \in V_n^r = \mathcal{L}(A)$ folgt sofort die Lösung

$$\mathbf{x} = A^{-1}\mathbf{y}, \quad \mathbf{y} \in \mathcal{L}(A. \tag{2.195}$$

Der Fall $m > n = r$: In diesem Fall gibt es mehr Gleichungen als Unbekannte; das Gleichungssystem ist überbestimmt. Im Allgemeinen wird man keinen Lösungsvektor $\mathbf{x}$ finden, der allen Gleichungen exakt genügt. Dies ist z.B. bei der multiplen Regression der Fall, da man üblicherweise eine größere Anzahl m von Fällen als unbekannte Regressionsparameter hat. Die (m, n)-Matrix $A = X$ der Prädiktoren hat aber im Allgemeinen den vollen Rang $r = n$, so dass man eine Lösung finden könnte, indem man von links mit A' multipliziert, so dass $A'A\mathbf{x} = A'\mathbf{y}$ folgt, und da $A'A$ den gleichen Rang wie A hat (s. (2.4), Seite 76) existiert die zu $A'A$ inverse Matrix $(A'A)^{-1}$, so dass

$$\mathbf{x} = (A'A)^{-1}A'\mathbf{y} \tag{2.196}$$

resultiert. Die Matrix $(A'A)^{-1}A'$ ist eine Pseudoinverse für A, s. Seite 102. Sind die Komponenten von $\mathbf{y}$ Messwerte, so sind sie üblicherweise durch Messfehler kontaminiert, so dass $\mathbf{y} \notin \mathcal{L}(A)$. (2.196) liefert dann keine Lösung $\mathbf{x}$, die allen m Gleichungen genügt. (2.196) ist die bekannte Kleinste-Quadrate-Schätzung $\hat{\mathbf{x}}$ für $\mathbf{x}$, vergl. (4.34), Seite 149, und Abschnitt 2.8.3.

Für den Fall $r < n$ liefert (2.194) den Lösungsraum.

Cramersche Regel: Diese Regel wird hier nur der Vollständigkeit wegen genannt. Es sei $A\mathbf{x} = \mathbf{y}$ ein Gleichungssystem, wobei A eine (n, n)-Matrix sei. Dann gilt für die j-te Komponente x_j von $\mathbf{x}$

$$x_j = \frac{|A_j|}{|A|}, \quad j = 1, \ldots, n \tag{2.197}$$

Dabei ist $|A|$ die Determinante von A, und $|A_j|$ ist die Determinante der Matrix A_j, die entsteht, indem man in A die j-te Spalte durch $\mathbf{y}$ ersetzt. Der Begriff der Determinante wird im Anhang, Abschnitt **??**, kurz eingeführt. Offenbar können die x_j nur berechnet werden, wenn $|A| \neq 0$; diese Bedingung setzt voraus, dass die Spaltenvektoren von A linear unabhängig sind.

2.10 Latente Variablen

Datenmatrizen X lassen sich oft "erklären", indem man etwa die Spaltenvektoren von X als Linearkombinationen der Vektoren einer (Teil-)Basis darstellt; diese Basisvektoren liefern u. U. eine Interpretation der Kovarianzen oder Korrelationen zwischen den gemessenen Variablen. Da man sich für eine bestimmte Basis oder Teilbasis aus der Menge der möglichen entscheiden muß, wählt man die Hauptachsentransformation. Die erste latente Variable erklärt dannn einen maximalen Anteil der Gesamtvarianz, die zweite erklärt dann den zweitgrößten Anteil, etc. Dieser Ansatz besteht in einer direkten Anwendung der SVD, s. Abschnitt 2.10.1.

Eine andere Fragestellung, die auf die Berechnung latenter Variablen führt, ist die Zuordnung von Fällen zu bestimmten Kategorien auf der Basis von Messungen einer Reihe von Variablen. Diese Messungen können wieder zu einer Matrix X zusammengefasst werden, und die latenten Variablen werden so bestimmt, dass die Projektionen der Fälle auf die latenten Variablen nach Maßgabe maximaler Trennung der Kategorien erfolgt. Der Standardansatz für diese Fragestellung wird in Abschnitt 2.10.2 vorgestellt.

2.10.1 Hauptkomponenten und SVD

Die Hauptkomponentenanalyse (engl. Principal Component Analysis (PCA); PCA ist die gängige Abkürzung) ist eine Methode, einen als eine (m, n)-Matrix X gegebenen Datensatz in einer der Faktorenanalyse ähnlichen Weise zu komprimieren; die Methode geht auf Karl Pearson (1901) und Harold Hotelling (1933) zurück. Gewöhnlich wird X als zentriert vorausgesetzt (spaltenzentriert, wenn die Spalten gemessene Variablen repräsentieren und man an der Analyse der Kovarianzen oder Korrelationen zwischen den Variablen interessiert ist, zeilenzentriert, wenn man an einer Analyse der Kovarianzen oder Korrelationen der "Fälle" interessiert

ist, – sofern diese Analyse Sinn macht[41]). Sie besteht im Wesentlichen aus einer Koordinatentransformation, nämlich dem Übergang von den ursprünglichen Koordinaten zu den Hauptachsen des Ellipsoids, das die Punktekonfiguration der Fälle definiert. Insbesondere Hotellings Ansatz bestand darin, die gesuchten "latenten Variablen" oder "latenten Dimensionen" so zu bestimmen, dass sich die Fälle auf der ersten latenten Variablen maximal unterscheiden, d.h. dass die Varianz der Koordinaten der Fälle auf der ersten latenten Dimension maximal ist, auf der zweiten latenten Dimension soll die Varianz der Koordinaten zweitmaximal sein, etc.

Die PCA ergibt sich unmittelbar aus der SVD: nach Satz 2.112 gilt $X = Q'\Lambda^{1/2}T'$, wobei Q und T die Eigenvektoren von XX' bzw. $X'X$ sind; Q und T sind also orthonormal. Dann folgt $XT = Q\Lambda^{1/2} = U$; oft wird U in L umbenannt, um den Ausdruck *latente Variable* zu repräsentieren. Die i-te Zeile von X enthält die Messungen des i-ten Falles für die n Variablen, und die i-te Zeile von L enthält die Koordinaten des i-ten Falles auf "latenten Variablen" oder "latenten Dimensionen", deren Orientierung durch die Hauptachsen der durch $X'X$ definierten Ellipsoide gegeben ist. Aus der Herleitung der SVD geht hervor, dass die Hotellingsche Forderung bezüglich der Koordinaten der Fälle durch die Wahl von L erfüllt wird; $\mathbf{L}_1'\mathbf{L}_1$ ist maximal (vergl. die Anmerkungen auf Seite 85). Darüber hinaus liefert die SVD die bei Hotelling nicht weiter diskutierte Beziehung zwischen der Repräsentation der Fälle einerseits und der Variablen andererseits (R- und Q-Analyse, Cattell (1966)). Bei der R-Analyse werden die Korrelationen zwischen den Variablen, bei der Q-Analyse werden die Korrelationen zwischen den Fällen analysiert. Die Frage ist, worin sich die latenten Dimensionen bei diesen beiden Analysearten unterscheiden, bzw. welche Beziehungen zwischen den Ergebnissen dieser Analysen existieren. Die SVD gibt diese Beziehungen in kompakter Form an, wie im Folgenden gezeigt wird.

Für die folgenden Betrachtungen wird angenommen, dass X spaltenzentriert oder sogar spaltenstandardisiert ist. $C = \frac{1}{m}X'X$ ist dann eine Kovarianz- bzw. eine Korrrelationsmatrix für die Variablen.

Die folgenden Gleichungen drücken zwei Varianten mit den im Zusammmenhang mit der PCA üblichen Bezeichnungen für die Matrizen aus:

$$X = Q\Lambda^{1/2}T' = LT', \quad L = Q\Lambda^{1/2} \tag{2.198}$$

$$= QA', \quad A = T\Lambda^{1/2} \tag{2.199}$$

Es werde $m > n$ vorausgesetzt, d.h. man habe mehr Fälle als Variablen; diese Voraussetzung ist für die folgenden Betrachtungen nicht essentiell, entspricht aber den normalerweise vorliegenden Bedingungen. Der Rang von X sei $r \leq \min(m, n)$; es wird der Einfachheit halber $r = n = \min(m, n)$ angenommen, weil im Allge-

[41] Sind die Fälle z.B. Personen, an denen verschiedene Variablen gemessen wurden, so bedeutet die Berechnung einer Korrelation zwischen Fällen das Mitteln über die Variablen, was eine sinnfreie Übung sein kann!

meinen die Eigenwerte *numerisch*[42] alle ungleich Null sind; der Fall, dass der wahre Wert von r kleiner als $\min(m,n)$ ist, wird später behandelt. Q bzw. L ist eine (m,n)-Matrix, T bzw. A ist eine (m,n)-Matrix, und Λ ist eine (m,n)-Diagonalmatrix. Die Spalten von Q bzw. L sind Basisvektoren für die Spaltenvektoren $\mathbf{x}_j$ von X, und die Zeilen von Q bzw. L repräsentieren die Fälle. Die Spaltenvektoren von T bzw. A sind Basisvektoren für die Zeilenvektoren von X.

Definition 2.20 *Es seien*

$$\boldsymbol{L}_k = \begin{pmatrix} \ell_{1k} \\ \ell_{2k} \\ \vdots \\ \ell_{mk} \end{pmatrix}, \quad \boldsymbol{a}_k = \begin{pmatrix} a_{1k} \\ a_{2k} \\ \vdots \\ a_{nk} \end{pmatrix},$$

$\boldsymbol{L}_k$ ist der k-te Spaltenvektor von L, $\boldsymbol{a}_k$ ist der k-te Spaltenvektor von A (s. (2.199)). Die Komponenten ℓ_{ik} von $\boldsymbol{L}_k$, $i = 1,\ldots,m$, heißen Faktorwerte *(factor scores) der Fälle auf der k-ten latenten Variablen, die Komponenten a_{jk} von $\boldsymbol{a}_k$, $j = 1,\ldots,n$, heißen* Ladungen *(factor loadings) der Variablen auf der k-ten latenten Variablen,.*

Wie oben schon angemerkt wurde soll die Bezeichung L für die Matrix $Q\Lambda^{1/2}$ an den Ausdruck 'latent' erinnern, A ist die in der Literatur übliche Bezeichung für die Matrix der Faktorladungen.

Anmerkung: Gelegentlich werden in der Literatur auch die Komponenten des Spaltenvektors $\mathbf{q}_k$ von Q als Faktorwerte bezeichnet; aus dem jeweiligen Kontext wird im Allgemeinen klar, was jeweils mit dem Ausdruck 'Faktorwert gemeint ist.

□

Aus (2.198) und (2.199) erhält man sofort die Beziehungen

$$X = LT' = QA' \tag{2.200}$$
$$X' = AQ' = TL' \tag{2.201}$$

Da T orthonormal ist, folgt aus (2.200) sofort der Hotellingsche Ansatz $XT = L$, d.h. die Spaltenvektoren von L sind Linearkombinationen der Spaltenvektoren $\mathbf{x}_j$ von X. Andererseits sind nach (2.200) die Spaltenvektoren $\mathbf{x}_j$ von X Linearkombinationen der Spaltenvektoren $\mathbf{L}_k$ von L bzw $\mathbf{q}_k$ von Q, und die Spaltenvektoren $\tilde{\mathbf{x}}_i$ von X' sind Linearkombinationen der Spaltenvektoren $\mathbf{a}_k$ von A bzw. der Spaltenvektkoren $\mathbf{t}_k$ von T; insbesondere gilt

$$\mathbf{x}_j = L\tilde{\mathbf{t}}_j = Q\tilde{\mathbf{a}}_j \tag{2.202}$$
$$\tilde{\mathbf{x}}_i = A\tilde{\mathbf{q}}_i = T\tilde{\mathbf{L}}_i \tag{2.203}$$

[42] Messungen haben stets nur eine endliche Genauigkeit, Berechnungen können stets nur mit endlich vielen Dezimalstellen durchgeführt werden, etc, so dass die Elemente x_{ij} von X i. A. von den "wahren" Werten abweichen.

wobei $\tilde{\mathbf{t}}_j$ der j-te Spaltenvektor von T' (Zeilenvektor von T) und $\tilde{\mathbf{q}}_i$ der i-te Spaltenvektor von Q' (Zeilenvektor von Q) ist. Die Spaltenvektoren $\mathbf{t}_k$ von T bzw. $\mathbf{a}_k$ von A repräsentieren, ebenso wie die Spaltenvektoren $\mathbf{L}_k$ bzw. $\mathbf{q}_k$, latente Variablen (auch: latente Dimensionen); auf Fragen der inhaltlichen Interpretation wird am Ende dieses Abschnitts eingegangen. Die Zeilenvektoren $\tilde{\mathbf{t}}_i$ bzw. $\tilde{\mathbf{a}}_i$ repäsentieren Koordinaten der Variablen auf den latenten Variablen, und die Zeilenvektoren $\tilde{\mathbf{L}}_i$ oder $\tilde{\mathbf{q}}_i$ von L bzw. Q repräsentieren die Fälle auf latenten Variablen.

Die Gleichung (2.203) beschreibt die in (2.82), Seite 69 angebene Rotation von achsenparallelen Ellipsoiden in orientierte Ellipsoide und umgekehrt. Denn (2.203) impliziert $\tilde{\mathbf{x}}_i' = \tilde{\mathbf{L}}_i' T'$, und die Multiplikation von rechts mit $X'X = T\Lambda T'$ und noch einmal mit $\tilde{\mathbf{x}}_i$ liefert

$$\tilde{\mathbf{x}}_i'(X'X)\tilde{\mathbf{x}}_i = \tilde{\mathbf{L}}_i' T'T\Lambda T'T\tilde{\mathbf{L}}_i = \tilde{\mathbf{L}}_i'\Lambda\tilde{\mathbf{L}}_i$$

d.h. T rotiert die Vektoren $\tilde{\mathbf{L}}_i$ des achsenparallelen Ellipsoids in die Vektoren $\tilde{\mathbf{x}}_i$ des orientierten Ellipsoids. In diesem Sinne ist die SVD äquivalent einer Hauptachsentransformation.

Die Faktorenscores ℓ_{ik} und die Ladungen a_{jk} sind Skalarprodukte

$$\ell_{ik} = \tilde{\mathbf{x}}_i'\mathbf{t}_k = \sum_{j=1}^{n} x_{ij}t_{jk} = \|\tilde{\mathbf{x}}_i\|\|\mathbf{t}_k\|\cos\theta_{ik}. \tag{2.204}$$

$$a_{jk} = \mathbf{x}_j'\mathbf{q}_k = \sum_{i=1}^{m} x_{ij}q_{ik} = \|\mathbf{x}_j\|\|\mathbf{q}_k\|\cos\phi_{jk}. \tag{2.205}$$

Je kleiner der Winkel θ_{ik} zwischen dem die k-te latente Dimension repräsentierenden Vektor $\mathbf{t}_k$ und dem Zeilenvektor $\tilde{\mathbf{x}}_i$ ist, desto größer ist ℓ_{ik}; der mögliche Maximalwert (für gegebene Längen $\|\tilde{\mathbf{x}}_i\|$ und $\|\mathbf{t}_k\|$) wird erreicht für $\theta_{ik} = 0$, also für $\cos\theta_{ik} = 1$. ℓ_{ik} reflektiert das Maß, in dem der i-te Fall durch die k-te Dimension bestimmt wird. Die Interpretation von a_{jk} ist analog.

Der Messwert x_{ij} für den i-ten Fall bei der j-ten Variablen ist

$$x_{ij} = \tilde{\mathbf{q}}_i'\tilde{\mathbf{a}}_j = \|\tilde{\mathbf{q}}_i\|\|\tilde{\mathbf{a}}_j\|\cos\varphi_{ij} = \sum_{k=1}^{n}\sqrt{\lambda_k}q_{ik}t_{kj} \tag{2.206}$$

und x_{ij} wird maximal (relativ zur Länge der Vektoren $\tilde{\mathbf{q}}_i$ und $\tilde{\mathbf{a}}_i$), wenn der Winkel φ_{ij} zwischen diesen Vektoren gleich Null ist, wenn also das Profil $\tilde{\mathbf{q}}_i$ der Werte des i-ten Falls auf den latenten Variablen proportional zum Profil $\tilde{\mathbf{a}}_j$ der Werte der Variablen auf den latenten Variablen ist. Natürlich kann man auch $x_{ij} = \tilde{\mathbf{L}}_i'\tilde{\mathbf{t}}_j$ betrachtet werden, – die Interpretation ist analog; es ist der Winkel zwischen den Vektoren $\tilde{\mathbf{q}}_i$ und $\tilde{\mathbf{t}}_j$, der für gegebene Längen der Vektoren den Wert von x_{ij} bestimmt.

Ladungen und die Korrelationen zwischen den Variablen: Die Skalierung $\mathbf{a}_k = \sqrt{\lambda_k}\mathbf{t}_k$ der Spaltenvektoren $\mathbf{t}_k$ von T erweist sich als vorteilhaft, wenn man

insbesondere an einer Analyse der Variablen interessiert ist. So folgt aus $X = QA'$ die Beziehung $X'X = AQ'QA'$, und wegen $Q'Q = I$ gilt

$$X'X = AA'. \tag{2.207}$$

Es gelte $X = Z$, d.h. die Matrix X sei spaltenstandardisiert. Dann ist

$$R = \frac{1}{m}X'X = \frac{1}{m}AA', \quad X = Z \tag{2.208}$$

die Matrix der Korrelationen zwischen den Variablen. Die Korrelation zwischen der j-ten und der k-ten Variablen ergibt sich als Skalarprodukt zwischen den Zeilenvektoren $\tilde{\mathbf{a}}_j$ und $\tilde{\mathbf{a}}_k$ von A:

$$r_{jk} = \frac{1}{m}\tilde{\mathbf{a}}_j'\tilde{\mathbf{a}}_k = \frac{1}{m}\sum_{u=1}^{n} a_{uj}a_{uk} = \frac{1}{m}\|\tilde{\mathbf{a}}_j\|\|\tilde{\mathbf{a}}_k\|\cos\theta_{jk}, \tag{2.209}$$

θ_{jk} der Winkel zwischen den Vektoren $\tilde{\mathbf{a}}_j$ und $\tilde{\mathbf{a}}_k$; je kleiner der Winkel, desto größer ist der Absolutbetrag $|r_{jk}|$. Für $j = k$ erhält man ($\cos\theta_{jj} = 1$ wegen $\theta_{jj} = 0$)

$$r_{jj} = \frac{1}{m}\|\tilde{\mathbf{a}}_j\|^2 = \frac{1}{m}\sum_{u=1}^{n} a_{ju}^2 = 1, \quad j = 1, \ldots, n \tag{2.210}$$

Für jede Variable ist die Summe $\|\mathbf{a}_j\|^2$ der Quadrate der Ladungen auf den latenten Dimensionen gleich m. Die Division durch m impliziert, dass die Variablen durch Punkte (Endpunkte der entsprechenden Vektoren) auf einer n-dimensionalen Hyperkugel mit dem Radius 1 repräsentiert werden. Ist insbesondere $n = 2$, so liegen die Punkte auf einem Kreis.

Es sei $\mathbf{a}_k$ der der k-te Spaltenvektor von A. Dann gilt

$$\|\mathbf{a}_k\|^2 = \sum_{jh=1}^{n} a_{jk} = \sum_{j=1}^{n} \lambda_k t_{jk}^2 = \lambda_k \sum_{j=1}^{n} t_{jk}^2 = \lambda_k, \tag{2.211}$$

denn $\sum_{j=1}^{n} t_{jk}^2 = 1$, da die Spaltenvektoren von T ja normiert sind. $\|\mathbf{a}_k\|^2$ ist die Summe der Quadrate der Ladungen der Variablen auf der k-ten latenten Dimension, – und sie ist gleich dem k-ten Eigenwert von $X'X$. Auch wenn X spaltenzentriert ist, so ist der Mittelwert der a_{jk} allerdings nicht gleich Null, so dass λ_k nicht proportional der Varianz der Ladungen auf der k-ten Dimension ist; der Proportionalitätsfaktor ist $1/m$. Andererseits ist die Matrix Q spaltenzentriert, wenn X spaltenzentriert ist: ist $\vec{1}$ ein m-dimensionaler Vektor, dessen Komponenten alle gleich 1 sind, so ist $\vec{1}'X = \vec{0}'$, d.h. die Spaltensummen von X sind alle gleich Null. Dann hat man aber

$$\vec{1}'X = \vec{1}'L\Lambda^{1/2}T' = \vec{0}',$$

wegen $\vec{1}'L = \vec{0}'$. Für den k-ten Spaltenvektor $\mathbf{L}_k$ erhält man dann

$$\|\mathbf{L}_k\|^2 = \sum_{i=1}^{m} \ell_{ik}^2 = \sum_{i=1}^{m} \lambda_k q_{ik}^2 = \lambda_k, \tag{2.212}$$

wegen $\sum_{i=1}^{m} q_{ik}^2 = 1$ (die Spalten von Q sind ja normiert). Da die Summe der ℓ_{ik} gleich Null ist, ist $\sum_i \ell_{ik}^2$ proportional zur Varianz der ℓ_{ik}. λ_k korrespondiert also zur Varianz der Koordinaten (Scores) der Fälle auf der k-ten latenten Dimension. Zusammen mit (2.211) hat man die Beziehung

$$\|\mathbf{L}_k\|^2 = \|\mathbf{a}_k\|^2 = \lambda_k, \quad k = 1, \ldots, n \tag{2.213}$$

Je größer also λ_k, desto mehr differenziert die k-te Dimension zwischen den Fällen, und um so größer sind die $|a_{jk}|$, die Absolutbeträge der Ladungen der Variablen auf der k-ten latenten Dimension.

Gesamtvarianz und Varianzanteile: Aus der elementaren Statistik ist bekannt, dass die Varianz einer Summe statistisch unabhängiger Variablen gleich der Summe der Varianzen dieser Variablen ist. Man kann die ℓ_{ik} als zufällige Werte auf unabhängigen (latenten) Variablen ansehen. Die Varianz auf der k-ten latenten Variablen ist λ_k/m. Dementsprechend kann

$$s_{tot}^2 = \frac{1}{m} \sum_{k=1}^{n} \lambda_k$$

als Gesamtvarianz (*tot* für 'total') der Daten angesehen werden. Dann ist

$$\pi_k = \frac{\lambda_k}{\sum_{j=1}^{n} \lambda_j} \tag{2.214}$$

der Anteil der Varianz der k-ten latenten Variablen an der Gesamtvarianz. π_k kann zur Diskussion der Frage, wieviele latente Variable zur Interpretation der Daten benötigt werden, herangezogen werden.

Die Matrix X kann auf der Basis der SVD über die dyadischen Produkte der Spaltenvektoren von Q und T ausgedrückt werden, denn die rechte Seite von (**??**) ist äquivalent zu

$$X = \sigma_1 \mathbf{q}_1 \mathbf{t}_1' + \sigma_2 \mathbf{q}_2 \mathbf{t}_2' + \cdots + \sigma_n \mathbf{q}_n \mathbf{t}_n' = \sum_{k=1}^{n} \sigma_k \mathbf{q}_k \mathbf{t}_k' \quad \sigma_k = \sqrt{\lambda_k}. \tag{2.215}$$

Zusammen mit (2.214) kann (2.215) benutzt werden, um den Wert des Ranges r von X abzuschätzen: Terme mit "hinreichend" kleinen λ_k-Werten können u.U. vernachlässigt werden. Man hat dazu den

Satz 2.36 *(Satz von Eckart & Young) Die Approximation*

$$X \approx X_r = Q_r \Lambda_r^{1/2} T_r' = \sum_{k=1}^{r} \sqrt{\lambda_k} \boldsymbol{q}_k \boldsymbol{t}_k', \quad r < n \tag{2.216}$$

approximiert X im Sinne der Methode der Kleinsten Quadrate.

Beweis: Bekannt wurde diese Aussage (samt Beweis) durch die Arbeit von Eckart & Young (1936); eine modernere Version des Beweises wird in Abschnitt 2.8.5 angeboten, vergl. insbesondere den Beweis zu Satz 2.32, S. 108. Dort wird der Begriff der Matrixnorm vorausgesetzt, vergl. Abschnitt 2.8.4. □

Fragen der Interpretation: Es wurde weiter oben gesagt, dass die Spaltenvektoren der Matrizen L und A – also die skalierten Versionen von A und T – "latente" Variablen repräsentieren. Rein formal sind diese Spaltenvektoren Basisvektoren für die Teilräume des $\mathbb{R}^m$ bzw. des $\mathbb{R}^n$, in denen die m-dimensionalen Vektoren $\mathbf{x}_j$ bzw. die n-dimensionalen Vektoren $\tilde{\mathbf{x}}_i$ liegen. Die Frage ist, ob man ihnen eine inhaltliche Bedeutung zuordnen kann, und wenn ja, ob die latenten Variablen für die $\mathbf{x}_j$ eine andere Bedeutung als die für die Vektoren $\tilde{\mathbf{x}}_i$ haben oder nicht.

Die Gleichungen (2.204) und (2.205) für die Koordinaten ℓ_{ik} des i-ten Falls auf der k-ten latenten Variablen und a_{jk} der j-ten gemessenen Variablen auf der k-ten latenten Variablen zeigen, dass es sich bei diesen Koordinaten um "Korrelationen" (um einen zwar saloppen, aber auf den Kern dieser Größen zielenden Aausdruck zu gebrauchen) eines Falles oder einer Variablen mit latenten Variablen handelt. ℓ_{ik} und a_{jk} werden maximal, wenn die Vektoren $\tilde{\mathbf{x}}_i$ und $\mathbf{t}_k$ einerseits und $\mathbf{x}_j$ und $\mathbf{q}_k$ parallel sind. In diesen Fällen entsprechen der i-Fall bzw. die j-te Variable den jeweiligen latenten Variablen und damit repräsentieren sie eine mögliche Interpretation der latenten Variablen. In der Praxis werden diese Maximalkorrelationen kaum vorkommen, aber man kann dann die Interpretation nach Maßgabe der Fälle oder Variablen vornehmen, die hoch mit den latenten Variablen "korrelieren". Im einen Fall wird die Interpretation in Termen hoch korrelierender Fälle, im anderen Fall durch hoch korrelierender Variablen erfolgen.

Die beiden Interpretationen werden nicht unabhängig voneinander sein. Aus den Gleichungen (2.198) und (2.199) lassen sich sofort die Gleichungen

$$L = Q\Lambda^{1/2} = XT, \quad A = T\Lambda^{1/2} = X'Q \tag{2.217}$$

herleiten, aus denen hervorgeht, dass Q und T bzw. L und A in Abhängigkeit voneinander definiert werden. Dann ist a_{jk} die Ausprägung der j-ten Variablen auf der k-ten latenten Dimension:

$$a_{jk} = \sqrt{\lambda_k} t_{jk} = x_{1j}q_{1k} + x_{2j}q_{2k} + \cdots x_{mj}q_{mk}, \tag{2.218}$$

d.h. a_{jk} ist eine Art gewogener Mittelwert der normierten Ausprägungen *der Fälle* auf der k-ten latenten Variablen; die "Gewichte" sind die Messungen x_{ij} für die j-te Variable., $i = 1, \ldots, m$. Man könnte $a_{jk} = \bar{q}_k$ schreiben um zu betonen, dass hier Werte derselben latenten Variablen gemittelt werden. Eine analoge Aussage gilt für die Scores ℓ_{ik}: Es ist $L = XT$, d.h. für die Komponenten ℓ_{ik} – der Score für den i-ten Fall auf der k-ten latenten Variablen – gilt

$$\ell_{ik} = \sqrt{\lambda_k} q_{ik} = x_{i1}t_{1k} + x_{i2}t_{2k} + \cdots + x_{in}t_{nk}. \tag{2.219}$$

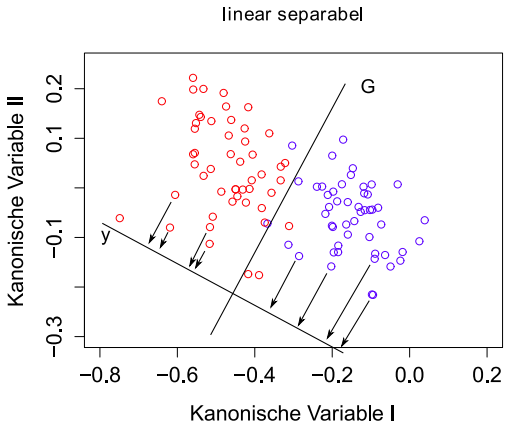
ℓ_{ik} ist demnach ein gewogener Mittelwert der $t_{1k}, \ldots, t_{nk}$, also den Ausprägungen der Variablen auf der k-ten latenten Variablen, mit den für den i-ten Fall spzifischen Gewichtungen x_{ij}, $j = 1, \ldots, n$. Die Ausprägung des i-ten Falls auf der k-ten latenten Dimension ist ein für den i-ten Fall spezifischer Mittelwert der Ausprägungen der Variablen auf der k-ten latenten Dimension. Die latenten Variablen sind Merkmale, aus denen sich die gemessenen Variablen zusammensetzen und die auch dazu dienen, die Fälle zu charakterisieren.

Beispiele findet man in http://www.uwe-mortensen.de/fakanalysews0506b.pdf, p. 91.

2.10.2 Diskriminieren und klassifizieren

Wie in den einführenden Bemerkungen schon angedeutet, soll eine Lösung für die Aufgabe, Fälle Gruppen oder Kategorien zuzuordnen gefunden werden. Betrachtet wird die Konfiguration der Fälle, und gesucht sind latente Variablen derart, dass die Projektion der Punkte auf die zu diesen Variablen korrespondierenden Koordinatenachsen ("Diskriminnanzfunktionen") maximal zwischen den Gruppen oder Kategorien trennt. s. Abbildung 10. Gegeben ist eine $(m \times n)$ Matrix X, de-

Abbildung 10: Klassifikation nach Fisher (1936) (I): Ω_1 blau, Ω_2 rot, eine mögliche Trennlinie G, eine Projektionsgerade Y



ren Zeilen Fälle und deren Zeilen Prädiktorvariablen repräsentieren. Eine – sagen wir: die erste – der gesuchten Achsen läßt sich durch einen m-dimensionalen Vektor $\mathbf{y}$ darstellen, dessen Komponenten die gesuchten Projektionen sind. $\mathbf{y}$ ist eine Linearkombination der Spaltenvektoren von X, d.h. es gilt allgemein $\mathbf{y} = X\mathbf{u}$. $\mathbf{u}$ ist ein Koeffizientenvektor, mit dem der gewünschte Vektor $\mathbf{y}$ erzeugt wird. Die Bestimmung von $\mathbf{y}$ ist demnach die Bestimmung von $\mathbf{u}$.

Da die Komponenten von $\mathbf{y}$ Koordinaten auf einer Koordintenachse darstel-

len, kann man sie zweifach indizieren: die j-te Komponente von $\mathbf{y}$ sei sei die Koordinate y_{ik} des i-ten Falls in der k-ten Kategorie. $\bar{y}_k$ sei der Mittelwert der Koordinaten der k-ten Gruppe. $\mathbf{u}$ soll so bestimmt werden, dass die Varianz der $\bar{y}_k$, $k = 1, \ldots, K$ (es gebe K Gruppen oder Kategorien) maximal wird relativ zur durchschnittlichen Varianz innerhalb der Gruppen. Die Varianz der $\mathbf{a}y_k$ wird durch eine Quadratsumme QS_{zw} (zw für "zwischen") definiert, und die durchschnittliche Varianz innerhalb der Gruppen wird durch eine Quadratsumme QS_{inn} (inn für "innerhalb") definiert. Es soll

$$\lambda = \frac{QS_{zw}}{QS_{inn}} \tag{2.220}$$

maximiert werden; λ ist als *Diskriminanzkoeffizient* bekannt. Die Varianz aller Komponenten von $\mathbf{y}$ wird durch eine Quadratsumme QS_{tot} (tot für "total") definiert, und wie aus der Varianzanalyse bekannt gilt

$$QS_{ges} = QS_{inn} + QS_{zw}. \tag{2.221}$$

Da $\mathbf{u}$ bestimmt werden muß, muß der Quotient (2.220) aus Funktion des unbekannten Vektors $\mathbf{u}$ angeschrieben werden. Wegen $\mathbf{y} = X\mathbf{u}$ muß also $\mathbf{y}$ durch $X\mathbf{u}$ ersetzt werden.

Zunächste eine kleine Vorbetrachtung. Es seien $\mathbf{u} = (u_1, \ldots, u_n)'$ und $\mathbf{x} = (x_1, \ldots, x_n)'$ zwei n-dimensionale Vektoren. Für das Quadrat des Skalarprodukts $(\mathbf{x}'\mathbf{u})^2$ gilt

$$(\mathbf{x}'\mathbf{u})^2 = (x_1u_1 + \cdots x_nu_n)^2 = (x_1u_1)^2 + \cdots + (x_nu_n)^2 + \sum_{i \neq j} x_ix_ju_iu_j. \tag{2.222}$$

Der Ausdruck rechts erinnert an eine quadratische Form (vergl. (2.79)), Seite 68). In der Tat läßt sich das Produkt u_iu_j in der Summe rechts als das (i, j)-te Element des dyadischen Produkts $\mathbf{u}\mathbf{u}'$ interpretieren:

$$\mathbf{u}\mathbf{u}' = \begin{pmatrix} u_1^2 & u_1u_2 & \cdots & u_1u_n \\ u_2u_1 & u_2^2 & \cdots & u_2u_n \\ \vdots & \vdots & \ddots & \vdots \\ u_nu_2 & u_nu_2 & \cdots & u_n^2 \end{pmatrix}$$

Man rechnet leicht nach, dass nun

$$(\mathbf{x}'\mathbf{u})^2 = \mathbf{x}'\mathbf{u}\mathbf{u}'\mathbf{x} \tag{2.223}$$

gilt.

Es sei X eine (m, n)-Matrix; die m Fälle seien in K Gruppen mit den umfängen $n_1, n_2, \ldots, n_K$ aufgeteilt, $m = \sum_{k=1}^{K} n_k$. Für jeden Fall werden Messwerte

bei insgesamt n Variablen bestimmt. Die Matrix X läßt sich dann wie in (2.224) anschreiben

$$\mathbf{y} = \begin{pmatrix} Y_{11} \\ Y_{21} \\ \vdots \\ Y_{n_1 1} \\ \hline Y_{12} \\ Y_{22} \\ \vdots \\ Y y_{n_2 2} \\ \hline \vdots \\ \hline Y_{1K} \\ Y_{2K} \\ \vdots \\ Y_{n_K K} \end{pmatrix}, \quad X = \begin{pmatrix} X_{111} & X_{112} & \cdots & X_{11p} \\ X_{211} & X_{212} & \cdots & X_{21p} \\ \vdots & \vdots & \cdots & \vdots \\ X_{n_1 11} & X_{n_1 12} & \cdots & X_{n_1 1p} \\ \hline X_{121} & X_{122} & \cdots & X_{12p} \\ X_{221} & X_{222} & \cdots & X_{22p} \\ \vdots & \vdots & \cdots & \vdots \\ X_{n_2 21} & X_{n_2 22} & \cdots & X_{n_2 2p} \\ \hline \vdots & \vdots & \cdots & \vdots \\ \hline X_{1K1} & X_{1K2} & \cdots & X_{1Kp} \\ X_{2K1} & X_{2K2} & \cdots & X_{2Kp} \\ \vdots & \vdots & \cdots & \vdots \\ X_{n_K K1} & X_{n_K K2} & \cdots & X_{n_K Kp} \end{pmatrix}, \tag{2.224}$$

Die Indizierung der Elemente von X bezieht sich (i) auf einen Fall in einer gegebenen Gruppe, (ii) auf die gegebene Gruppe, und (iii) auf eine Variable. Es soll eine Linearkombination $\mathbf{y}$ der Spaltenvektoren von X bestimmt werden, die eine möglichst gute Separierung der Gruppen gestattet. Dies bedeutet, dass ein Vektor $\mathbf{u}$ gesucht wird derart, dass

$$\mathbf{y} = X\mathbf{u}. \tag{2.225}$$

Man muß natürlich spezifizieren, was mit "möglichst gute Separierung der Gruppen" gemeint ist. Dem Ansatz (2.225) zufolge wird für jeden Fall eine Komponente von $\mathbf{y}$ bestimmt; die Indizierung der Komponenten von $\mathbf{y}$ werde dabei wie in (2.224) angegeben vorgenommen, so dass gruppenspezifische Mittelwerte $\bar{y}_k$ der Komponenten von $\mathbf{y}$ bestimmt werden können:

$$\bar{y}_k = \frac{1}{n_k} \sum_{i=1}^{n_k} Y_{ik}, \quad k = 1, \ldots, K \tag{2.226}$$

Die "möglichst gute Separierung der Gruppen" soll nun bedeuten, dass die $\bar{y}_k$ sich maximal voneinander unterscheiden sollen. Dies bedeutet, dass die Varianz der $\bar{y}_k$ so groß wie möglich sein soll, was natürlich die Einführung geeigneter Nebenbedingungen erfordert (die maximale Varianz wäre ohne Nebenbedingungen unendlich), auf die später eingangen wird. Wie aus der Varianzanalyse bekannt läßt sich nun die Gesamtvarianz, bzw. die ihr entsprechende Quadratsumme QS_{ges}, in eine Quadratsumme QS_{inn} und eine Quadratsumme QS_{zw} "zwischen" den Gruppen

aufteilen; diese entspricht der Varianz der Mittelwerte der Gruppen. Man hat

$$QS_{inn} = \sum_{k=1}^{K}\sum_{i=1}^{n_k}(Y_{ik}-\bar{y}_k)^2, \quad QS_{zw} = \sum_{k=1}^{K} n_k(\bar{y}_k-\bar{y})^2 \tag{2.227}$$

$$QS_{ges} = \sum_{k=1}^{K}\sum_{i=1}^{n_k}(Y_{ik}-\bar{y})^2 \tag{2.228}$$

$$\tag{2.229}$$

und es gilt (2.221). wie man leicht nachrechnet. Die Komponente Y_{ik} von $\mathbf{y}$ ist das Skalarprodukt der Zeilenvektoren $\tilde{\mathbf{x}}_{ik}$ und $\mathbf{u}$, also $Y_{ik} = \tilde{\mathbf{x}}'_{ik}\mathbf{u}$. Der Index i bezeichnet stets den i-ten Fall in der k-ten Gruppe. Gleichung (2.226) bedeutet dann

$$\bar{y}_k = \frac{1}{n_k}\sum_{i=1}^{n_k}\tilde{\mathbf{x}}'_{ik}\mathbf{u}.$$

Es werde $y_{ik} = Y_{ik} - \bar{y}_k$ gesetzt. Dann ist

$$QS_{inn} = \sum_{k=1}^{K}\sum_{i=1}^{n_k} y_{ik}^2 = \sum_{k=1}^{K}\sum_{i=1}^{n_k}(\tilde{\mathbf{x}}'_{ik}\mathbf{u})^2 = \sum_{k=1}^{K}\sum_{i=1}^{n_k}(\mathbf{u}'\tilde{\mathbf{x}}_{ik})^2,$$

d.h. (vergl. (2.222))

$$QS_{inn} = \sum_{k=1}^{K}\sum_{i=1}^{n_k}\mathbf{u}'\tilde{\mathbf{x}}_{ik}\tilde{\mathbf{x}}'_{ik}\mathbf{u} = \mathbf{u}'\underbrace{(\sum_{k=1}^{K}\sum_{i=1}^{n_k}\tilde{\mathbf{x}}_{ik}\tilde{\mathbf{x}}'_{ik})}_{W}\mathbf{u}. \tag{2.230}$$

$$QS_{inn} = \mathbf{u}'W\mathbf{u}. \tag{2.231}$$

Für QS_{zw} findet man

$$QS_{zw} = \sum_{k=1}^{K} n_k(u_1(\bar{x}_{k1}-\bar{x}_1)+\cdots+u_p(\bar{x}_{kp}-\bar{x}_p))^2,$$

und

$$QS_{zw} = \sum_{k=1}^{K} n_k\mathbf{u}'(\bar{\mathbf{x}}_{k\cdot}-\bar{\mathbf{x}})(\bar{\mathbf{x}}_{k\cdot}-\bar{\mathbf{x}})'\mathbf{u} = \mathbf{u}'\underbrace{(\sum_{k=1}^{K} n_k(\bar{\mathbf{x}}_{k\cdot}-\bar{\mathbf{x}})(\bar{\mathbf{x}}_{k\cdot}-\bar{\mathbf{x}})')}_{B}\mathbf{u},$$

d.h.

$$QS_{zw} = \mathbf{u}'B\mathbf{u} \tag{2.232}$$

Die Maximierung von QS_{zw}, d.h. der Maximierung der Gruppenmittelwerte $\bar{y}_k$, kann nun als Maximierung des Quotienten

$$\lambda(\mathbf{u}) := \frac{QS_{zw}}{QS_{inn}} = \frac{\mathbf{u}'B\mathbf{u}}{\mathbf{u}'W\mathbf{u}} \tag{2.233}$$

definiert werden. Offenbar entspricht $\lambda(\mathbf{u})$ dem aus der Varianzanalyse bekannten F-Wert, der hier allerdings als Funktion von $\mathbf{u}$ maximiert werden soll. Andererseits ist der Quotient auf der rechten Seite der auf Seite 95, Gleichung (2.147), definierte generalisierte Rayleigh-Quotient.

Eine Möglichkeit, den Maximalwert von $\lambda(\mathbf{u})$ zu bestimmen, besteht darin, $\lambda(\mathbf{u})$ nach $\mathbf{u}$ zu Differenzieren und die Ableitung gleich Null zu setzen. Eine andere besteht darin, $\lambda(\mathbf{u})$ in einen Rayleigh-Quotienten (vergl. (2.105), Seite 81) umzuformen und diesen zu maximieren. rrr

Dazu muß man nur berücksichtigen, dass B und W symmetrische Matrizen sind. Für W hat man die Darstellung $W = P\Lambda P'$, wobei P die Matrix der Eigenvektoren von W ist und Λ die Diagonalmatrix der von Null verschiedenen Eigenwerte von W. Dann ist $W^{1/2} = P\Lambda^{1/2}$. Weiter sei $\mathbf{v} = W^{1/2}\mathbf{u}$; dann ist $\mathbf{u} = W^{-1/2}\mathbf{v}$ und $\lambda(\mathbf{u})$ kann in der Form

$$\lambda = \frac{\mathbf{u}'B\mathbf{u}}{\mathbf{u}'W^{1/2}W^{1/2}\mathbf{u}} = \frac{\mathbf{v}'W^{-1/2}BW^{-1/2}\mathbf{v}}{\mathbf{v}'\mathbf{v}}$$

geschrieben werden. Setzt man $A = W^{-1/2}BW^{-1/2}$, so erhält man

$$\lambda = \frac{\mathbf{v}'A\mathbf{v}}{\mathbf{v}'\mathbf{v}}, \tag{2.234}$$

d.h. λ entspricht einem Rayleigh-Quotienten. Nach dem Satz von Courant-Fisher wird λ maximal, wenn $\mathbf{v} = \mathbf{t}$, $\mathbf{t}$ der Eigenvektor von A, der zum maximalen Eigenwert λ_1 von A korrespondiert, d.h. es muß $A\mathbf{t} = \lambda_{\max}\mathbf{t}$ und damit

$$W^{-1/2}BW^{-1/2}\mathbf{t} = \lambda_{\max}\mathbf{t} \tag{2.235}$$

gelten. Multiplikation von links mit $W^{-1/2}$ liefert dann

$$W^{-1}BW^{-1/2}\mathbf{t} = \lambda_{\max}W^{-1/2}\mathbf{t}.$$

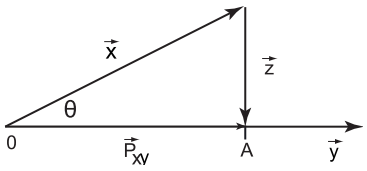
$\mathbf{t}$ ist ein spezieller Vektor für $\mathbf{v} = W^{1/2}\mathbf{u}$; setzt man $\mathbf{t} = W^{1/2}\mathbf{u}_t$, so folgt $\mathbf{t}W^{-1/2} = \mathbf{u}_t$ und man erhält

$$W^{-1}B\mathbf{u}_t = \lambda_{\max}\mathbf{u}_t. \tag{2.236}$$

$\mathbf{u}_t$ ist also ein Eigenvektor von $W^{-1}B$; $\lambda_{\max}$, der maximale Wert von λ, ist der zugehörige Eigenwert. $\mathbf{y}_t = X\mathbf{u}_t$ ist der Vektor, dessen Komponenten die Projektionen der Fälle auf eine Diskriminanzdimension sind, auf der die Gruppenmittelmittelwerte maximal separiert sind.

Die Gleichung (2.236) charakterisiert das generalisierte Eigenwertproblem, vergl. Gleichung (2.146), Seite 95, – man muß Gleichung (2.236) nur von links mit W multipliplizieren, und (2.234) ist ein generalisierter Rayleigh-Quotient, vergl. Gleichung (2.147), ebenfalls Seite 95. Existiert für (2.236) mehr als nur ein

Abbildung 11: Orthogonale Projektion des Vektors $\mathbf{x}$ auf einen Vektor $\mathbf{y}$ bzw. auf eine Gerade



Eigenvektor $\mathbf{u}_l$, so gibt es mehr als nur einen Vektor $\mathbf{y}$, d.h. mehr als nur eine Achse, die zwischen den Klassen oder Gruppen diskriminiert.

Sind $\mathbf{y}_j = X\mathbf{u}_j$ und $\mathbf{y}_k = X\mathbf{u}_k$ zwei verschiedene Vektoren,so gilt

$$\mathbf{y}_j'\mathbf{y}_k = 0, \quad j \neq k \tag{2.237}$$

Beweis: Es ist

$$\mathbf{y}_j'\mathbf{y}_k = \mathbf{u}_j'X'X\mathbf{u}_k = \mathbf{v}_j W^{-1/2}(X'X)W^{-1/2}\mathbf{v}_k = \mathbf{v}_j W^{-1/2}WW^{-1/2]}\mathbf{v}_k = \mathbf{v}_j'\mathbf{v}_k = 0,$$

denn $\mathbf{v}_j$ und $\mathbf{v}_k$ sind Eigenvektoren der symmetrischen Matrix $W^{-1/2}BW^{-1/2}$ und deswegen orthonormal, und W ist eine Schätzung für $X'X$, Gleichung (2.230).

□

2.11 Projektionen

2.11.1 Orthogonale Projektion eines Vektors auf einen anderen

Die Projektion eines Vektors auf einen anderen spielt in vielen Anwendungen eine wichtige Rolle. Um die Idee zu illustrieren, wird die Projektion eines Vektors $\mathbf{y}$ auf einen Vektor $\mathbf{x}$ betrachtet, vergl. Abbildung 11. $\mathbf{x}$ und $\mathbf{y}$ schließen den Winkel θ ein, und es ist

$$\vec{P}_{yx} = a\mathbf{y}, \quad a \in \mathbb{R} \tag{2.238}$$

der Vektor, der sich ergibt, wenn $\mathbf{x}$ auf $\mathbf{y}$ projiziert wird; in Abbildung 11 ist $a < 1$, aber $a > 1$ ist möglich. Nun ist einerseits

$$\mathbf{z} = \vec{P}_{yx} - \mathbf{x} = a\mathbf{y} - \mathbf{x}, \tag{2.239}$$

und da $\mathbf{z}$ senkrecht auf $\mathbf{y}$ steht muß $\mathbf{z}'\vec{P}_{yx} = (\vec{P}_{yx} - \mathbf{x})'\vec{P}_{yx} = 0$ gelten, so dass $\vec{P}_{yx}'\vec{P}_{yx} = \mathbf{x}'\vec{P}_{yx}$ folgt, d.h. es gilt $a^2\|\mathbf{y}\|^2 = a\mathbf{x}'\mathbf{y}$, so dass man

$$a = \frac{\mathbf{x}'\mathbf{y}}{\|\mathbf{y}\|^2} \Rightarrow \vec{P}_{xy} = \frac{\mathbf{x}'\mathbf{y}}{\|\mathbf{y}\|^2}\mathbf{y} \tag{2.240}$$

erhält.

Man kann die Länge von $\vec{P}_{xy}$ als Koordinate der Projektion des Endpunkts von $\mathbf{x}$ auf eine Koordinatenachse $\mathbf{y}$ interpretieren. Diese Koordinate ist dann durch

$$\|\vec{P}_{xy}\| = a\|\mathbf{y}\| \tag{2.241}$$

gegeben.

2.11.2 Projektionsmatrizen

Definition 2.21 *Eine Matrix P heißt* Projektionsmatrix*, wenn sie (i) symmetrisch ist, d.h. es gilt $P' = P$, und wenn sie (ii) idempotent ist, d.h. es gilt $PP = P$.*

Folgerung: Es sei P eine (m, n) Projektionsmatrix und I sei die (m, n)-Identitätsmatrix. Dann ist $(I - P)$ ebenfalls eine Projektionsmatrix. Denn

$$(I - P)' = I' - P' = I - P, \quad (I - P)(I - P) = I - 2P + PP = I - P.$$

Weiter gilt der

Satz 2.37 *P sei eine Projektionsmatrix. Dann hat P die Eigenwerte $\lambda = 0$ und $\lambda = 1$.*

Beweis: Für die Eigenwerte und Eigenvektoren von P gilt $P\mathbf{v} = \lambda\mathbf{v}$. Da P symmetrisch ist, folgt, dass alle Eigenwerte größer oder gleich Null sind. Dann hat man wegen der Idempotenz von P

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

wegen $PP = P$. Es folgt $\lambda\mathbf{v} = \mathbf{v}$ und damit $(\lambda^2 - \lambda)\mathbf{v} = \vec{0}$. Da $\mathbf{v}$ ein Eigenvektor ist, muß $\mathbf{v} \neq \vec{0}$ sein, also folgt $\lambda^2 - \lambda = 0$ bzw. $\lambda^2 = \lambda$ bzw $\lambda = \sqrt{\lambda}$. Eine Lösung ist sicherlich $\lambda = 0$. Eine andere Lösung $\lambda \neq 0$ ergibt sich, wenn man $\lambda^2 = \lambda$ durch λ dividiert: es folgt $\lambda = 1$. Eine weitere Lösung $\lambda \neq 0$ existiert nicht. Denn angenommen, es existiert ein $0 < \lambda = a \in \mathbb{R}$, $a \neq 1$. Für $a < 1$ folgt $a > \sqrt{a}$ und für $a > 1$ folgt $a < \sqrt{a}$, entgegen der Forderung $a = \sqrt{a}$. Also gibt es außer $\lambda = 1$ und der "trivialen" Lösung $\lambda = 0$ keine anderen Eigenwerte. □

Anmerkung: P sei eine (m, n)-Projektionsmatrix. Dann existieren n Eigenvektoren, – einer zum Eigenwert 0 und $n-1$ mit dem korrespondierenden Eigenvektor $\lambda = 1$. Da der Rang einer symmetrischen Matrix gleich der Anzahl der von Null verschiedenen Eigenwerte ist (Satz 2.16, Seite 75), hat P den Rang $\mathrm{rg}(P) = n-1$.

□

Beispiel 2.10 Das lineare Modell Es werde das Allgemeine Lineare Modell (ALM)

$$\mathbf{y} = X\mathbf{b} + \mathbf{e} \tag{2.242}$$

betrachtet. Der Parametervektor $\mathbf{b}$ wird mit der Methode der Kleinsten Quadrate geschätzt; man findet (s. Abschnitt 4.4.3 im Anhang); die Schätzung ist durch

$$\hat{\mathbf{b}} = (X'X)^{-1}X'\mathbf{y} \tag{2.243}$$

gegeben, und man hat[43]

$$\hat{\mathbf{y}} = X\hat{\mathbf{b}} = X(X'X)^{-1}X'\mathbf{y} = P\mathbf{y}. \tag{2.244}$$

Man hat demnach

$$\mathbf{y} = \hat{\mathbf{y}} + \hat{\mathbf{e}} = P\mathbf{y} + \hat{\mathbf{e}}, \tag{2.245}$$

und da $\mathbf{y} - P\mathbf{y} = \hat{\mathbf{e}}$ folgt

$$(I - P)\mathbf{y} = \hat{\mathbf{e}}, \tag{2.246}$$

wobei $\hat{\mathbf{e}}$ der Fehlervektor ist, wenn $\mathbf{b}$ durch die KQ-Schätzung $\hat{\mathbf{b}}$ ersetzt wird. Offenbar gilt

$$\hat{\mathbf{y}}'\hat{\mathbf{e}} = 0, \tag{2.247}$$

denn

$$\begin{aligned} \hat{\mathbf{y}}'\hat{\mathbf{e}} &= (P\mathbf{y})'(\mathbf{y} - \hat{\mathbf{y}}) = \mathbf{y}'P'\mathbf{y} - \mathbf{y}'P'\hat{\mathbf{y}} \\ &= \hat{\mathbf{y}}'P\mathbf{y} - \mathbf{y}'PP\mathbf{y} = \hat{\mathbf{y}}'P\mathbf{y} - \hat{\mathbf{y}}'P\mathbf{y} = 0 \end{aligned}$$

Demnach steht $\hat{\mathbf{e}}$ senkrecht auf $\hat{\mathbf{y}}$.

Die in (2.244) eingeführte Matrix $P = X(X'X)^{-1}X'$ ist eine Projektionsmatrix, denn es gilt

$$(X(X'X)^{-1}X')' = X(X'X)^{-1}X', \quad \text{(Symmetrie)} \tag{2.248}$$

$$X(X'X)^{-1}X'X(X'X)^{-1}X' = X(X'X)^{-1}X', \quad \text{(Idempotenz)} \tag{2.249}$$

Nach (2.244) gilt $\hat{\mathbf{y}} = P\mathbf{y}$; $\hat{\mathbf{y}}$ ist also eine Projektion von $\mathbf{y}$ auf die lineare Hülle $\mathcal{L}(X)$ mit $\hat{\mathbf{y}} \in \mathcal{L}(X)$, und $\mathbf{y} \perp \mathcal{L}(X)$. $\|\hat{\mathbf{e}}\|$ ist die kürzeste Distanz zwischen dem Endpunkt von $\mathbf{y}$ und $\mathcal{L}(X)$.

Nach Gleichung (2.244) gilt $\hat{\mathbf{y}} = P\mathbf{y}$; die Komponenten von $\mathbf{y}$ sind die tatsächlich gemessenen Werte, die Komponenten von $\hat{\mathbf{y}}$ sind die auf der Basis der Prädiktoren vorhergesagten Werte. P heißt deshalb auch *Einflußmatrix* (influence matrix). Die Gleichung $\hat{\mathbf{y}} = P\mathbf{y}$ gibt dann an, wie die gemessenen Werte die vorhergesagten Werte beeinflussen. So ist der Wert der i-ten Komponente (des

[43] Wegen dieser Beziehung – $\hat{\mathbf{y}} = P\mathbf{y}$ – findet man gelegentlich auch den Ausdruck *Hat-Matrix* für die Projektionsmatrix, wegen $\hat{\mathbf{y}} = P\mathbf{y}$, weil sie dem $\mathbf{y}$ einen Hut (rngl. hat) aufsetzt. Im Deutschen spricht man eher von einem Dach, das dem $\mathbf{y}$ aufgesetzt wird.

i-ten Falls) von $\hat{\mathbf{y}}$ durch das Skalarprodukt des i-ten Zeilenvektors von P mit den Datenvektor $\mathbf{y}$ gegeben:

$$\hat{y}_i = p_{i1}y_1 + p_{i2}y_2 + \cdots + p_{im}y_m. \tag{2.250}$$

Das Element p_{ij} läßt sich direkt interpretieren als das Ausmaß, den die Messung y_j via die KQ-Schätzungen der Regressionsparameter auf auf $\hat{y}_i$ hat. Man spricht von der *Hebelwirkung* (leverage) der Messung y_i auf die Schätzungen $\hat{y}_i$ (vergl Abb. 5, S. 66). Damit lassen sich Ausreißer identifizieren, oder, anders formuliert, der Effekt von Ausreißern läßt sich damit charakterisieren. Hierzu wird insbesondere das Element p_{ii} betrachtet. Da eine Projektionsmatrix idempotent und symmetrisch ist, gilt $PP = P$ und damit ist p_{ii} das Skalarprodukt der i-ten Zeile von P und der i-ten Spalte von P, so dass

$$p_{ii} = \sum_{j=1}^{m} p_{ij}^2 = p_{ii}^2 + \sum_{i \neq j} p_{ij}^2 \tag{2.251}$$

Man sieht leicht, dass diese Beziehung nicht gelten kann, wenn die $p_{ij} > 1$ sein können, d.h. es folgt, dass

$$0 \leq p_{ij} \leq 1. \tag{2.252}$$

Weiter gilt, dass wegen der Symmetrie von P die Summe der Eigenwerte gleich der Summe der Diagonalelemente p_{ii} ist, und diese Summe ist glelich der Summe der Eigenwerte. Da die Eigenwerte entweder den Wert 0 oder 1 haben, folgt

$$\sum_{i=1}^{m} p_{ii} = n \tag{2.253}$$

und n ist der Rang der (m, n)-Datenmatrix X, – diese muß vollen Rang haben, da sonst die Inverse $(X'X)^{-1}$ nicht existieren würde, und dann würde P nicht existieren. Aus (2.251) folgt, dass $p_{ii} = 0$ oder $p_{ii} = 1$, wenn $p_{ij} = 0$ für alle i und j. Für $p_{ii} = 0$, so wird $\hat{y}_i$ durch keine andere Beobachtung y_j beeinflußt. Gilt andererseits $p_{ii} = 1$, so gilt $p_{ii} = y_i$, – in diesem Fall passt das Regressionsmodell perfekt, es ist fehlerfrei. Hoaglin & Welsch (1978) liefern Beispiele für derartige Analysen. □

2.11.3 Projektionen auf Hauptachsen

Es sei X eine (m, n)-Matrix von Messwerten; x_{ij} sei der Messwert des i-ten Objects ("Person") für die j-te Variable ("Test"), $1 \leq i \leq m$ und $1 \leq j \leq n$. Für X gilt die Singularwertzerlegung $X = Q\Lambda^{1/2}T' = LP'$ mit $L = Q\Lambda^{1/2}$. Wegen der Orthonormalität von T folgt $XT = L$. Das Element ℓ_{ik} von L ist die Koordinate des i-ten Falls auf der k-ten latenten Dimension und ergibt sich als Skalarprodukt des i-ten Zeilenvektors $\tilde{\mathbf{x}}_i'$ von X ($\tilde{\mathbf{x}}_i$ ist Spaltenvektor von X') und der k-ten Spalte $\mathbf{t}_k$ von T, d.h. es ist

$$\ell_{ik} = \tilde{\mathbf{x}}_i'\mathbf{t}_k. \tag{2.254}$$

$\mathbf{t}_k$ definiert die Orientierung der k-ten Hauptachse eines Ellipsoids, und ℓ_{ik} ist die Länge des Vektors $\vec{P}_{ik}$, der sich als Projektion von $\tilde{\mathbf{x}}_i$ auf die k-te Hauptachse des Ellipsoids ergibt, das die Punktekonfiguration der Fälle repräsentiert. $\vec{P}_{ik}$ entspricht dem Vektor $\vec{P}_{xy}$ in Abbildung 11. Dem vorangegangenen Abschnitt zufolge kann $\vec{P}_{ik} = a_{ik}\mathbf{t}_k$, $a_{ik} \in \mathbb{R}$, geschrieben werden, wobei der Faktor a_{ik} indiziert wurde um anzuzeigen, dass er für $\tilde{\mathbf{x}}_i$ und $\mathbf{t}_k$ charakteristisch ist. Nach (2.240) gilt nun

$$a_{ik} = \frac{\tilde{\mathbf{x}}_i' \mathbf{t}_k}{\|\mathbf{t}_k\|^2} = \tilde{\mathbf{x}}_i' \mathbf{t}_k, \tag{2.255}$$

da ja $\|\mathbf{t}_k\| = 1$, und nach (2.241) hat man dann

$$\|\vec{P}_{ik}\| = a_{ik}\|\mathbf{t}_k\| = \tilde{\mathbf{x}}_i' \mathbf{t}_k, \tag{2.256}$$

d.h. wegen (2.254) gilt

$$\ell_{ik} = \|\vec{P}_{ik}\|, \tag{2.257}$$

so dass die Koordinate ℓ_{ik} die Länge der Projektion des Vektors $\tilde{\mathbf{x}}_i$ auf die k-te Hauptachse des Ellipsoids ist.

3 Funktionenräume und PCA

3.1 Einführung

Vielfach ist die Untersuchung zeitlicher oder räumlicher Verläufe von Interesse; hier werden nur zeitliche Verläufe betrachtet, viele der zur Analyse dieser Verläufe eingeführten Begriffsbildungen übertragen sich auf räumliche Verläufe. Man betrachtet *Funktionenräume* $\mathcal{F}$, d.h. Mengen von Funktionen, die über einem bestimmten Bereich D definiert sind. Linearkombinationen von Funktionen lassen sich analog zu den bisher betrachteten Linearkombinationen definieren: sind f, g Elemente eines Funktionenraums $\mathcal{F}$, so soll auch $\lambda f + \mu g \in \mathcal{F}$ gelten; ist außerdem für alle $f, g \in \mathcal{F}$ außerdem das Skalarprodukt zweier Funktionen erklärt,

$$\langle f, g \rangle = f'g = \int_D f(t)g(t)dt, \quad f, g \in \mathcal{F} \tag{3.1}$$

(d.h. existiert das Integral), so hat man über

$$\|f\|^2 = \int_D |f(t)|^2 dt < \infty \tag{3.2}$$

auch eine Norm $\|f\|$ erklärt und $\mathcal{F}$ ist ein Vektorraum. Es sei $\{\phi_1, \phi_2, \ldots\}$ eine Menge von Funktionen aus $\mathcal{F}$ derart, dass

$$f(t) = \sum_{i=1}^{\infty} a_i \phi_i(t) \tag{3.3}$$

gilt, wobei die a_i Koeffizienten sind, die für f spezifisch sind. Die ϕ_i heißen dann *Basisfunktionen*. Es kann sein, dass unendlich viele Basisfunktionen zur Darstellung von f benötigt werden; dann heißt der Vektorraum *unendlich-dimensional*. Wie Basisvektoren sind Basisfunktionen linear unabhängig, sie können darüber hinaus auch orthogonal sein. So sind die Basisfunktionen ϕ_i, $i = 1, 2, 3, \ldots$ orthonormal, wenn

$$\langle \phi_i, \phi_j \rangle = \int_D \phi_i(t)\phi_j(t)dt = \delta_{ij} = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases} \tag{3.4}$$

erfüllt ist; δ_{ij} heißt *Kronecker-Delta*.

Beispiel 3.1 Fourier-Reihen[44] Eine bekannte Reihenentwicklung einer Funktion $f(t)$ ist die Fourier-Reihe

$$f(t) = a_0 + \sum_{n=1}^{\infty}(a_n \cos(nt) + b_n \sin(nt)), \tag{3.5}$$

wobei die Koeffizienten a_n und b_n durch

$$a_0 = \frac{1}{2\pi}\int_{-\infty}^{\infty} f(t)dt \tag{3.6}$$

$$a_n = \frac{1}{\pi}\int_{-\infty}^{\infty} f(t)\cos(nt)dt \tag{3.7}$$

$$b_n = \frac{1}{\pi}\int_{-\infty}^{\infty} f(t)\sin(nt)dt \tag{3.8}$$

gegeben sind. Die cos- und sin-Funktionen sind orthogonal:

$$\int_{-T/2}^{T/2} \cos(m\omega_0 t)\cos(n\omega_0 t)dt = \begin{cases} 0, & m \neq n \\ T/2, & m = n \neq 0 \end{cases} \tag{3.9}$$

$$\int_{-T/2}^{T/2} \sin(m\omega_0 t)\sin(n\omega_0 t)dt = \begin{cases} 0. & m \neq n \\ T/2, & m = n \neq 0 \end{cases} \tag{3.10}$$

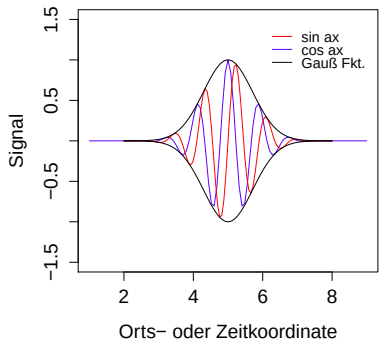
$$\int_{-T/2}^{T/2} \sin(m\omega_0 t)\cos(n\omega_0 t)dt = 0, \quad \text{für alle } m \text{ und } n \tag{3.11}$$

Man hat also eine Menge orthogonaler Basisfunktionen.

Ein wichtiges Anwendungsgebiet der Fourier-Entwicklung ist die Signalanalyse; hier eignet sich die Fourier-Entwicklung insbesondere dann, wenn das Signal $f(t)$ stationär ist. Für nicht-stationäre Signale sind andere Entwicklungen etwa auf der Gabor-Wavelet-Basis von größerem Nutzen (s. Abbildung 12 und die folgende Erläuterung). □

[44] Joseph Fourier, (1768 – 1830), französischer Mathematiker und Physiker

Abbildung 12: Gabor-Funktion



cos- und sin-Funktionen haben einen Nachteil: sie sind auf $(-\infty, \infty)$ definiert. Es ist aber oft notwendig, örtlich und zeitlich begrenzte Signale zu repräsentieren. Für diesen Zweck haben sich andere Basisfunktionen als geeigneter erwiesen. So können Gauß-Funktionen $\exp(-(t-t_n)^2$, $n = 1, 2, \ldots$ verwendet werden (Gishtasby & O'Neill (1994), Calcaterra (2008)), oder Gabor-Funktionen:

$$f_n(t) = \exp(-\frac{(x-x_n)^2}{2\sigma^2})\phi(k\omega_o t), \tag{3.12}$$

wobei ϕ eine Sinus- oder Kosinus-Funktion ist (Gabor (1948)), vergl. Abbildung 12. Weitere Basisfunktionen sind Polynome (Hermite- oder Laguerre-Polynome); die Diskussion der verschiedenen Möglichkeiten geht weit über den Rahmen dieses Skripts hinaus.

Unter Umständen lassen sich die Basisfunktionen auch empirisch bestimmen. Dies geschieht bei der Karhunen-Loéve-Analyse, auf die im folgenden eingegangen werden soll.

3.2 Karhunen-Loève-Entwicklung und PCA

Die KL-Entwicklung dient der Charakterisierung und Interpretation stochastischer Prozesse. Man stelle sich eine Untersuchung vor, bei der das Ergebnis eines experimentellen Durchgangs ein bestimmter Werteverlauf $\omega \in \Omega$ innerhalb eines Zeitintervalls $D = [a, b]$; statt eines Zeitintervalls kann natürlich auch ein Ortsbereich betrachtet werden, im Folgenden werden allerdings nur zeitliche Intervalle betrachtet. Ω ist der Stichprobenraum, d.h. die Menge der möglichen Verläufe. Wie bei der Betrachtung zufälliger Veränderlicher werden 'Ereignisse', d.h. Klassen bestimmter Verläufe, durch Teilmengen von Ω, definiert; diese Teilmengen bilden eine Sigma-Algebra Σ, und die Wahrscheinlichkeiten, mit denen Ereignis-

se eintreten, werden durch ein Wahrscheinlichkeitsmaß P festgelegt. Mit (Ω, Σ, P) ist dann der Wahrscheinlichkeitsraum der Untersuchung festgelegt.

Jedem $\omega \in \Omega$ wird nun eine Funktion der Zeit zugeordnet: $\omega \mapsto X(t, \omega)$, $t \in D$. $X(t, \omega)$ als Funktion der Zeit bildet den Verlauf ω ab. $X(t, \omega)$ heißt auch *Pfad* oder *Trajektorie* des Prozesses. Die Schreibweise $X(t, \omega)$ ist üblich, kann aber verwirrend sein, da $X(t, \omega)$ einen Wert des Verlaufs zu einem Zeitpunkt t meinen könnte. Papoulis (1968), p. 280, macht diesen Punkt explizit, indem er die vier möglichen Interpretationen von $X(t, \omega)$ auflistet:

1. $X(t, \omega)$ kann eine ganze Familie von Funktionen der Zeit bedeuten,
2. $X(t, \omega)$ kann eine einzelne Funktion der Zeit repräsentieren,
3. Für fixen Wert von t kann $X(t, \omega)$ eine zufällige Veränderliche bedeuten,
4. Schließlich kann $X(t, \omega)$ für fixen Wert von t und festes $\omega \in \Omega$ eine einzelne Zahl repäsentieren.

Was jeweils gemeint ist wird durch den jeweiligen Zusammenhang bestimmt. Ein stochastischer Prozess ist dann eine Familie $X_t = \{X(t, \omega)\}_{t \in D}$ von zufälligen Funktionen (Pfaden, Trajektorien) über dem Zeitbereich D. Um die Notation zu vereinfachen wird im Folgenden einfach $X(t)$ statt $X(t, \omega)$ geschrieben, und der Prozess wird kurz mit $X_t = \{X(t)\}_{t \in D}$ bezeichnet.

Erwartungswert und Varianz eines stochastischen Prozesses X_t sind durch

$$\mathbb{E}(X_t) = \int_\Omega X(t, \omega) dP(\omega) = m(t), \quad Var(X_t) = \mathbb{E}[(X_t - \mathbb{E}(X_t))^2] \tag{3.13}$$

definiert; für jeden Wert $t \in D$ wird über alle möglichen Trajektorien $X(t)$ gemittelt. $m(t)$ wird auch die Mittelwertsfunktion des Prozesses genannt.

Es sei Y_t ein stochastischer Prozess mit der Mittelwertsfunktion $\mathbb{E}(Y_t)$. Dann heißt

$$X_t = \{X_t\}_{t \in D}, \quad X_t = Y_t - m(t) \tag{3.14}$$

zentrierter stochastischer Prozess.

Definition 3.1 *Der stochastische Prozess heißt* stetig im quadratischen Mittel*, wenn*

$$\mathbb{E}[(X_{t+\varepsilon} - X_t)^2] = 0. \tag{3.15}$$

Es sei X_t ein zentrierter stochastischer Prozess; dann ist

$$R_X(s, t) = \mathbb{E}(X(s)X(t)), \quad s, t \in D \tag{3.16}$$

die *Autokorrelationsfunktion* von X_t.

Satz 3.1 *Es sei X_t ein stochastischer Prozess mit der Autokorrelationsfunktion $R_X(s, t)$. X_t ist stetig im quadratischen Mittel genau dann, wenn R_X stetig auf $D = [a, b] \times [a, b]$ ist.*

Beweis: Wong (1971). □

Der folgende Satz geht auf Loève (1945) Karhunen (1947) zurück.

Satz 3.2 *Es sei* $X_t = \{X(t)|t \in [a,b]\}$ *mit* $\mathbb{E}(X_t) = 0$ *für alle* $t \in [a,b]$ *und stetiger Kovarianzfunktion* $C(s,t)$*, wobei die Funktionen* $X(t)$ *quadratintegrierbar seien. Dann gilt*

$$\hat{X}(t) = \sum_{i=1}^{n} a_i \phi_i(t), \tag{3.17}$$

und die

$$a_i = \int_a^b X(t)\phi_i(t) \tag{3.18}$$

sind zufällige Veränderliche[45] *mit*

$$\mathbb{E}(a_i) = 0, \quad \mathbb{E}(a_i a_j) = \delta_{ij}\lambda_i \tag{3.19}$$

mit

$$\delta_{ij} = \begin{cases} 0, & i \neq j \\ 1, & i = j \end{cases} \qquad \textit{(Kronecker-Delta)}$$

sind; die ϕ_i *sind die orthonormalen Eigenfunktionen von* $C(s,t)$*, mit den* λ_i *als dazu korrespondierenden Eigenwerten. Weiter gilt*

$$\lim_{n\to\infty} \hat{X}(t) = X(t). \tag{3.20}$$

Beweis: Der Ansatz (3.17) gilt für eine willkürliche Funktion $f(t)$, also insbesondere für eine zufällige Funktion $X(t)$ aus der Menge X_t; da $X(t)$ zufällig ist, ist a_i eine zufällige Veränderliche. Es gilt

$$\mathbb{E}(a_i) = \mathbb{E}\left[\int_a^b X_t \phi_i(t)\right] = \int_a^b \mathbb{E}[X_t]\phi_i(t) = 0, \tag{3.21}$$

da ja $\mathbb{E}[X_t] = 0$. Nach Mercers Theorem gilt $C(s,t) = \sum_{i=1}^{\infty} \lambda_i \phi_i(s)\phi_i(t)$ und ϕ_i

[45] a_i ist zufällig weil $X(t)$ eine zufällige Funktion aus der Menge der zufälligen Funktionen ist, die den stochastischen Prozess X_t definieren.

und ϕ_j sind orthonormal, $i \neq j$. Dann folgt

$$\begin{aligned}
\mathbb{E}[a_i a_j] &= \mathbb{E}\left[\int_a^b \int_a^b X_s X_t \phi_i(s)\phi_j(t) ds dt\right] \\
&= \int_a^b \int_a^b \mathbb{E}[X_s X_t]\phi_i(s)\phi_j(t) ds dt \\
&= \int_a^b \int_a^b k(s,t)\phi_i(s)\phi_j(t) ds dt \\
&= \int_a^b \phi_i(s) \left(\int_a^b k(s,t)\phi_j(t) dt\right) ds \\
&= \lambda_i \int_a^b \phi_i(s)\phi_j(s) ds \\
&= \delta_{ij}\lambda_i, \qquad (3.22)
\end{aligned}$$

δ_{ij} das Kronecker-Delta. Es sei

$$\varepsilon_n(t) = \mathbb{E}\left[\left(X(t) - \sum_{i=1}^n a_i \phi_i(t)\right)^2\right]. \qquad (3.23)$$

Dann folgt

$$\varepsilon_n(t) = \mathbb{E}[X^2(t)] - 2\mathbb{E}\left[X(t)\sum_{i=1}^n a_i\phi_i(t)\right] + \mathbb{E}\left[a_i a_j \phi_i(t)\phi_j(t)\right] \qquad (3.24)$$

Es ist $\mathbb{E}(X^2(t)] = C(t,t)$, und

$$\begin{aligned}
\mathbb{E}\left[X(t)\sum_{i=1}^n a_i\phi_i(t)\right] &= \mathbb{E}\left[X(t)\sum_{i=1}^n \int_D X(s)\phi_i(s) ds \phi_i(t)\right] \\
&= \sum_{i=1}^n \left(\int_D \mathbb{E}[X(s)X(t)]\phi_i(s) ds\right)\phi_i(t) \\
&= \sum_{i=1}^n \left(\int_D C(s,t)\phi_i(s) ds\right)\phi_i(t) \\
&= \sum_{i=1}^n \lambda_i \phi_i^2(t),
\end{aligned}$$

und schließlich folgt analog

$$\mathbb{E}\left[\sum_{i=1}^n a_i a_j \phi_i(t)\phi_j(t)\right] = \sum_{i=1}^n \lambda_i \phi_i^2(t).$$

Also folgt

$$\varepsilon_n(t) = C(t,t) - \sum_{i=1}^n \lambda_i \phi_i(t)\phi_i(t),$$

und wegen Mercers Theorem folgt

$$\lim_{n\to\infty} \varepsilon_n(t) = 0$$

d.h. $\hat{X}(t) \to X(t)$. □

$X(t)$ **als Zeitreihe:** Erhebt man $X(t)$ als kontinuierliche Funktion über $D = [a, b]$, so muß man die Integralgleichung

$$\int_D k(s,t)\phi_i(t)dt = \lambda_i\phi_i(s)$$

für $i = 1, 2, \ldots, k$ lösen, d.h. die Eigenfunktionen $\phi_i(t)$ für alle $t \in [a, b]$ bestimmen. Das ist im Allgemeinen sehr aufwendig. Einfacher wird die Aufgabe, wenn $X(t)$ für diskrete Werte $t_1, \ldots, t_N$ bestimmt wird. Man erhält dann den Vektor

$$X = (X_1, \ldots, X_N),'$$

für den die (m, n)-Matrix C der Autokorrelationen berechnet werden kann. Man erhält man dann die Gleichung

$$C\vec{\phi}_j = \lambda_j\vec{\phi}_j, \quad j = 1, \ldots, N \tag{3.25}$$

$\vec{\phi}_j$ ist jetzt ein N-dimensionaler (Eigen-)Vektor mit λ_j als zugehörigem Eigenwert. $\vec{\phi}_j$ repräsentiert die entsprechende Eigenfunktion an den Stellen $t_1, \ldots, t_N$, und natürlich sind die Eigenvektoren $\vec{\phi}_j$ paarweise orthonormal. Gesucht ist die Entwicklung für $X = (X(t_1), X(t_2) \ldots, X(t_N))'$ als Reihe:

$$X = \sum_{j=1}^{N} a_j\phi_j(t),$$

mit $\phi_j = (\phi_j(t_1), \ldots, \phi(t_N))'$. Die Koeffizienten a_j erhält man durch orthogonale Basisentwicklung:

$$\phi_j'X = a_j, \quad j = 1, \ldots, N \tag{3.26}$$

Natürlich ist es von Interesse, weniger als N Eigenfunktionen zu wählen, – gesucht ist ja die sparsamste Repräsentation der Daten. Also betrachtet man

$$\hat{X} = \sum_{j=1}^{r} a_j\phi_j(t), \quad r < N \tag{3.27}$$

Die Güte der Abschätzung $\hat{X}$ läßt sich durch den Quotienten

$$\frac{\sum_{j=1}^{r} \lambda_j}{\sum_{j=1}^{N} \lambda_j} \geq \alpha \tag{3.28}$$

bewerten, wobei α ein festgesetzter Anteil an erklärter Varianz ist.

4 Anhang

4.1 Elementarmatrizen und elementare Operationen

Es werden zuerst die *elementaren Zeilen- oder Spaltenumformungen* ener beliebigen (m, n)-Matrix U eingeführt:
1. Die Multiplikation eines beliebigen Zeilenvektors von $\tilde{\mathbf{u}}_i$ mit einer Zahl $0 \neq \lambda \in \mathbb{R}$,
2. Addition des Vektors $c\tilde{\mathbf{u}}_i$ zu einem beliebigen Zeilenvektor $\tilde{\mathbf{u}}_k$ von U,
3. Vertauschung von irgendzwei Zeilenvektoren $\tilde{\mathbf{u}}_i$ und $\tilde{\mathbf{u}}_k$ von U.

Die elementaren Spaltenumformungen sind analog definiert (man muß nur den Ausdruck Zeilenvektor' durch den Ausdruck 'Spaltenvektor' ersetzen.

Satz 4.1 *Die Anwendung elementarer Umformungen auf die Matrix M verändert nicht den Rang r von M.*

Beweis: Einen sehr ausführlichen Beweis für diesen Satz findet man in Sperner, Band 11, p. 29. (1961). Knapper läßt sich die Aussage beweisen, wenn vom Begriff der Elementarmatrix Gebrauch gemacht wird, vergl. Fischer (1997), Abchnitt 2.7.

Definition 4.1 *Die*[46] *(n, n)-Matrix E_{ij} enthalte nur Nullen bis auf das (i, j)-te Element, das gleich 1 ist. Dann heißt E_{ij} eine* Standardmatrix.

Definition 4.2 *Es sei I_n die (n, n)-Einheitsmatrix. Eine* Elementarmatrix *entsteht, wenn auf I_n eine der möglichen elementaren Umformungen angewendet wird; dabei entstehen drei Typen von Elementarmatrizen:*
1. $I_n \to Q_i^j(\lambda)$: das (i, j)-te Element von I_n wird durch $\lambda \in \mathbb{R}$ ersetzt, d.h.[47]

$$Q_i^j(\lambda) = I_n + \lambda E_{ij} \tag{4.1}$$

2. $I_n \to P_i^j$: die i-te Zeile von I_n wir mit der j-ten Zeile vertauscht, d.h.

$$P_i^j = I_n - E_{ii} - E_{jj} + E_{ij} + E_{ji}, \quad i \neq j. \tag{4.2}$$

Die Subtraktion von E_{ii} und E_{jj} bewirkt die Ersetzung der Einsen an der i-ten und j-ten Diagonalzelle von I_n, und die Addition von E_{ij} und E_{ji} bewirkt die Ersetzung der Null durch eine Eins an der (i, j)-ten und der (j, i)-ten Position von I_n.
3. $I_n \to S_i(\lambda)$: die Eins in der i-ten Diagonalzelle von I_n wird durch $1 \neq \lambda \in \mathbb{R}$ ersetzt:

$$S_i(\lambda) = I_n + \lambda E_{ii} - E_{ii} = I_n + (\lambda - 1)E_{ii}. \tag{4.3}$$

[46] Im Folgenden wird die in https://de.wikipedia.org/wiki/Elementarmatrix vorgestellte Notation übernommen.
[47] Die Bezeichnung der Elementarmatrizen entspricht der in Fischer (1997), p.155 gewählten.

Statt der elementaren Zeilenumformungen kann man die analog definierten Spaltenumformungen vornehmen.

Man überprüft nun direkt die Aussagen über elementare Umformungen einer (m, n)-Matrix A:
1. A_I entstehe aus$A$ durch Multiplikation der i-ten Zeile von A mit $\lambda \in \mathbb{R}$. Dann gilt

$$A_i = S_i(\lambda)A.$$

2. A_{II} entstehe durch der j-ten Zeile zur i-ten Zeile. Dann gilt

$$A_{II} = Q_j^i A.$$

3. A_{III} entsehe aus A durch Addition des λ-fachen der j-ten Zeile zur i-ten Zeile. Dann gilt

$$A_{III} = Q_j^i(\lambda)A.$$

3. A_{IV} entstehe aus A durch Vertauschen der i-ten mit der j-ten Zeile. Dann gilt

$$A_{IV} = P_i^j A.$$

Will man Spalten- statt Zeilenumformungen vornehmen, muß A statt von links mit einer Elementarmatrix von rechts multipliziert werden.

Die Elementarmatrizen sind invertierbar und ihre Inversen sind wieder Elementarmatrizen, d.h. es gelten die Gleichungen

$$(S_i(\lambda))^{-1} = S_i(\frac{1}{\lambda}), \quad (Q_i^j)^{-1} = Q_i^j(-1) \tag{4.4}$$

$$(Q_i^j(\lambda))^{-1} = Q_i^j(-\lambda), \quad (Pi^j)^{-1} = P_i^j \tag{4.5}$$

Zum Beweis multipliziere man die rechten Seiten der Gleichungen mit den linken: es ergibt sich stets die Einheitsmatrix.

C. F. Gauß hat ein allgemeines Verfahren gefunden, mit dem man lineare Gleichungssysteme $A\mathbf{x} = \mathbf{y}$ systematisch lösen kann. Dazu wird die Matrix A durch sukzessive elementare Umformungen z. B. in eine *obere Dreiecksmatrix* B überführt:

$$B = \begin{pmatrix} b_{11} & \cdots & b_{1n} \\ & \ddots & \vdots \\ 0 & & b_{nn} \end{pmatrix} \tag{4.6}$$

Da die elementaren Umformungen durch Elementarmatrizen repräsentiert werden, heißt dies, dass Elementarmatrizen $B_1, \ldots, B_r$ existieren derart, dass

$$B = B_r B_{r-1} \cdots B_1 A. \tag{4.7}$$

Dieser Sachverhalt führt auf eine Methode zur Berechnung der inversen A^{-1}, sofern diese existiert. Darüber hinaus läßt sich von B der Rang der Matrix ablesen: er ist gleich der Anzahl der Zeilen von B, die nicht nur Nullen aufweisen.

4.1.1 Lineare Gleichungen und Gauß-Algorithmus

Gegeben sei ein lineares Gleichungssystem $A\mathbf{x} = \mathbf{y}$; gesucht ist eine möglichst systematische Art und Weise, einen Lösungsvektor $\mathbf{x}$ zu bestimmen. In diesem Abschnitt soll das gaußsche Eliminationsverfahren (auch Gauß-Algorithmus vorgestellt werden. Die Idee des Verfahrens ist, die Koeffizientenmatirx A in eine Dreiecksmatrix zu transformieren, , anhand der serh schnell eine Lösung gefunden werden kann, sofern eine Lösung existiert. Zur Illustration werde ein System mit drei Unbekannten betrachtet:

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 &= y_1 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 &= y_2 \\ a_{31}x_1 + a_{32}x_2 + a_{33}x_3 &= y_3 \end{aligned} \tag{4.8}$$

Bei der Bestimmung der Dreiecksmatrix mit den Elementen $\tilde{a}_{ij}$ wird auch der Vektor $\mathbf{y}$ in einen Vektor $\tilde{\mathbf{y}}$ transformiert:

$$\begin{aligned} \tilde{a}_{11}x_1 + \tilde{a}_{12}x_2 + \tilde{a}_{13}x_3 &= \tilde{y}_1 \\ \tilde{a}_{22}x_2 + \tilde{a}_{23}x_3 &= \tilde{y}_2 \\ \tilde{a}_{33}x_3 &= \tilde{y}_3 \end{aligned} \tag{4.9}$$

Für dieses System findet man sofort die Lösung für x_3, nämlich $x_3 = \tilde{y}_3/\tilde{a}_{33}$, die man dann in die zweite Gleichung einsetzen kann, die dann nur noch die Unbekannten x_1 und x_2 enthält und die damit nach x_2 aufgelöst werden kann, und die Lösungen für x_3 und x_2 können dann in die erste Gleichung eingesetzt werden, um x_1 zu bestimmen.

Um die Koeffizienten $\tilde{a}_{ij}$ zu bestimmen, macht man von den *elementaren Umformungen* Gebrauch. Demnach bleibt eine Gleichung ja korrekt, wenn man beide Seiten mit einem Faktor multipliziert. Ebenso kann man man zwei Gleichungen addieren; es entsteht wieder eine gültige Gleichung. Diese Operationen muß man so anwenden, dass man von den a_{ij} zu den $\tilde{a}_{ij}$ gelangt.

Die Koeffizientenmatrix in (4.9) ist eine Dreiecksmatrix

$$\tilde{A} = \begin{pmatrix} \tilde{a}_{11} & \tilde{a}_{12} & \cdots & \tilde{a}_{1n} \\ 0 & \tilde{a}_{22} & \cdots & \tilde{a}_{2n} \\ & & \ddots & \\ 0 & 0 & \cdots & \tilde{a}_{nn} \end{pmatrix}. \tag{4.10}$$

Die elementaren Umformungen, die von A zu $\tilde{A}$ führen, verändern den Rang nicht, d.h. $\tilde{A}$ hat denselben Rang wie A. Berechnet man die Determinante von $\tilde{A} - \lambda I_n$,

so erhält man

$$|\tilde{A}| = \begin{vmatrix} \tilde{a}_{11} - \lambda & \tilde{a}_{12} & \cdots & \tilde{a}_{1n} \\ 0 & \tilde{a}_{22} - \lambda & \cdots & \tilde{a}_{2n} \\ & & \ddots & \\ 0 & 0 & \cdots & \tilde{a}_{nn} - \lambda \end{vmatrix} = (\tilde{a}s_{11} - \lambda)(\tilde{a}_{22} - \lambda) \cdots (\tilde{a}_{nn} - \lambda) = 0. \tag{4.11}$$

Daraus folgt unmittelbar, dass die $\tilde{a}_{jj}$ die Eigenwerte von $\tilde{A}$ sind.

Die Anwendung der elementaren Umformungen macht man sich am besten durch ein konkretes Beispiel klar:

$$\begin{aligned} 2x_1 - x_2 + 3x_3 &= 1 \quad (I) \\ 3x_1 + x_2 - 2x_3 &= 0 \quad (II) \\ x_1 + x_2 + x_3 &= 3 \quad (III) \end{aligned}$$

Die Koeffizientenmatrix ist

$$\left(\begin{array}{ccc|c} 2 & -1 & 3 & 1 \\ 3 & 1 & -2 & 0 \\ 1 & 1 & 1 & 3 \end{array}\right)$$

Man beginnt, indem man die erste Gleichung (I) noch einmal anschreibt, die zweite durch $3III - II$ ersetzt und die dritte durch $2III - I$

$$\begin{aligned} 2x_1 - x_2 + 3x_3 &= 1 \\ 0 + 2x_2 + 5x_3 &= 9 \\ 0 + 3x_2 + x_3 &= 5 \end{aligned}$$

Die Koeffizientenmatrix ist hier

$$\left(\begin{array}{ccc|c} 2 & -1 & 3 & 1 \\ 0 & 2 & 5 & 9 \\ 0 & 3 & 1 & 5 \end{array}\right)$$

Ersetzt man die Gleichung (4.12) durch 3(4.12) - 2(4.12), so erhält man

$$\begin{aligned} 2x_1 - x_2 + 3x_3 &= 1 \\ 0 + 2x_2 + 5x_3 &= 9 \\ 0 + 0 + 17x_3 &= 17 \end{aligned}$$

mit der Koeffizientenmatrix

$$\left(\begin{array}{ccc|c} 2 & -1 & 3 & 1 \\ 0 & 2 & 5 & 9 \\ 0 & 0 & 17 & 17 \end{array}\right) \tag{4.12}$$

Hieraus ergeben sich direkt die Lösungen: $17x_3 = 17 \Rightarrow x_3 = 1$, $2x_2 + 5x_3 = 2x_2 + 5 = 9 \Rightarrow x_2 = 4/2 = 2$ und $2x_1 - 2 + 3 = 2x_1 + 1 = 1 \Rightarrow x_1 = 0$.

Man kann die ursprüngliche Matrix A und die transformierte Matrix $\tilde{A}$ einander gegenüber stellen:

$$\begin{pmatrix} 2 & -1 & 3 \\ 3 & 1 & -2 \\ 1 & 1 & 1 \end{pmatrix}, \quad \begin{pmatrix} 2 & -1 & 3 \\ 0 & 2 & 5 \\ 0 & 0 & 17 \end{pmatrix}$$

und sieht die Transformation von A in eine Dreiecksmatrix direkt. Die hier vorgeführte Transformation hat einen gewissen ad hoc-Charakter und mag insofern nicht zufriedenstellend erscheinen, und natürlich existieren kanonische Darstellungen des Algorithmus (Golub & van Loan (2013), Kapitel 3), auf die hier aber nicht eingegangen werden soll bzw. kann, da diese Darstellungen auf der einen Seite recht lang sind und auf der anderen nicht benötigt werden, wenn man nur an der Anwendung interessiert ist, was im Allgemeinen der Fall sein wird; Statistikpakete enthalten entsprechende Module. Hier soll nur festgestellt werden, dass A und $\tilde{A}$ natürlich denselben Rang haben.

Im folgenden Abschnitt wird anhand der SVD von A der Teilraum $\mathbf{L}_{n-r}$ näher bestimmt.

4.1.2 Die Cholesky-Zerlegung

Die Anwendung des Gauß-Algorithumus impliziert, dass die Koeffizientenmatrix A des Gleichungssystems $A\mathbf{x} = \mathbf{y}$ in eine Dreiecksform überführt wird. Für den Spezialfall einer symmetrischen, positiv-definiten Matrix läßt sich zeigen, dass A stets als Produkt zweier Dreiecksmatrizen darstellbar ist; dieser Sachverhalt bedeutet, dass der Aufwand für das Lösen des Gleichungssystems drastisch reduziert wird; dies ist die Cholesky-Zerlegung[48]. In diesem Skript wird auf numerische Fragen kaum eingegangen, da die Programmpakete für multivariate Verfahren im Allgemeinen effiziente Algorithmen enthalten, um die sich der Anwender nicht weiter kümmern muß. In einigen theoretischen Herleitungen wird aber auf die Cholesky-Zerlegung eingegangen, so dass in diesem Absatz kurz auf sie eingegangen weden soll.

Satz 4.2 *Es sei A eine symmetrische, positiv-definite Matrix. Dann existiert eine untere Dreiecksmatrix L derart, dass*

$$A = LL'. \tag{4.13}$$

Beweis: Statt eines allgemeinen Beweises wird der Satz anhand einer (3×3)-Matrix illustriert; das angewendetete Prinzip übertzrägt sich sofort auf den all-

[48] André-Louis Cholesky (1875 – 1918), französischer Mathematiker

gemeinen $(n \times n)$-Fall. Nach Behauptung existieren also L und L^* derart, dass

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = \begin{pmatrix} L_{11} & 0 & 0 \\ L_{21} & L_{22} & 0 \\ L_{31} & L_{32} & L_{33} \end{pmatrix} \begin{pmatrix} L_{11} & L_{21} & L_{31} \\ 0 & L_{22} & L_{32} \\ 0 & 0 & L_{33} \end{pmatrix}$$

Rechnet man das Produkt auf der rechten Seite aus, so erhält man

$$A = LL' = \begin{pmatrix} L_{11}^2 & L_{11}L_{21} & L_{11}L_{31} \\ L21L_{11} & L_{21}^2 + L_{22}^2 & L_{21}L_{31} + L_{22}L_{22}L_{32} \\ L_{31}L_{11} & L_{31}L_{21} + L_{32}L_{22} & L_{31}^2 + L_{32}^2 + L_{33}^2 \end{pmatrix}.$$

Das Element a_{ij} von A ist dann gleich dem Element in der i-ten Zeile und j-ten Spalte der Matrix auf der rechten Seite, also

$$a_{11} = L_{11}^2, \quad a_{22} = L_{21}^2 + L_{22}^2, \quad a_{33} = L_{31}^2 + L_{32}^2 + L_{33}^2,$$

und

$$a_{21} = L21L_{11}, \quad a_{23} = L_{21}L_{31} + L_{22}L_{22}L_{32}, a_{31} = L_{31}L_{11}$$

und wegen der vorausgesetzten Symmetrie von A gilt $a_{ij} = a_{ji}$, so dass nicht mehr Elemente bestimmt werden müssen. Wegen der vorausgesetzten Positiv-Definitheit gilt $a_{11} > 0$, also folgt $L_{11} = \sqrt{a_{11}}$. Dann folgt $L_{21} = a_{21}/L_{11} = a_{21}/\sqrt{a_{11}}$ und $L_{31} = a_{31}/\sqrt{a_{11}}$, etc. Auf diese Weise lassen sich die L_{ij} aus den a_{ij} berechnen. Das Prinzip kann leicht auf den Fall allgemeiner $(n \times n)$-Matrizen übertragen werden. □

Die LDL-Zerlegung: Eine etwas verallgemeinerte Form der Cholesky-Zerlegung ist die *LDL-Zerlegung* einer symmetrischen, positiv-definiten Matrix:

$$A = LDL', \tag{4.14}$$

wobei L eine untere Dreiecksmatrix und D eine Diagonalmatrix ist. Man kann $D = D^{1/2}D1/2$ schreiben, wobei $D^{1/2}$ ei8ne Diagonalmatrix ist, deren Diagonalelemente die Wurzeln aus den Diagonalelementen von D ist, d.h. die Diagonalelemente von D müssen größer als Null sein. Setzt man $G = LD^{1/2}$, so hat man

$$A = GG'. \tag{4.15}$$

Es sei $A\mathbf{x} = \mathbf{y}$ ein Gleichungssystem. Setzt man $G\mathbf{z} = \mathbf{y}$, $G'\mathbf{x} = \mathbf{z}$, so hat man $GG'\mathbf{x} = \mathbf{y}$. Man löst also das Gleichungssystem, indem man zunächst $G\mathbf{z} = \mathbf{y}$ und dann $G'\mathbf{x} = \mathbf{z}$ löst; da G und G' in Dreiecksform sind, sind diese beiden Gleichungssysteme schnell und einfach zulösen. Man findet für die Elemente g_{ik} von G

$$g_{ik} = \begin{cases} 0, & i < k \\ \sqrt{a_{ik} - \sum_{j=1}^{k-1} g_{kj}^2}, & i = k \\ \frac{1}{a_{ik}} \left(a_{ik} - \sum_{j=1}^{k-1} g_{ij} g_{kj} \right), & i > k \end{cases} \tag{4.16}$$

Eine ausführliche Darstellung von Lösungen von linearen Gleichungssytemen durch Rückführung auf Dreiecksmatrizen wird in Golub & vanLoan (2013), Kap. 3 und 4, gegeben.

4.2 Steinerscher Austauschsatz und das Fundamental-Lemma

Satz 4.3 *Es sei V ein Vektorraum mit der Basis $\mathcal{B} = (\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n)$. Für $\vec{0} \neq \boldsymbol{v} \in V$ gibt es dann die Darstellung*

$$\boldsymbol{v} = a_1\boldsymbol{v}_1 + \cdots + a_n\boldsymbol{v}_n \tag{4.17}$$

Jeder Vektor $\boldsymbol{v}_j \in \boldsymbol{B}$ mit $a_j \neq 0$ kann durch $\boldsymbol{v}$ ausgetauscht werden, und

$$\{\boldsymbol{v}_1, \boldsymbol{v}_{j-1}, \boldsymbol{v}, \boldsymbol{v}_{j+1}, \ldots, \boldsymbol{v}_n\} \tag{4.18}$$

ist ebenfalls eine Basis von V.

Beweis: Da $\mathbf{v} \neq \vec{0}$ sind nicht alle $a_j = 0$. Da die Anordnung der $\mathbf{v}_j$ in $\mathbf{B}$ beliebig ist, kann angenommen werden, dass $a_1 \neq 0$, so dass $\mathbf{v}_1$ durch $\mathbf{v}$ ersetzt werden kann. Dann gilt

$$b_1\mathbf{v} + b_2\mathbf{v}_2 + \cdots + b_n\mathbf{v}_n = \vec{0}.$$

Da zu zeigen ist, dass $\{\mathbf{v}, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ eine Basis ist, muß gezeigt werden, dass die $\mathbf{v}, \mathbf{v}_2, \ldots, \mathbf{v}_n$ linear unabhängig sind, d.h. die Darstellung des Vektors $\vec{0}$ nur möglich ist, wenn $b_1 = \cdots = b_n = 0$. Setzt man den obigen Ausdruck der Linearkombination von $\mathbf{v}$ ein und vereinfacht, so erhält man den Ausdruck

$$b_1a_1\mathbf{v}_1 + (b_1a_2 + b_2)\mathbf{v}_2 + \cdots + (b_1a_n + b_n)\mathbf{v}_n = \vec{0}.$$

Wegen der linearen Unabhängigkeit der $\mathbf{v}_j$ folgt dann aber

$$b_1a_1 = b_1a_2 + b_2 = \cdots = b_1a_n + b_n = 0.$$

Da $a_1 \neq 0$ folgt $b_1 = 0$ und also $b_1 = \cdots = b_n = 0$, d.h. die Vektoren $\{\mathbf{v}, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ sind linear unabhängig.

Jetzt muß nur noch gezeigt werden, dass $\{\mathbf{v}, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ ein Erzeugendensystem für V ist. Dazu sei $\mathbf{w} \in V$ ein beliebiger Vektor aus V. Da $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ nach Voraussetzung eine Basis ist, existieren Koeffizienten $c_1, \ldots, c_n \in \mathbb{R}$ derart, dass

$$\mathbf{w} = c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n.$$

Aus (4.17) folgt

$$\mathbf{v}_1 = \frac{1}{a_1}(\mathbf{v} - a_2\mathbf{v}_2 - \cdots - a_n\mathbf{v}_n)$$

Setzt man diesen Ausdruck für $\mathbf{v}_1$ in den Ausdruck für $\mathbf{w}$ ein, so sieht man, dass $\mathbf{w}$ als Linearkombination der $\{\mathbf{v}, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ dargestellt werden kann, so dass diese Vektoren ebenfalls eine Basis für V bilden. □

Der folgende Satz wird gelegentlich als *Fundamental-Lemma der Linearen Algebra* bezeichnet (etwa in Koecher (1997), p. 21).

Satz 4.4 *Der Vektorraum V habe eine Basis $\mathcal{B} = (\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n)$. Dann ist jede Teilmenge $M \subset V$ mit $m > n$ Elementen linear abhängig. Zwei verschiedene Basen von V haben stets dieselbe Anzahl von Elementen aus V.*

Beweis: Es sei $\mathcal{B} = (\mathbf{v}_1, \ldots, \mathbf{v}_n)$ eine Basis von V, und es sei $M = \{\mathbf{w}_1, \ldots, \mathbf{w}_m\} \subset V$, $m > n$. Es werde angenommen, die $\mathbf{w}_1, \ldots, \mathbf{w}_m$ seien linear unabhängig. Nach dem Austauschsatz 4.3 kann einer der Basisvektoren aus $\mathcal{B}$ etwa durch den Vektor $\mathbf{v} = \mathbf{w}_1$ ausgetauscht werden. Man erhält dann die Basis $\mathbf{w}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$, und $\mathbf{w}_2$ kann als Linearkombination

$$\mathbf{w}_2 = b_1\mathbf{w}_1 + a_2\mathbf{v}_2 + \ldots + a_n\mathbf{v}_n$$

ausgedrückt werden. So fährt man weiter fort, indem man $\mathbf{v}_2$ durch $\mathbf{w}_2$ austauscht, etc, so dass man schließlich die $\mathbf{v}_1, \ldots, \mathbf{v}_n$ durch die $\mathbf{w}_1, \ldots, \mathbf{w}_n$ ersetzt hat. Aber $m > n$, und $\mathbf{w}_m$ kann nun als Linearkombination der $\mathbf{w}_1, \ldots, \mathbf{w}_n$ ausgerückt werden. Aber das heißt, dass die $\mathbf{w}_1, \ldots, \mathbf{w}_m$ insgesamt linear abhängig sind, entgegen der Annahme ihrer linearen Unabhängigkeit. Es können nur n der m Elemente $\mathbf{w}_j$ linear unabhängig sein, und das heißt, dass alle Basen von V dieselbe Anzahl von Elementen haben müssen. □

4.3 Zur Berechnung von Ellipsen für eine Punktekonfiguration

Gegeben seien zwei konzentrische Kreise mit den Radien a und $b < a$. Die Gerade $\overline{MB}$ schneidet den kleineren Kreis im Punkt A. Für die Gerade $\overline{BC}$ gilt $\overline{BC} = y + d$ mit $y = \overline{CP}$, und es ist $\overline{MC} = x$. t ist der Winkel, den die Gerade $\overline{MA}$ bzw. $\overline{MB} = a$ mit der x-Achse bildet. Die Position des Punktes P hängt vom Wert des Parameters (Winkels) t ab , d.h. $P = P(t)$.

Behauptung: Die Punkte $P(t)$, $0 \leq t \leq 2\pi$, liegen auf einer Ellipse $\mathcal{E}$ der Form

$$\mathcal{E}: \quad \frac{x^2}{a^2} + \frac{y^2}{b^2} = 1, \quad x = x(t),\ y = y(t) \tag{4.19}$$

Beweis: Für alle $t \in [0, 2\pi]$ gilt nach dem Satz des Pythagoras $a^2 = x^2 + (y+d)^2$, so dass $y + d = \sqrt{a^2 - x^2}$. Nach dem Strahlensatz gilt

$$\frac{y}{b} = \frac{\sqrt{a^2 - x^2}}{a},$$

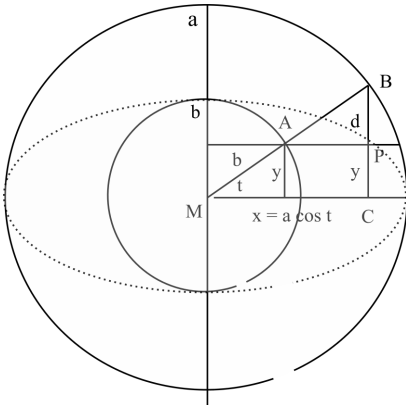
so dass

$$\frac{y^2}{b^2} = \frac{a^2 - x^2}{a^2} = 1 - \frac{x^2}{a^2},$$

woraus sofort (4.19) folgt. □

Parameterdarstellung: Bei dieser Darstellung wird die Ellipse (4.19) als Menge der Punkte $P = P(t)$ für $0 \leq t \leq 2\pi$ definiert. Aus der Abbildung 13 liest man

Abbildung 13: Kreise mit den Radien a bzw. $b < a$ und zugehörige Ellipse (Menge der Punkte P) mit den Halbachsen a und b



direkt ab, dass $\cos t = x/a$ ist, d.h. es ist $x = x(t) = a\cos t$. Weiter ist offenbar $\sin t = y/b$, so dass $y = y(t) = b\sin t$. Der Punkt P hat die Koordinaten x und y. Mithin folgt für die Ellipse $\mathcal{E}$

$$\mathcal{E} = \{\mathbf{p}(t)|\mathbf{p}(t) = (x(t), y(t))' = (a\cos t, b\sin t)\}. \tag{4.20}$$

$\mathbf{p}(t)$ der Vektor mit dem Anfangspunkt M und dem Endpunkt $P(t)$.

Die Gleichung (4.19) impliziert, dass die Länge der ersten Hauptachse durch $x_0 = a$ gegeben ist (man setzt $y = 0$), und die Länge der zweiten Hauptachse ist $y_0 = b$ (man setzt und $x = 0$). Gesucht sind Werte für a und b derart, dass die erste Hauptachse der Ellipse der maximalen Ausdehnung der Punktekonfiguration entspricht; die k-te Spalte $\mathbf{L}_k$ hat die Länge $\sqrt{\lambda_k}$). Diese Forderung führt zu

$$a = \sqrt{\lambda_1}, \quad b = \sqrt{\lambda_2}. \tag{4.21}$$

In Bezug auf (4.20) ist diese Wahl für a und b sofort evident: für $t = 0$ (die Orientierung des Vektors entspricht der der x-Achse) ist $\|\mathbf{p}(t)\| = \sqrt{\lambda_1}$, und für $t = \pi/2$ ergibt sich $\|\mathbf{p}(t)\| = \sqrt{\lambda_2}$.

Setzt man $\mathbf{y} = (y_1, y_2)'$ mit $x = y_1$, $y = y_2$, so ist (4.19) äquivalent zu

$$\mathbf{y}'\Lambda^{-1}\mathbf{y} = 1. \tag{4.22}$$

Ist C eine Kovarianz- oder Korrelationsmatrix, so gilt $C = T\Lambda T'$ und

$$C^{-1} = (T\Lambda T')^{-1} = (T')^{-1}\Lambda^{-1}T^{-1} = T\Lambda^{-1}T' = \frac{y_1^2}{\lambda_1} + \frac{y_2^2}{\lambda_2}$$

d.h. C und C^{-1} haben dieselben Eigenvektoren, aber die Eigenwerte von C^{-1} sind die Reziprokwerte der Eigenwerte von C, und es folgt

$$y_{01} = \sqrt{\lambda_1}, \quad y_{02} = \sqrt{\lambda_2}.$$

4.4 Die Differentiation von Vektoren

4.4.1 Die allgemeine Differentiationsformel

Ein Vektor ist durch seine Komponenten festgelegt. Man kann dann fragen, wie sich der Vektor verändert, wenn man seine Komponenten verändert. Solche Veränderungen lassen sich oft durch einen Differentialquotienten beschreiben. So sei etwa $\mathbf{x} = (x_1, x_2, \ldots, x_n)'$. Dabei wird stillschweigend angenommen, dass keine Komponente von der anderen abhängt. Man definiert nun den Differentialquotienten von $\mathbf{x}$ in Bezug auf die j-te Komponente x_j durch

$$\frac{d\mathbf{x}}{dx_j} = \begin{pmatrix} dx_1/dx_j \\ \vdots \\ dx_j/dx_j \\ \vdots \\ dx_n/dx_j \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix} = \mathbf{e}_j. \tag{4.23}$$

Es gibt noch einen zweiten Fall, bei dem die Komponente von einer Variablen, etwa der Zeit t, abhängen, so dass

$$\mathbf{x}(t) = \begin{pmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_n(t) \end{pmatrix}$$

geschrieben wird. Man kann dann die Veränderung von $\mathbf{x}$ mit t durch den Vektor

$$\frac{d\mathbf{x}(t)}{dt} = \begin{pmatrix} dx_1(t)/dt \\ dx_2(t)/dt \\ \vdots \\ dx_n(t)/dt \end{pmatrix}$$

ausdrücken. Dieser Fall wird im Folgenden nicht behandelt.

Der Fall (4.23) tritt u.a. dann auf, wenn eine Größe in Abhängigkeit von einem Vektor $\mathbf{b} = (b_1, \ldots, b_p)'$ von Parametern maximiert oder minimiert werden soll.

Es sei $\mathbf{y} = A\mathbf{x}$; für eine gegebene Matrix A hängt $\mathbf{y}$ von $\mathbf{x}$ ab, so dass man $\mathbf{y} = \mathbf{y}(\mathbf{x})$ schreiben kann. Weiter ist $A\mathbf{x} = x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n$, so dass

$$\frac{\partial \mathbf{y}}{\partial x_j} = \frac{\partial A\mathbf{x}}{\partial x_j} = \mathbf{a}_j, \quad j = 1, \ldots, n$$

Dementsprechend erhält man

$$\frac{d\mathbf{y}}{d\mathbf{x}} = [\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n] = A. \tag{4.24}$$

4.4.2 Die Differentiation quadratischer Formen

Es wird die quadratische Form $Q(\mathbf{x}) = \mathbf{x}'C\mathbf{x}$ betrachtet, wobei $C' = C$ eine (n, n)-Matrix ist und $\mathbf{x} \in \mathbb{R}^n$ n-dimensionale Vektoren sind. Es soll bezüglich $\mathbf{x}$ differenziert werden. Man differenziert zunächst nach einer Komponente x_j von $\mathbf{x}$ und findet (Anwendung der Kettenregel)

$$\frac{\partial Q}{\partial x_j} = \mathbf{e}_j'C\mathbf{x} + \mathbf{x}'C\mathbf{e}_j \tag{4.25}$$

$\mathbf{e}_j$ der j-te Einheitsvektor. $\mathbf{e}_j'C\mathbf{x}$ und $\mathbf{x}'C\mathbf{e}_j$ sind Skalarprodukte und es folgt $\mathbf{e}_j'C\mathbf{x} = \mathbf{x}'C\mathbf{e}_j$, wegen $C' = C$, so dass

$$\frac{\partial Q}{\partial x_j} = 2\mathbf{e}_j'C\mathbf{x}. \tag{4.26}$$

Fasst man die $\mathbf{e}_j'$ zur Einheitsmatrix zusammen, so erhält man

$$\frac{\partial Q}{\partial \mathbf{x}} = 2C\mathbf{x}. \tag{4.27}$$

wegen $C' = C$. Damit hat man auch die Ableitung von $\|\mathbf{x}\|^2 = \mathbf{x}'\mathbf{x}$ nach $\mathbf{x}$ gefunden, denn mit $C = I$, I die Einheitsmatrix, folgt aus (4.27)

$$\frac{\partial \mathbf{x}'\mathbf{x}}{\partial \mathbf{x}} = 2I\mathbf{x} = 2\mathbf{x}. \tag{4.28}$$

Die Maximierung von $\mathbf{x}'A\mathbf{x}$ unter derNebenbedingung $\mathbf{x}'\mathbf{x} = a \in \mathbb{R}$. Die allgemeine Theorie der Extremwertbestimmung unter Nebenbedingungen kann in Abschnitt 4.4.4 nachgelesen werden.

Zu maximieren sei

$$Q(\mathbf{x}) = \mathbf{x}'A\mathbf{x} + \lambda(\mathbf{x}'\mathbf{x} - a), \quad \mathbf{x}'\mathbf{x} - a = 0, \quad A' = A \tag{4.29}$$

Dann ist

$$\frac{dQ}{d\mathbf{x}} = 2A\mathbf{x} - 2\lambda\mathbf{x},$$

so dass

$$A\mathbf{x} = \lambda\mathbf{x}, \tag{4.30}$$

d.h. $Q(\mathbf{x})$ wird maximal, wenn $\mathbf{x}$ ein Eigenvektor von A ist und $\lambda \in \mathbb{R}$ der zugehörige Eigenwert ist.

4.4.3 Die Kleinste-Quadrate-Schätzung für das Lineare Modell

Es sei X eine (m, n)-Matrix mitvollem Rang $r = n \leq m$, $\mathbf{b} \in \mathbb{R}^n$ sei ein unbekannter Vektor von (Regressions-)Parametern und $\mathbf{y}, \mathbf{e} \in \mathbb{R}^m$ seinen Vektoren; es gelte

$$\mathbf{y} = X\mathbf{b} + \mathbf{e} \tag{4.31}$$

$\mathbf{e}$ ist ein Fehlervektor, und $\mathbf{b}$ soll so bestimmt werden, dass der Fehler minimalisiert wird. Dies soll heißen, dass $\mathbf{e}'\mathbf{e} = \|\mathbf{e}\|^2$ minimal werden soll; $\mathbf{e}$ soll so kurz wie möglich werden. Es ist

$$\mathbf{e}'\mathbf{e} = (\mathbf{y} - X\mathbf{b})'(\mathbf{y} - X\mathbf{b}) = \mathbf{y}'\mathbf{y} - \mathbf{y}'X\mathbf{b} - \mathbf{b}'X'\mathbf{y} + \mathbf{b}'X'X\mathbf{b} \tag{4.32}$$

Nun ist $X\mathbf{b}$ ein Vektor, so dass $\mathbf{y}'X\mathbf{b}$ und $\mathbf{b}'X'\mathbf{y}$ Skalarprodukte sind und $\mathbf{b}'X'X\mathbf{b}$ ist eine quadratische Form, so dass

$$\mathbf{e}'\mathbf{e} = \mathbf{y}'\mathbf{y} - 2\mathbf{b}'X'\mathbf{y} + \mathbf{b}'X'X\mathbf{b}$$

Nach (4.24) ist $d(X\mathbf{b})/d\mathbf{b} = X$, und nach (4.27) ist $d(\mathbf{b}'X'X\mathbf{b})/d\mathbf{b} = 2X'X\mathbf{b}$. $\mathbf{e}'\mathbf{e}$ wird minimal für $\mathbf{b} = \hat{\mathbf{b}}$ derart, dass

$$\left.\frac{d(\mathbf{e}'\mathbf{e})}{d\mathbf{b}}\right|_{\mathbf{b}=\hat{\mathbf{b}}} = 2(X'\mathbf{y} - (X'X)\mathbf{b}) = 0 \tag{4.33}$$

so dass $X'\mathbf{y} = (X'X)\hat{\mathbf{b}}$, d.h.

$$\hat{\mathbf{b}} = (X'X)^{-1}X'\mathbf{y}, \tag{4.34}$$

Die Projektionsmatrix: Es sei $\hat{\mathbf{y}} = X\hat{\mathbf{b}}$, so dass

$$\mathbf{y} = \hat{\mathbf{y}} + \hat{\mathbf{e}} = X\hat{\mathbf{b}} + \hat{\mathbf{e}}, \tag{4.35}$$

wobei $\hat{\mathbf{e}}$ der Vektor der Fehler ist, wenn $\mathbf{y}$ durch $\hat{\mathbf{y}}$ "vorhergesagt" wird. Man hat dann wegen (4.34)

$$\hat{\mathbf{y}} = X\hat{\mathbf{b}} = X(X'X)^{-1}X'\mathbf{y} = P\mathbf{y}; \tag{4.36}$$

darin ist

$$P = X(X'X)^{-1}X' \tag{4.37}$$

die *Projektionsmatrix*. Offenbar gilt

$$P' = P \tag{4.38}$$

$$PP = P \tag{4.39}$$

denn

$$PP = X(X'X)^{-1}X'X(X'X)^{-1}X' = X(X'X)^{-1}X' = P,$$

d.h. P ist symmetrisch und idempotent.

Es gilt

$$\hat{\mathbf{y}}'\hat{\mathbf{e}} = 0, \tag{4.40}$$

d.h. $\hat{\mathbf{y}}$ und $\hat{\mathbf{e}}$ sind orthogonal. Denn wegen (4.38) und (4.39) gilt

$$\hat{\mathbf{y}}'\hat{\mathbf{e}} = (P\mathbf{y})'\hat{\mathbf{e}} = \mathbf{y}'P'(\mathbf{y} - \hat{\mathbf{y}}) = \mathbf{y}'P\mathbf{y} - \mathbf{y}'PP\mathbf{y} = 0.$$

Einige Implikationen: Es sei $X = Q\Sigma T'$, $\Sigma = \Lambda^{-1/2}$, die SVD von X. Dann ist $X'X = T\Lambda T'$ und

$$(X'X)^{-1} = (T\Lambda T')^{-1} = (T')^{-1}\Lambda^{-1}T^{-1} = T\Lambda^{-1}T', \tag{4.41}$$

denn $T^{-1} = T'$ wegen der Orthonormalität von T. (4.34) impliziert dann

$$\hat{\mathbf{b}} = T\Lambda^{-1}T'T\Lambda^{1/2}Q'\mathbf{y} = T\Lambda^{-1/2}Q'\mathbf{y}. \tag{4.42}$$

Aber es ist $\mathbf{y} = X\mathbf{b} + \mathbf{e} = Q\Lambda^{1/2}Q'\mathbf{b} + \mathbf{e}$, so dass aus (4.42)

$$\hat{\mathbf{b}} = T\Lambda^{-1/2}Q'(Q\Lambda^{1/2}Q'\mathbf{b} + \mathbf{e}) = \mathbf{b} + T\Lambda^{-1/2}Q'\mathbf{e} \tag{4.43}$$

folgt, und schreibt man $T\Lambda^{-1/2}Q'$, indem man das dyadische Produkt anwendet, so ergibt sich (vergl. (2.215), Seite 120, wo der Fall $X = Q\Lambda^{1/2}T'$ betrachtet wird)

$$\hat{\mathbf{b}} = \mathbf{b} + \left(\sum_{k=1}^{n} \frac{\mathbf{t}_k\mathbf{q}_k'}{\sqrt{\lambda_k}}\right)\mathbf{e} \tag{4.44}$$

Der wahre Parametervektor $\mathbf{b}$ und die Kleinste-Quadrate-Schätzung $\hat{\mathbf{b}}$ unterschieden sich also um den Vektor $T\Lambda^{-1/2}Q'\mathbf{e} = (\sum_{k=1}^{n} \mathbf{t}_k\mathbf{q}_k'/\sqrt{\lambda_k})\mathbf{e}$. Der ist um so größer, je größer einerseits die Komponenten von $\mathbf{e}$ sind, und je kleiner andererseits die λ_k sind. Es sei $C = X'X$. Aus $C = T\Lambda T'$ folgt, dass die Spaltenvektoren von C als Linearkombinationen der Spalten von $T\Lambda$ dargestellt werden können: Es sei $\mathbf{c}_j$ die j-te Spalte von C, und t_{ik} sei das i-te Element des k-ten Eigenvektors in T. Dann ist der j-te Spaltenvektor von T' durch $(t_{1j}, \ldots, t_{kj}, \ldots, t_{nj})'$ gegeben und man hat

$$\mathbf{c}_j = t_{1j}\lambda_1\mathbf{t}_1 + t_{2j}\lambda_2\mathbf{t}_2 + \cdots + t_{nj}\lambda_n\mathbf{t}_n. \tag{4.45}$$

Sind nur die ersten Eigenwerte $\lambda_1, \lambda_2, \ldots$ "groß" und sind die restlichen "klein", so werden die Kovarianzen bzw. Korrelationen in den Spalten von C nur durch die ersten Eigenvektoren "erklärt", die restlichen spielen eine geringere Rolle, d.h. die Messwerte werden durch wenige latente Variablen (repräsentiert durch die Eigenvektoren von C) bestimmt, – was hohe Korrelationen (Absolutbetrag) zwischen den Variablen bedeutet. In der multiplen Regression sind die Variablen die Prädiktoren für $\mathbf{y}$. Korrelierte Prädiktoren bedeuten also die Existenz kleiner Eigenwerte und damit große Differenzen zwischen $\mathbf{b}$ und der Schätzung $\hat{\mathbf{b}}$, – und damit ungenaue Voraussagen für $\mathbf{y}$. Die Prädiktoren sollten daher möglichst unkorreliert sein. Eine Möglichkeit, den Effekt von Korrelationen zwischen den Prädiktoren zu reduzieren, besteht wieder darin, für X die SVD $Q\Sigma T'$ einzusetzen:

$$\mathbf{y} = Q\Sigma T'\mathbf{b} + \mathbf{e} = Q\Sigma(T'\mathbf{b}) + \mathbf{e} = L\boldsymbol{\beta} + \mathbf{e}, \quad L = Q\Sigma \tag{4.46}$$

mit L als neuer Prädiktormatrix und $\boldsymbol{\beta} = T'\mathbf{b}$ als neuem Parametervektor; die Spaltenvektoren von L sind orthgonal und also unkorreliert. Im Skriptum über Regressionsverfahren wird dieser Ansatz ausführlicher diskutiert.

4.4.4 Extrema unter Nebenbedingungen

Es sei $f(x_1, \ldots, x_n)$ eine Funktion der Variablen $x_1, \ldots, x_n$. Gesucht sind diejenigen Werte x_{0j} von x_j, $j = 1, \ldots, n$, für die f ein Maximum annimmt, wobei aber die Nebenbedingung $g(x_1, \ldots, x_n) = k$, k ein Konstante berücksichtigt werden sooo, d.h. der Vektor $\mathbf{x}_0 = (x_{01}, \ldots, x_{0n})'$ soll so bestimmt werden, dass auch $g(x_{01}, \ldots, sx_{0n}) = k$ erfüllt ist. Man kann i.A. g so definieren, dass $k = 0$ gesetzt werden kann.

Der Einfachheit halber werden die Überlegungen zur Maximierung unter Nebenbedingungen für den Fall $n = 2$ durchgeführt; das Resultat überträgt sich unmittelbar auf den Fall $n > 2$. Dazu wird $x = x_1$, $y = x_2$ gesetzt. Es soll also $f(x, y)$ unter der Nebenbedingung $g(x, y) = 0$ maximiert werden (oder allgemein ein Extremwert bestimmt werden).

$g(x, y) = 0$ bedeutet, dass es eine Funktion $y = g(x)$ gibt, so dass $f(x, y) = f(x, g(x))$ und $g(x, g(x)) = 0$ geschrieben werden kann. Geometrisch beschreibt $f(x, y)$ eine Fläche im 3-dimensionaolen Raum und $g(x, y) = 0$ beschreibt eine Kurve in der $X \times Y$-Ebene. Die Nebenbedingung $g = 0$ bedeutet nun, dass man $f(x, y)$ nur für die diejenigen Punkte (x, y) berechnet, die auf der Kurve $g(x, y) = 0$ liegen. Für diese Kurve werde $f_g = f(x, y|g(x, y) = 0)$ geschrieben. Die Menge der Punkte (x, y), für die $f(x, y) = k$ gilt, definiert eine Höhenlinie von $f(x, y)$. Dann existiert eine Konstante $k = c$, die die Kurve f_g genau dort berührt, wo diese ihr Maximum annimmt.

Man hat die Ableitungen

$$\frac{\partial f(x, g(x))}{\partial x} = \frac{\partial f}{\partial x} + \frac{\partial f}{\partial g}\frac{dg(x)}{dx} = f_x + f_y g',$$

wobei die Kettenregel angewendet wurde. Analog dazu erhält man für g

$$\frac{dg}{dx} = \frac{\partial g}{\partial x} + \frac{\partial g}{\partial y}\frac{dg(x)}{dx} = g_x + g_y g'.$$

Die Extremwerte werden bestimmt, indem man die entsprechenden Ableitungen gleich Null setzt. Dementsprechend erhält man die Gleichungen

$$f_x + f_y g' = 0 \tag{4.47}$$
$$g_x + g_y g' = 0 \tag{4.48}$$

Die bisher hergeleiteten Ableitungen enthalten noch die Ableitung g' von g. Um das Extremum zhu bestimmen, eliminiert man g' am besten, da die Bestimmung von g' kompliziert sein kann. Man hat nun $g' = -f_x/f_y = -g_x/g_y$; diese Beziehung bedeutet, dass die *Gradientenvektoren* $(f_x, f_y)'$ und $(g_x, g_y)'$ dieselbe Orientierung haben, d.h. sie unterscheiden sich allenfalls in ihrer Länge, so dass man

$$\begin{pmatrix} f_x \\ f_y \end{pmatrix} = \lambda \begin{pmatrix} g_x \\ g_y \end{pmatrix} \tag{4.49}$$

schreiben kann. $\lambda \in \mathbb{R}$ ist ein neuer, freier Parameter, der sogenante *Lagrange-Faktor* oder *Lagrange-Multiplikator*. Er drückt einfach aus, dass man nur etwas über die Orientierung, nicht aber über die Länge der Gradientenvekotren am Ort des Maximums weiß. Die Vektorgleichung (4.49) zusammen mit der Bedingung $g(x, y) = 0$ führt sofort auf ein System von drei Gleichungen mit den drei Unbekannten x, y und λ:

$$\begin{aligned} f_x - \lambda g_x &= 0 && (4.50)\\ f_y - \lambda g_y &= 0 && (4.51)\\ g(x, y) &= 0 && (4.52) \end{aligned}$$

Diese Überlegungen müssen nicht immer explizit durchgeführt werden, denn sie implizieren die Möglichkeit, von vornherein die *Lagrange-Funktion* $L(x, y, \lambda)$ aufzustellen:

$$L(x, y, \lambda) = f(x, y) + \lambda g(x, y), \quad g(x, y) = 0. \tag{4.53}$$

Man findet den Extremwert, indem man L partiell nach x, nach y und nach λ differenziert und die entstehenden partiellen Ableitungen gleich Null setzt.

Die drei Gleichungen (4.50), (4.51) und (4.52) heißen zusammen die *Lagrangesche Multiplikatorenregel*, nach dem Mathematiker und Astronomen Jean-Louis Lagrange (1736 – 1813), der diese Regel 1788 herleitete.

Beispiel 4.1 Gegeben sei die Funktion $f(x, y) = 6 - x^2 - \frac{1}{3}y^2$ und die Nebenbedingung $x + y = 2$, die in der Form $x + y - 2 = 0$ angeschrieben werden kann. Dann ist

$$f_x = -2x,\ f_y = -\frac{2}{3}y,\ g_x = 1,\ g_y = 1,$$

und man erhält das Gleichungssystem

$$\begin{aligned} -2x + \lambda 1 &= 0\\ -\frac{2}{3}y + \lambda 1 &= 0\\ x + y - 2 &= 0, \end{aligned}$$

woraus $x = 1/2$, $y = 3/2$ und $\lambda = -1$ folgt. □

Beispiel 4.2 (Satz von Courant-Fisher). Es sei A eine symmetrische, positivdefinite $n \times n$-Matrix mit den Eigenwerten $\lambda_1 \geq \lambda_2 \leq \cdots \geq \lambda_n$. Dann gilt

$$\max_{\mathbf{x} \neq \vec{0}} \frac{\mathbf{x}' A \mathbf{x}}{\mathbf{x}' \mathbf{x}} = \max_j \lambda_j = \lambda_1, \tag{4.54}$$

und der Vektor $\mathbf{x}$, für den das Maximum angenommen wird, ist der zu λ_1 korrespondierende Eigenvektor $\mathbf{t}_1$ von A. Weiter gilt

$$\min_{\mathbf{x} \neq \vec{0}} \frac{\mathbf{x}' A \mathbf{x}}{\mathbf{x}' \mathbf{x}} = \min_j \lambda_j = \lambda_n \tag{4.55}$$

und der Vektor $\mathbf{x}$, der $\mathbf{x}'A\mathbf{x}$ minimalisiert, ist der zu λ_n korrespondierende Eigenvektor von A.

Beweis: Als Nebenbedingung werde $\mathbf{x}'\mathbf{x} = 1$ gesetzt. Dann ist die Funktion

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

zu maximieren. Man erhält sofort

$$\frac{\partial Q}{\partial \mathbf{x}} = 2A\mathbf{x} - 2\lambda\mathbf{x},$$

und man erhält als Lösung $\mathbf{u}$ für $\partial Q/\partial\mathbf{x} = 0$ die Gleichung $A\mathbf{u} = \lambda\mathbf{u}$ ($\mathbf{u}$ ist der Vektor, für den $\partial Q/\partial\mathbf{x} = 0$ gilt). Der Rayleigh-Quotient wird maximal, wenn $\mathbf{x} = \mathbf{u}$ der erste Eigenvektor von A ist. □

4.5 Alternativer Beweis von Satz 2.34

Es sei $\mathbf{a}_{zi}$ der i-te Zeilenvektor von A, $i = 1, \ldots, m$. Es sei $A\mathbf{x} = \vec{0}$; dies bedeutet, dass die Skalarprodukte $\mathbf{a}'_{ui}\mathbf{x} = 0$ für alle i, d.h. $\mathbf{x}$ ist orthogonal zu allen Zeilenvektoren von A. Für die $\mathbf{x}$ mit $A\mathbf{x} = \mathbf{y} \neq \vec{0}$ gilt diese Aussage nicht, d.h. $\mathbb{R}^n$ wird in zwei Teilmengen U und V zerlegt:

$$U = \{\mathbf{x} \in \mathbb{R}^n | A\mathbf{x} = \vec{0}\} = \text{kern}(A), \quad V = \{\mathbf{x} \in \mathbb{R} | A\mathbf{x} = \mathbf{y} \neq \vec{0}\}.$$

$U = \text{kern}(A)$ ist ein Teilraum des $\mathbb{R}^n$. V ist ebenfalls ein Teilraum des $\mathbb{R}^n$, denn es sei $A\mathbf{x}_1 = \mathbf{y}_1$, $A\mathbf{x}_2 = \mathbf{y}_2$, $\mathbf{x}_1, \mathbf{x}_2 \notin \text{kern}(A)$. Dann ist, für $\lambda, \mu \in \mathbb{R}$ beliebig, $A\lambda\mathbf{x}_1 = \lambda\mathbf{y}_1$, $A\mu\mathbf{x}_2 = \mu\mathbf{y}_2$ und $A(\lambda\mathbf{x}_1 + \mu\mathbf{x}_2) = \lambda\mathbf{y}_1 + \mu\mathbf{y}_2 = \mathbf{y} \in \mathcal{L}(A)$, mithin ist $\mathbf{x} = \lambda\mathbf{x}_1 + \mu\mathbf{x}_2 \in V$. Offenbar ist $U \cap V = \emptyset$, denn ein Vektor $\mathbf{x} \in \mathbb{R}^n$ kann nicht zugleich in U und in V sein. Also folgt $U + V = \mathbb{R}$ und es folgt

$$\dim(U + V) = \dim(U) + \dim(V) = \dim \mathbb{R}^n = n. \tag{4.56}$$

Zur Bestimmung von $\dim(U)$ werde

$$A\mathbf{x} = x_1\mathbf{a}_1 + \cdots + x_r\mathbf{a}_r + x_{r+1}\mathbf{a}_{r+1} + \cdots + x_n\mathbf{a}_n = \vec{0}$$

betrachtet. Ohne Beschränkung der Allgemeinheit kann angenommen werden, dass die Spaltenvektoren $\mathbf{a}_1, \ldots, \mathbf{a}_r$ die linear unabhängigen Vektoren von A sind. Schreibt man

$$x_1\mathbf{a}_1 + \cdots + x_r\mathbf{a}_r = -(x_{r+1}\mathbf{a}_{r+1} + \cdots + x_n\mathbf{a}_n),$$

und berücksichtigt man, dass die $\mathbf{a}_{r+k}$ für $k = 1, \ldots, n - r$ Linearkombinationen der $\mathbf{a}_j$, $j = 1, \ldots, r$ sind, so dass

$$\mathbf{a}_{r+k} = \lambda_{k1}\mathbf{a}_1 + \cdots + \lambda_{kr}\mathbf{a}_r = \sum_{j=1}^{r} \lambda_{kj}\mathbf{a}_j \tag{4.57}$$

gelten muß, so hat man

$$\sum_{j=1}^{r} x_j \mathbf{a}_j = -\sum_{k=1}^{n-r} x_{r+k} \sum_{j=1}^{r} \lambda_{kj} \mathbf{a}_j. \tag{4.58}$$

Über den Vektor $\mathbf{x}$ ist bisher keine weitere Annahme gemacht worden außer, dass $A\mathbf{x} = \vec{0}$ gelten soll. Man kann also insbesondere

$$\mathbf{x} = \mathbf{x}_{r+k} = (x_1, \ldots, x_r, 0, \ldots, 0, 1, 0, \ldots, 0)', \quad k = 1, \ldots, n-r$$

setzen, wobei die 1 an der $(r+k)$-ten Stelle stehen soll, also

$$\begin{aligned} \mathbf{x}_{r+1} &= (x_1, \ldots, x_r, 1, 0, \ldots, 0)', \\ \mathbf{x}_{r+2} &= (x_1, \ldots, x_r, 0, 1, 0, \ldots, 0)', \\ \mathbf{x}_{r+3} &= (x_1, \ldots, x_r, 0, 0, 1, 0, \ldots, 0)' \\ &\vdots \end{aligned}$$

Die Gleichung (4.58) nimmt dann die Form

$$\sum_{k=1}^{r} x_{r+k} \sum_{j=1}^{r} \mu_{kj} \mathbf{a}_j = \sum_{j=1}^{r} \lambda_{kj} \mathbf{a}_j = -\sum_{j=1}^{r} x_j \mathbf{a}_j.$$

so dass

$$\sum_{j=1}^{r} \lambda_{kj} \mathbf{a}_j + \sum_{j=1}^{r} x_j \mathbf{a}_j = \sum_{j=1}^{n} (x_j + \lambda_j) \mathbf{a}_j = 0.$$

Da die $\mathbf{a}_j$ linear unabhängig sind folgt $x_j + \lambda_j = 0$ oder $-\lambda_j = x_j$ für alle j. es ist also

$$\mathbf{x}_k = (-\lambda_{k1}, \ldots, -\lambda_{kr}, 0, \ldots, 0, 1, 0, \ldots, 0)', \quad k = 1, \ldots, n-r \tag{4.59}$$

wobei die 1 an der $(r+k)$-ten Stelle steht. In der Tat ist

$$A\mathbf{x}_k = -\lambda_{k1}\mathbf{a}_1 - \cdots - \lambda_{kr} + \mathbf{a}_{r+k} = 0$$

wegen (4.57).

Die $\mathbf{x}_k$ sind linear unabhängig, wie man sofort sieht, denn

$$\mu_1 \mathbf{x}_1 + \cdots + \mu_k \mathbf{x}_k = 0$$

impliziert $\mu_i = 0$ für $i = 1, \ldots, n-r$. Sie bilden damit eine Basis für einen $(n-r)$-dimensionalen Teilraum. Damit ist gezeigt, dass $\text{kern}(A)$ mindestens $(n-r)$-dimensional ist. Die Frage ist, ob die Dimensionalität von $\text{kern}(A)$ nicht größer ist. Das ist aber nicht möglich, da ja bereits $\text{rg}(A) = r$ angebommen wurde, der Rang von $\text{kern}(A)$ kann also nicht größer als $n-r$ sein. Wegen (4.56) (= Dimensionssatz) folgt weiter, dass $\text{rg}(V) = \mathcal{L}(A) = r$ ist. □

Für $r = n$ folgt demnach $\text{rg}[\text{kern}(A)] = 0$, d.h. in diesem Fall hat $A\mathbf{x} = \vec{0}$ nur eine Lösung: $\mathbf{x} = \vec{0}$. Dies ist evident, denn in diesem Fall sind die $\mathbf{a}_1, \ldots, \mathbf{a}_n$ linear unabhängig und $\sum_j x_j \mathbf{a}_j = \vec{0}$ nur dann, wenn $x_1 = \cdots = x_n = 0$. □

4.6 Vektortransformationen und Abbildungen

Dieser Abschnitt enthält einige grundsätzliche Betrachtungen über Produkte von Matrizen und Vektoren, die einerseits das Verständnis der Vektor- und Matrixrechung vertiefen, andererseits für das Verständnis der unmittelbaren Anwendung der Matrixrechung auf Fragen der multivariaten Statistik nicht unbedingt notwendig sind und deshalb übersprungen werden können.

Das Produkt $X\mathbf{u} = \mathbf{v}$, X eine (m, n)-Matrix, $\mathbf{u}$ ein n-dimensionaler Vektor, $\mathbf{v}$ ein m-dimensionaler Vektor kann als Abbildung f: $\mathbf{u} \mapsto \mathbf{v}$ eines n-dimensionalen Vektors auf einen m-dimensionalen Vektor verstanden werden, wobei die Abbildung f durch die Matrix X definiert wird. Das Gleiche gilt für das Produkt $\mathbf{u}'X = \mathbf{v}'$, wenn $\mathbf{u}$ ein m-dimensionaler und $\mathbf{v}$ ein n-dimensionaler Vektor ist. Viele Sachverhalte der Vektor- und Matrixalgebra lassen sich sehr elegant als Eigenschaften von Abbildungen ausdrücken.

Eine Abbildung f einer Menge $\mathcal{M}$ in eine Menge $\mathcal{N}$ ordnet jedem Element aus $\mathcal{M}$ *genau einem* Element aus $\mathcal{N}$ zu:

$$f : \mathcal{M} \to \mathcal{N}, \quad x \mapsto y = f(x), \quad x \in \mathcal{M},\, y \in \mathcal{N} \tag{4.60}$$

Man schreibt gelegentlich auch

$$f(\mathcal{M}) = \mathcal{N}. \tag{4.61}$$

$f(\mathcal{M})$ heißt das *Bild von* $\mathcal{M}$ in $\mathcal{N}$, und $\mathcal{M}$ ist das *Urbild* von $f(\mathcal{M})$. Man schreibt auch $\mathrm{Im} f = \mathcal{N}$[49].

Eine spezielle Abbildung ist die *Identität* oder identische Abbildung

$$\mathrm{id}(\mathcal{M}) = \mathcal{M}. \tag{4.62}$$

Die Einheitsmatrix I_n der Spalten bzw. Zeilen aus den n-dimensionalen Einheitsvektoren $\mathbf{e}_1, \ldots, \mathbf{e}_n$ bestehen, spezifiziert die identische Abbildung, denn sicherlich gilt

$$I_n\mathbf{x} = \mathbf{x}, \quad \mathbf{x} \in \mathbb{R}^n. \tag{4.63}$$

Für eine Teilmenge von Abbildungen existiert die *inverse Abbildung* f^{-1}:

$$f(\mathcal{M}) = \mathcal{N}, \quad f^{-1}f(\mathcal{M}) = \mathcal{M} = f^{-1}(\mathcal{N}). \tag{4.64}$$

Wenn f durch eine Matrix M definiert ist, so bedeutet die Existenz der inversen Abbildung f^{-1} die Existenz einer inversen Matrix M^{-1}. Es wird deutlich werden, dass inverse Matrizen M^{-1} für eine Matrix M nur für spezielle Matrizen existieren.

Die Forderung, dass einem Element $x \in \mathcal{M}$ nur *ein* Element $y \in \mathcal{N}$ zugeordnet wird schließt nicht aus, das verschiedenen Elementen $x, x' \in \mathcal{M}$ der gleiche Wert

[49] Im wohl als Abkürzung des lat. *imago* für 'Bild'.

$y \in \mathcal{N}$ zugeordnet werden kann. In diesem Fall kann von einem Element $y \in \mathcal{N}$ nicht eindeutig auf das Element $x \in \mathcal{M}$ mit $f(x) = y$ zurückgeschlossen werden.

Mit der Schreibweise $f(\mathcal{M})$ ist nicht ein einzelnes Element gemeint, sondern die Menge der Werte, die man erhält, wenn man f für alle Werte aus X bestimmt, also

$$f(\mathcal{M}) = \{f(x), x \in \mathcal{M}\}. \tag{4.65}$$

Offenbar gilt $f(\mathcal{M}) \subseteq \mathcal{N}$.

Definition 4.3 *Es sei $f : \mathcal{M} \to \mathcal{N}$. Dann ist f*

1. injektiv*, wenn aus $x, x' \in \mathcal{M}$ und $f(x) = f(x')$ folgt, dass $x = x'$ (und damit $f(x) \neq f(x') \Rightarrow x \neq x'$). Es kann $f(\mathcal{M}) \subset \mathcal{N}$ gelten, d.h. $f(\mathcal{M})$ kann eine echte Teilmenge von $\mathcal{N}$ sein.*
2. surjektiv*, wenn zu jedem $y \in \mathcal{N}$ ein $x \in \mathcal{M}$ existiert derart, dass $y = f(x)$. Es gilt $f(\mathcal{M}) = \mathcal{N}$.*
3. bijektiv*, wenn f sowohl injektiv als auch surjektiv ist. Es gilt $f(\mathcal{M}) = \mathcal{N}$.*
4. Die Menge kern$(f) = \{x \in \mathcal{M} | f(x) = 0\}$ heißt Kern *der Abbildung f; man schreibt für den Kern auch kern$(f) = f^{-1}(\vec{0})$.*
5. Es sei $f(\mathcal{M}) = \mathcal{N}$, d.h. $f(\boldsymbol{x}) = \boldsymbol{y}$ für $\boldsymbol{x} \in \mathcal{M}$, $\boldsymbol{y} \in \mathcal{N}$. Dann heißt $f^{-1}(\boldsymbol{y}) = \{\boldsymbol{x} \in \mathcal{M} | f(\boldsymbol{x}) = \boldsymbol{y}\}$ die Faser *über $\boldsymbol{y} \in \mathcal{N}$.*

Anmerkung: Die Schreibweise $f^{-1}(f)$ für den Kern einer Abbildung f ergibt sich aus der in 4. gegebenen Definition: ist $\mathbf{x} \in \text{kern} f$, so gilt $f(\mathbf{x}) = \vec{0}$. Aus der Definition der Inversen folgt dann $\mathbf{x} = f^{-1}(\vec{0})$. Die Definition des Kerns setzt wie die Defintion der Faser offenbar voraus, dass die Inverse existiert. □

Beispiele: $f : \mathbb{R} \to \mathbb{R}$, $x \mapsto ax + b$ für $a, b \in \mathbb{R}$ fest gewählte Konstante. $\mathbb{R}$ bezeichnet die Menge der reellen Zahlen. f ist sicher injektiv, denn $f(x) = f(x')$ impliziert $ax + b = ax' + b$ und damit $x = x'$, wie man leicht nachrechnet. f ist auch surjektiv, denn für $y = ax + b$ existiert genau ein $x = (y - b)/a$ derart, dass $y = f(x)$. Da f sowohl injektiv wie surjektiv ist, ist f auch bijektiv.

Mit $\mathbb{R} \times \mathbb{R} = \mathbb{R}^2$ wird die Menge der Paare (x, y), $x, y, \in \mathbb{R}$, bezeichnet, allgemein mit

$$\mathbb{R}^m = \underbrace{\mathbb{R} \times \mathbb{R} \times \cdots \times \mathbb{R}}_{m-\text{mal}}$$

Die Menge der m-tupel $(x_1, x_2, \ldots, x_m)$, $x_j \in \mathbb{R}$, d.h. der m-dimensionalen Vektoren. $\mathbb{R}^n$ ist dann die Menge der n-dimensionalen Vektoren, etc. Mit $\mathbb{R}^{m,n}$ wird die Menge der (m, n)-Matrizen bezeichnet. Alle diese Definitionen übertragen sich auf $\mathcal{C}$, die Menge der komplexen Zahlen $x + iy$, $x, y \in \mathbb{R}$ und $i = \sqrt{-1}$.

Die Schreibweise $M \in \mathbb{R}^{m,n}$ bedeutet, dass M eine (m, n)-Matrix ist. Die Schreibweise $f : V_m \to V_n$ bedeutet dann, dass f eine Abbildung der m-dimensionalen Vektoren in die Menge der n-dimensionalen Vektoren ist. Wenn $V_m = \mathbb{R}^m$, $V_n = \mathbb{R}^n$ kann man auch $f : \mathbb{R}^m \to \mathbb{R}^n$ schreiben.

Da $M\mathbf{x} = \mathbf{y}$ mit $\mathbf{x} \in V_m$, $\mathbf{y} \in V_n$, folgt, dass f durch eine Matrix $M \in \mathbb{R}^{m,n}$ definiert ist.

Definition 4.4 *Es seien V und W Vektorräume und $f: V \to W$ sei eine Abbildung von V ind W.*
1. f heißt linear *bzw.* homomorph*, wenn*

$$f(\lambda \boldsymbol{v} + \mu \boldsymbol{w}) = \lambda f(\boldsymbol{v}) + \mu f(\boldsymbol{w}) \tag{4.66}$$

für $\lambda, \mu \in \mathbb{R}$ und für alle $\boldsymbol{v} \in V$ und $\boldsymbol{w} \in W$.
2. f heißt isomorph*, wenn f bijektiv ist; man sagt auch, f definiere einen* Homomorphismus *bzw.* Isomorphismus *wenn f bijektiv ist.*
3. f definiert einen Endomorphismus*, wenn $V = W$, und*
4. einen Automorphismus*, wenn f bijektiv ist und außerdem $V = W$ gilt.*

Es sei $M \in \mathbb{R}^{m,n}$; M definiert eine lineare, also homomorphe Abbildung, denn $M\mathbf{x} = \mathbf{y}$ erfüllt die Bedingungen einer linearen Abbildung. für $m \neq n$ ist f offenbar weder ein Endomorphismus noch ein Automorphismus.

Dem Begriff des Kerns in der allgemeinen Definition 4.3 von Abbildungen entspricht für $f \in \mathbb{R}^{m,n}$ der Nullvektor $\vec{0}$.

Satz 4.5 *Es sei $f(\mathcal{M}) = \mathcal{N}$. Dann gilt*
1. f ist surjektiv genau dann, wenn Im $f = \mathcal{N}$
2. f ist injektiv genau dann, wenn kern$f = \{\vec{0}\}$.
3. f sei injektiv und die Vektoren $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \in \mathcal{M}$ seien linear unabhängig. Dann sind auch die Bilder $f(\boldsymbol{x}_1), \ldots, f(\boldsymbol{x}_n)$ linear unabhängig.

Anmerkung: Die Schreibweise kern$f = \{\vec{0}\}$ bedeutet, dass kern$f$ nur das eine Element $\vec{0}$ enthält. □

Beweis: $\Rightarrow$ für 1. und 2. folgt sofort aus der Definition von injektiv und surjektiv. Um $\Leftarrow$ zu sehen, beztrachte man zwei Vektoren $\mathbf{u}, \mathbf{v} \in \mathcal{M}$ mit $\mathbf{u} \neq \mathbf{v}$, aber $f(\mathbf{u}) = f(\mathbf{v})$. Wegen der Linearität von f folgt dann

$$f(\mathbf{v}) - f(\mathbf{u}) = f(\mathbf{v} - \mathbf{u}) = \vec{0},$$

d.h. es gilt $\mathbf{v} - \mathbf{u} \in$ kern$f$.

Um 3. einzusehen sei angenommen, dass

$$\lambda_f(\mathbf{x}_1) + \cdots \lambda_n f(\mathbf{x}_n) = \vec{0}$$

gilt. Es wurde vorausgesetzt, dass f injektiv ist. Daraus folgt, dass

$$\lambda_1 \mathbf{x}_1 + \cdots \lambda_n \mathbf{x}_n = \vec{0}$$

gelten muß, denn $\lambda_1 \mathbf{x}_1 + \cdots \lambda_n \mathbf{x}_n$ ist ja das Urbild von f. Da die $\mathbf{x}_1, \ldots, \mathbf{x}_n$ als linear unabhängig vorausgesetzt wurden, muß $\lambda_1 = \cdots = \lambda_n = 0$ gelten, und dann folgt sofort, dass auch die $f(\mathbf{x}_j)$ linear unabhängig sind. □

Definition 4.5 *Es sei f eine Abbildung $f: \mathbb{R}^m \to \mathbb{R}^n$; dann heißt die Dimensionalität des Bildes* Imf *der* Rang*; man schreibt $rg(f) = \dim \in f$.*

f sei durch eine Matrix $A \in \mathbb{R}^{m,n}$ definiert, so dass $A : \mathbb{R}^m \to \mathbb{R}^n$ und $\mathbf{x} \mapsto \mathbf{y} = A\mathbf{x}$. Dann ist $f = A$. $\mathbf{e}_1, \ldots, \mathbf{e}_m$ ist die kanonische Basis von $\mathbb{R}^m$, d.h. die n m-dimensionalen Spaltenvektoren von A können als Linearkombinationen $A\mathbf{e}_1, A\mathbf{e}_2, \ldots, A\mathbf{e}_n$ geschrieben werden. Dann ist das Bild der durch A definierten Abbildung die lineare Hülle

$$\text{Im}A = \mathcal{L}(A\mathbf{e}_1, A\mathbf{e}_2, \ldots, A\mathbf{e}_n).$$

Demnach wird ImA auch der *Spaltenraum* von A bezeichnet. Der Begriff des Ranges einer Matrix A wird in Abschnitt 2.4 noch ausführlich diskutiert.

Beispiel 4.3 Es sei $f : \mathbb{R}^3 \to \mathbb{R}^4$; für $\mathbf{x} \in \mathbb{R}^3$ und $\mathbf{y} \in \mathbb{R}^4$ soll also $f(\mathbf{x}) = \mathbf{y}$ gelten; insbesondere sei f durch

$$f(\mathbf{x}) = \begin{pmatrix} x_1 + 2x_2 \\ 0 \\ \frac{1}{2}x_1 + x_2 \\ 0 \end{pmatrix}$$

definiert. Gesucht ist die zu f gehörige Matrix $M = A$ sowie der Kern von f.

Der Kern von f ist diejenige Menge von Vektoren $\mathbf{x}$, für die $f(\mathbf{x}) = \vec{0}$. Für dies Komponenten dieser Vektoren $\mathbf{x}$ muß also gelten

$$\begin{aligned} x_1 + 2x_2 &= 0 \\ \frac{1}{2}x_1 + x_2 &= 0, \end{aligned}$$

d.h. $x_1 = -2x_2$. Der Kern ist dann

$$kern(f) = \{(x_1, x_2, x_3)' | x_1 = -2x_2\} = \{(-2x_2, x_2, x_3)'\}.$$

Es gilt

$$A\mathbf{x} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \\ a_{41} & a_{42} & a_{43} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} = \mathbf{y}.$$

Es gibt also 12 Elemente a_{ij}, die zu bestimmen sind, wobei allerdings nur bestimmte Relationen zwischen den Komponenten gegeben sind, die aus dem Spezialfall $A\mathbf{x} = \vec{0}$ folgen. Wie die Diskussion linearer Gleichungssysteme zeigen wird, läßt sich aus diesen Bedingungen keine eindeutige Lösung für die a_{ij} ableiten.

Andererseits ist das Bild von f eine Linearkombination der Spalten von A, und damit folgt

$$\begin{pmatrix} x_1 + 2x_2 \\ 0 \\ \frac{1}{2}x_1 + x_2 \\ 0 \end{pmatrix} = \frac{x_1}{2} \begin{pmatrix} 2 \\ 0 \\ 1 \\ 0 \end{pmatrix} + x_2 \begin{pmatrix} 2 \\ 0 \\ 1 \\ 0 \end{pmatrix} = (\frac{x_1}{2} + x_2) \begin{pmatrix} 2 \\ 0 \\ 1 \\ 0 \end{pmatrix},$$

d.h. die Spalten von A haben die Form $(2c, 0, c, 0)'$ mit $c \in \mathbb{R}$ (Barrantes Campos (2012), p. 231). □

Wenn also eine Matrix $M \in \mathbb{R}^{m,n}$ eine Abbildung definiert, so kann man fragen, ob sie injektiv, surjektiv oder bijektiv ist. Die Abbildung ist injektiv, wenn aus $M\mathbf{x} = \mathbf{u}$ und $X\mathbf{y} = \mathbf{v}$ und $\mathbf{u} = \mathbf{v}$ folgt, dass $\mathbf{x} = \mathbf{y}$ ist, und aus $\mathbf{u} \neq \mathbf{v}$ folgt $\mathbf{x} \neq \mathbf{y}$. Die Frage nach der Injektivität ist also eine Frage nach der Eindeutigkeit der Abbildung. M definiert eine surjektive Abbildung, wenn für jeden Vektor $\mathbf{u} \in V_n$ eine Vektor $\mathbf{x} \in V_m$ existiert derart, dass $M\mathbf{x} = \mathbf{u}$, d.h. die Frage nach der Surjektivität ist die Frage, ob durch M alle Elemente von V_n bestimmt werden. M definiert eine bijektive Abbildung, wenn M eine sowohl injektive wie auch surjektive Abbildung definiert. Dies ist die Frage, ob eine surjektive Abbildung auch eindeutig ist. Offenbar hängen diese Eigenschaften von der Struktur der Matrix M ab. Was mit dem Begriff der Struktur einer Matrix genau gemeint ist, wird im Folgenden entwickelt.

Beispiel 4.4 Es sei

$$T = \begin{pmatrix} t_{11} & t_{12} \\ t_{21} & t_{22} \end{pmatrix} = \begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix}. \tag{4.67}$$

T definiert eine Abbildung $\mathbb{R}^2 \to \mathbb{R}^2$:

$$T\mathbf{x} = \begin{pmatrix} x_1\cos\phi + x_2\sin\phi \\ x_1\sin\phi - x_2\cos\phi \end{pmatrix} = x_1 \begin{pmatrix} \cos\phi \\ \sin\phi \end{pmatrix} + x_2 \begin{pmatrix} \sin\phi \\ -\cos\phi \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}.$$

T definiert die Rotation eines Vektors $\mathbf{x}$ um einen Winkel ϕ. Dadurch werden die Elemente von $\mathbb{R}^2$ auf Elemente von $\mathbb{R}^2$ abgebildet, – $\mathbf{y}$ ist ja wieder ein Element von $\mathbb{R}^2$. Die Abbildung ist sicher injektiv und surjektiv, also bijektiv und damit umkehrbar, d.h. man kann einen Vektor $\mathbf{y} \in \mathbb{R}^2$ wählen und in "zurückdrehen", so dass man wieder bei $\mathbf{x}$ landet. Die Abbildung bzw. Matrix, die diese inverse Rotation bewirkt, wird mit T^{-1} bezeichnet. □

4.7 Determinanten

Der Begriff der Determinante ist bereits auf Seite 27 im Zusammenhang mit einem linearen Gleichungssystem aufgetaucht, mit dem die Koeffizienten von Linearkombinationen bestimmt wurden aufgetaucht: die Ausdrücke für die Koeffizienten sind Quotienten, bei denen Zähler und Nenner Zahlen sind, die sich als

Funktionen ganzer Matrizen ergeben. Der Ausdruck 'Determinante' bringt zum Ausdruck, dass es vom Wert bestimmter Determinanten abhängt, ob das Gleichungssystem eine Lösung hat oder nicht, sie "determiniert" in diesem Sinne die Existenz einer Lösung. Tatsächlich liegen die Anfänge der Determinantentheorie wohl bei Gerolamo Cardano (1501 – 1576), der den Spezialfall einer $(2,2)$-Matrix behandelte; allgmeiner wurden sie dann von Gottfried Wilhelm Leibniz (1646 – 1716) im Zusammenhang mit Gleichungssytemen mit $n > 2$ Unbekannten untersucht. Es war allerdings der Genfer Mathematiker Gabriel Cramer (1704 – 1752), der die allgemeine Verwendung von Determinanten zur Lösung linearer Gleichungssysteme formulierte (Cramersche Regel), und die eigentliche Entwicklung der Determinantentheorie fand erst im 19-ten Jahrhundert statt. Die heutige axiomatische Definition des Begriffs der Determinante wurde von Karl Theodor Wilhelm Weierstraß (1815 – 1897) eingeführt.

Definition 4.6 *Es sei* $A = [\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n]$ *eine quadratische Matrix, d.g. die Spaltenvektoren* $\boldsymbol{a}_j$ *von* A *seien* n*-dimensional. Die* Determinante $|A|$ *bzw.* $\det(A)$ *von* A *ist eine Zahl (im Folgenden werden nur Matrizen mit reellen Elementen betrachtet, so dass die Determinante eine reelle Zahl ist), wobei* $|A|$ *die folgenden Eigenschaften hat:*
1. $|A|$ *ist* multilinear*, d.h. es gilt mit* $j \neq k$
(i) $\det(\boldsymbol{a}_1, \ldots, \boldsymbol{u} + \boldsymbol{v}, \ldots, \boldsymbol{a}_n) = \det(\boldsymbol{a}_1, \ldots, \boldsymbol{u}, \ldots, \boldsymbol{a}_n) + \det(\boldsymbol{a}_1, \ldots, \boldsymbol{v}, \ldots, \boldsymbol{a}_n)$,
(ii) $\det(\boldsymbol{a}_1, \ldots, \lambda\boldsymbol{a}_j, \ldots, \boldsymbol{a}_n) = \lambda \det(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_j, \ldots, \boldsymbol{a}_n)$
2. $|A|$ *ist* alternierend*, d.h. gilt* $\boldsymbol{a}_j = \boldsymbol{a}_k$ *für irgendzwei Spaltenvektoren, so gilt* $|A| = 0$.
3. $\det(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n) = 1$.
Dann heißt $\det(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n)$ *die* Determinante *der Matrix* $A = [\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n]$.

Betrachtet man statt der Spaltenvektoren die Zeilenvektoren einer Matrix A,so ändert dies nicht den Wert der Determinante von A.

Permutationen: Eine Permutation der Zahlen $1, 2, \ldots, n$ ist eine mögliche Anordnung dieser Zahlen: für die Zahlen $\{1, 2, 3\}$ ist $(3, 1, 2)$ eine mögliche Permutation. Für n Zahlen $1, 2, \ldots, n$ enthält die Menge $\mathcal{P}$ der möglichen Permutationen $n! = 1 \cdot 2 \cdot \cdots \cdot n$ Elemente. Es sei $\tau = (i_1, i_2, \ldots, i_n)$ ein Element von $\mathcal{P}$, wobei die i_k, $k = 1, \ldots, k$ Elemente aus $\{1, 2, \ldots, n\}$ sind. $\tau(i)$ und $\tau(k)$ definierten eine *Transposition*, wenn $\tau(i) = k$ und $\tau(k) = i$. Jedem $\tau \in \mathcal{P}$ kannn ein Vorzeichen zugeordnet werden: unterscheiden sich $\tau(i)$ und $\tau(j)$ durch nur eine Transposition, so ist $\tau(j) = -1\tau(i)$. Benötigt man m Transpositionen, um $\tau(i)$ in die Folge $\{1, 2, \ldots, n\}$ zu transponieren, so ist das Vorzeichen von $\tau(i)$ durch $\operatorname{sgn}(\tau) = (-1)^m$ gegeben. Für jede Permutation τ existiert eine Inverse τ^{-1} derart, dass $\tau^{-1}\tau = \{1, 2, \ldots, n\}$; τ^{-1} ist die auf τ angewandte Permutation, die die Permutaiton τ wieder in die "natürliche" Anordnung $1, 2, \ldots, n$ überführt. Die

Determinante $|A|$ einer (n,n)-Matrix $A=(a_{ij})$ ist dann durch

$$|A| = \sum_{\tau\in\mathcal{P}} \operatorname{sgn}(\tau) a_{\tau(1),1} a_{\tau(2),2} \cdots a_{\tau(n),n} \tag{4.68}$$

gegeben (s. a. Leibnizregel und Laplacescher Entwicklungssatz weiter unten).

Folgerungen:

1. *Die Matrix B gehe aus der Matrix A durch Vertauschung zweier Spalten hervor. Dann gilt $|B| = -|A|$.*

   Die Aussage ergibt sich aus der sukzessiven Anwendung der definierenden Eigenschaften der Determinante, insbesondere
   a. Addition des Vektors an j-ter Stelle zum Vektor an der k-ten Stelle,
   b. Subtraktion des neuen Vektors an der k-ten Stelle vom Vektor an der j-ten Stelle,
   c. Addition des Vektors an der j-ten Stelle zum Vektor an der k-ten Stelle,
   d. Ersetzung des Vektors an der j-ten Stelle durch sein Negatives.

2. *Es sei $A=[\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n]$. Dann gilt*

   $$|[\lambda_1\mathbf{a}_1, \lambda_2\mathbf{a}_2, \ldots, \lambda_n\mathbf{a}_n]| = \lambda_1\lambda_2\cdots\lambda_n|A|. \tag{4.69}$$

   Für $\lambda_1, \cdots = \lambda_n = \lambda$ ist dann

   $$|[\lambda\mathbf{a}_1, \lambda\mathbf{a}_2, \ldots, \lambda\mathbf{a}_n]| = |\lambda A| = \lambda^n|A| \tag{4.70}$$

   Die Aussagen folgen sofort aus Definition 4.6, 1. (ii).

3. *Die Matrix B entstehe aus A, indem der Vektor $\boldsymbol{a}_j$ durch $\boldsymbol{a}_j + \lambda\boldsymbol{a}_k$ ersetzt wird, wobei $\boldsymbol{a}_k$ ebenfalls ein Vektor von A ist und $\lambda\in\mathbb{R}$. Dann gilt $|A| = |B|$.*

   Denn

   $$\begin{aligned}\det(\mathbf{a}_1, \cdots, \mathbf{a}_j + \lambda\mathbf{a}_k, \ldots, \mathbf{a}_k, \ldots, \mathbf{a}_n) &= |A| + \lambda\det(\mathbf{a}_1, \ldots, \mathbf{a}_k, \ldots, \mathbf{a}_k, \ldots, \mathbf{a}_n)\\ &= |A|,\end{aligned}$$

   denn $\det(\mathbf{a}_1, \ldots, \mathbf{a}_k, \ldots, \mathbf{a}_k, \ldots, \mathbf{a}_n) = 0$ wegen Folgerung 1

4. *$|A| = [\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n]$ verändert ihren Wert nicht, wenn einer der Vektoren $\boldsymbol{a}_j$ durch eine Linearkombination der übrigen ersetzt wird.*

   Diese Folgerung ist eine offenkundige Verallgemeinerung von Folgerung 3.

5. *Ist einer der Vektoren von A der Nullvektor, so ist $|A| = 0$.*

   Die Matrix B ergebe sich aus der Marix A, indem der Vektor $\mathbf{a}_j$ durch $\lambda\mathbf{a}_j$ ersetzt wird; Für $\lambda = 0$ ist $\lambda\mathbf{a}_j = \vec{0}$, und gleichzeitig gilt dann gilt $|B| = \lambda|A| = 0\cdot|A| = 0$.

6. *Es sei* $A = [\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n]$ *und die* $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ *seien linear abhängig. Dann gilt* $|A| = 0$.

   Es sei $\mathbf{a}_j$ eine Linearkombination der übrigen, d.h es sei

   $$\mathbf{a}_j = \sum_{k=1, k\neq j}^{n} a_k \mathbf{a}_k.$$

   Dann kann $\mathbf{a}_j$ durch

   $$\mathbf{a}_j - \sum_{k=1, k\neq j}^{n} a_k \mathbf{a}_k = \vec{0}$$

   ersetzt werden, wodurch sich der Wert der Determinante nicht ändert. Wenn aber einer der Vektoren der Nullvektor ist, so ist nach Folgerung 5 die Determinante gleich Null.

   Anmerkung: Es sei $B = \lambda A$, $\lambda \in \mathbb{R}$. Findet man numerisch, dass $|B| \approx 0$, so kann man nicht folgern, dass die Spalten- oder Zeilenvektoren von B linear abhängig sind. Denn es sei $0 < \lambda| < 1$, Dann wird λ^n "klein" und nach (**??** muß $|B|$ dann ebenfalls "klein" sein, obwohl $|A|$ nicht "klein" ist.

7. *Sind zwei der Vektoren von* A *identisch, so ist* $|A| = 0$.

   Die Vektoren von A sind in diesem Fall linear abhängig, und nach Folgerung 6 ist dann $|A| = 0$.

8. *Es sei* $C = [\boldsymbol{c}_1, \ldots, \boldsymbol{c}_n]$ *und die* $\mathbf{c}_j$ *seien linear abhängig, so dass nach Folgerung 6* $|C| = 0$ *folgt. Es sei* $\mathcal{L}_C = \mathcal{L}(\boldsymbol{c}_1, \ldots, \boldsymbol{c}_n)$ *die lineare Hülle der* $\mathbf{c}_j$*. Die Vektoren* $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ *seien Elemente aus* $\mathcal{L}_C$ *und es sei* $A = [\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n]$*. Dann gilt* $|A| = 0$.

   Die $\mathbf{a}_j$ sind Linearkombinationen der $\mathbf{c}_j$, und da die $\mathbf{c}_j$ linear abhängig sind, sind auch die $\mathbf{a}_j$ linear abhängig, so dass nach Folgerung 6 $|A| = 0$ sein muß.

9. *Es sei* A *eine* (n, n)*-Matrix und* A' *sei die Transponierte von* A*. Dann gilt*

   $$|A'| = |A|. \tag{4.71}$$

   Man hat die **Leibniz-Formel**

   $$\begin{aligned} |A'| &= \sum_{\tau\in\mathcal{P}} \operatorname{sgn}(\tau) a_{1,\tau(1)} a_{2,\tau(2)} \cdots a_{n,\tau(n)} \\ &= \sum_{\tau\in\mathcal{P}} \operatorname{sgn}(\tau) a_{\tau^{-1}(1),1} \cdots a_{\tau^{-1}(n),n} \end{aligned}$$

   Es sei $\sigma \in \mathcal{P}$ mit $\sigma = \tau^{-1}$; dann gilt $\operatorname{sgn}(\tau) = \operatorname{sgn}(\sigma)$. Da über alle Elemente von $\mathcal{P}$ summiert wird, folgt

   $$|A'| = \sum_{\sigma\in\mathcal{P}} \operatorname{sgn}(\sigma) a_{\sigma(1),1} \cdots a_{\sigma(n),n} = |A|.$$

10. **Produktsatz** *Es seien A und B zwei (n,n)-Matrizen. Dann gilt*

$$|AB| = |A||B|. \tag{4.72}$$

Es sei $C = AB$; der j-te Spaltenvektor $\mathbf{c}_j$ von C ist dann durch $\mathbf{c}_j = A\mathbf{b}_j$ gegeben, $\mathbf{b}_j$ der j-te Spaltenvektor von B, $j = 1, \ldots, n$. Dann ist

$$|C| = |[\sum_{j=1}^{n} b_{j1}\mathbf{a}_j, \sum_{j=1}^{n} b_{j2}\mathbf{a}_j, \ldots, \sum_{j=1}^{n} b_{jn}\mathbf{a}_j]|.$$

Die Aussage (4.72) folgt dann aus der wiederholten Anwendung der Eigenschaft 1. (i) der Definition 4.6 und der Folgerung 4.69.

11. *Es sei A eine (n,n)-Matrix mit vollem Rang. Dann gilt*

$$|A^{-1}| = \frac{1}{|A|}. \tag{4.73}$$

Da A vollen Rang hat existiert die Inverse A^{-1}, so dass $A^{-1}A = I$, I die Einheitsmatrix. (4.72) impliziert dann $|A^{-1}A| = |A^{-1}||A| = |I| = 1$ (s. Definition 4.6, 3.). (4.73) folgt unmittelbar.

12. **Cramersche Regel**[50] *Es sei A eine (n,n)-Matrix, und $\boldsymbol{b}$ und $\boldsymbol{x}$ seien n-dimensionale Vektoren. Sind A und $\boldsymbol{b}$ gegeben und ist $\boldsymbol{x}$ unbekannt, so ist $A\boldsymbol{x} = \boldsymbol{b}$ ein lineares Gleichungssystem. Es sei $A_j(\boldsymbol{b})$ die Matrix, die aus A entsteht, wenn der j-te Spaltenvektor von A durch $\boldsymbol{b}$ ersetzt wird, und $I_j(\boldsymbol{x})$ sei die Matrix, wenn in der Einheitsmatrix I die j-te Spalte durch $\boldsymbol{x}$ ersetzt wird. Dann folgt die* Cramersche Regel*,*

$$x_j = \frac{|A_j(\boldsymbol{b})|}{|A|}. \tag{4.74}$$

wobei x_j die j-te Komponente von $\boldsymbol{x}$, also die j-te Unbekannte ist.

Man macht sich zunächst klar, dass $|I_j(\mathbf{x})| = x_j$ gilt: Für den Fall $n = 2$ hat man

$$\begin{vmatrix} x_1 & 0 \\ x_2 & 1 \end{vmatrix} = x_1 \cdot 0 - 0 \cdot x_2 = x_1, \quad \begin{vmatrix} 1 & x_2 \\ 0 & x_2 \end{vmatrix} = 1 \cdot x_2 - 0 \cdot x_1 = x_2.$$

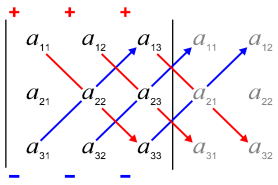
Den Fall $n = 3$ illustriert man anhand der Sarrusschen Regel (s. unten), und für $n > 3$ kann man den Laplaceschen Entwicklungssatz (s. unten) heranziehen.

Weiter ist

$$A_j(\mathbf{b}) = [\mathbf{a}_1, \ldots, \mathbf{a}_{j-1}, \mathbf{b}, \mathbf{a}_{j+1}, \ldots, \mathbf{a}_n],$$

[50] Gabriel Cramer (1704 – 1752), Genfer Mathematiker

Abbildung 14: Zur Regel von Sarrus



und weiter ist $A^{-1}\mathbf{a}_k = \mathbf{e}_k$ der k-te Einheitsvektor (wegen $A^{-1}A = I$). Dann folgt

$$A^{-1}A_j(\mathbf{b}) = [\mathbf{e}_1, \ldots, \mathbf{e}_{j-1}, A^{-1}\mathbf{b}, \mathbf{e}_{j+1}, \ldots, \mathbf{e}_n] = I_j(\mathbf{b}) = x_j,$$

und

$$|A^{-1}A_j(\mathbf{b})| = |A^{-1}||A_j(\mathbf{b})| = x_j,$$

wegen (4.72), und wegen (4.73) folgt (4.74).

13. **Regel von Sarrus** *Es sei also A eine $(3,3)$-Matrix. Sarrus*[51] *fand eine Regel, nach der sich diese Formel für* $\det(A)$ *leicht finden läßt: man hängt an die Matrix die ersten beiden Spalten noch einmal an und verbindet dann zuerst die Diagonale von A, d.h. die Elemente a_{11}, a_{22}, a_{33}: sie bilden das erste Produkt, s. Abbildung 14. Dann verbindet man die Elemente a_{12}, a_{23} und a_{31}: sie bilden das zweite Produkt, und schließlich noch die Elemente a_{13}, a_{21} und a_{32}, sie bilden das dritte Produkt. Die ersten drei Produkte werden addiert. Dann folgen drei Produkte mit negativem Vorzeichen: man verbindet die Elemente von a_{13} bis a_{32}, deren Produkt das erste mit negativem Vorzeichen ist, dann folgt das Produkt $a_{11}a_{23}a_{12}$ und schließlich das Produkt $a_{12}a_{21}a_{33}$, vergl. Abbildung*[52] *14. Für*

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

hat man also

$$|A| = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31} - a_{11}a_{23}a_{12} - a_{12}a_{21}a_{33}. \tag{4.75}$$

Die Regel von Sarrus ist sehr hilfreich, wenn man die Determinante einer $(3,3)$-Matrix "per Hand" berechnen muß, aber leider gilt sie *nur* für den Fall $n = 3$; für alle $n > 3$ gilt sie nicht.

[51] Pierre Frédéric Sarrus (1798 – 1861), französischer Mathematiker
[52] kopiert aus dem Wikipedia-Eintrag 'Regel von Sarrus'

14. **Entwicklung einer Determinante nach einer Spalte bzw. Zeile:** *Der in (4.75) – in (4.75) ist $n = 3$ – gegebene Ausdruck für* $\det(A)$ *ist durch eine Summe von Produkten aus jeweils n Elementen von A gegeben; der allgemeine Ausdruck ist nach (4.68)*

$$|A| = \sum_{\tau \in \mathcal{P}} sgn(\tau) a_{\tau(1),1} a_{\tau(2),2} \cdots a_{\tau(n),n}$$

τ ist eine der n! Anordnungen (Permutationen) von n Elementen aus A, $\tau(k) = j$ ist eine der Zahlen $j \in \{1, 2, \ldots, n\}$. Die Anordnung der Elemente innerhalb eines Produktes in (4.75) ist eine Konsequenz der Sarusschen Regel; da es auf die Anordnung der Elemente innerhalb eines Produktes nicht ankommt, kann diese so anordnen, dass die Komponente von $\boldsymbol{a}_1$ an erster, die von $\boldsymbol{a}_2$ an zweiter und die von $\boldsymbol{a}_3$ an dritter Stelle steht:

$$\begin{aligned} |A| &= a_{11}a_{22}a_{33} + a_{31}a_{12}a_{23} + a_{21}a_{32}a_{13} \\ &\quad -a_{31}a_{22}a_{13} - a_{11}a_{12}a_{23} - a_{21}a_{12}a_{33} \end{aligned} \tag{4.76}$$

Diese Anordnung entpricht der "kanonischen" Anordnung in (4.68), wie sie oben noch einmal angegeben wurde: die zweiten Indices zeigen immer die Reihenfolge 1, 2, 3.

Inspiziert man die Produkte, so fällt auf, dass die jeweils erste Zahl in einem Produkt, etwa a_{11}, die folgenden Zahlen so bestimmt, dass sie aus einer Restmatrix A_{11} kommen, die entsteht, wenn man die dem ersten Index entsprechende Zeile und die dem zweiten Indexes entpsrechhende Spalte aus A streicht. Für a_{11} erhält man die Restmatrix

$$A_{11} = \begin{pmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{pmatrix}, \quad |A_{11}| = a_{22}a_{33} - a_{23}a_{32} \tag{4.77}$$

Man sieht in (4.76), dass die Komponente a_{11} von $\mathbf{a}_1$ zweimal auftritt: $a_{11}a_{22}a_{33}$ und $a_{11}a_{12}a_{23}$. Die Teilprodukte $a_{22}a_{33}$ und $a_{12}a_{23}$ sind gerade die Summanden, die die Determinante von A_{11} definieren. Die Teilsumme $a_{11}a_{22}a_{33} - a_{11}a_{12}a_{23}$ in (4.76) läßt sich also zu

$$a_{11}a_{22}a_{33} - a_{11}a_{12}a_{23} = a_{11}|A_{11}| \tag{4.78}$$

zusammenfassen. Die zweite Komponente a_{21} von $\mathbf{a}_1$ taucht in (4.76) ebenfalls in zwei Summanden (Produkten) auf, und zwar als Differenz

$$a_{21}a_{32}a_{13} - a_{21}a_{12}a_{33} = -a_{21}|A_{21}|, \tag{4.79}$$

wobei jetzt

$$A_{21} = \begin{pmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{pmatrix}, \quad |A_{21}| = a_{12}a_{33} - a_{13}a_{32} \tag{4.80}$$

ist. A_{21} entsteht durch das Streichen der zweiten Zeile und der ersten Spalte aus A. Die Determinante $|A_{21}|$ geht hier mit negativem Vorzeichen ein, um die Differenz $a_{21}a_{32}a_{13} - a_{21}a_{12}a_{33}$ darzustellen. Allgemein wird das Vorzeichen für eine Differenz durch die Indices von A_{ji} gemäß $(-1)^{j+i}$ bestimmt: ist $j+i$ eine gerade Zahl, so ist das Vorzeichen positiv, ist $j+i$ ungerade, so ist das Vorzeichen negativ. Betrachtete man also die dritte Komponente a_{31} von $\mathbf{a}_1$, so findet man die Differenz $a_{31}a_{12}a_{23} - a_{31}a_{22}a_{13}$ im Ausdruck (4.76) für $|A|$. Für diese Differenz wird die dritte Zeile und wieder die erste Spalte aus A gestrichen; man erhält

$$A_{31} = \begin{pmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{pmatrix}, \quad |A_{31}| = a_{12}a_{23} - a_{13}a_{23}, \tag{4.81}$$

so dass

$$a_{31}a_{12}a_{23} - a_{31}a_{22}a_{13} = a_{31}|A_{31}|. \tag{4.82}$$

Hier ist das Vorzeichen wieder positiv, weit $1+3$ eine gerade Zahl ist. Die Determinante $|A|$ ist die Summe der Differenzen (4.78), (4.79) und (4.82):

$$\begin{aligned} |A| &= a_{11}|A_{11}| - a_{21}|A_{21}| + a_{31}|A_{31}| \\ &= (-1)^{1+1}a_{11}|A_{11}| + (-1)^{2+1}a_{21}|A_{21}| + (-1)^{3+1}a_{31}|A_{31}| \\ &= \sum_{k=1}^{3}(-1)^{k+1}a_{k1}|A_{k1}| \end{aligned} \tag{4.83}$$

(4.83) ist die *Entwicklung von $|A|$ nach der ersten Spalte von A*, eine andere Bezeichnung ist **Laplacescher Entwicklungssatz**[53]. Die Matrizen A_{ji} heißen *Minoren* von A: sie entstehen durch Streichen einer Spalte und einer Zeile von A, sie sind also $(n-1, n-1)$-Matrizen. Die Determinanten $|A_{ji}|$ der Minoren heißen *Kofaktoren* von A. Sie sind die Faktoren der Komponenten a_{kj} des Spaltenvektors $\mathbf{a}_j$, nach dem $|A|$ entwickelt wird.

15. **Zur Berechnung von A^{-1}:** *A habe vollen Rang, so dass die Inverse A^{-1} existiert. Es ist $A^{-1}A = AA^{-1} = I$. Es sei $\boldsymbol{y}_j$ der j-te Spaltenvektor von A^{-1}; dann folgt $A\boldsymbol{y}_j = \boldsymbol{e}_j$, $\boldsymbol{e}_j$ die j-te Spalte der Einheitsmatrix I. Die Beziehung $A\boldsymbol{y}_j = \boldsymbol{e}_j$ ist ein Gleichungssystem mit $\boldsymbol{y}_j$ als unbekanntem Vektor, so dass nach der Cramerschen Regel die i-te Komponente von $\boldsymbol{y}_j$ durch*

$$y_{ij} = \frac{|A_j(\boldsymbol{e}_i)|}{|A|} \tag{4.84}$$

gegeben ist. Es sei A_{ij} die Matrix, die aus A durch Streichen der i-ten Zeile und der j-ten Spalte entsteht. Dann gilt

$$|A_j(\boldsymbol{e}_i)| = (-1)^{i+j}|A_{ij}| := \alpha_{ij}. \tag{4.85}$$

[53] Pierre Simon Laplace (1749 – 1827), französischer Mathematiker

$|A_{ij}|$ *heißt der* Kofaktor *oder* Minor *des Elements* a_{ij} *von* A*. Die Matrix* $Adj(A)$ *der* α_{ij}*-Werte heißt* Adjunkte *der Matrix* A*, und (4.84) impliziert dann*

$$A^{-1} = \frac{1}{|A|}\mathrm{Adj}(A). \tag{4.86}$$

Zu zeigen ist (4.85). $A_j(\mathbf{e}_i)$ entsteht, wenn die j-te Spalte von A durch en Einheitsvektor $\mathbf{e}_i$ ersetzt wird. Man kann die Determinante $|A_j(\mathbf{e}_i)|$ bestimmen, indem man sie nach der j-ten Spalte entwickelt:

$$|A_j(\mathbf{e}_i)| = \sum_{k=1}^{n}(-1)^{k+j}e_{ij}|A_{ij}|.$$

Aber $e_{ij} = 0$ für alle $i \neq k$, so dass nur ein Summand – der für $k = i$ – übrig bleibt, also hat man (4.85).

16. *Es sei* A *eine orthonormale* (n, n)*-Matrix mit vollem Rang. Dann gilt*

$$|A| = 1. \tag{4.87}$$

Es ist

$$(\det(A))^2 = \det(A)\det(A) = \det(A')\det(A) = \det(A'A) = \det(I) = 1,$$

und es ist $\sqrt{(\det(A))^2} = \det(A) = \sqrt{1} = 1$.

17. *Es sei* $A = diag(a_{11}, a_{22}, \ldots, a_{nn})$ *eine Diagonalmatrix. Dann gilt*

$$|A| = |\mathrm{diag}(a_{11}, a_{22}, \ldots, a_{nn})| = \prod_{i=1}^{n} a_{ii}, \tag{4.88}$$

A kann in der Form $A = [a_{11}\mathbf{e}_1, a_{22}\mathbf{e}_2, \ldots, a_{nn}\mathbf{e}_n]$ geschrieben werden. Nach Folgerung 2, Gleichung (4.69) gilt dann

$$|A| = a_{11}a_{22}\cdots a_{nn}|I| = \prod_{i=1}^{n} a_{ii}.$$

18. **Ähnliche Matrizen** *Es seien* M *und* X *Matrizen mit vollem Rang, wobei* M *symmetrisch sei. Die Matrix* N *sei durch*

$$X^{-1}MX = N$$

definiert. Die Matrizen M *und* N *heißen dann* ähnlich*, und es gilt*

$$|X^{-1}MX| = |N|. \tag{4.89}$$

Insbesondere gilt

$$|N| = \prod_{i=1}^{n} \lambda_i, \tag{4.90}$$

wobei $\lambda_1, \ldots, \lambda_n$ die (positiven) Eigenwerte von M sind.

Nach dem Produktsatz und nach Folgerung 12, Gleichung (4.73) gilt

$$|X^{-1}MX = |X^{-1}M||X| = |X^{-1}||M||X| = \frac{1}{|X|}|M||X| = |M| = |N|.$$

Insbesondere sei $X = P$, P die Matrix der (orthognormalen) Eigenvektoren von M. Dann gilt $P'MP = \Lambda = \text{diag}(\lambda_1, \ldots, \lambda_n)$, λ_j, $j = 1, \ldots, n$ die zu den Eigenvektoren korrespondierenden Eigenwerte von M. Man hat dann $MP = P\Lambda$ und der Produktsatz liefert

$$|PM| = |P||M| = |P||\Lambda|.$$

Da P aber orthonormal ist folgt $|P| = 1$, so dass $|M| = |\Lambda| = \prod_{i=1}^{n} \lambda_i$ nach Folgerung 4.96, Gleichnung (4.88).

19. **(Charakteristisches Polynom)** *Es sei A eine (n, n)-Matrix; $\boldsymbol{v}$ sei ein Eigenvektor von A und λ sei der zugehörige Eigenwert, so dass*

$$A\boldsymbol{v} = \lambda\boldsymbol{v}$$

gilt. Es folgt

$$A\boldsymbol{v} - \lambda\boldsymbol{v} = (A - \lambda I)\boldsymbol{v} = \vec{0}.$$

Die Gleichung

$$|A - \lambda I| = \lambda^n - a_1\lambda^{n-1} + a_2\lambda^{n-2} + \cdots + (-1)^n a_n = 0 \tag{4.91}$$

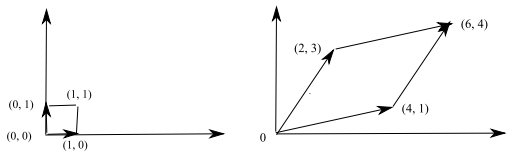
heißt charakteristische Gleichung *von A, das Polynom auf der rechten Seite heißt* charakteristisches Polynom *von A. Dabei gilt*

$$a_1 = \sum_{i=1}^{n} a_{ii}, \quad a_n = |A|, \tag{4.92}$$

d.h. a_1 ist die Spur von A. Die Nullstellen des Polynoms sind gleich den Eigenwerten von A.

Das charakteristische Polynom kann z.B. durch die Anwendung des Laplaceschen Entwicklungssatzes auf $|A - \lambda I|$ hergeleitet werden; die Koeffizienten a_i ergeben sich aus den Elementen der Matrix $A - \lambda I$. Die Herleitung soll hier übergangen werden. Die Bestimmung der Eigenwerte und der Eigenvektoren wird im Allgemeinen von Programmen übernommen (Golub & van Loan (2013)).

Abbildung 15: Abbildung eines Quadrats auf ein Parallelogramm



4.7.1 Transformationen und Volumen

Matrizen definieren Transformationen bzw. Abbildungen von Vektoren: ist A eine (m, n)-Matrix und $\mathbf{x}$ ein n-dimensionaler Vektor, so ist $\mathbf{y} = A\mathbf{x}$ ein m-dimensionaler Vektor. $\mathbf{y}$ ist eine Linearkombination der Spaltenvektoren von A, und man kann sagen, dass $\mathbf{x}$ durch A in $\mathbf{y}$ transformiert wird, oder dass der Vektor $\mathbf{x}$ durch A auf den Vektor $\mathbf{y}$ abgebildet wird. im Folgenden spielen insbesondere quadratische Matrizen eine Rolle; sie bilden n-dimensionale Vektoren auf n-dimensionale Vektoren ab; diese Matrizen definieren einen Endomorphismus, also eine Abbildung eines n-dimensionalen Vektorraums in sich selbst (s. Abschnitt 4.5). Es sei A also eine (n, n)-Matrix. Man kann eine Menge $\mathcal{G}$ von Vektoren $\mathbf{x}$ auswählen und die Menge $\mathcal{G}^* = \{\mathbf{y} | \mathbf{y} = A\mathbf{x},\ \mathbf{x} \in \mathcal{G}\}$ bestimmen. Hat die Matrix vollen Rang, so existiert die Determinante von A, $|A| \neq 0$. Es wird nun die geometrische Bedeutung von $|A|$ betrachtet.

Der Einfachheit halber wird eine $(2, 2)$-Matrix betrachtet. Es sei insbesondere $\mathcal{G}$ ein Quadrat mit den Seitenlängen gleich 1. Dann ist $\mathcal{G}^*$ durch ein Parallelogramm gegeben, s. Abbildung 15. Das Quadrat ist durch die Einheitsvektoren $\mathbf{e}_1 = (1, 0)'$ und $\mathbf{e}_2 = (1, 0)'$ definiert. Die Matrix A ist durch

$$A\mathbf{x} = \mathbf{y}, \quad A = [\mathbf{a}_1, \mathbf{a}_2] = \begin{pmatrix} 4 & 2 \\ 1 & 3 \end{pmatrix} \tag{4.93}$$

gegeben. Wählt man $\mathbf{x}_1 = \mathbf{e}_1$ und $\mathbf{x}_2 = \mathbf{e}_2$, so erhält man

$$A\mathbf{e}_1 = \mathbf{a}_1 = \begin{pmatrix} 4 \\ 1 \end{pmatrix}, \quad A\mathbf{e}_2 = \mathbf{a}_2 = \begin{pmatrix} 2 \\ 3 \end{pmatrix}, \tag{4.94}$$

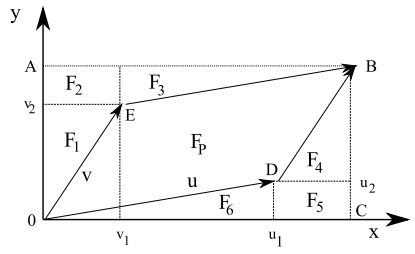
also gerade die Vektoren, die das Parallelogramm definieren. Wählt man $\mathbf{x} = \vec{0}$, so erhält man $\mathbf{y} = A\mathbf{x} = \vec{0}$; man bildet also den Punkt $(0, 0)$ im linken Koordinatensystem auf den Punkt $(0, 0)$ im rechten Koordinatensystem ab. Der Punkt $(6, 4)$ des Parallelogramms ergibt sich durch die Summe der Vektoren $\mathbf{y}_1 = (4, 1)$ und $\mathbf{y}_2 = (2, 3)$:

$$\mathbf{y}_1 + \mathbf{y}_2 = \begin{pmatrix} 4 \\ 1 \end{pmatrix} + \begin{pmatrix} 2 \\ 3 \end{pmatrix} = \begin{pmatrix} 6 \\ 4 \end{pmatrix}.$$

Es läßt sich nun zeigen, dass der Flächheninhalt des Parallelogramms durch die Determinante $|A|$ von A gegeben ist.

Das Quadrat hat sicherlich den Flächeninhalt 1. Der Flächeninhalt des Parallelogramms ergibt sich aus der bekannten Formel $F = hg$, wobei h die Höhe des Paralelogramms und g die Länge des Vektors $\mathbf{a}_2 = (3,2)'$ ist. h ist die Länge des Vektors $\mathbf{z}$ in Abbildung 11, Seite 127, und nach Gleichung (2.239), Seite 127, läßt sich $h = \|\mathbf{z}\|$ ausrechnen. Für die Zwecke dieses Abschnitts ist es aber sinnvoller, den Flächeninhalt ohne diese Berechnungen zu bestimmen, s. Abbildung 16.

Abbildung 16: Durch die Vektoren $\mathbf{a}_1 = (u_1, u_2)'$, und $\mathbf{a}_2 = (v_1, v_2)'$ definiertes Parallelogramm



Satz 4.6 *Es sei* $|A| = a_{11}a_{22} - a_{12}a_{21}$. *Der Flächeninhalt* F_p *des durch die Vektoren* $\boldsymbol{u}$ *und* $\boldsymbol{v}$ *definierten Parallelogramms (s. Abb. 16) ist durch*

$$F_p = u_1v_2 - u_2v_1 = a_{11}a_{22} - a_{12}a_{21} \tag{4.95}$$

gegeben.

Beweis: Offenbar ist F_P gleich der Fläche $F_{tot} = (u_1+v_1)(u_2+v_2)$ des Rechtecks 0ABC minus der Summe der Flächen F_1 bis F_6. Es ist

$$F_1 = F_4 = \frac{1}{2}v_1v_2 \tag{4.96}$$

$$F_2 = F_4 = v_1u_2 \tag{4.97}$$

$$F_3 = F_6 = \frac{1}{2}u_1u_2 \tag{4.98}$$

Einsetzen der Ausdrücke für F_1 bis F_6 liefert sofort

$$F_p = F_{tot} - (F_1 + \cdots + F_6) = |u_1v_2 - v_1u_2|, \tag{4.99}$$

also den Absolutbetrag der Determinante von A. □

Die Aussage, dass der Flächeninhalt des durch A repräsentierten Parallelogramms gleich dem Absolutbetrag der Determinante von A ist, kann auf allgemeine (n,n)-Matrizen verallgemeinert werden, wobei noch zu klären ist, warum vom *Absolutbetrag* der Determinante die Rede ist. Im vorangegangenen Beispiel ist die Determinante positiv, und Flächeninhalte bzw. allgemein Volumen sind

stets positive Zahlen. Vertauscht man allerdings die Spalten von A, so dass die Matrix $B = [\mathbf{a}_2, \mathbf{a}_1]$ entsteht, so findet man, dass die Determinante

$$|B| = a_{12}a_{21} - a_{11}a_{22} = -(a_{11}a_{22} - a_{12}a_{21}) = -|A|$$

negativ ist. In diesem Fall muß der Absolutbetrag von $|B|$ genommen werden, wenn von einem Volumen bzw. Flächeninhalt gesprochen werden soll. Auf die Bedeutung des Vorzeichens einer Determinante wird weiter unten eingegangen.

Das Parallelogramm entsteht, wenn alle Vektoren innerhalb einschließlich der Ränder des Quadrats in Abbildung 15 mit der Matrix A transformiert werden. Man kann nun ein beliebiges Gebiet G aus der xy-Ebene definieren und alle Vektoren $\mathbf{x}$ aus G mit A transformieren und erhält eine Menge $\mathcal{G}^*$ von Vektoren $\mathbf{y}$,

$$\mathcal{G}^* = \{\mathbf{y} | \mathbf{y} = A\mathbf{x}, \mathbf{x} \in G\}.$$

Die Frage ist, was über das Volumen von $\mathcal{G}^*$ gesagt werden kann.

Zunächst werde statt eines Quadrats mit der Seitenlänge 1 ein Quadrat mit der Seitenlänge $a \neq 1$ betrachtet. Dieses Quadrat ist das Gebiet G und hat den Flächeninhalt $F_G = a^2$. Die Einheitsvektoren $\mathbf{e}_j$ gehen dann über in die Vektoren $a\mathbf{e}_j$, die nun auf die Vektoren

$$Aa\mathbf{e}_1 = a\mathbf{a}_1 = \begin{pmatrix} 4a \\ 1a \end{pmatrix}, \quad Aa\mathbf{e}_2 = a\mathbf{a}_2 = \begin{pmatrix} 2a \\ 3a \end{pmatrix}, \tag{4.100}$$

abgebildet werden; es entsteht ein Parallelogramm, dessen Seitenlänge und Diagonale um den Faktor a verändert werden. Der Flächeninhalt des Parallelogramms beträgt nun nach (4.99) (die Komponenten der Vektoren $\mathbf{u}$ und $\mathbf{v}$ werden alle mit dem Faktor a multipliziert)

$$F_p = F_{\mathcal{G}^*} = a^2|u_1v_2 - v_1u_2| = F_G|A|. \tag{4.101}$$

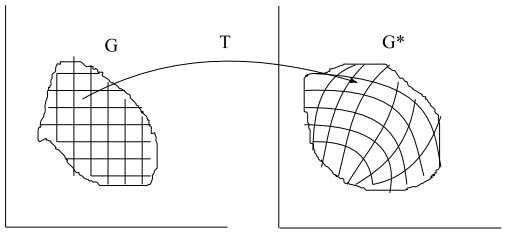
Für die Determinante $|A|$ folgt daraus sofort

$$|\det(A)| = \frac{F_{\mathcal{G}^*}}{F_G}, \tag{4.102}$$

d.h. im allgemeinen Fall ist $|A|$ gleich dem *Verhältnis* der Flächeninhalte bzw. im mehr als 2-dimensionalen Fall der Volumina der Gebiete $\mathcal{G}^*$ und F_G. Die Redeweise, dass $|A|$ gleich dem Volumen eines Parallelepipeds ist, impliziert den Fall $F_G = 1$.

Das Gebiet G ist aber nicht notwendig ein Quadrat, sondern kann irgendein zusammenhängedes Gebiet sein (Abbildung 17). T in Abb. 17 kann irgendeine Transformation sein, die lineare Transformation, wie sie durch eine (n, n)-Matrix gegeben ist, ist ein Spezialfall. Das Volumen des Gebiets G kann abgechätzt werden, indem man G in Quadrate unterteilt und die Summe der Fläche dieser Quadrate bildet. Die Approximation ist um so genauer, die kleiner die Quadrate

Abbildung 17: Allgemeine Gebietstransformation



und je größer ihre Anzahl. Im Grenzfall geht man zu infinitesimalen Flächen über und bestimmt die Fläche (das Volumen) von G durch ein Gebietsintegral. Analog dazu verfährt man mit der Fläche bzw. dem Volumen von $\mathcal{G}^* = T(G)$. Die zu den Quadraten in G korrespondierenden Flächen in $\mathcal{G}^*$ können im Limit durch infinitesímale Quadrate angenähert werden, die eine lineare Transformation A der Quadrate in G sind, d.h. A ist eine Matrix, und man findet analog zu (4.101)

$$\text{vol}(G) = |\det(A)|\text{vol}(\mathcal{G}^*) \tag{4.103}$$

oder

$$|\det(A)| = \frac{\text{vol}(G)}{\text{vol}(\mathcal{G}^*)}. \tag{4.104}$$

Einen exakten Beweis findet man z.B. in Courant II (1963); einen moderner formulierten . allerdings dann auch abstrakteren Beweis findet man in Fischer (1997). Eine für die multivariate Statistik relevante Anwendung dieses Ergebnisses wird im folgenden Abschnitt betrachtet.

4.7.2 Transformationen und Orientierung

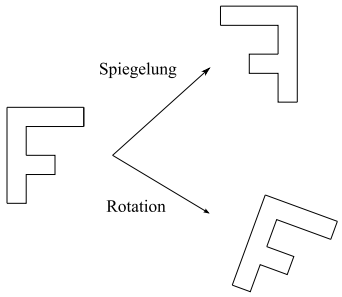
Zum Abschluß soll noch kurz auf die Bedeutung des Vorzeichens einer Determinante eingegangen werden. Dazu werden die beiden Matrizen[54]

$$A = \begin{pmatrix} -1 & -\frac{1}{4} \\ 1 & -\frac{1}{2} \end{pmatrix}, \quad B = \begin{pmatrix} -\frac{1}{4} & 1 \\ 1 & \frac{1}{2} \end{pmatrix}. \tag{4.105}$$

Man rechnet leicht nach, dass $|A| = 3/4$ und $|B| = -9/8$. Die Determinanten sind ungleich Null, weshalb die Spaltenvektoren $\mathbf{a}_1$ und $\mathbf{a}_2$ von A linear unabhängig sind, ebenso die Spaltenvektoren $\mathbf{b}_1$ und $\mathbf{b}_2$ von B, d.h. sowohl $\{\mathbf{a}_1, \mathbf{a}_2\}$ wie $\{\mathbf{b}_1, \mathbf{b}_2\}$ sind Basen des $\mathbb{R}^2$, also kann man mit beiden Basen alle Vektoren des $\mathbb{R}^2$ erzeugen. Wie gezeigt wurde, kann man $|A|$ als Volumen bzw. als Quotien von Volumina deuten, aber $|B|$ nicht, da Volumina nicht negativ sein können.

[54] Fischer (1997), p. 203)

Abbildung 18: Transformation: Spiegelung (Änderung der Orientierung) und Rotation



Allgemein gilt für Matrizen, dass sie Vektoren transformieren, d.h. gilt $A\mathbf{x} = \mathbf{y}$, so transformiert A den Vektor $\mathbf{x}$ in den Vektor $\mathbf{y}$, und natürlich gilt die analoge Aussage für B. Die Transformation bewirkt allgemein eine Rotation und eine Veränderung der Länge von $\mathbf{x}$, oder eines von beiden (und in speziellen Fällen gar keine Veränderung). Nun können die Vektoren $\mathbf{x}_1, \ldots, \mathbf{x}_p$ eine Figur abbilden, eben z.B. ein F. Die Matrix A rotiert alle $\mathbf{x}_j$ und verändert die Längen, und die Figur, hier das F, wird als Ganzes "nur" rotiert. Wendet man dagegen die Matrix B auf die $\mathbf{x}_j$ an, so wird das F nicht nur rotiert und eventuell verzerrt, sondern darüber hinaus gespiegelt, d.h. es wird die Orientierung der Figur oder allgemein der Vektorkonfiguration verändert (Abbildung 18). Die Spiegelung oder Veränderung der Orientierung wird durch das Vorzeichen der Determinante, hier von B, angezeigt. Man kann sagen, dass der Betrag der Determinante von B, $|\det(B)|$ ein Volumen repräsentiert, und das Vorzeichen der Determinante zeigt an, ob die Orientierung der Konfiguration durch die Transformation geändert wird.

Der Begriff der Orientierung kann mithilfe des *Vektorprodukts* illustriert werden, das im $\mathbb{R}^3$ definiert ist. Gegeben seien drei 3-dimensionale Vektoren $\mathbf{x}$, $\mathbf{y}$, und $\mathbf{z}$. Die Vektoren $\mathbf{x}$ und $\mathbf{y}$ liegen in einer Ebene (d.h. sie definieren eine Ebene im $\mathbb{R}^3$, und $\mathbf{z}$ steht senkrecht auf dieser Ebene. Ohne Beschränkung der Allgemeinheit kann man annehmen, dass diese Ebene durch die kanonischen Einheitsvektoren $\mathbf{e}_1$ und $\mathbf{e}_2$ definiert wird; dann steht $\mathbf{e}_3$ senkrecht auf dieser Ebene, hat also dieselbe Orientierung wie $\mathbf{z}$. Das Vektorprodukt ist dann definiert durch

$$\mathbf{x} \times \mathbf{y} = \begin{vmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 \\ x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \end{vmatrix} \tag{4.106}$$

Entwickelt man diese Determinante nach der ersten Zeile, so erhält man

$$\begin{aligned}\mathbf{z} = \mathbf{x} \times \mathbf{y} &= \mathbf{e}_1 \begin{vmatrix} x_2 & x_3 \\ y_2 & y_3 \end{vmatrix} - \mathbf{e}_2 \begin{vmatrix} x_1 & x_3 \\ y_1 & y_3 \end{vmatrix} + \mathbf{e}_3 \begin{vmatrix} x_1 & x_2 \\ y_1 & y_2 \end{vmatrix} \\ &= (x_2y_3 - x_3y_2)\mathbf{e}_1 + (x_3y_1 - x_1y_3)\mathbf{e}_2 + (x_1y_2 - x_2y_1)\mathbf{e}_3 \\ &= \begin{pmatrix} x_2y_3 - x_3y_2 \\ x_3y_1 - x_1y_3 \\ x_1y_2 - x_2y_1 \end{pmatrix} \end{aligned} \tag{4.107}$$

$\mathbf{x} \times \mathbf{y}$ ist ein Vektor, der orthogonal zu den Vektoren $\mathbf{x}$ und $\mathbf{y}$ ist; so ist

$$\mathbf{x}'\mathbf{z} = \mathbf{x}'(\mathbf{x} \times \mathbf{y}) = x_1x_2y_3 - x_1x_3y_2 + x_2x_3y_1 - x_1x_2y_3 + x_3x_1y_2 - x_3x_2y_1 = 0,$$

und analog dazu findet man $\mathbf{y}'(\mathbf{x} \times \mathbf{y}) = 0$. Also steht $\mathbf{z}$ senkrecht auf $\mathbf{x}$ und $\mathbf{y}$ und damit auf der durch eiese beiden Vektoren definierten Ebene.

Man rechnet leicht nach, dass

$$\mathbf{x} \times \mathbf{y} = -(\mathbf{y} \times \mathbf{x}). \tag{4.108}$$

Es läßt sich nun zeigen, dass dem Vektor $\mathbf{z}$ eine Drehrichtung zugeordnet werden kann; dieser Sachverhalt ergibt sich aus der Analyse der Bewegung eines Massepunktes auf einer Kreisbahn; die Details werden hier übergangen (man findet die Analyse Lehrbüchern dr Physik, z.B. in Westphal (1959), p. 16)). Zerlegt man die bei Drehbewegungen auftretenden Kräfte, so zeigt $\mathbf{z}$ die Orientierung der resultierenden Kraft und $\|\mathbf{z}\|$ ist die Größe der Kraft. Ist $\|\mathbf{z}\|$ die Kraft, mit der man eine Schraube in ein Holzstück dreht, so entspricht die Orientierung der üblichen Schraubenbewegung; $-\mathbf{z}$ entpricht dann der Orientierung der Drehung beim Herausschrauben der Schraube.

Um den Begriff der Orientierung formal fassen zu können, wird der Begriff der *geordneten Basis* eingeführt; dies ist eine Basis $\mathbf{b}_1, \ldots, \mathbf{b}_n$, bei der es auf die Anordnung der $\mathbf{b}_j$ ankommt. Nun seien $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ und $\mathcal{C} = \{\mathbf{c}_1, \ldots, \mathbf{c}_n\}$ zwei geordnete Basen eine n-dimensionalen Vektorraumes V. Dann existiert stets eine Matrix A, die die Basis $\mathcal{B}$ in die Basis $\mathcal{C}$ überführt[55], so dass mit $B = [\mathbf{b}_1, \ldots, \mathbf{b}_n]$ und $C = [\mathbf{c}_1, \ldots, \mathbf{c}_n]$ die Beziehung $C = AB$ gilt. Dann heißen die Basen $\mathcal{B}$ und $\mathcal{C}$ *gleichorientiert*, wenn $|A| > 0$ ist. Gilt diese Bedingung *nicht*, so heißen sie *verschieden orientiert*; dann gilt $|A| < 0$.

Vertauscht man in einer Matrix zwei Vektoren miteinander, so ändert sich das Vorzeichen der Determinante. Hat man also $\mathcal{B}_1 = \{\mathbf{b}_1, \ldots, \mathbf{b}_j, \mathbf{b}_{j+1}, \ldots, \mathbf{b}_n\}$ und $\mathcal{B}_2 = \{\mathbf{b}_1, \ldots, \mathbf{b}_{j+1}, \mathbf{b}_j, \ldots, \mathbf{b}_n\}$, so sind $\mathcal{L}(\mathcal{B}_1$ und $\mathcal{L}(\mathcal{B}_2)$ verschieden orientierte Vektorräume. insbesondere heißt eine geordnete Basis *positiv orientiert*, wenn sie dieselbe Orientierung wie die kanonische Basis $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ hat; andernfalls heißt sie *negativ orientiert*.

[55] Warum?

4.7.3 Die Transformation zufälliger Veränderlicher

Integrale lassen sich oft leichter bestimmen, wenn man die Integrationsvariable durch eine andere Variable ersetzt. Um die Substitution von Variablen diskutieren zu können, wird der Begriff der Differenzierbarkeit benötigt:

Definition 4.7 *Eine Funktion $f(x)$ mit einer Variablen heißt differenzierbar in x, wenn eine Zahl $f'(\xi)$ existiert derart, dass*

$$f(x+\xi) = f(x) + f'(x)\xi + o(\xi), \quad \lim_{\xi \to 0} \frac{o(\xi)}{\|\xi\|} = 0. \tag{4.109}$$

gilt. $o(\xi)$ ist eine Funktion voin ξ, die schneller gegen Null geht als ξ. Dabei ist

$$f'(x = \lim_{\xi \to 0} \frac{f(x+\xi) - f(x)}{\xi} \tag{4.110}$$

Der mehrdimensionale Fall ist analog definiert:

Definition 4.8 *Es sei $U \subset \mathbb{R}^n$ eine offene Menge und es sei $\boldsymbol{f}$ eine Abbildung $\boldsymbol{f}$: $U \to \mathbb{R}^m$. in $\boldsymbol{x} \in U$* total differenzierbar *(auch einfach* differenzierbar*), wenn eine weitere lineare Abbildung A: $\mathbb{R}^n \to \mathbb{R}^m$ gibt derart, dass in einer Umgebung von $\boldsymbol{x}$*

$$\boldsymbol{f}(\boldsymbol{x}+\boldsymbol{\xi}) = \boldsymbol{f}(\boldsymbol{x}) + A\boldsymbol{\xi} + o(\|\boldsymbol{\xi}\|) \tag{4.111}$$

gilt.

Hier ist $A = (a_{ij})$ eine (m,n)-Matrix und $\boldsymbol{\xi}$ ein n-dimensionaler Vektor; weiter ist $\mathbf{f} = (f_1, \ldots, f_n)'$ ein n-dimensionaler Vektor. Offenbar ist differenzierbar, wenn die Komponenten f_i differenzierbar sind.

Satz 4.7 *Es seien $\boldsymbol{f}$ und A wie in Definition 4.8 definiert. Dann gilt (i) $\boldsymbol{f}$ ist in $\boldsymbol{x}$ stetig, (ii) die Komponenten f_i von $\boldsymbol{f}$ sind in $\boldsymbol{x}$ partiell differenzierbar, d.h. es gilt*

$$\frac{\partial f_i}{\partial x_j} = a_{ij}. \tag{4.112}$$

Beweis: (vergl. Forster (2), p. 78) Es ist

$$\lim_{\boldsymbol{\xi} \to \vec{0}} A\boldsymbol{\xi} = \mathbf{f}(\boldsymbol{\xi}),$$

und dies heißt, dass $\mathbf{f}$ in $\mathbf{x}$ stetig ist. Weiter gilt

$$f_i(\mathbf{x}+\boldsymbol{\xi}) = f_i(\mathbf{x}) + \sum_{i=1}^{n} a_{ij}\xi_j + o(\|x_i\|), \quad i = 1, \ldots n$$

d.h.

$$f_i(\boldsymbol{\xi} + h\mathbf{e}_j) = f_i(\mathbf{x}) + ha_{ij} + o(\|h\mathbf{e}_j\|),$$

so dass

$$\frac{\partial f_i}{\partial x_j} = \lim_{h\to 0} \frac{f_i(\mathbf{x} + h\mathbf{e}_j) - f_i(\mathbf{x})}{h} = a_{ij} + o(\|h\mathbf{e}_j\|) = a_{ij}.$$

□

Die Implikation des Satzes ist, dass die Matrix A durch die Abbildung $\mathbf{f}$ eindeutig bestimmt ist.

Definition 4.9 *Die Matrix A heißt das* Differential *oder* Funktionalmatrix *oder* Jacobi-Matrix[56] *der Abbildung* $\boldsymbol{f}$ *im Punkt* $\boldsymbol{x}$.

Im mehrdimensionalen Fall wird also der Differentialquotient durch eine Matrix von partiellen Differentialquotienten ersetzt. Man hat die Schreibweisen

$$(D\mathbf{f})(\mathbf{x}) := J_{\mathbf{f}} := \frac{\partial(f_1, \ldots, f_n)}{\partial(x_1, \ldots, x_n)}(\mathbf{x}) := \left(\frac{\partial f_i}{\partial x_j}\right), \tag{4.113}$$

wobei ":=" bedeutet, dass der Ausdruck links von diesem Zeichen durch den rechts davon stehenden definiert wird. In ausgeschriebener Form hat die Funktional- oder Jacobi-Matrix die Gestalt

$$J_{\mathbf{f}} = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1} & \frac{\partial f_n}{\partial x_2} & \cdots & \frac{\partial f_n}{\partial x_n} \end{pmatrix} \tag{4.114}$$

Die Differenzierbarkeit von bedeutet nach (4.111), dass in einer hinreichend kleinen Umgebung von $\mathbf{x}$ linear approximiert werden kann. $A\boldsymbol{\xi}$ ist ein Vektor, der die infinitesimale Veränderung von $\mathbf{f}$ im Punkt $(\mathbf{x})$ angibt.

Die Kettenregel: Im mehrdimensionalen Fall gilt

Satz 4.8 *Es seine* $U \subset \mathbb{R}^n$ *und* $V \subset \mathbb{R}^m$ *offene Mengen und* $\boldsymbol{f}$: $U \to \mathbb{R}^m$, $\boldsymbol{g}$: $V \to \mathbb{R}^k$ *offene Mengen und es gelte* f: $U \to \mathbb{R}^m$, $g : V \to \mathbb{R}^k$, $(\mathbf{U}) \subset \mathbf{V}$. $\boldsymbol{f}$ *sei in* $\boldsymbol{x}$ *und* $\boldsymbol{g}$ *sei in* $\boldsymbol{y} := \boldsymbol{f}(\boldsymbol{x})$ *differenzierbar. Es sei* $\boldsymbol{g}\,o\,\boldsymbol{f}$: $U \to \mathbb{R}^k$ *die aus* $\boldsymbol{g}$ *und* $\boldsymbol{f}$ *zusammengesetzte Abbildung; sie sei in* $\boldsymbol{x}$ *differenzierbar und es gilt*

$$D(\boldsymbol{g}\,o\,\boldsymbol{f})(\boldsymbol{x}) = (D\boldsymbol{g})\boldsymbol{f}(\boldsymbol{x}) \cdot D\boldsymbol{f}(\boldsymbol{x}). \tag{4.115}$$

Beweis: Siehe z.B. Forster (2), p.80.

Zur einführenden Illustration werde die Substitutionsregel bei der Integration einer Funktion von nur einer Veränderlichen erinnert. So sei $f(x)$ eine Funktion

[56] Carl Gustav Jacobi (1804 – 1851), Mathematiker

auf einem Intervall $I \subseteq \mathbb{R}$. Eine neue Variable u kann durch eine auf dem Intervall I invertierbare Funktion $\psi(u) = x$ eingeführt werden. Ist $F(x)$ die Stammfunktion von f, d.h. gilt $F(x) = \int f(x)dx$[57]. So kann F als Verkettung $Fo\varphi = F(\varphi(u))$ gesehen werden. Man kann dann

$$F(\varphi(u)) = G(u). \tag{4.116}$$

iund die Kettenregel führt auf

$$\frac{dF(\varphi(u))}{du} = \frac{dF(\varphi)}{d\varphi}\frac{d\varphi}{du} = f(\varphi)\varphi' = g(u). \tag{4.117}$$

und

$$\begin{aligned} \int_a^b f(\varphi(u))\varphi'(u)du &= F(\varphi(b)) - F(\varphi(a)) \\ &= \int_{\varphi(a)}^{\varphi(b)} f(x)dx, \quad x = \varphi(u) \end{aligned} \tag{4.118}$$

Diese Aussage "ist die Grundlage für die Einführung neuer Veränderlicher in ein Integral" (Courant I (1961), p. 183). Wie Courant I, p. 184, ausführt, folgt man aber der Gleichung (4.118) von rechts nach links: gegeben ist eine zu integrierende Funktion $f(x)$, Die Aufgabe, das Integral zu bestimmen, kann u. U. vereinfacht werden, indem man für x die Funktion $x = \varphi(u)$ und damit die neue Integrationsvariable u einführt. Man bestimmt also

$$G(u) = \int f(\varphi(u))\varphi'(u)du$$

und ersetzt dann u wieder durch x. Dazu muß φ umkehrbar sein, d.h. es muß die inverse Funktion $\psi(x) = u = \varphi^{-1}(x)$ existieren mit $\psi'(x) \neq 0$ für alle $x \in I$. Als *Grundformel* (Courant I, p. 184) erhält man dann

$$\int f(x)dx = \int f(\varphi(u))\varphi'(u)du, \quad u = \psi(x) \tag{4.119}$$

Man bestimmt also das Integral auf der linken Seite der Gleichung, indem man das Integral auf der rechten Seite findet und anschließend u über die Abbildung $u = \psi(x)$ wieder einführt. Es sei angemerkt, dass

$$\varphi' = \frac{1}{\psi'} \tag{4.120}$$

gilt; hier wird klar, warum $\psi' \neq 0$ für $x \in I$ gefordert werden muß.

[57]Diese Schreibweise mag etwas lax sein, weil die Laufvariable x nicht vom Argument x in $F(x)$ unterschieden wird, ist aber als möglichst einfache, intuitive Schreibweise zu lesen (vergl. Courant I (1961), p. 182)

Es ist wichtig, sich die Bedeutung des Faktors φ' in Gleichung (4.118) noch einmal klar machen: offenbar genügt es nicht, x einfach durch $\varphi(u)$ zu ersetzen, es muß eben noch der Faktor $\varphi'(u)$ hinzugefügt werden. Im Vergleich mit der rechten Seite von (4.118) erhält man die Beziehung

$$dx = \varphi'(u)du \tag{4.121}$$

Die Transformation impliziert i. A., dass x und u und damit auch dx und du auf verschiedenen Skalen laufen[58] Bei nichtlinearen Funktionen φ bzw. ψ wird die Beziehung zwischen dx und dy aber nicht nur durch eine Konstante bestimmt; $\varphi'(u)$ kann ja eine Funktion von u sein.

In (4.119) repräsentiert $f(x)dx$ ein "infinitesimales Rechteck" der Breite dx und der Höhe $f(x)$. Beim Integral auf der rechten Seite werden aber nicht die infinitesimalen Rechtecke $f(\varphi(u))du$ aufsummiert, sondern die Rechtecke $f(\varphi(u))\varphi'(u)du$, was man so lesen kann, dass der Verlauf von f durch den u.U. von u abhängigen Faktor φ' modifiziert wird. Ist $\varphi' = a$ eine Konstante, so werden die $f(\varphi(u))$-Werte vergrößert, wenn $a > 1$, und verkleinert, wenn $a < 1$ ist. Die Veränderung der Skala geht also mit einer Modifikation des Verlaufs von $f(\varphi(u))$ einher.

Im Übrigen kann man die infinitesimalen Flächen $f(x)dx$ etc als Spezialfall des allgemeinen Volumenbegriffs auffassen; diese Sichtweise wird weiter unten noch elaboriert.

Beispiel 4.5 Eine einfache, illustrierende Anwendung ist die Integration der Funktion $f(x) = \sin(2x)$. Man kann hier die Transformation $u = \psi(x) = 2x$ einführen; dann ist

$$\frac{du}{dx} = \frac{d\psi(x)}{dx} = \psi'(x) = 2, \quad \text{d.h. } du = 2dx, \text{bzw. } dx = \frac{1}{2}du$$

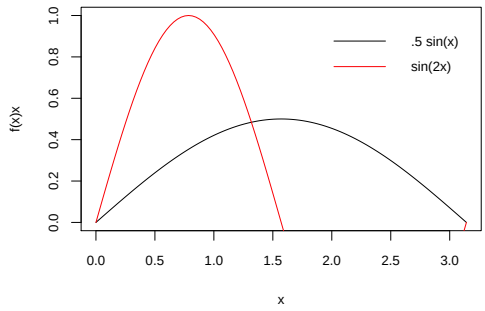
einer Veränderung auf der x-Skala entspricht eine doppelt so große Veränderung auf der u-Skala, und einer Veränderung auf der u-Skala entspricht die Hälfte dieser Verändrung auf der x-Skala. Diese Beziehungen übertragen sich auf die infinitesimalen Größen dx und du. Sind a und b die Grenzen des Integrals auf der u-Skala, so erhält man wegen $u/2 = x$ die Grenzen $a/2$ und $b/2$ auf der x-Skala; speziell für $a = 0$ und $b = \pi$ erhält man also

$$\begin{aligned} \int_0^{\pi/2} \sin(2x)dx &= \int_0^{\pi} \sin(u)\frac{1}{2}du = \frac{1}{2}\int_0^{\pi} \sin(u)du \\ &= -\frac{1}{2}\cos(\pi) + \frac{1}{2}\cos(0) = 1. \end{aligned} \tag{4.122}$$

Abbildung 19 illustriert das Prinzip: Die Funktion $\sin(2x)$ "überdeckt" nur die

[58] Ein einfacher Fall sind die Celsius- und Fahrenheitskalen für die Temperatur, die durch eine lineare Transformation aufeinander bezogen sind: $F = aC + b$ mit $a = 1.8$, $b = 32$. Der Unterschied von einem Grad Celsius entspricht einem Unterschied von 1.8 Grad F. Differenziert man, so erhält man $dF/dC = a$, d.h. $dF = a \cdot dC$. d.h. der Faktor a bezieht auch die Differentiale aufeinander. Da dx und dy infinitesimale *Differenzen* sind, verschwindet die additive Konstante b.

Abbildung 19: Skalen und Fläche unter Funktionen



Hälfte des Intervalls, den die Funktion $.5\sin(u)$ überdeckt, und damit die Flächen unter den Funktionen – also die Integrale über die entsprechenden Intervalle – gleich groß sind, müssen die Funktionswerte von $\sin(u)$ in diesem Fall mit dem Faktor $1/2$ multipliziert werden. Eine intuitive Plausibiltätsbetrachtung mag hier hilfreich sein: Geht man von der Riemannschen Definition des Integrals als Grenzwert einer Summe von Rechtecken aus, so gilt speziell für eine Zerlegung des Integrationsintervalls $[a, b]$ in gleichgroße Intervalle $\Delta x = (b - a)/n$

$$\int f(x)dx = \lim_{\Delta x_i \to 0} \lim_{n \to \infty} \sum_{i=1}^{n} f(t_i)\Delta x, \quad t_i \in [x_i - x_{i-1}) \tag{4.123}$$

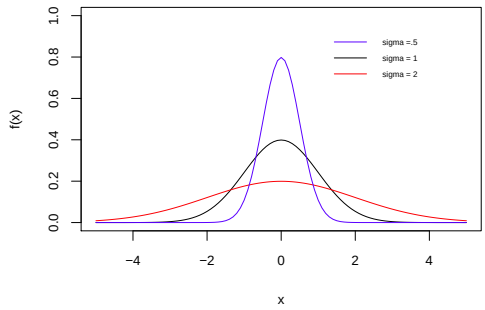
d.h. für $\Delta x \to dx$ für $n \to \infty$. Gilt nun $f(t_i)\Delta x = f(u_i)\varphi'\Delta x$, so erhält man im Limit für die Rechtecke $du = \varphi' dx$, und φ' kann als Dehnung oder Verkürzung von Δx oder als Streckung oder Stauchung von $f(t_i)$ interpretiert werden, so dass $f(u)\varphi' du = f(x)dx$ gilt. So haben die Rechtecke von $\sin(u)$ nur die Breite $du = dx/2$, haben also nur den halben Flächeninhalt. Da man den Faktor $1/2$ auch auf die Höhe der Rechtecke beziehen kann, kann man ebensogut sagen, dass die Rechtecke für $\sin(u)$ nur die halbe Höhe der Rechtecke für die Funktion $\sin(2x)$ haben. □

Beispiel 4.6 Es sei Z eine $N(0, 1)$-verteilte zufällige Veränderliche, d.h. Z sei standardnormalverteilt, so dass $\mathbb{E}(Z) = 0$ und $Var(Z) = 1$. Es sei nun $X = \varphi(Z) = \sigma Z + \mu$. Die Verteilung von X wird sich also über die Verteilung von Z ergeben, denn $\varphi^{-1}(X) = (X - \mu)/\sigma = Z$. Dementsprechend ergibt sich der folgende Ansatz

$$P(X \leq x) = P(\sigma Z + \mu \leq x) = P(Z \leq \frac{x - \mu}{\sigma}),$$

so dass die Verteilungsfunktion F von X auf G, die Verteilungsfunktion von Z

Abbildung 20: Skalen und Fläche: Gauss-Funktionen für verschiedene σ- (sigma)Werte unter Funktionen



bezogen wird. Für die Dichten folgt

$$f(x) = \frac{dF(x)}{dx} = \frac{dG((x-\mu)/\sigma)}{dx} = f_Z\left(\frac{x-\mu}{\sigma}\right)\frac{1}{\sigma},$$

wobei der Faktor $1/\sigma$ durch Anwendung der Kettenregel entsteht. Hier wird also der Verlauf von f um den konstanten Faktor $1/\sigma$ modifiziert: ist $\sigma > 1$, wird der Verlauf flacher, dafür wird der Bereich, in dem $f > \varepsilon$ für ein $\varepsilon > 0$ ist, größer; für $\sigma < 1$ gilt die Umkehrung. Wegen $X = \sigma Z + \mu$ gilt $\Delta Z = \sigma \Delta Z$, d.h. eine Veränderung von Z um ΔZ bedeutet eine σ-fache Veränderung Δx von X. Setzt man $(x-\mu)/\sigma$ für z in die Dichtefunktion $f_Z(z) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}z^2)$ für Z ein und multipliziert, mit $1/\sigma$, so erhält man die Dichtefunktion für X:

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right). \tag{4.124}$$

Allgemein gelte nun $X = \varphi(Z)$, wobei φ^{-1} für alle X existiere und von Null verschieden sei. Dann ergibt sich

$$P(X \leq x) = P\varphi(Z) \leq x) = P(Z \leq \varphi^{-1}(x)),$$

so dass

$$f(x) = f_Z(\varphi^{-1}(z))\frac{d\varphi^{-1}(x)}{dx} = f_Z(\psi(x))\frac{d\psi(x)}{dx} \tag{4.125}$$

f_Z ist hier wieder in der in (4.119) angegebenen allgemeinen Form. Während eine lineare Transformation von Z wieder auf die Gaußsche Dichte (4.124) führt, muß dies bei einer nichtlinearen Transformation φ nicht mehr der Fall sein. □

Der mehrdimensionale Fall: Es seien $\mathbf{x} = (x_1, \ldots, x_m)'$ und $\mathbf{y} = (y_1, \ldots, y_n)'$ zufällige Vektoren, d.h. die x_i und y_i seien zufällige Veränderliche, und es gelte

$\mathbf{y} = A\mathbf{x}$, A eine (n,n)-Matrix mit vollem Rang, so dass die Inverse A^{-1} existiert. Gesucht ist die Verteilung von $\mathbf{y}$, wenn die von $\mathbf{x}$ gegeben ist. Der Übersichlichkeit halber wird das vorgehen am Fall $n = 2$ illustriert.

Gegeben sei ein Doppelintegral einer Funktion mit zwei unabhängigen Veränderlichen; eine Transformation der beiden Variablen kann oft die Aufgabe, das Integral zu bestimmen, erleichtern. Dazu werden die Funktionen

$$x = \varphi(u,v), \quad y = \psi(u,v) \tag{4.126}$$

eingeführt. Die Jacobi-Matrix ist

$$J = \frac{\partial(x,y)}{\partial(u,v)} = \begin{pmatrix} \frac{\partial\varphi}{\partial u} & \frac{\varphi}{\partial v} \\ \frac{\partial\psi}{\partial u} & \frac{\partial\psi}{\partial v} \end{pmatrix} = \begin{pmatrix} \varphi_u & \varphi_v \\ \psi_u & \psi_v \end{pmatrix}. \tag{4.127}$$

Dann gilt der

Satz 4.9 Transformationssatz

$$\iint_G f(x,y)dxdy = \iint_{G*} f(\varphi(u,v),\psi(u,v))|\det(J_{\varphi\psi})|dudv \tag{4.128}$$

Beweis: Statt eines ausführlichen Beweises wird hier eine eher intuitive Betrachtung vorgestellt (Courant II, p. 231), einen vollständigen Beweis findet man zB in Courant II, pp. 222 – 225. Man denke sich das Gebiet $\mathcal{G}^*$ in Rechtecke aufgeteilt: Geraden u_i parallel zur v-Achse und Geraden v_j parallel zur u-Achse. Die Differenz zwischen u_i und u_{i+1} betrage $u_{i+1} - u_i = \Delta u$, die zwischen v_{j+1} und v_j sei Δv für alle i und j. Das i-te Rechteck G_i hat dann die Fläche $\Delta u \Delta v$. in der xy-Ebene entspricht g_i ein Teilgebiet G_i^*, definiert durch die Abbildungen $\varphi(u_i,v_j)$, $\psi(u_i,v_j)$ etc. Im allgemeinen Fall sind die Gebiete G_i^+ nicht mehr notwendig Rechtecke, denn die Verbindungslinien zwischen den Eckpunkten können nun krummlinig sein. Sind φ und ψ allerdings linear, so sind die G_i^* Parallelogramme. Man betrachte nun die Determinante

$$D := \begin{vmatrix} \varphi(u_i+h,v_j) - \varphi(u_i,v_j) & \varphi(u_i,v_j+k) - \varphi(u_i,v_j) \\ \psi(u_i+h,v_j) - \psi(u_i,v_j) & \psi(u_i,v_j+k) - \psi(u_i,v_j) \end{vmatrix}$$

Für hinreichend kleine Werte von h und k kann man dafür

$$\begin{vmatrix} \varphi_u(u_i,v_j) & \varphi_v(u_i,v_j) \\ \psi_u(u_i,v_j) & \psi_v(u_i,v_j) \end{vmatrix} hk \approx hkD \tag{4.129}$$

schreiben; hkD approximiert die Fläche von G_i^*. $f(\varphi(u_i,v_j),\psi(u,v))Dkh$ definiert dann ein Volumenelement, und die Summe über alle diese Elemente approximiert dann das Integral und man hat

$$\lim_{h\to 0,k\to 0} \sum_{i,j} f(\varphi(u_i,v_j),\psi(u_i,v_j))Dkh = \iint_{\mathcal{G}^*} f(\varphi(u,v),\psi(u,v))Ddudv.$$

□

Um das folgende Beispiel vorzubereiten, wird das folgende Lemma bewiesen:

Lemma: A habe vollen Rang. Dann gilt

$$(A^{-1})'A^{-1} = (AA')^{-1}. \tag{4.130}$$

Beweis: Es ist[59] $(AA')^{-1} = (A')^{-1}A^{-1}$. Multiplikation von links mit A' liefert $A'(AA')^{-1} = A^{-1}$, und die Transponierung der Gleichung führt auf $((AA')^{-1})'A = (A^{-1})'$. Multiplikation von rechts mit A^{-1} liefert schließlich $(AA')^{-1} = (A^{-1})'A^{-1}$, und das war zu zeigen. □

Beispiel 4.7 Die multivariate Normalverteilung Es sei $\mathbf{z} = (z_1, \ldots, z_n)'$ ein zufälliger Vektor, dessen Komponenten stochastisch unabhängig $N(0,1)$-verteilt seien, d.h. es soll

$$f(\mathbf{z}) = \frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{1}{2}\mathbf{z}'\mathbf{z}\right). \tag{4.131}$$

gelten. Weiter sei $\mathbf{x} = A\mathbf{z}+\boldsymbol{\mu}$, d.h. $\mathbf{x}$ und $\boldsymbol{\mu}$ seien ebenfalls n-dimensionale Vektoren und A sei eine (n,n)-Matrix mit vollem Rang. Dann ist die Verteilung von $\mathbf{x}$ durch

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}\sqrt{|\Sigma|}} \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\right). \tag{4.132}$$

gegeben.

Beweis: Es ist $\mathbf{x} - \boldsymbol{\mu} = A\mathbf{z}$, so dass

$$(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\mathbf{u})' = A\mathbf{z}\mathbf{z}'A'.$$

Dann ist $\mathbb{E}[(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})'] = A\mathbb{E}(\mathbf{z}\mathbf{z}')A'$, und

$$\mathbb{E}(\mathbf{z}\mathbf{z}') = (\mathbb{E}(z_i z_j)), \quad \mathbb{E}(z_i z_j) = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases},$$

wegen der vorausgesetzten Unabhängigkeit der Komponenten von $\mathbf{z}$, d.h. $\mathbb{E}(\mathbf{z}\mathbf{z}') = I$ die Einheitsmatrix, so dass

$$\boldsymbol{\Sigma} = \mathbb{E}[(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})'] = AA', \tag{4.133}$$

d.h. AA' ist gleich der Matrix der Varianzen und Kovarianzen der Komponenten von $\mathbf{x}$. Da A vollen Rang hat, existiert die zu A inverse Matrix A^{-1}, so dass

$$\mathbf{z} = A^{-1}(\mathbf{x}-\boldsymbol{\mu})$$

folgt, und nach (4.131) gilt

$$f(\mathbf{z}'\mathbf{z}) = f((\mathbf{x}-\boldsymbol{\mu})'(A^{-1})'A^{-1}(\mathbf{x}-\boldsymbol{\mu})|A^{-1}|$$

[59] Wegen $(AB)^{-1} = B^{-1}A^{-1}$

denn $d\mathbf{x}/d\mathbf{z} = A^{-1}$, und wegen (4.130) und nach dem Produktsatz für Determinanten gilt mit $\mathbf{\Sigma} = AA'$

$$|\Sigma| = |AA'| = |A||A'| = |A|^2$$

so dass $|A| = |\mathbf{\Sigma}|^{1/2}$, also erhält man für die Dichte von $\mathbf{x}$ die Dichte (4.132). □

Kommentar: $|\Sigma|^{1/2}$ ist gleich der Funktionaldeterminante, die dem φ' im 1-dimensionalen Fall entspricht. □

Literatur

[1] Barrantes Campos, H.: Elementos de álgebra lineal. *Editorio Unniversidad a Distancia*, San José, 2012

[2] Calcaterra, C. (2008) Linear combinations of Gaussinas with a single variance are dense in L^2. *Proceeding of the World Congress on Engineering*, Vol II, WCE 2008

[3] Cadima, J., Jolliffe, I. (2009) On relationships between uncentered and column-cedntered principal component analysis. *Pak. J. Statistics*, 25(4), 473 – 503

[4] Cattell, R.B. The data box: its orderung of total resources in terms of possible relational systems. In: Cattell, R.B. (ed): Handbook of multivariate experimental psychology. Chicago 1966

[5] Courant, R.: Vorlesungen über Differential- und Integralrechung, Erster Band: Funktionen mehrerer Veränderlicher. Springer Verlag Berlin 1961

[6] Courant, R.: Vorlesungen über Differential- und Integralrechung, Zweiter Band: Funktionen einer Veränderlichen. Springer Verlag Berlin 1963

[7] Dorier, J.L. (1995) A general outline of the genesis of vector space theory. *Historia Mathematica* 22, 227–261

[8] Eckart, C., Young, G. (1936), The approximation of one matrix by another of lower rank. *Psychometrika*, 1 (3): 211–8. doi:10.1007/BF02288367.

[9] Fischer, G.: Lineare Algebra. Braunschweig Wiesbaden 1997

[10] Forster, G.: Analysis 2 – Differentialrechnung im $\mathbb{R}^n$, gewöhnliche Differentialgelichungen. Springer- Sprektum, Wiesbaden 2017

[11] Forster, G.: Analysis 3 – Maß- und Integrationstheorie, Integralsätze im $\mathbb{R}^n$ und Anwendungen. Springer- Sprektum, Wiesbaden 2017

[12] Gabor, D. (1948) A new microscopic principle. *Nature*, 161, 777–778

[13] Gabriel, K.R. (1971) The biplot display of matrices with application to principal component analyis. *Biometrica* 58 (3), 453 – 467

[14] Golub, G.H., van Loan, C. F.: Matrix computations. Baltimore 2013

[15] Goshtaby, A., O'Neill, W. (1994) Curve fitting by a sum of Gaussians. *CVGIP: Graphical Models and Image Processing*, 56(4). 281–288

[16] Hoaglin, D,C, Welsch, R.E. (1978) The hat matrix in Regression and ANOVA. *The American Statistician*, 32(1), 17 – 22

[17] Honeine, P. (2014) An eigenanalysis of data centering in machine learning. *Statistics Machine Learning (stat.ML)* arXiv:1407.2904fl, 10 Jul 2014.

[18] Hotelling, H. (1933). Analysis of a Complex of Statistical Variables Into Principal Components, *Journal of Educational Psychology* , 24, 417–441 und 498-520. (10.97/year)

[19] Karhunen, K. (1947), Über lineare Methoden in der Wahrscheinlichkeitsrechnung. *Ann. Acad. Sci. Fennicae. Ser. A. I. Math.-Phys.*, 37, 1–79.

[20] Koecher, M.: Lineare Algebra und analytische Geometrie. Springer-Verlag, Berlin, Heidelberg, New York 1997

[21] Lange, M. (2009) A Tale of two Vectors. *Dialectica*, 63(4), 397–431

[22] Loève, M. (1978), Probability theory. Vol. II, 4th ed. Graduate Texts in Mathematics, 46. Springer-Verlag.

[23] Lorenz, F.: Lineare Algebra I, II. Mannheim, 1988

[24] Mardia, K.V., Kent, J. T., Bibby, J.M.: Multivariate Analysis. Academic Press, London, New York, Toronto 1979

[25] Papoulis, A.: Probability, random variables, and stochastic processes. Tokyo 1965

[26] Pearson, K. (1901) On lines and planes of closest fit to systems of points in space, *Philosophical Magazine, Series 6*, 2(11), pp. 559-572.

[27] Seber, G.A.F.: Linear regression analysis. John Wiley & Sons, New York, London, Sydney Toronto 1977

[28] Shaw-Taylor, J. Christianini, N.: Kernel Methods for Pattern Analysis. Cambridge University Press, Cambridge 2004

[29] Stewart, G. W.: Introduction to Matrix Computations. Academic Press, New York 1973

[30] Stewart, G. W. (1993) On the early history of the singular value decomposition. *SIAM Review*, 35 (4), 551 – 566

[31] Wong, E.: Stochastic processes in information and dynamical ssystems. New York 1971

Index

[Normalenvektor, 33
Ähnlichkeit, 20

Abbildung
    bijektiv, 156
    Bild von, 155
    homomorphe, 157
    identische, 155
    injektiv, 156
    inverse, 155
    isomorphe, 157
    Kern einer, 156
    lineare, 157
    Rang der Abbildung, 158
    surjektiv, 156
absolute Homogenität, 29
achsenparallel, 68
Adjunkte
    einer quadratischen Matrix, 167
algebraische Struktur, 29
Algorithmus, Gauß-, 140
Allgemeines Lineares Modell, 129
assoziativ, 48
Assoziativität, 48
Ausreißer, 130
Autokorrelation, 134
Automophismus, 157

basic structure, 84
Basis
    eines Vektorraums, 32
    geordnete, 174
    gleichorientierte, 174
    orthonormale, 34
Basisentwicklung eines Vektors
    orthonormale, 36
Basisfunktion, 31, 132
Basiswechsel, 34, 100

Cauchy-Schwarzsche Ungleichung, 20
charakteristische Gleichung, 168
charakteristisches Polynom, 168
Cholesky-Zerlegung, 142
cosine similarity, 20
Cramersche Regel, 28, 115

Definitheit, 29
Determinante, 27, 61
diagonalisierbar, 91
Diagonalmatrix, 43
Differential, 176
differenzierbar, total, 175
Dimension, 41
Dimensionalitätsatz, 39
Diskriminanzkoeffizient, 123
Distanz
    Dreiecksungleichung, 8
    euklidische, 30
    Minkowski, 30
    Reflexivität, 8
Dreiecksmatrix, 139
Dreiecksungleichung, 29

Ebene
    in einem Vektorraum, 33
Eckart & Young
    Satz von, 120
Eigenraum, 93
Eigenvektor, 63, 69
    Links-, 89
    Rechts-, 89
Eigenwert
    komplexer, 92
    Nullstellen eines Polynoms, 92
Einflußmatrix, 129
Einheitsvektor
    kanonischer, 12
Einheitsvektoren, 18
Einsvektor, 12
elementare Umformungen, 138
Elementarmatrix, 138
Eliminationsverfahren, gaußsches, 140
Ellipsoid, 66
Endomorphismus, 157

Endomporphismus, 169
Erwartungswert, 51
Erzeugendensystem, 31

Faktorwerte, 117
Faser, 156
Fundamental-Lemma, 144
Funktionalmatrix, 176
Funktionenraum, 131

Gebietsintegral, 172
Gesamtvarianz der Daten, 120
Gleichung, charakteristische, 88
Gradientenvektor, 151
Gram-Matrix, 68
Grundformel einf. einer neuen Variablen, 177
Grundstruktur einer Matrix, 84

Hat-Matrix, 129
Hauptachsentransformation, 72
Hauptraum, 94
Hauptvektor, 94
Hebelwirkung, 66, 130
Homomorphismus, 157

idempotent, 97
Identität, 155
influence matrix, 129
Inverse, 59, 150
    generalisierte, 102
    Links-, 58
    Moore-Penrose, 102
    Pseudo, 102
    Rechts-, 58
Isomorphismus, 157

Jacobi-Matrix, 176

Kern, 57, 112
Kettenregel, 176
Kodimension, 41
Kofaktor, 167
Kofaktoren, 166
kollinear, 27
kommutativ, 48
Komplement
    orthogonales, 39
Koordinaten bezüglich einer Basis, 32
Korrelation
    kanonische, 95
Kosinus-Ähnlichkeit, 20
Kovarianzen
    von Linearkombinationen unabhängiger Variablen, 52
Kreuzproduktmatrizen, 57
    Eigenwerte, 76
    Rang von, 58

längeninvariant, 62
Längenskalierung, 49
Ladung, 36
Ladungen, 117
Lagrange
    -Faktor, 152
    -Funktion, 152
    -Multplikator, 152
    sche Multiplikatorenregel, 152
latente Dimension, 116
latente Variable, 22, 116
LDL-Zerlegung, 143
Leibniz-Formel, 162
leverage, 66, 130
Lineare Gleichungssysteme, 25
lineare Hülle, 31
lineare Hülle (Gleich'syst.), 112
Linearkombination, 14
Linkseigenvektoren, 84
Linksinverse, 58

Matrix
    ähnliche, 167
    Wurzel einer symmetrischen, 79
    assoziierte, 93
    charakterische Gleichung, 88
    Diagonal-, 43
    Dreiecks, 139
    gestürzte, transponierte, 42
    hermitesch, 93
    imaginär, 93

inverse, 59
inverse (2-dim), 60
konjugierte, 92
Präzisions, 61
schief-symmetrisch, 93
spaltenstandardisiert, 65
spaltenzentriert, 65
symmetrische, 43
zufällige, 50
Matrixnorm, 105
Matrizen
ähnliche (Fußnote), 69
ähnliche, 91
Matrizenkalkül, 7
Metrik, 8
City-Block, 10
euklidisch, 9
euklidische, 9, 30
Manhattan, 10
Minkowski, 9, 30
Minor, 166, 167

negativ semidefinit, 67
Norm
p-, 106
-Matrix, 105
allgemein, 29
definierende Eigenschaften, 29
euklidische, 104
Frobenius-, 105
Hilbert-Schmidt-, 105
induzierte, 106
Maximum, 105
Schur-, 105
Vektor-, 17
Normalverteilung
multivariate, 182
Nullität, 57
Nullmatrix, 54
Nullraum, 57
Nullvektor, 12

Operator
linearer, 52
orientiert
negativ, 174
positiv, 174
Orientierung
eines Vektorraums, 172
Orientierungsinvarianz, 63
orthogonales Komplement, 39
Orthogonalität, 21
Orthonormalbasis (ONB), 34
orthonormale Basisentwicklung, 36

Permutation, 160
Pfad, 134
Polynom, charakteristisches, 70, 88
positiv semidefinit, 67, 76
Prädiktoren, 14
Präzisionsmatrix, 61
Produkt
dyadisches, 16, 108
dyadisches und Kovarianzmatrizen, 51
inneres, 16
tensorielles, 16
Produktsatz, 163
Projektion
Vektor auf einen anderen, 127
Projektionsmatrix, 128, 149
Pythagoras, 18

quadratische Form, 67

Rang
einer Vektormenge, 41
voller, 41
Rayleigh-Koeffizient, 81
Rayleigh-Quotient, 81
generalisierter, 95
Rechtseigenvektoren, 84
Rechtsinverse, 58
Rotation, 34, 64
Rotationsmatrix, 62

Sarrussche Regel, 164
Satz
Eckart-Young, 108

von Courant-Fischer, 81
von Schmidt-Mirsky, 108, 109
Schur-Norm, 105
Schwarzsche Ungleichung, 20
singulär, 70
Singularwert, 84
Singularwertzerlegung, 83
Skalar, 17
Skalarprodukt, 16
Skalarprodukt als Ähnlichkeitsmaß, 21
Spaltenrang, 53
Spaltenraum, 54, 158
Spaltenvektoren, 42
Spektraldarstellung, 75
Standardmatrix, 138
stetig (im quadrat. Mittel), 134
stochastischer Prozess
zentrierter, 134
Subaddditivität, 29

Teilräume
Dimensionsformel, 41
Durchschnitt, 40
Kombination von, 39
orthogonale, 38
orthogonale Summe, 39
Summe zweier, 40
Vereinigung, 40
Teilraum
r-dimensionaler, 33
Dimension des, 41
invarianter, 94
orthogonales Komplement, 39
orthonormale Basis, 38
Teilraum, Teilvektorraum, 31
Teilraum, trivialer, 31
Teilvektorraum, 31
Trajektorie, 134
Transformationssatz, 181
Transposition, 160

Umformungen, elementare, 140
Unterraum, 41

Vektor
abbildung, 44
charakteristischer, 63, 69
Eigen-, 63
Eins-, 12
latenter, 69
nicht messfehlerfreier, 25
Norm, 18
normiert, 18
Null-, 12
transformation, 44
Zeilen-, 17
zufälliger, 15, 50
zuordnung, 44
Vektornorm, 104, 106
Vektorprodukt, 173
Vektorraum, 29
n-dimensiónaler, 33
euklidischer, 30
normierter, 105

Wert
charakteristischer, 69
latenter, 69

Zeilenrang, 53
Zeilenraum, 54
Zeilenvektor, 17
Zeilenvektoren, 42
Zentrierungsmatrix, 97
Zufallsmatrix, 50
Zufallsvektor, 50